Integrative approaches and complementary techniques for complex analysis

Arnaud.Poterszman@igbmc.fr

The challenge of myriads of complexes

Protein act as complexes. From analysis of ~ 6200 yeast proteins

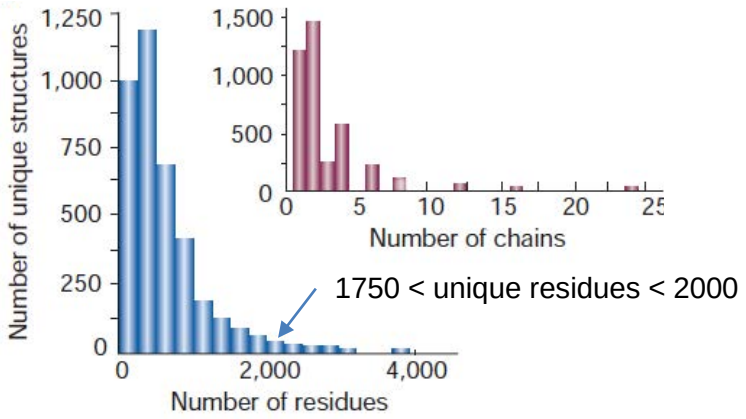
30 000 binary interactions (by focused small scale experiments)

Affinity purification of 1732 proteins → 232 complexes composed of
7.5 proteins per complex

9 partners per protein and 3.6 partners per domain (not all direct or at the same time)

Distribution of protein complexes in the PDB (intervals of 250 residues)

Under-representation of structures from large complexes in view of the estimated average of 7.5 protein per complex



Many macromolecules are recalcitrant to main structural biology methods

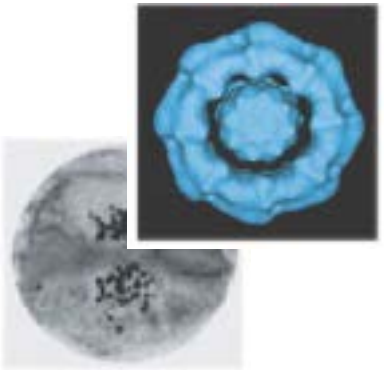
Structural data on complex systems is often limited to isolated subunits and their domains or to low resolution envelopes by SAXS or EM.

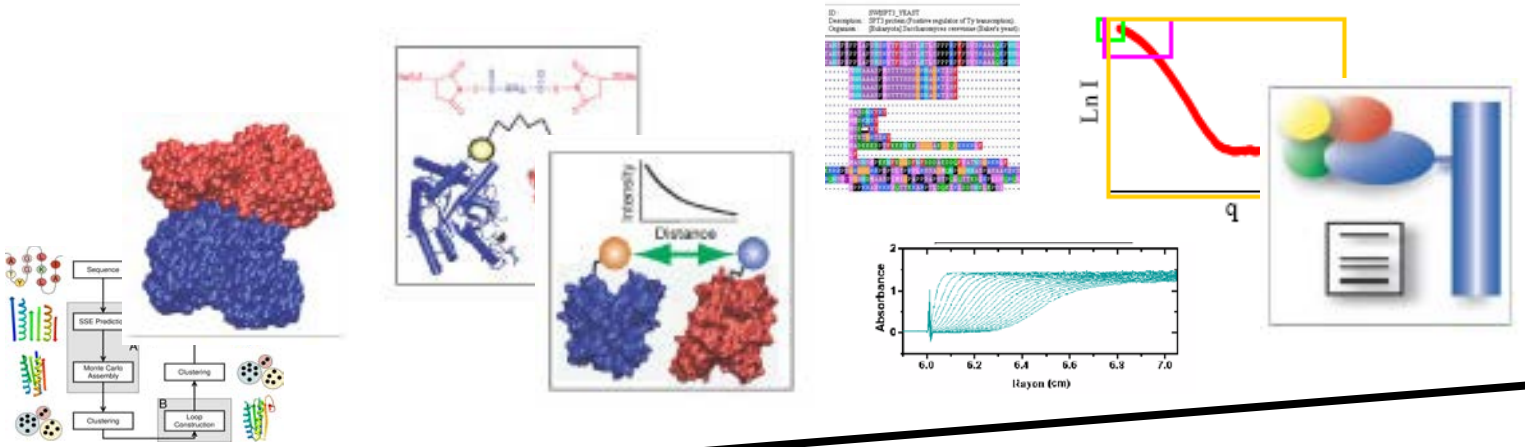
Difficulty to express/reconstitute (incomplete bochemical characterization) and poorly abundant

Conformational heterogenity (prevents cristallisation and high resolution cryo-EM or do not stay intact during analysis (dissocation and/or aggregation on the EM grid)

Integrative determination of macromolecular structures

Aim: combine heterogenous data (information of any source) and propose hybrid models to provide the best possible description of the system (a set of models consitent with available data).

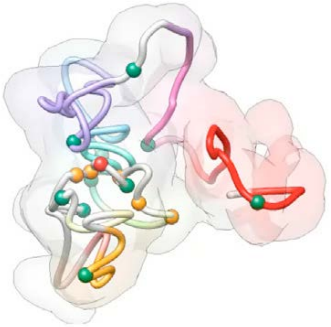




Atomic positions | Residue positions | Member orientations | Member positions

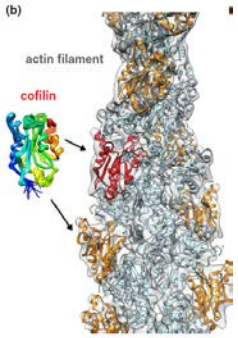
Integrative determination of macromolecular structures

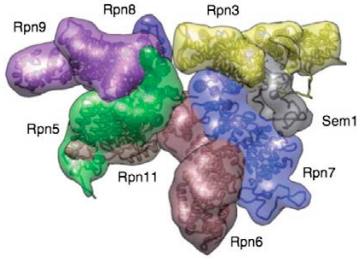
Aim: combine heterogenous data (information of any source) and propose hybrid models to provide the best possible description of the system (a set of models consitent with available data).

