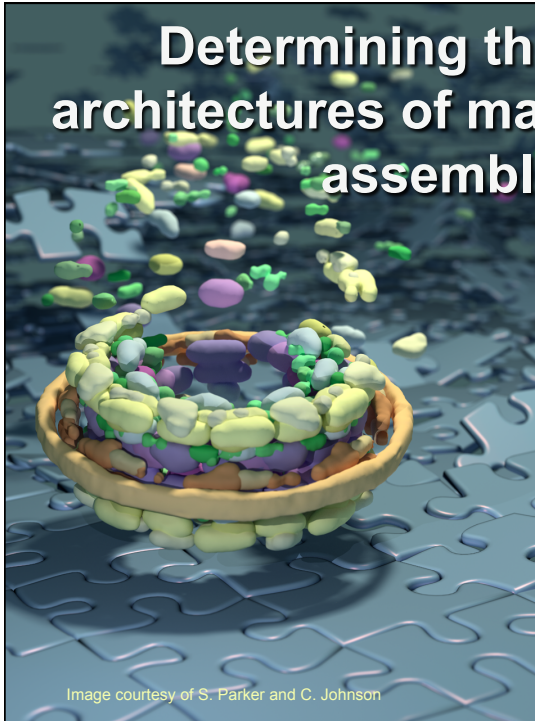


Determining the subunit architectures of macromolecular assemblies

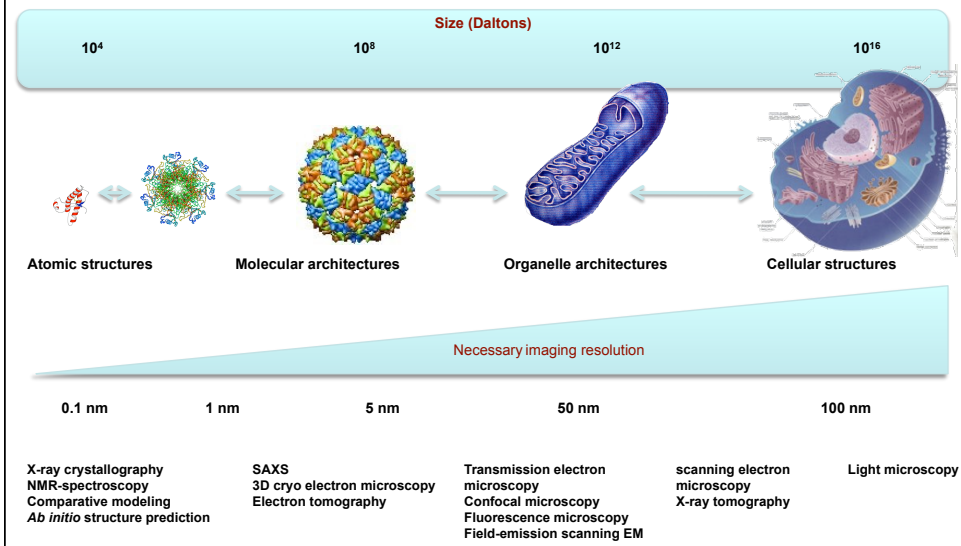


Frank Alber
 Molecular & Computational
 Biology
 University of Southern
 California (USC)
 Los Angeles

 alber@usc.edu

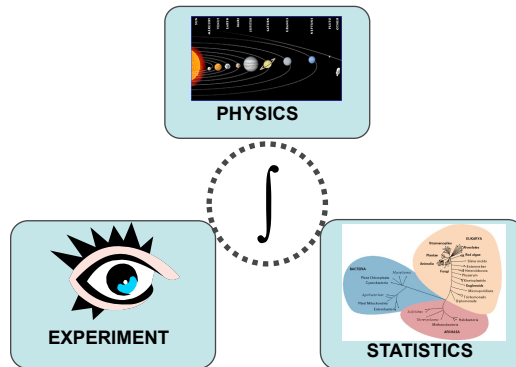
Image courtesy of S. Parker and C. Johnson

Structure at different levels of organization



Integrative approach

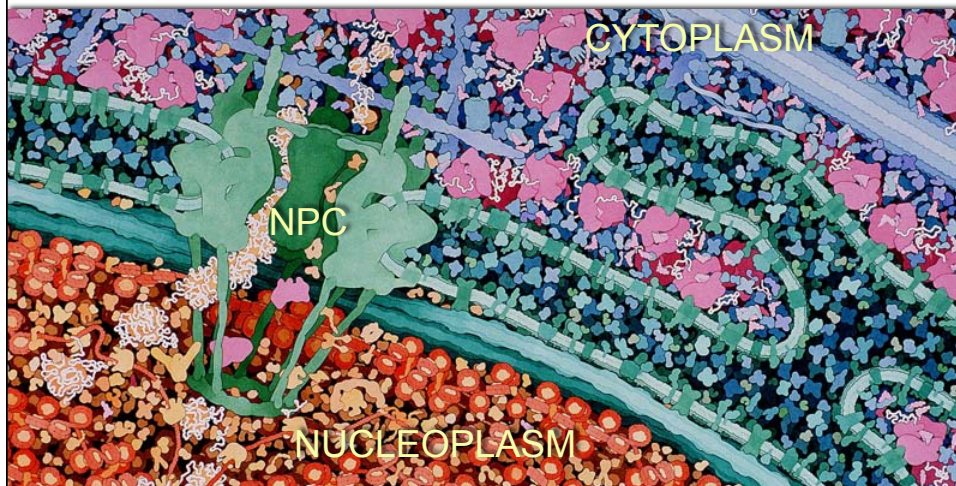
Use structural **information** from any
source: measurement, first principles, rules,
resolution: low or high resolution
to obtain the set of **all** models that are consistent with it.



Maximize **efficiency**, **accuracy**, **resolution**, and **completeness**
of the structural coverage of proteins and their assemblies.

9/30/08

Cellular complexity

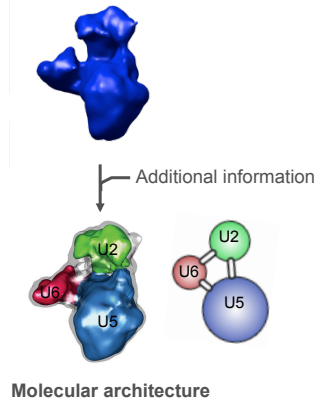
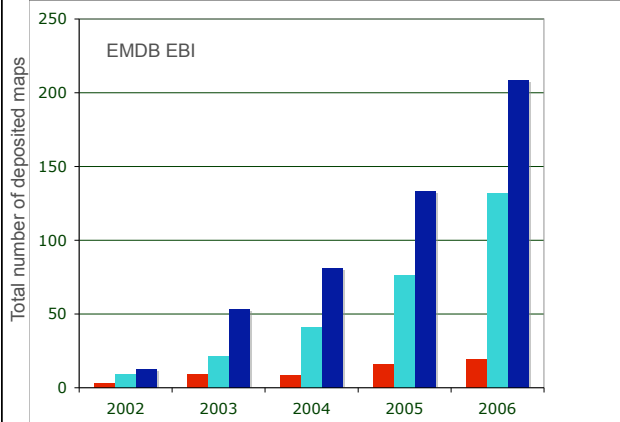


Painting from David Goodsell, Scripps Institute

9/30/08

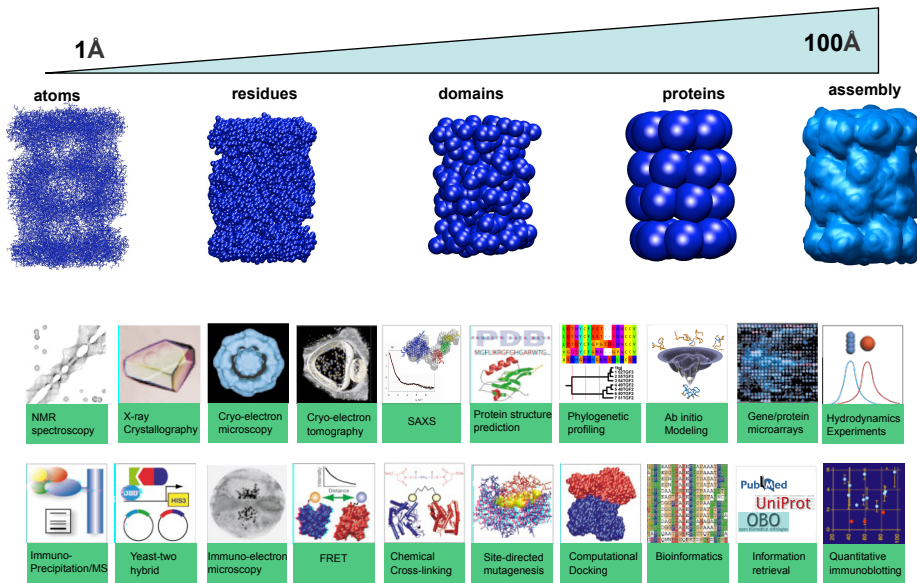
Challenges in assembly structure characterization

Deposited electron microscopy maps



9/30/08

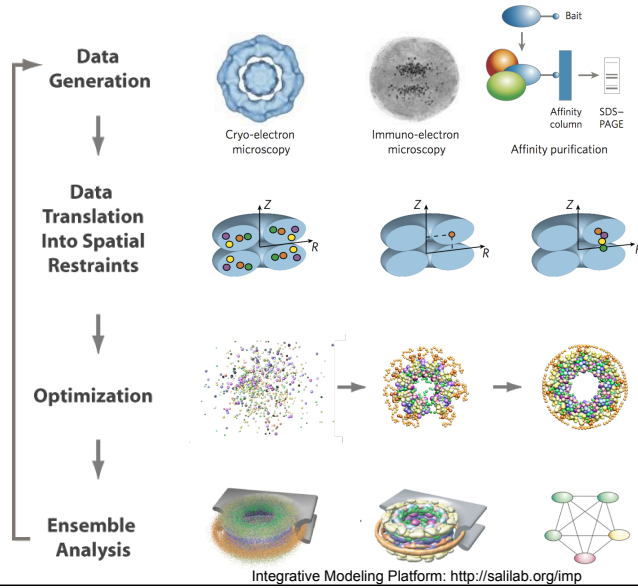
Resolution of spatial data



9/30/08

Characterizing Structures by Satisfaction of Spatial Restraints

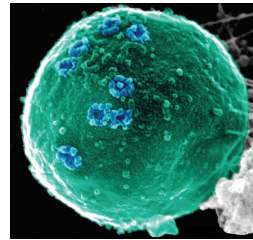
Alber *et al.* *Nature* 450, 683-694, 2007.
Robinson, Sali, Baumeister. *Nature* 450, 974-982, 2007.