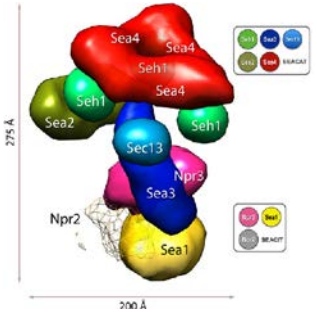


Atomic, multi-scale and topologic models

Genomic region from 3C data







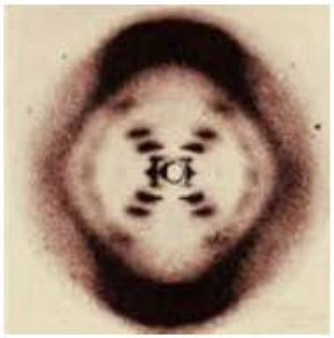
Atomic positions | Residue positions | Member orientations | Member positions

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

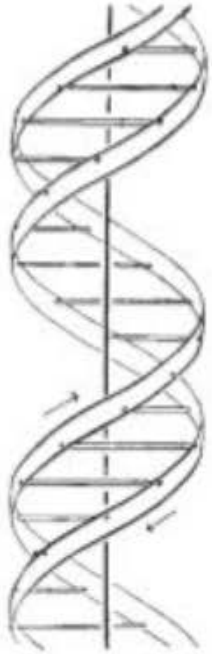
A Structure for Deoxyribose Nucleic Acid

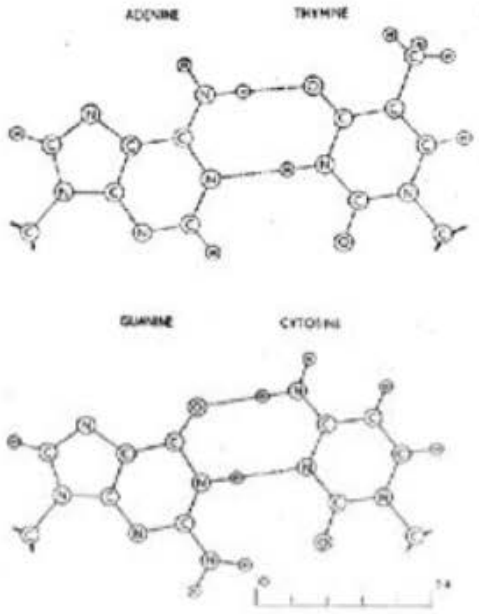
J. D. WATSON
F. H. C. CRICK

No. 4356 April 25, 1953
NATURE



X-ray diffraction





Composition
Stoichiometry
Chemical complementarity

To understand and modulate cellular processes, we need their models.
These are best generated by considering all available information.

Integrative approaches

1/ Principles of integrative structure determination

2/ The example of the Nuclear Pore Complex

3/ A few experimental methods for complex analysis

Integrating modelling platforms

Integrative modeling platform (IMP)
Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, et al. Putting the pieces together:integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol 2012;10, e1001244.

Inferential Structure Determination (ISD) framework
Rieping W, Nilges M, Habeck M. ISD: a software package for Bayesian NMR structure calculation. Bioinformatics 2008;24:1104–5.

HADDOCK
van Zundert GC1, Rodrigues JP1, Trellet M2, Schmitz C3, Kastritis PL4, Karaca E4, Melquiond AS5, van Dijk M6, de Vries SJ7, Bonvin AM1. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016 Feb 22;428(4):720-5.

RNABuilder
Flores SC, Sherman MA, Bruns CM, Eastman P, Altman RB. Fast flexible modeling of RNA structure using internal coordinates. IEEE/ACM Trans Comput Biol Bioinform 2011;8:1247–57.

Integrative Modeling Platform (IMP)