02/15/2007

Nuclear Pore Complex

1. Structure
2. Evolution
3. Mechanism of transport
4. Mechanism of assembly



Andrej Sali (UCSF)
Damien Devos (EMBL Heidelberg)

Mike Rout (Rockefeller University)
Svetlana Dokudovskaya, Liesbeth Veenhoff
Orit Karni-Schmidt, Julia Kipper, Tari Suprpto,
Julia Kipper, Rosemary Williams

Brian Chait (Rockefeller University)
Wenzhu Zhang

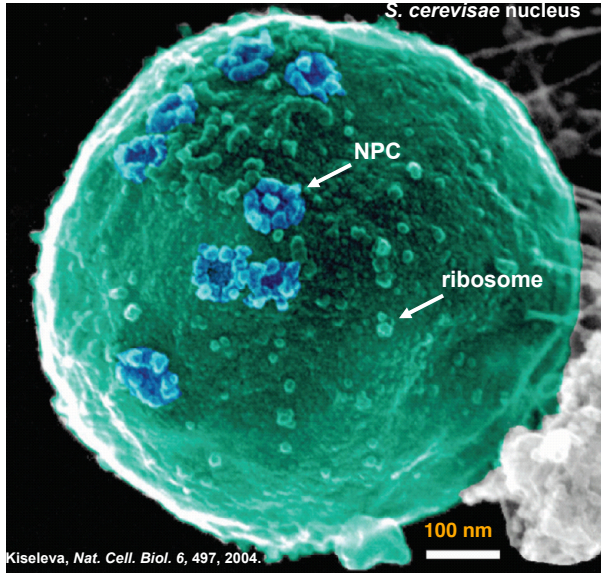


Alber *et al.* *Nature* 450, 683-694, 2007
Alber *et al.* *Nature* 450, 695-701, 2007
Devos *et al.* *PLoS Biology* 12, 1-9, 2004
Devos *et al.* *PNAS* 14, 2172-2177, 2006



9/30/08

Nuclear Pore Complex (NPC)



field emission scanning EM image

Consists of broadly conserved **nucleoporins** (nups).

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

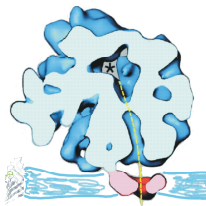
Mediates all known nuclear **transport**, via cognate transport factors.

9/30/08

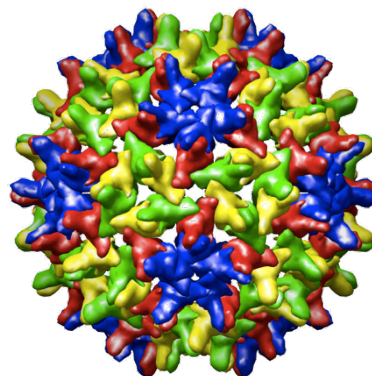
Objects actively transported by the nuclear pore complex



>8 nm diameter

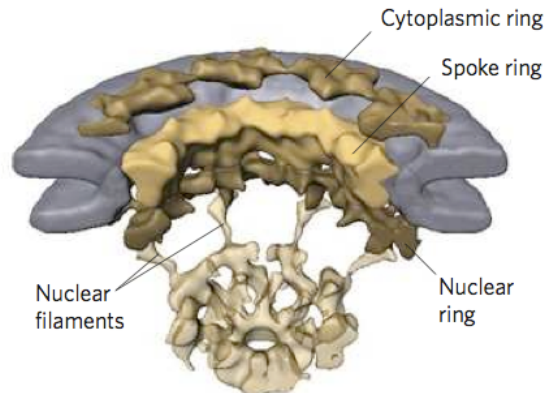


20 nm diameter



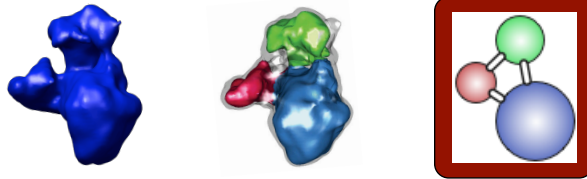
32-36 nm diameter

What is known about the NPC structure?

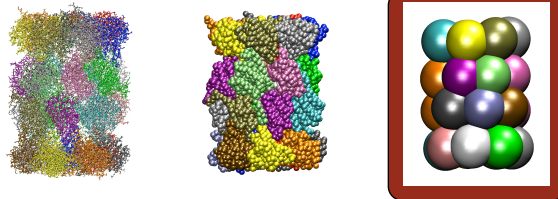


Beck, Lucic, Foerster, Baumeister, Medalia, Nature, 449 2007.

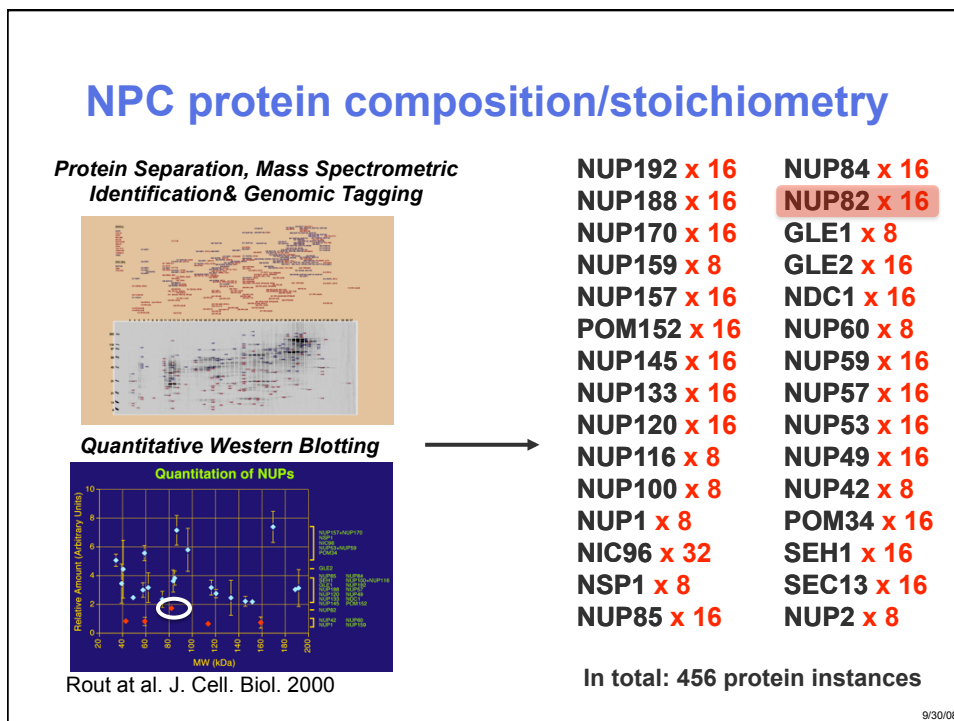
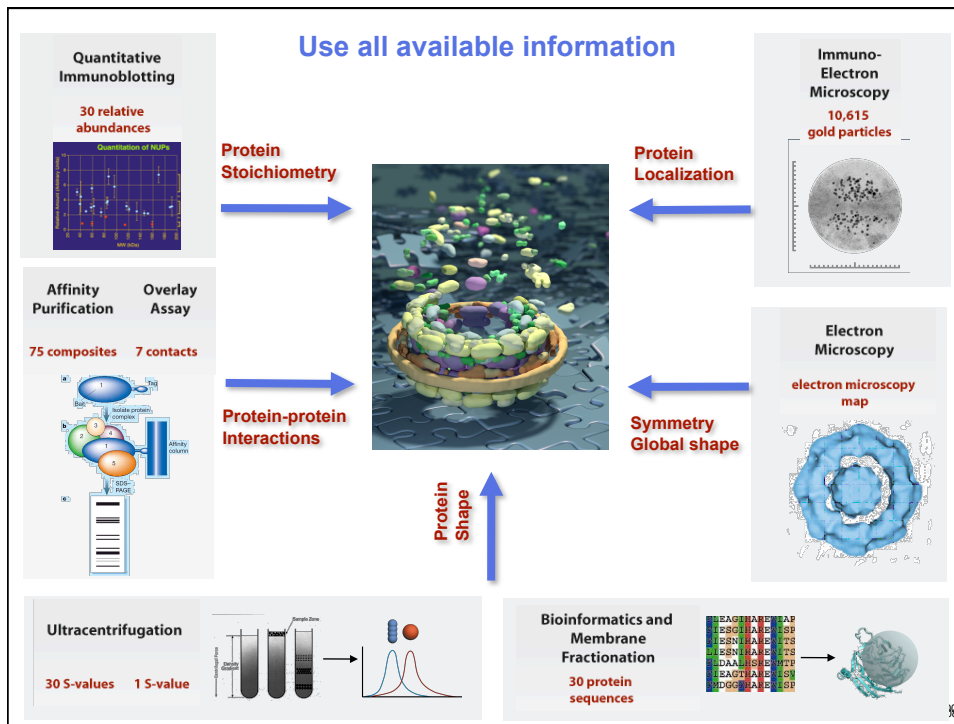
9/30/08



How to determine the **subunit architecture** of the NPC
(*ie*, the configuration of the nups)?



9/30/08

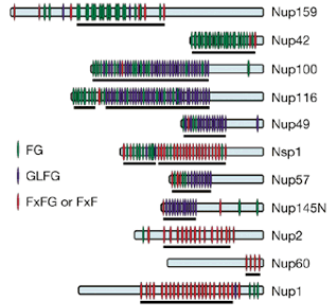


Representation

FG Repeat Nups in Yeast

- ⇒ Disordered Filaments
- ⇒ Found at NPC Filaments
- ⇒ Karyopherin Docking Sites

Protein	Mw (kDa) (Protein +PrA)	S _{0.0001M}	Bead number	Bead representation	Bead radius (nm)
Nup192	217	9.1	2		3.0
Nup188	214	9.3	2		3.0
Nup170	195	9.0	2		2.9
Nup159	185	5.1	11		1.6
Nup157	183	7.1	3		2.5
Pom152	178	-	10		1.6
Pom152*	155	4.8			
Nup133	159	7.6	2		2.7
Nup120	146	8.0	2		2.6
Nup116*	100	3.6	13		1.4
Nup1	140	-			
Nup1*	120	3.7	9		1.5
Nup100	126	4.4	13		1.4
Nic96	122	6.3	2		2.4
Nsp1	112	3.5	12		1.3
Nup85/Seh1	150	6.8			
Nup85	111	-	3		2.0
Nup84	110	4.9	3		2.0
Nup82	108	5.3	2		2.3
Nup145C	107	6.7	2		2.3
Ndc1	100	-	2		2.2
Gle1	88	5.9	2		2.1
Nup60	85	3.8	4		1.6
Nup59	85	4.2	4		1.6
Nup57	83	4.1	3		1.8
Nup53	79	4.1	3		1.7
Nup145N	86	3.7	6		1.5
Nup49	75	3.9	3		1.7
Nup42	69	3.0	5		1.4
Gle2	66	4.6	1		2.3
Seh1	65	3.4	1		2.2
Pom34	60	-	3		1.5
Sec13	59	4.2	1		2.1



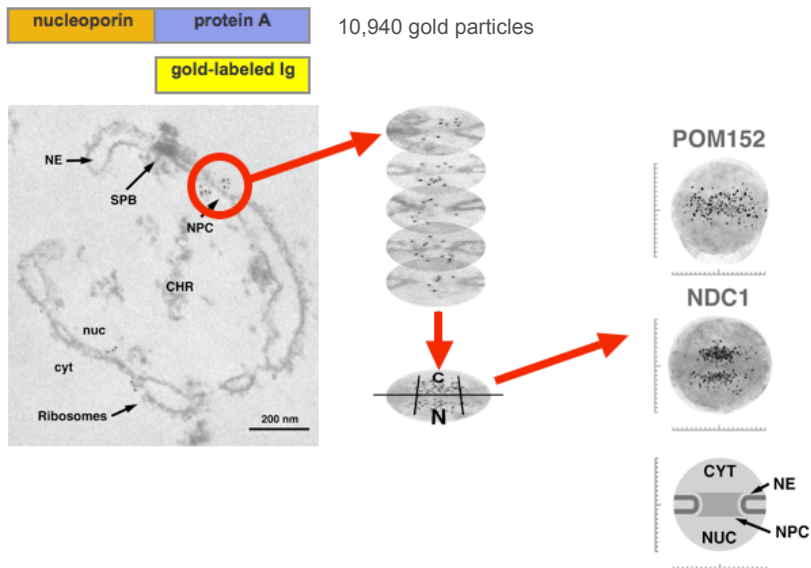
Consensus sequence of FG repeat region of Nsp1p:

```

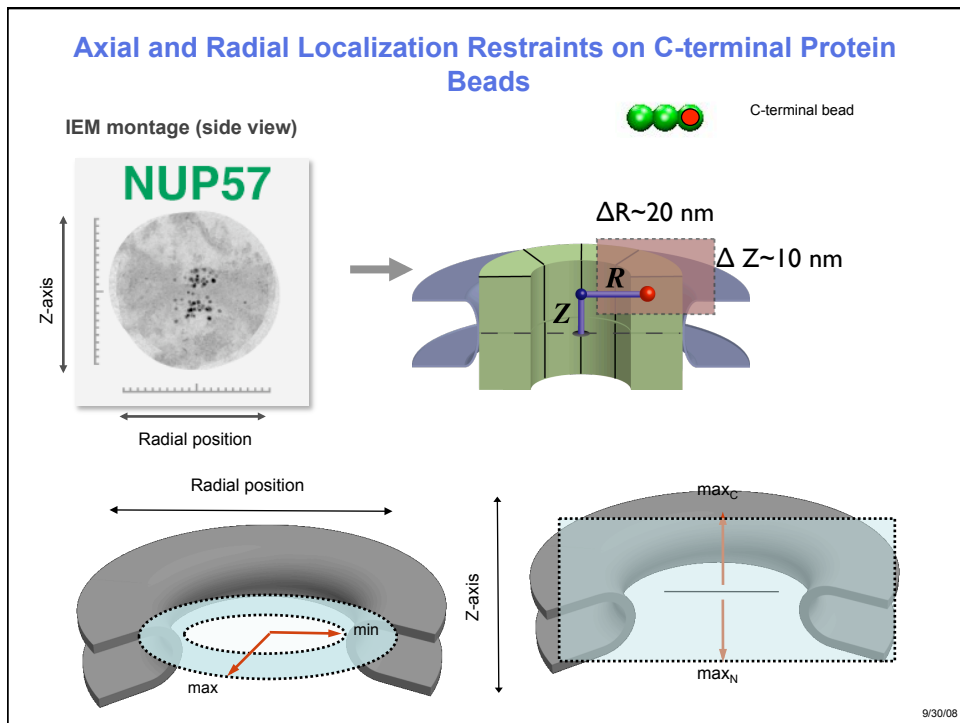
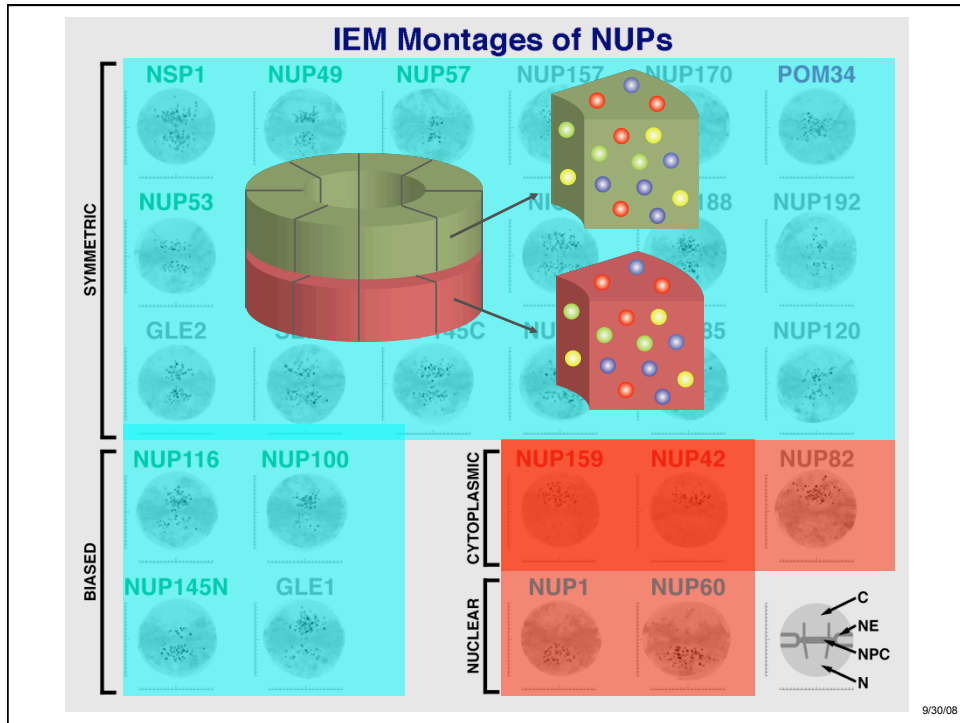
...PSFSFGAKSDENKAGATSK
PAFSFGAKPEEKDDNSSK
PAFSFGAKSNEKDQGTAK
PAFSFGAKPAEKNNNETSK...
    
```