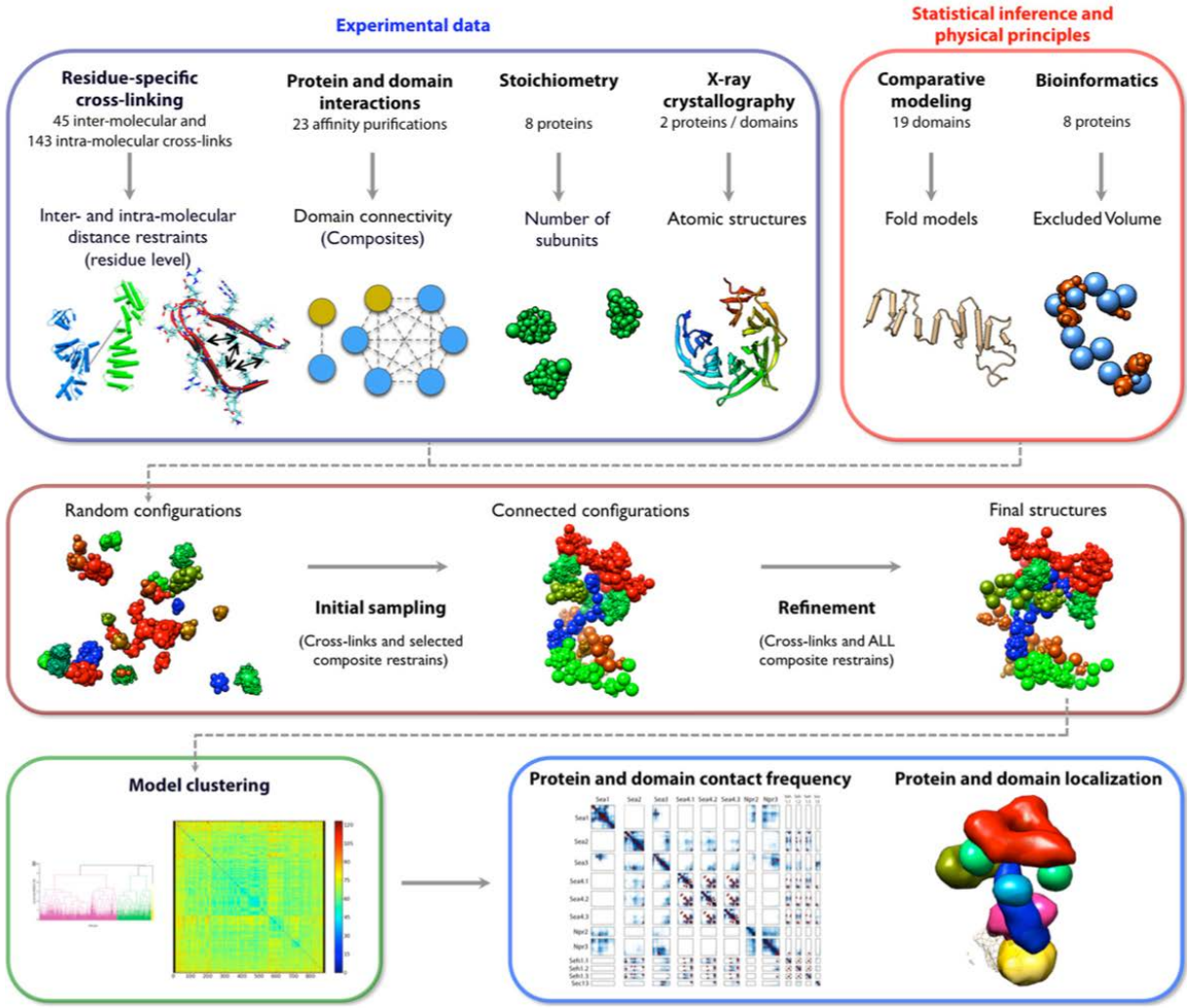
http://integrativemodeling.org

1/ Gathering data

2/ Representing and translating data into restraints

3/ Sampling good scoring configurations

4/ Analysis and assesment



A Sali

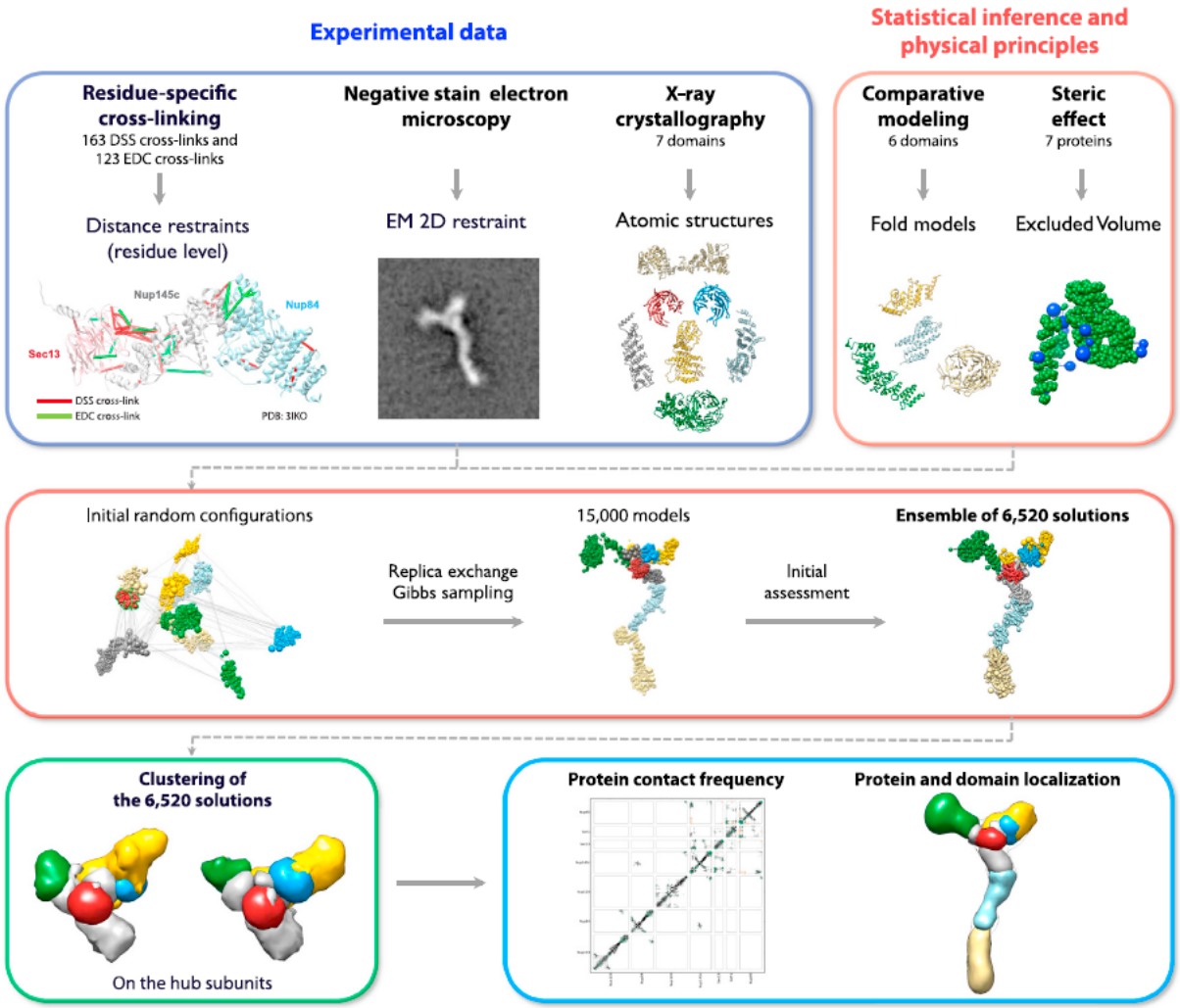
The four stages of integrated modelling

1/ Gathering data

2/ Representing and translating data into restraints

3/ Sampling good scoring configurations

4/ Analysis and assesment



Integrative structure modeling of RNA Polymerase II stalk

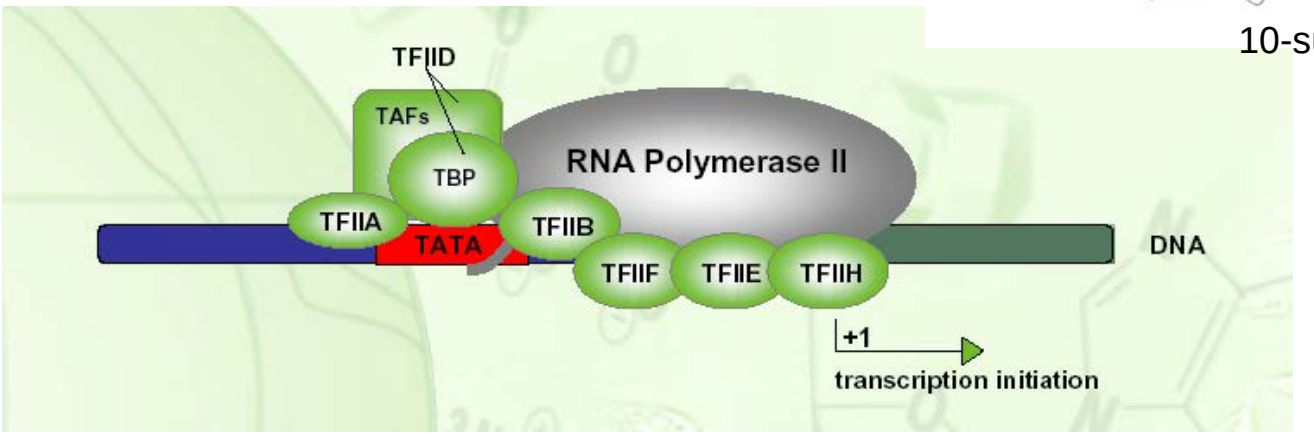
RNA Pol II is a eukaryotic complex that catalyzes DNA transcription to synthesize mRNA strands Eukaryotic RNA polymerase II contains 12 subunits, Rpb1 to Rpb12