9/30/08

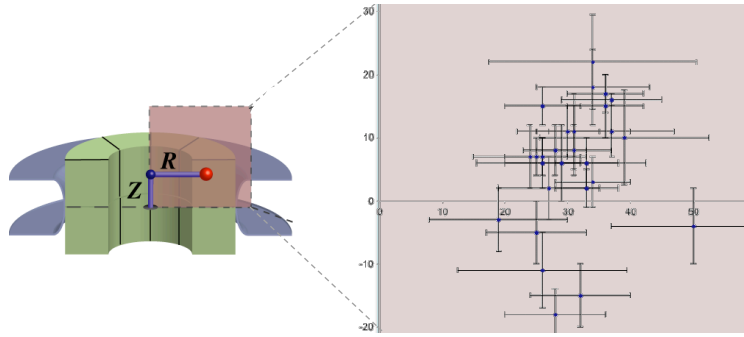
Immuno-electron microscopy



9/30/08



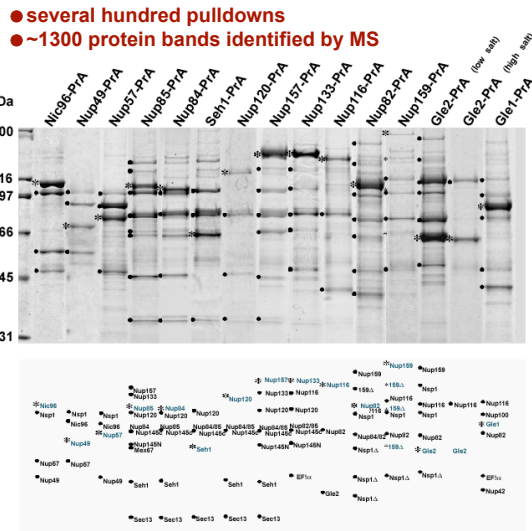
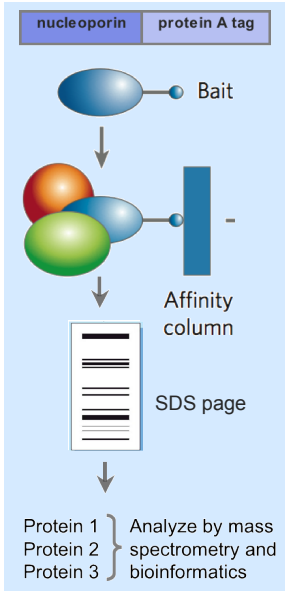
2 dimensional localization



9/30/08

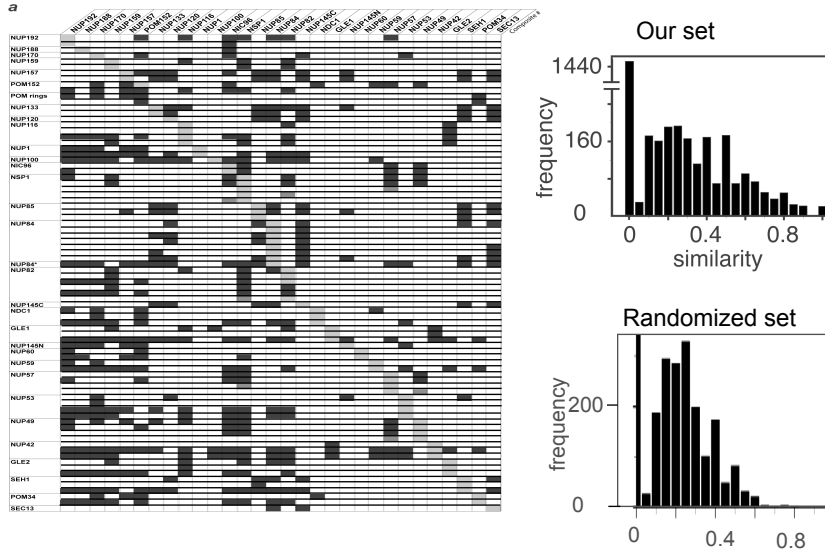
Affinity purifications:

Tagging, Immunoprecipitation and Analysis of Protein Subcomplexes



9/30/08

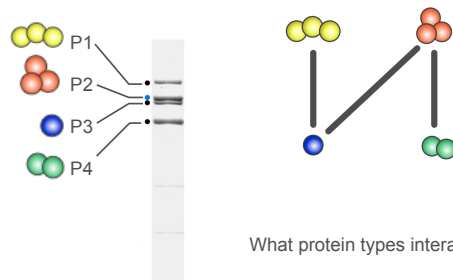
Affinity purifications: Tagging, Immunoprecipitation and Analysis of Protein Subcomplexes



9/30/08

Complex connectivity

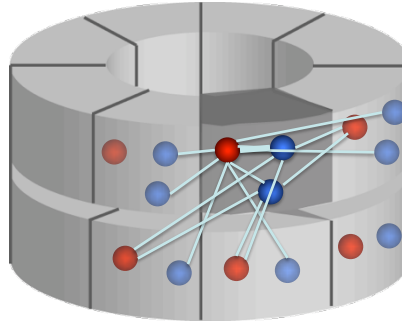
at least three $(n-1)$ interactions must be present connecting four (n) protein types



9/30/08

Protein interactions

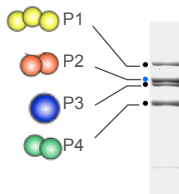
Which protein pair interacts ?



How do proteins interact?

9/30/08

Pulldowns are informative structurally, but subject to assignment ambiguity



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

01/08/2008

Conditional restraints

Assignment of restraint becomes part of the optimization process



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

9/30/08

Subcomplex shape

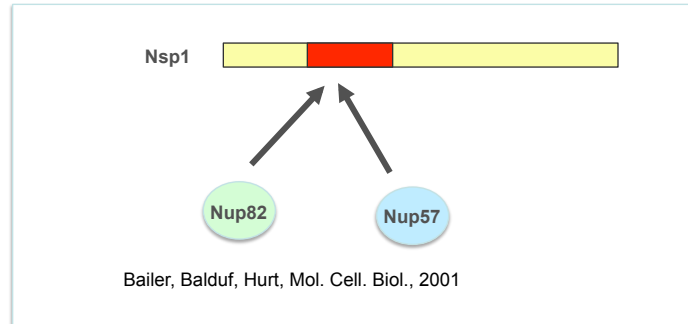
- Hydrodynamics – complex diameter restraint



- Cryo-EM
- SAXS

9/30/08

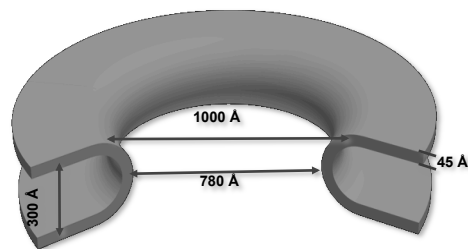
Competitive protein binding



Context dependent spatial restraint

9/30/08

Nuclear Envelope Provides a Mould for NPC



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

Trans-membrane regions:

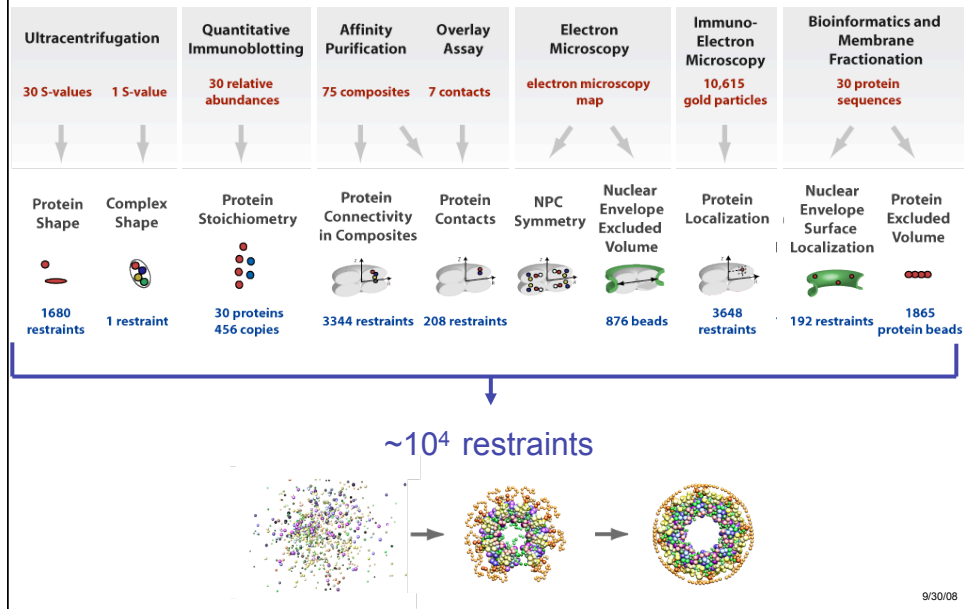
Pom152

Pom34

Ndc1

*restrained to NE surface

Use all spatial information



Optimization

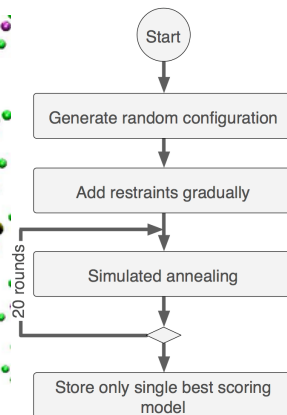
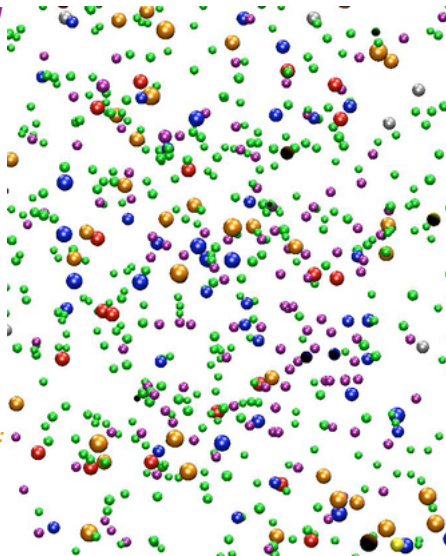
Membrane spanning proteins:
Pom152 Pom34
Ndc1

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



9/30/08

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

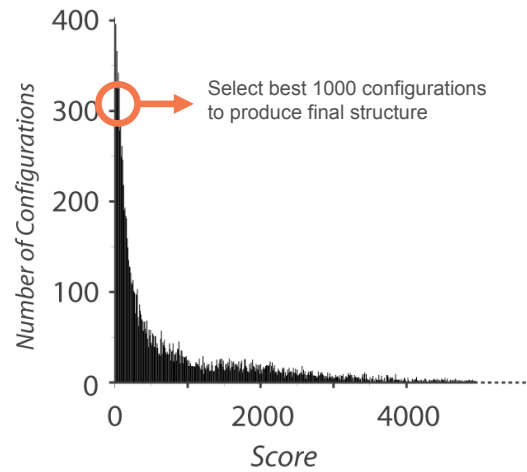
Nup84 complex:

Nup84 Seh1
Nup85 Sec13
Nup120 Nup145C
Nup133

Large Core proteins:

Nup192 Nup170
Nup188 Nup157

Nup82
Nic96



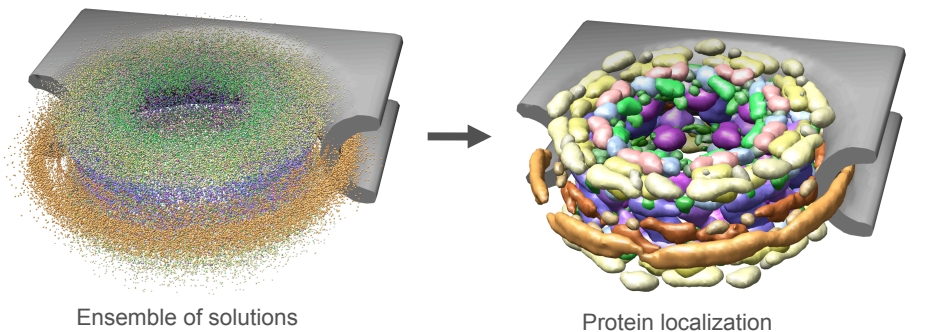
9/30/08

The molecular architecture of the NPC (the configuration of the nups)

9/30/08

Protein Localization Probability

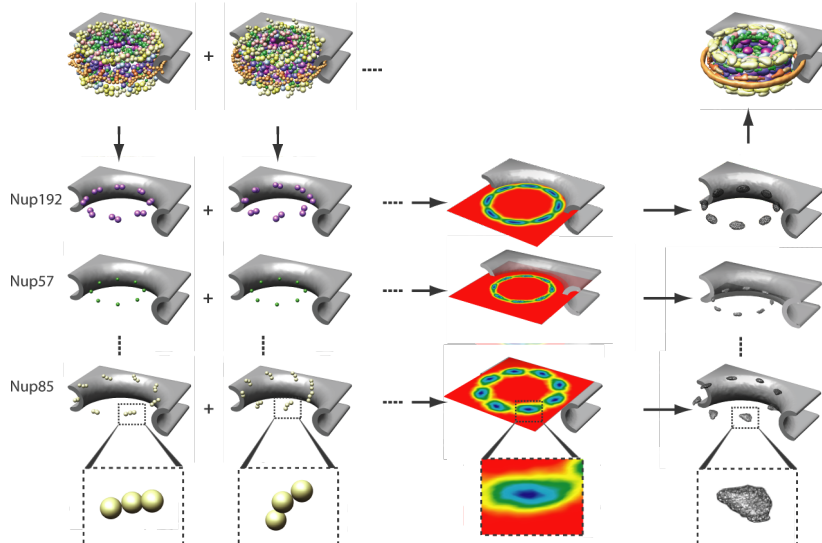
Calculated from the structural superposition of the ensemble of models that satisfy all input restraints



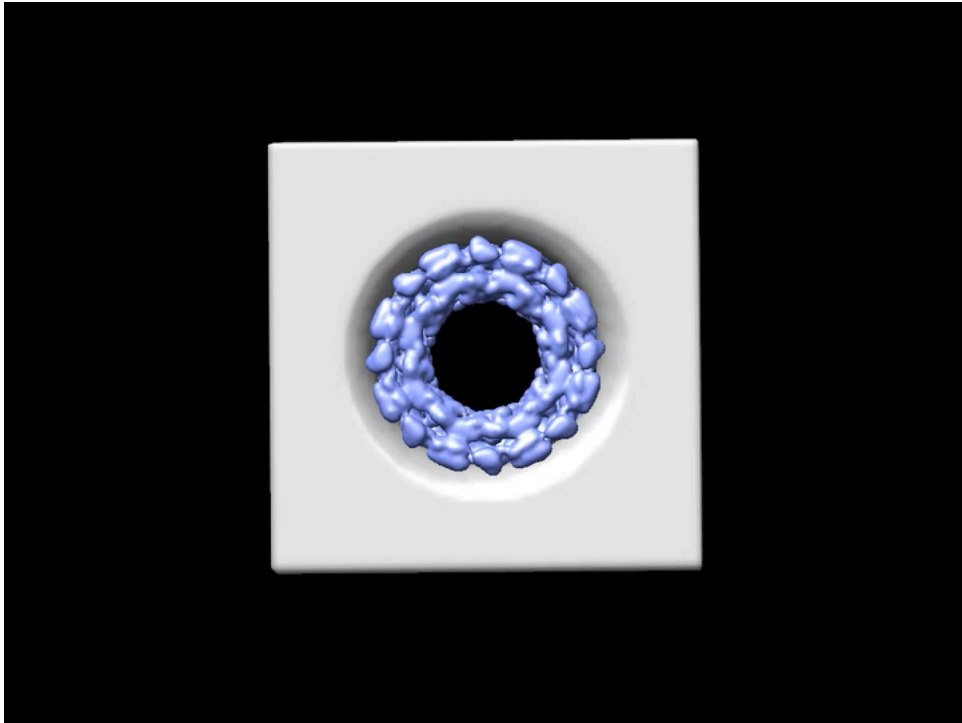
can see position of every NPC protein

9/30/08

Protein Localizations

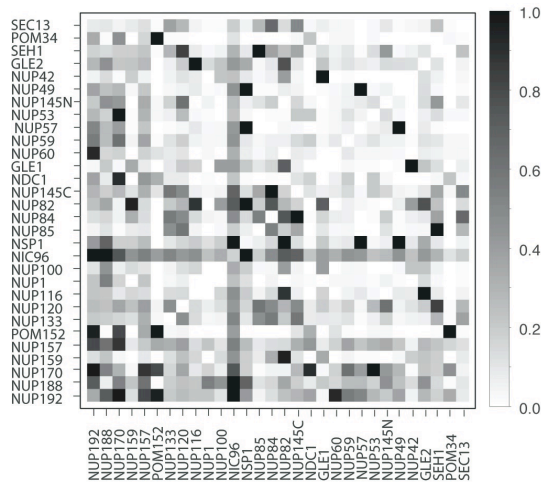


9/30/08



Protein Contacts

There is enough information to describe possible protein interactions

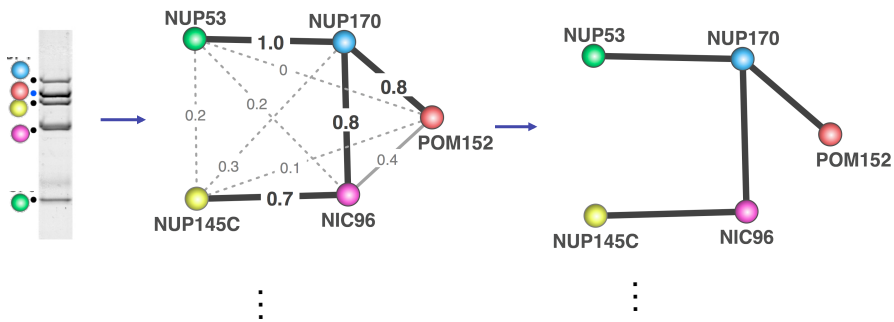


from the ensemble of all structures that satisfy all restraints

9/30/08

Protein contacts

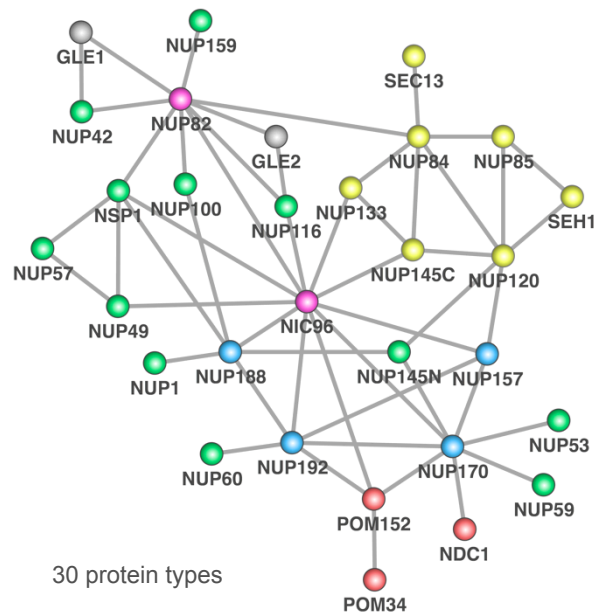
How many times is a protein contact present in the ensemble ?



80 % of all protein pairs can be assigned as a clear contact or non-contact

9/30/08

Protein contacts

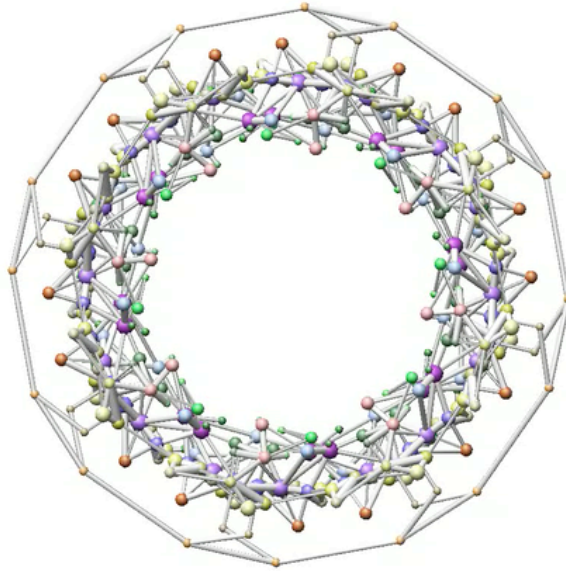


30 protein types

9/30/08

Protein configuration

456 protein instances and their spatial relationship



9/30/08

How accurate is the structure of the NPC?