The yeast RNA Pol II dissociates into a 10-subunit core and a Rpb4/Rpb7 heterodimer

Rpb4 and Rpb7 are conserved from yeast to humans, and form a stalk-like protrusion extending from the main body of the RNA Pol II complex

Rpb4/Rpb7

10-subunits core



Integrative structure modeling

1/ Gathering data

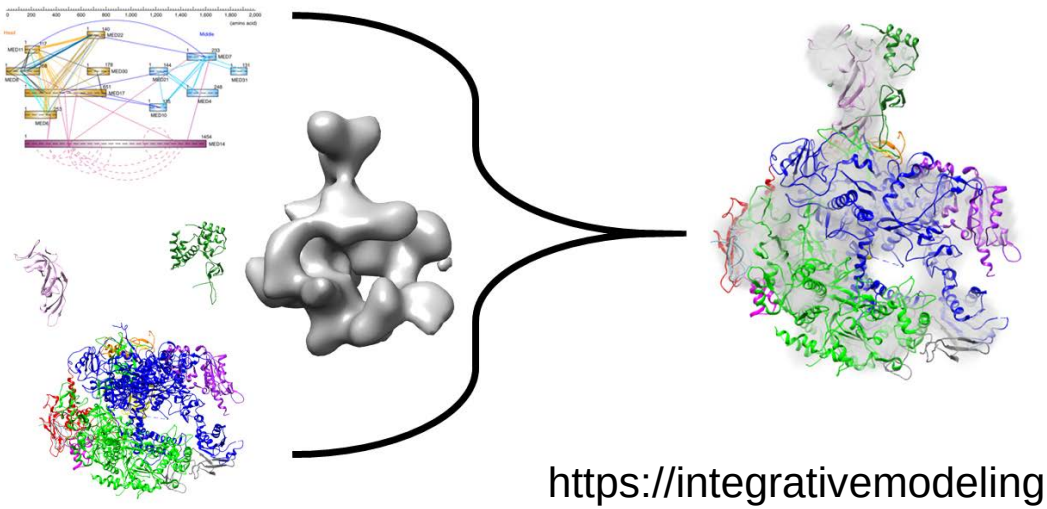
2/ Representing and translating data into restraints

3/ Sampling good scoring configurations

4/ Analysis and assesment

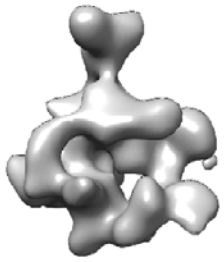
Determine the localization of two subunits of the yeast RNA Polymerase II, Rpb4 and Rpb7 (stalk), hypothesizing that we already know the structure of the remaining 10-subunit complex based on:

- chemical cross-linking coupled with mass spectrometry (CX-MS),
- negative-stain electron microscopy (EM),
- X-ray crystallography data



https://integrativemodeling.org/

Integrative structure modeling of RNA Polymerase II stalk



Gathering data

Represent and translate data into restraints

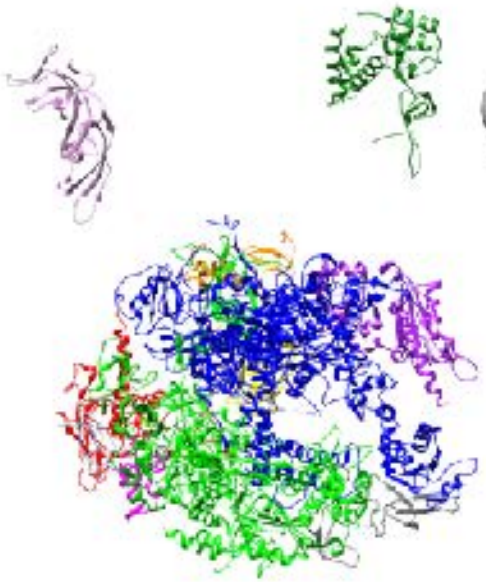
Sampling good scoring configurations

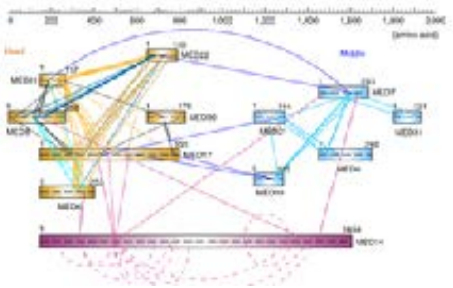
Analysis and assesment

Experimental map of entire complex at 20.9Å resolution

RX-ray structures of the 10-subunit core of RNA Pol II and of parts of Rbp4 and Rbp7

Chemical cross-linking coupled with mass spectrometry (CX-MS)





https://integrativemodeling.org/

Integrative structure modeling

Representation:

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

Experimental map of entire complex at 20.9Å resolution (represented with Gaussan Mixture Models (GMMs))

Macromolecules are represented using high and low resolution spherical beads and 3D gaussians (1 aa/bead and 20 aa/bead). Multi-scale representation

Gathering data

↓

Represent and translate data into restraints

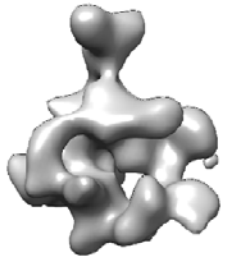
↓

Sampling good scoring configurations

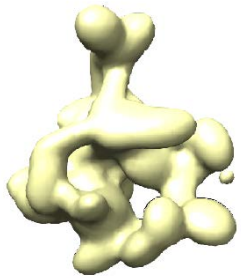
↓

Analysis and assesment

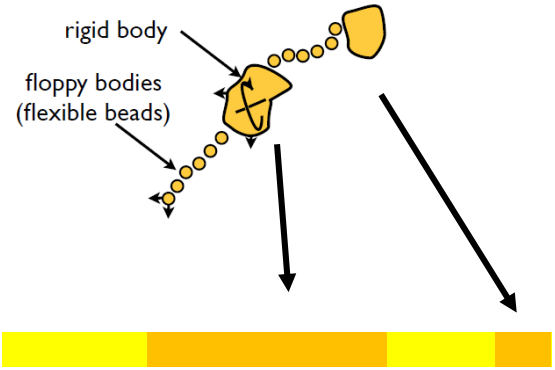
- Missing (unresolved) parts are modelled by low resolution beads
- Resolved regions as rigid bodies, allow unresolved regions to move (floppy bodies)



Experimental map



50 Gaussian mixture model



https://integrativemodeling.org/

Integrative structure modeling

Gathering data

↓

Represent and translate data into restraints

↓

Sampling good scoring configurations

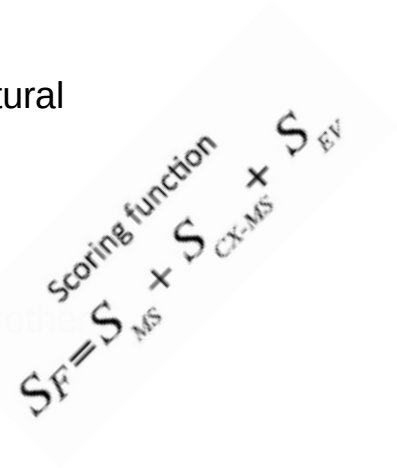
↓

Analysis and assesment

Define a scoring function, by which the individual structural models will be scored based on the input data

A simple sum of individual restraints

Each restraint maps to one of our input experiments or other physical/statistical information



Sequence connectivity restraint: residues that are adjacent in sequence will also be close in space due to the peptide bond

Excluded volume restraint: one protein cannot occupy the same space as another

EM restraints: A density overlap function to compare the GMM approximation of our model (em_components) with that of the EM map (target_gmm_file)

No electrostatics or stereochemistry; very different to a typical molecular mechanics simulation