Assessing the well-scoring models

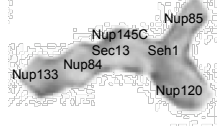
1. Self-consistency of independent experimental data.
2. Structural similarity among the configurations in the ensemble that satisfy the input restraints.
3. Simulations where a native structure is assumed, corresponding restraints simulated from it, and the resulting calculated structure compared with 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.

9/30/08

Assessment:

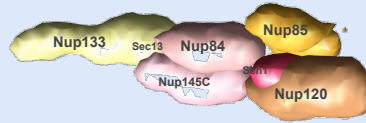
Experimental spatial data about the modeled structure that were not used in the calculation of the model

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

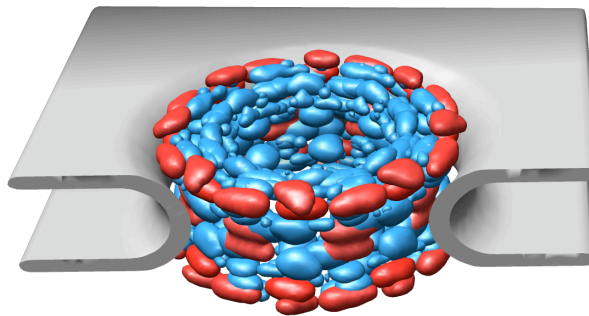


Our Structure

9/30/08

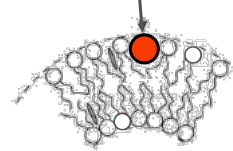
Nup133 senses membrane curvature

Amphipathic α -helical (ALPS-like) motive that binds (stabilizes) curved membranes



Nup133

α -helix



Drin et al.
Nat Struct Mol Biol 2007

predicted to be present
also in 4 other nups

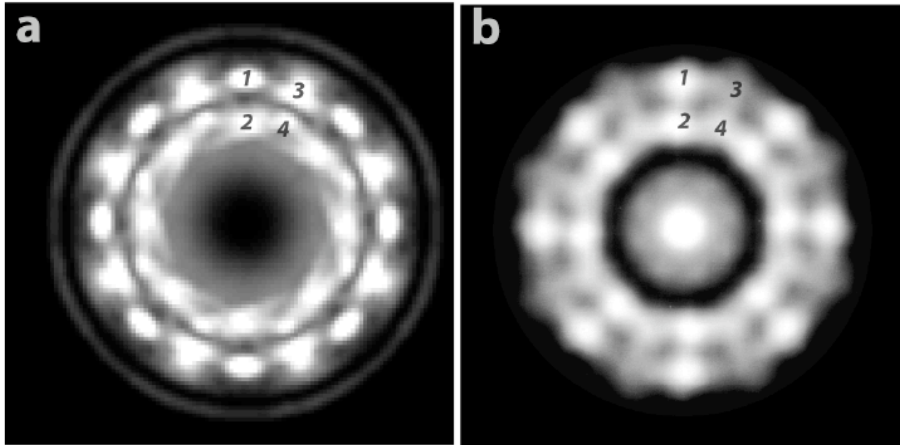
Nup133, Nup120, Nup85, Nup170, Nup188

9/30/08

Density projections

Our structure

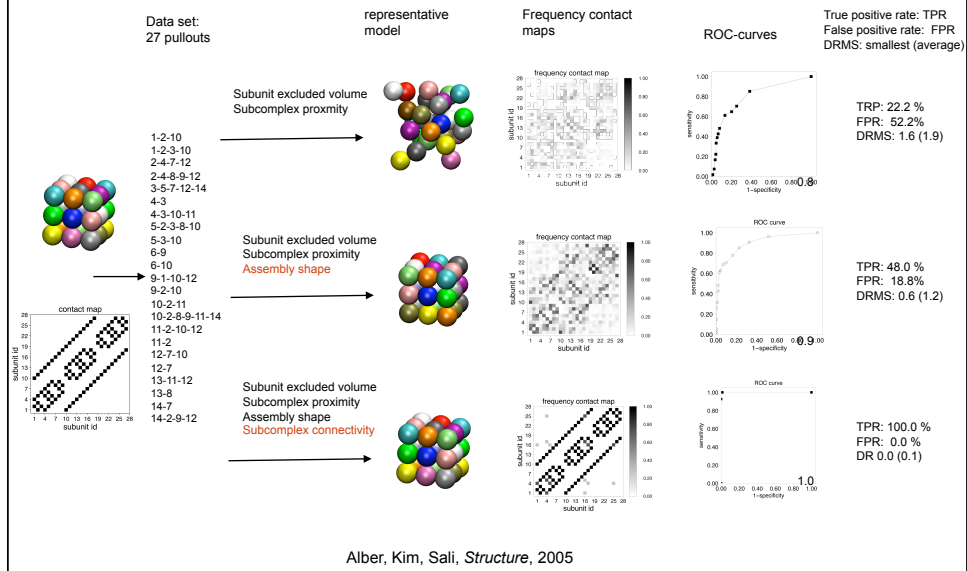
Cryo-EM density reconstruction



Courtesy of Chris Akey
Yang et al, *Mol. Cell.* 1998.

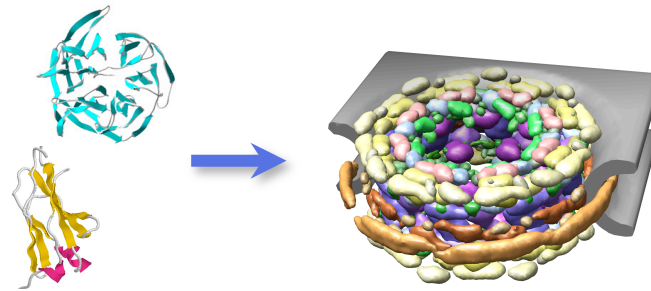
9/30/08

Simplified test case



Towards a higher resolution structure of NPC?

Characterize structures of the individual subunits, then fit them into the current low-resolution structure (aided by cross-linking information and cryoEM maps of subcomplexes).

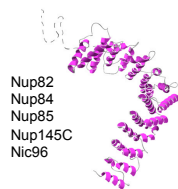


9/30/08

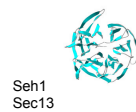
Fold Prediction

Devos, Dokudavskaya, Alber, Williams, Chait, Sali, Rout. *PLoS Biology* 12, 1, 2004
 Devos, Dokudavskaya, Williams, Alber, Eswar, Chait, Rout, Sali, *PNAS* 14, 2172, 2006.

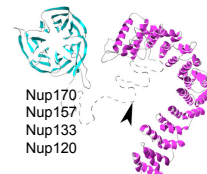
α -solenoid



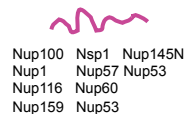
β -propeller



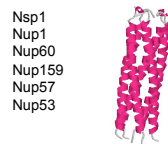
Clathrin-like



unstructured-FG repeat regions



Coiled-coiled



IgG-fold



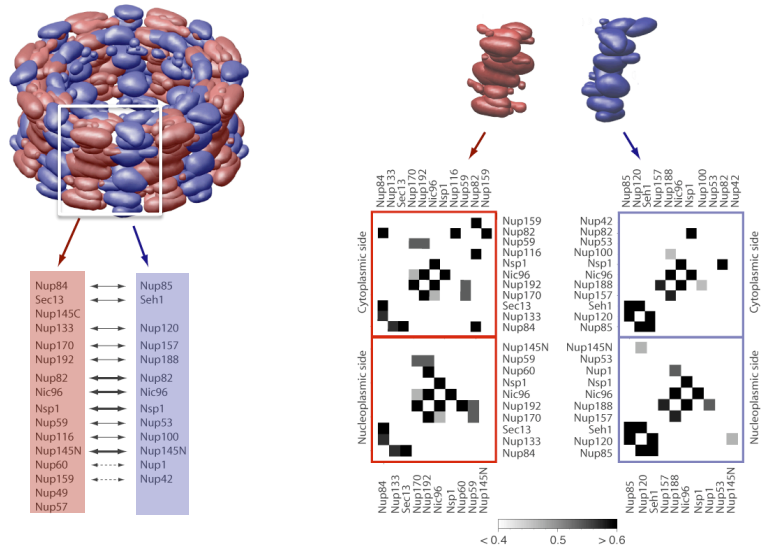
Trans-membrane helices



These fold assignments cover all 44 domains and 95% of the NPC residues.

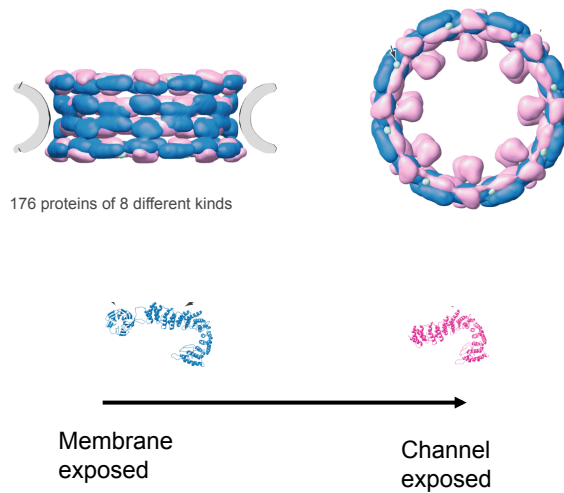
9/30/08

Proteins can be grouped into a relatively small number of modules that are structurally similar

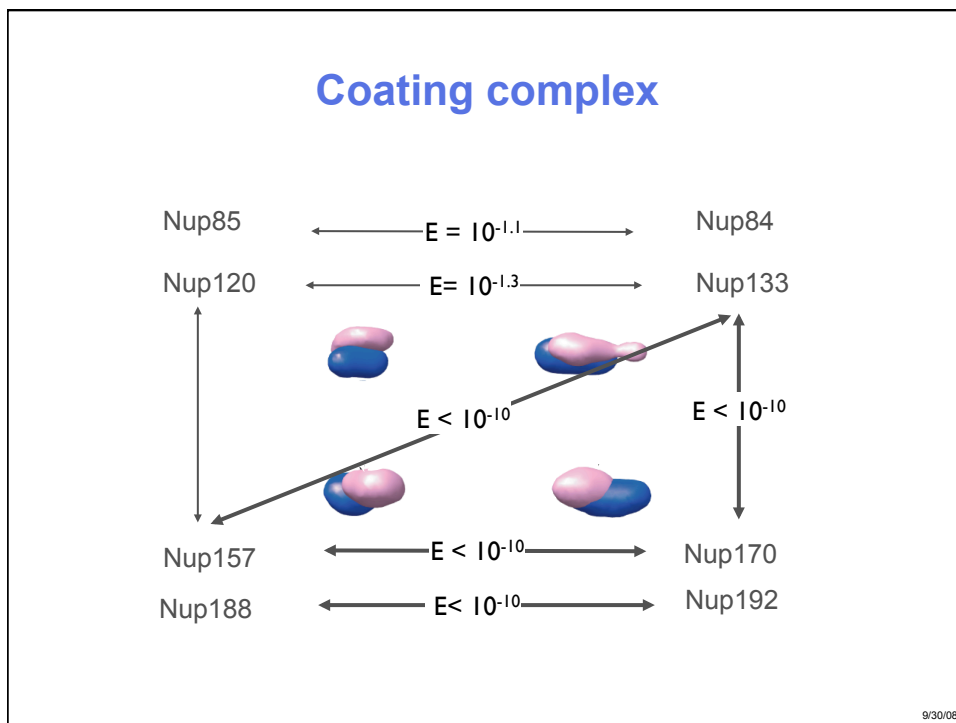
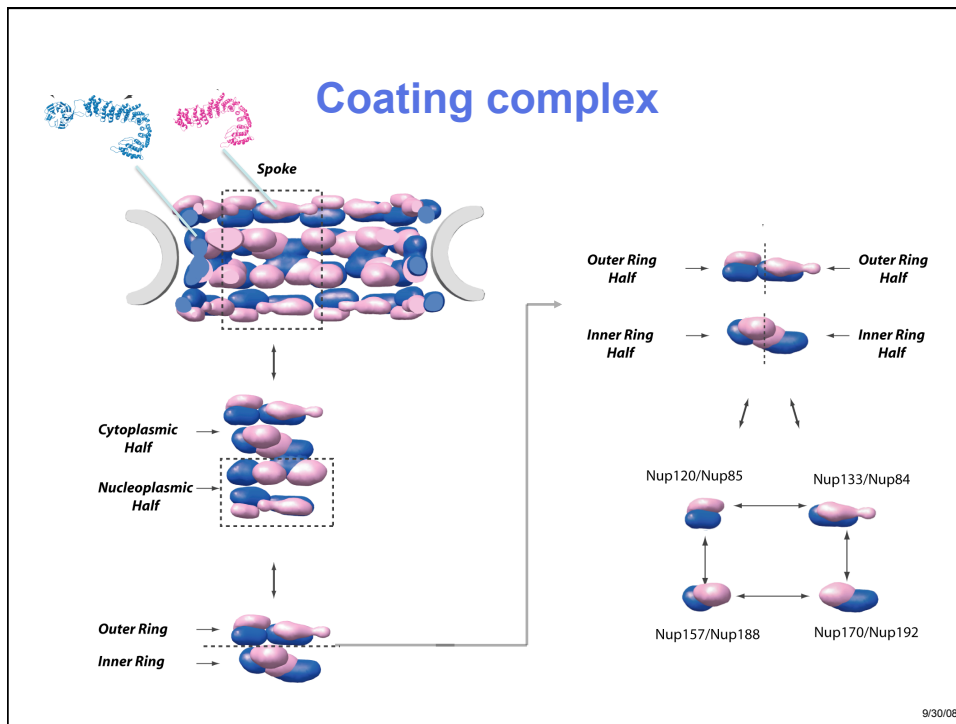


9/30/08

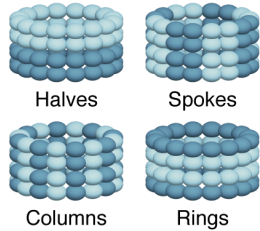
Coating complex



9/30/08



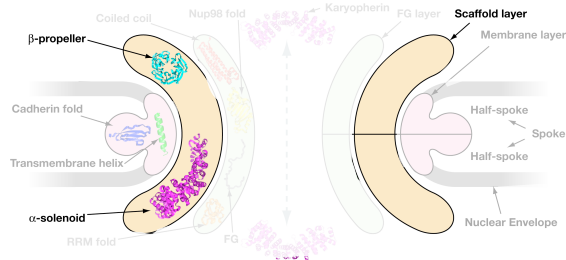
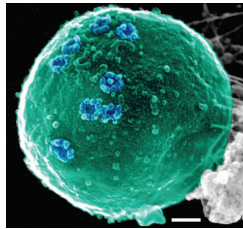
Simplicity of the fold composition and modular architecture of the NPC



Only 8 fold types account for 95% of the NPC residues (3 fold types cover ~85% of the NPC).

Evolved by extensive intra- and inter-gene duplication.

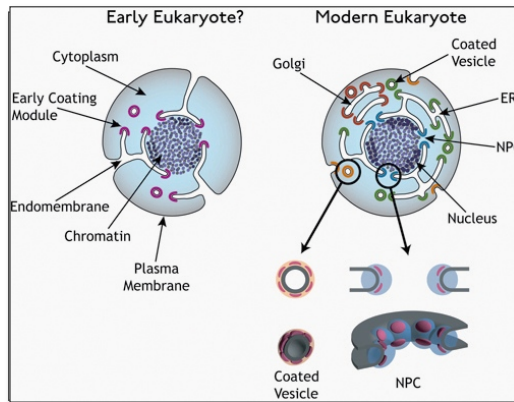
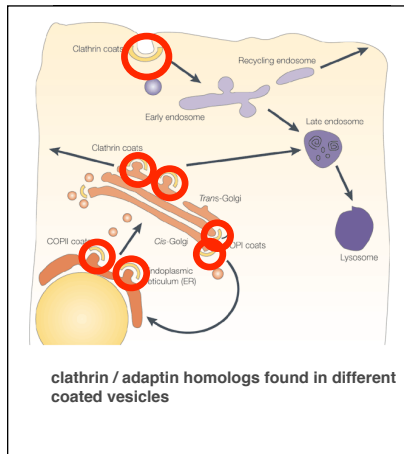
Minimal layered architecture, given the demands of function.



9/30/08

Membrane coating complex: Structure and origin

Devos, Dokudavskaya, Alber, Williams, Chait, Sali, Rout. *PLoS Biology* 12, 1, 2004
Devos, Dokudavskaya, Williams, Alber, Eswar, Chait, Rout, Sali, *PNAS* 14, 2172, 2006.



Coated Vesicles and NPCs originated in a common precursor:
Protocoatmer hypothesis.

Present day "intermediates"?

9/30/08

Divergent vs convergent?

- **Divergent**
 - Fold types similarities
 - Combination of folds in complex
 - Combination of domain folds in proteins
 - Size similarities
 - Function similarities
- **Convergent**
 - Lack of sequence similarity
 - Different adaptor between COPII and COPI/CCV

Makes a lot of sense

9/30/08

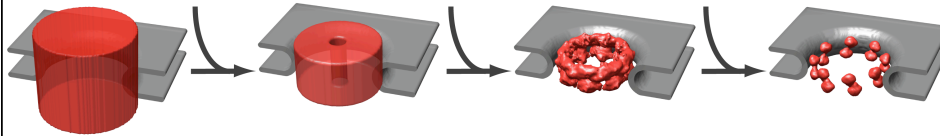
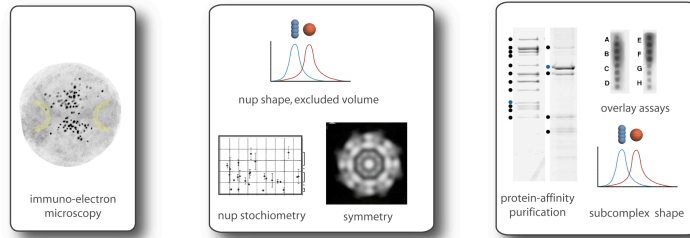
Benefits of data integration



© 2007, The New York Times

9/30/08

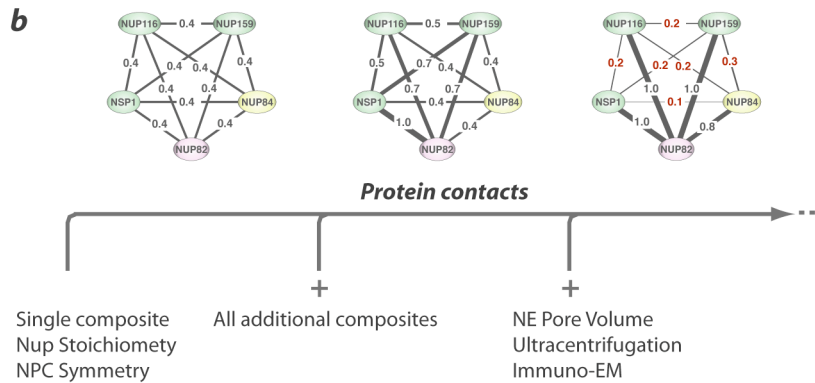
More information results in higher precision of nup localization