Integrative structure modeling

Gathering data

↓

Represent and translate data into restraints

↓

Sampling good scoring configurations

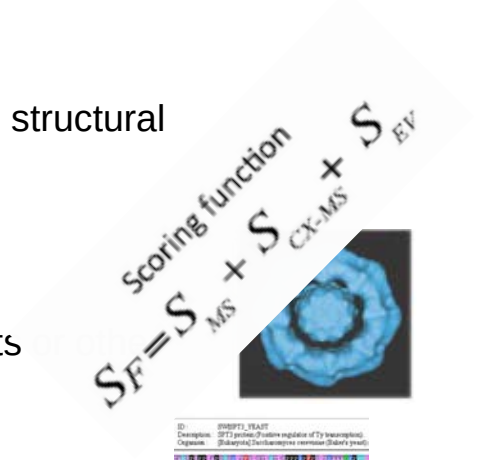
↓

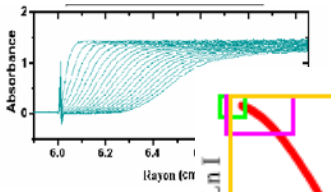
Analysis and assesment

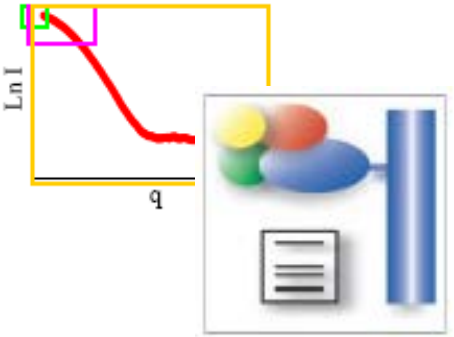
Define a **scoring function**, by which the individual structural models will be scored based on the input data

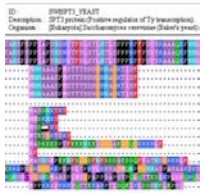
A simple sum of individual restraints

Each restraint maps to one of our input experiments physical/statistical information



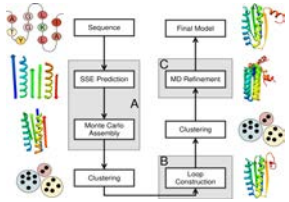


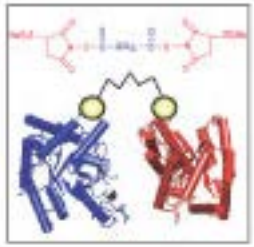


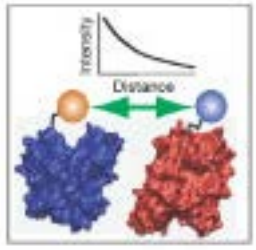


Scoring:

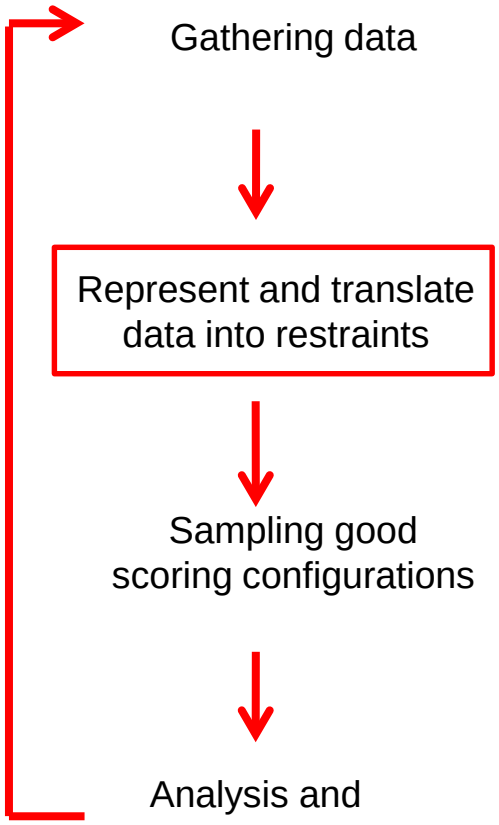
- Density maps
- EM images
- Proteomics
- FRET
- Chemical and Cys cross-linking
- Homology-derived restraints
- SAXS
- H/D exchange
- Second harmonic generation
- Native mass spectrometry
- Genetic interactions
- Statistical potentials
- Molecular mechanics forcefields
- Bayesian scoring
- Library of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







Integrative structure modeling



Define a scoring function, by which the individual structural models will be scored based on the input data

A simple sum of individual restraints

Each restraint maps to one of our input experiments or other physical/statistical information

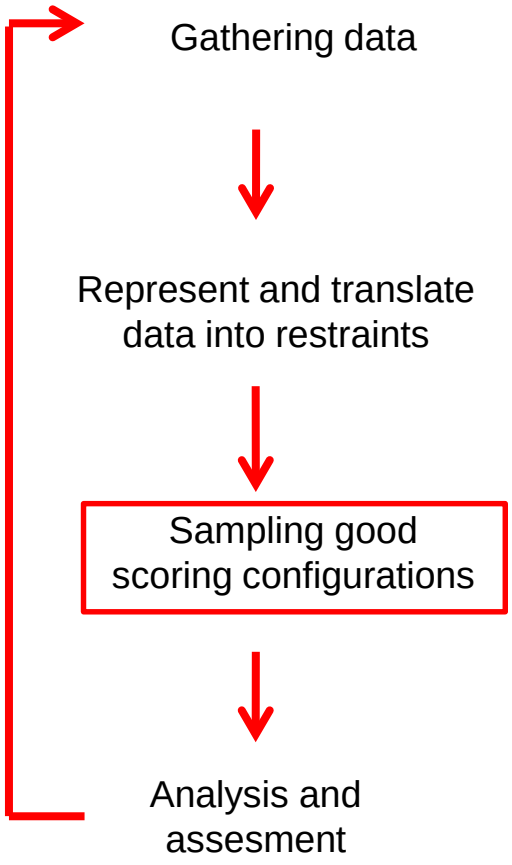
Sequence connectivity restraint: residues that are adjacent in sequence will also be close in space due to the peptide bond

Excluded volume restraint: one protein cannot occupy the same space as another

EM restraints: A density overlap function to compare the GMM approximation of our model (em_components) with that of the EM map (target_gmm_file)

Cross-linking restrains: protein and residue numbers for each of the two linked residues (cross linker length,

Integrative structure modeling

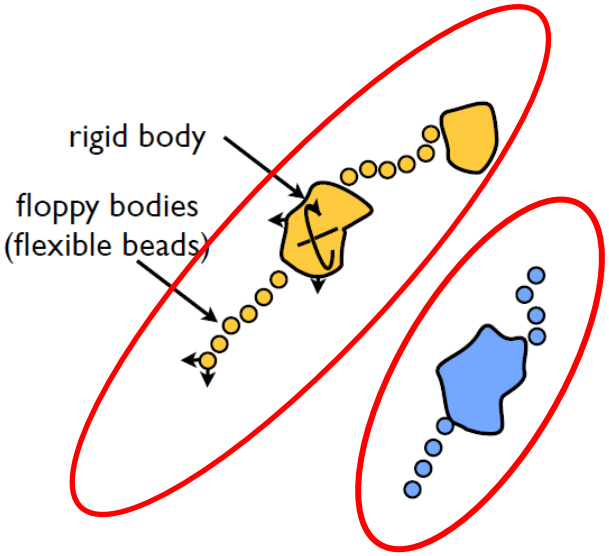


Here Monte Carlo is used to sample (not minimize) system (generate many models that satisfy the data)

Need to define a set of movers: rigid_bodies defines the components that will be moved as rigid bodies (in this case, the parts of Rpb4 and Rpb7 for which we have atomic structure). Unstructured regions will move as flexible beads.

Sampling:

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



Srb (super rigid body)

https://integrativemodeling.org/

Integrative structure modeling

Gathering data

↓

Represent and translate data into restraints

↓

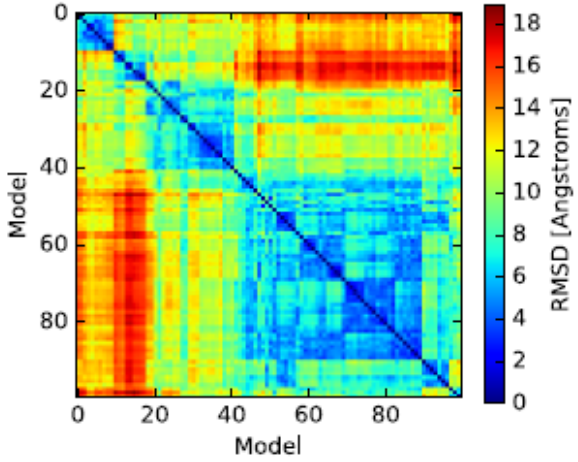
Sampling good scoring configurations

↓

Analysis and assesment

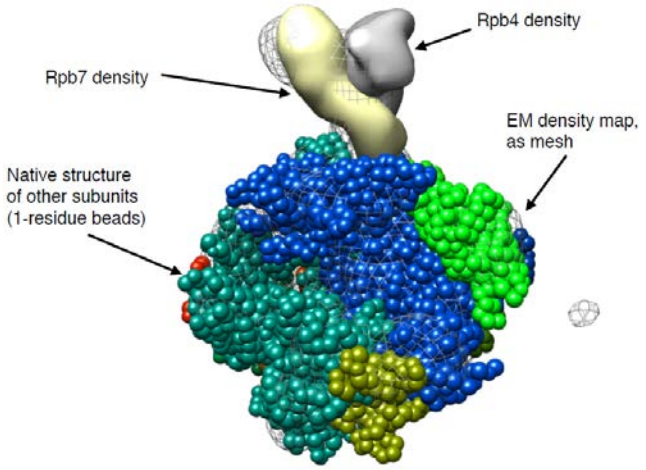
Cluster (group by similarity) the sampled models to determine high-probability configurations.

- Chose a reference and align (superpose) all structures
- Calculate distances between structures (RMSD)
- Calculate localization densities for selected subunits



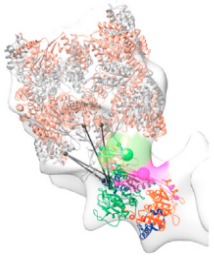
Analysis:

Clustering
Chimera
Pymol
PDB files
Density maps

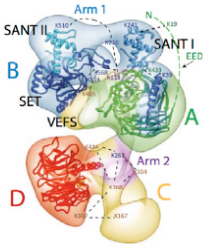


https://integrativemodeling.org/

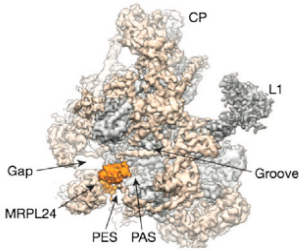
Examples of Recently Determined Integrative Structures