Position of 16 copies of Nup192

9/30/08

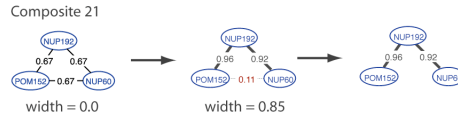
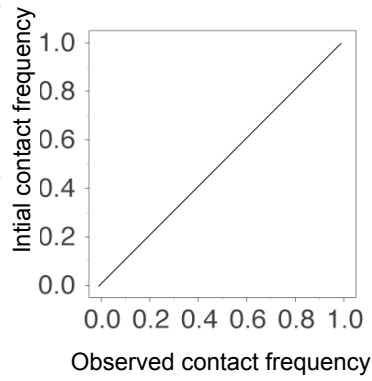
Protein contact prediction



9/30/08

Protein contacts

A



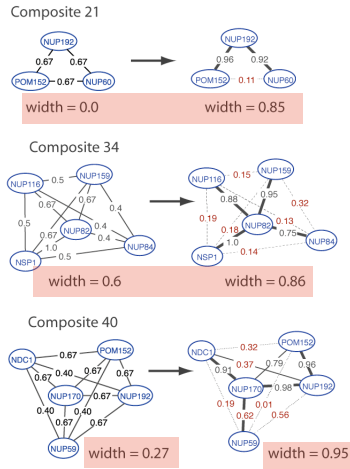
width: the difference between largest and smallest frequency value in a pulldown

$$P_{a \rightarrow b} = \sum_{k=a+1}^N \binom{N}{k} r^k (1-r)^{N-k}$$

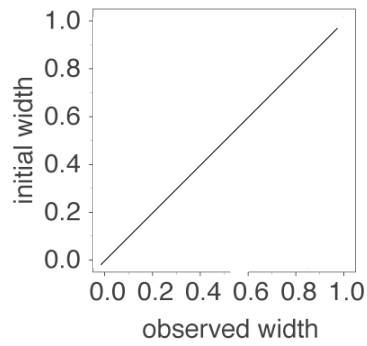
r = reference frequency from protein contact a-b
 N = total number of models
 k = number of models with specific protein contact a-b

9/30/08

Protein contacts



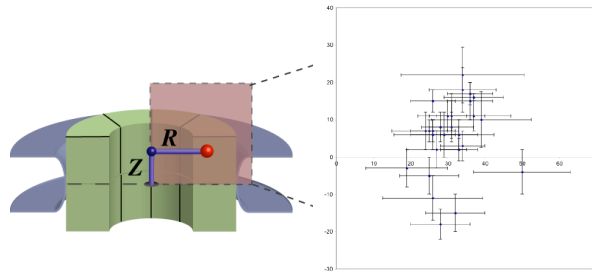
B



Composites contact frequencies are split into low and high frequency values

9/30/08

2 dimensional localization



9/30/08

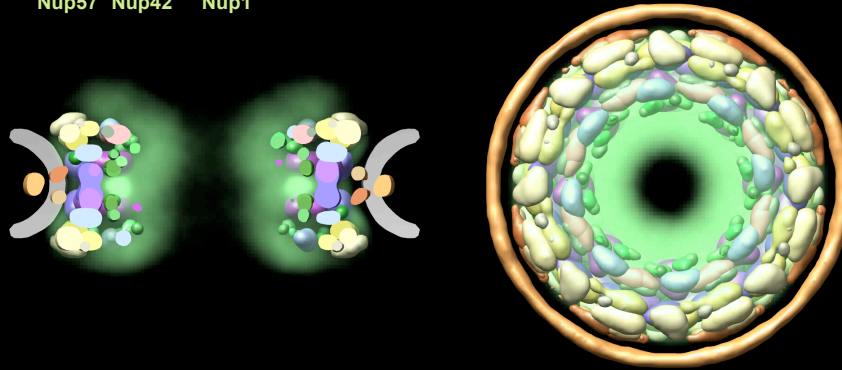
Cargo transport through the NPC

9/30/08

NPC function: selective permeability

Unstructured regions of FG nups contain "FG" repeats:
 FG / FxF / FxFG / GLFG / SAFG / PSFG / SAFGxPSFG / FN

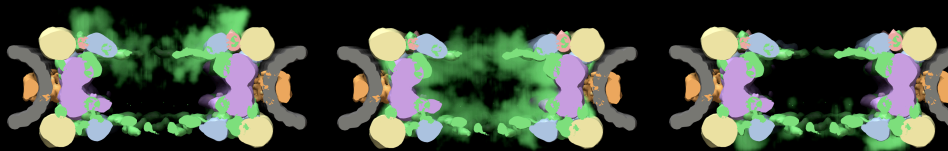
Nsp1	Nup159	Nup60
Nup53	Nup116	Nup145N
Nup59	Nup100	Nup49
Nup57	Nup42	Nup1



FG proteins form a selective barrier

Unstructured regions of FG repeat proteins

Cytosolic side



Gene knock-outs

only on cytosolic side

Nup159
 Nup116
 Nup100
 Nup42

Nuclear side

on cytosolic and nuclear side

Nsp1
 Nup53
 Nup59
 Nup57
 Nup49

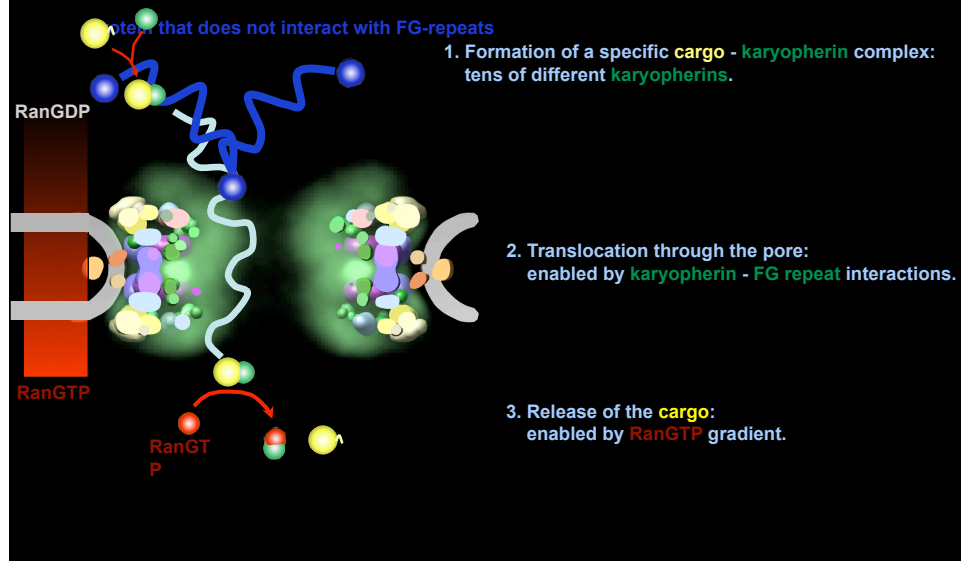
Gene knock-outs

only on nuclear side

Nup1
 Nup53
 Nup145N

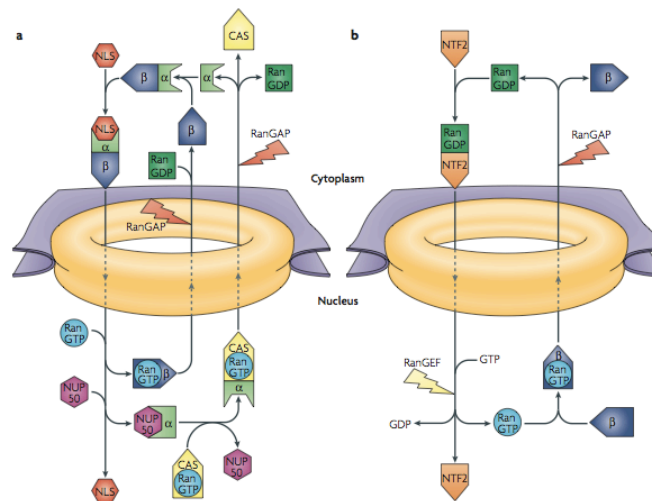
Strawn et al., Nat. Cell. Biol., 2004

Virtual Gating: Simplified import pathway



Nucleo-cytoplasmic transport of macromolecules

Spatial-temporal choreography of protein interactions and their dissociations



Stewart, Nat. Rev. Mol. Cell. Biol. 2007

9/30/08

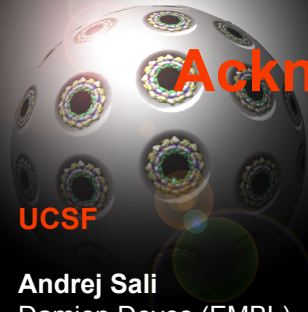
Concluding Remarks

9/30/08

Summary

- Using an integrative approach we determined an architectural map of the NPC using a diverse set of proteomic and biophysical data.
- There are protein configurations that satisfy all experimentally determined restraints.
- These structures are supported by similarity to each other, toy model feasibility, and “external” data.
- The structural description inspired hopefully testable hypotheses about NPC evolution and transport.
- The low-resolution structure provides a starting point for a higher resolution characterization of the assembly (eg, EM, SAXS, x-ray, cross-linking).

9/30/08



Acknowledgments

UCSF

Andrej Sali
Damien Devos (EMBL)

Rockefeller U

Michael Rout
Sveta Dokudovskaya
Liesbeth Veenhoff
Tari Suprpto
Julia Kipper
Rosemary Williams
Orit Karni-Schmidt

NIH
NSF
The Sandler Family Foundation
Human Frontiers Science Program
SUN
IBM
Intel
Structural Genomix

Brian Chait
Wenzhu Zhang