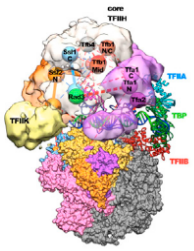
INO80



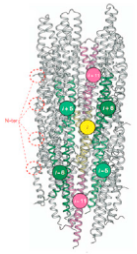
PRC2



39S ribosome subunit

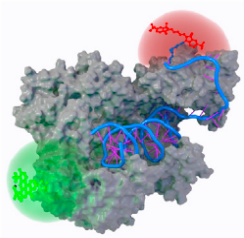


RNA pol II

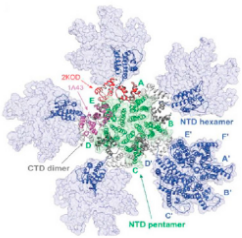


type III secretio system needle

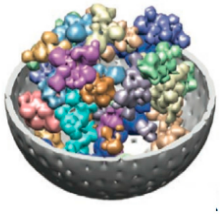
desulfurase-Isu-frataxin



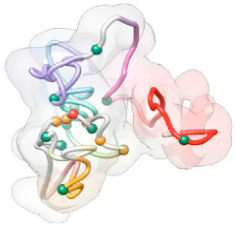
HIV-1 reverse transcriptase:DNA



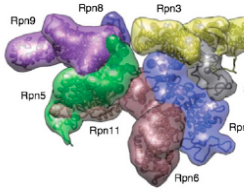
HIV capsid



genome architecture

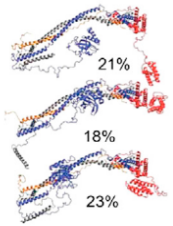


α-globin gene domain

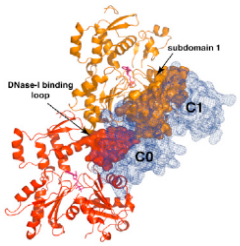


proteosomal lid

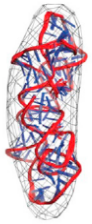
SAGA transcriptional coactivator



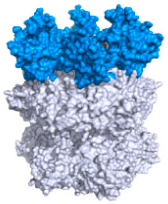
ESCRT-I complex



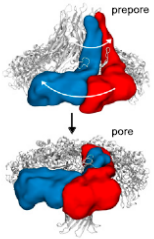
actin – myosin binding protein C



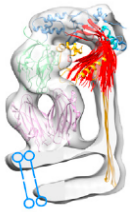
$[\Psi^{CD}]_2$



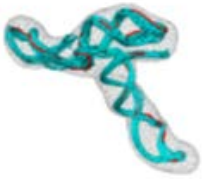
KaiB-KaiC complex



pore-forming toxin aerolysin



pleurotolysin



RNA ribosome-binding element

https://integrativemodeling.org/

Topological models from MS-based hybrid analysis

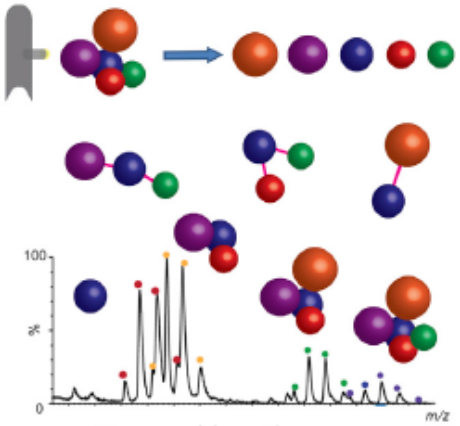
Politis et al. 2015 (Robinson and Sali's labs)

Gathering information

(ia) Protein purification
- Isolate protein complex

(ib) Proteomics
- Identify protein subunits

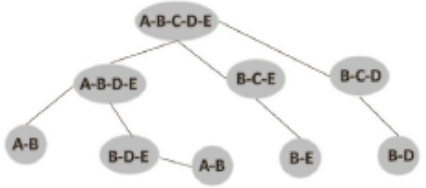
(ic)a: MS of intact complex
(id)b: MS/MS and/or disruption in solution
(ie)c: Chemical cross-linking MS
- List of subcomplexes
- Stoichiometry
- Pairwise interactions



Restraints

(ii) Representation & scoring function

Dissasembly pathway



Sampling

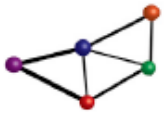
(iii) Sample models that fit the data
- Scoring function-guided sampling

Scoring function

$$S_F = S_{MS} + S_{CX\text{-}MS} + S_{EV}$$

Analysis

(iv) Analysis & interpretation
- Protein contacts & configurations



Topological Models of Heteromeric Protein Assemblies from Mass Spectrometry: Application to the Yeast eIF3:eIF5 Complex

Argyris Politis,[1,2,5] Carla Schmidt,[1,5] Elina Tjioe,[3] Alan M. Sandercock,[4,6] Keren Lasker,[3,7] Yuliya Gordiyenko,[1,8] Daniel Russel,[3] Andrej Sali,[3,*] and Carol V. Robinson[1,*]

A mass spectrometry–based hybrid method for structural modeling of protein complexes

Argyris Politis[1,4,5], Florian Stengel[2,5], Zoe Hall[1], Helena Hernández[1], Alexander Leitner[2], Thomas Walzthoeni[2], Carol V Robinson[1] & Ruedi Aebersold[2,3]

Native MS

Cross-link MS

H-DX MS

Quantitative MS

The main limitation: sample preparation

buffer exchange or desalting procedure

Ultra centrifugation

- micro-concentrators :
  - Microcon, Centricon, Amicon (Millipore)
  - Vivaspin (Vivasciences)

- centrifuge
- 5 to 7 cycles at least
- takes time but very efficient procedure !

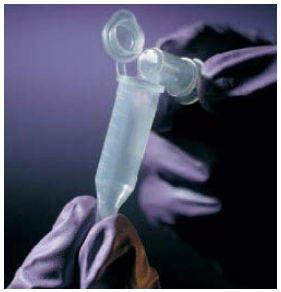
Size exclusion chromatography

- gel filtration colums :
  - NAP-10 et NAP-5 (GE Healthcare)

- Often 2 runs with a concentration step in between
- takes less time but dilutes the sample

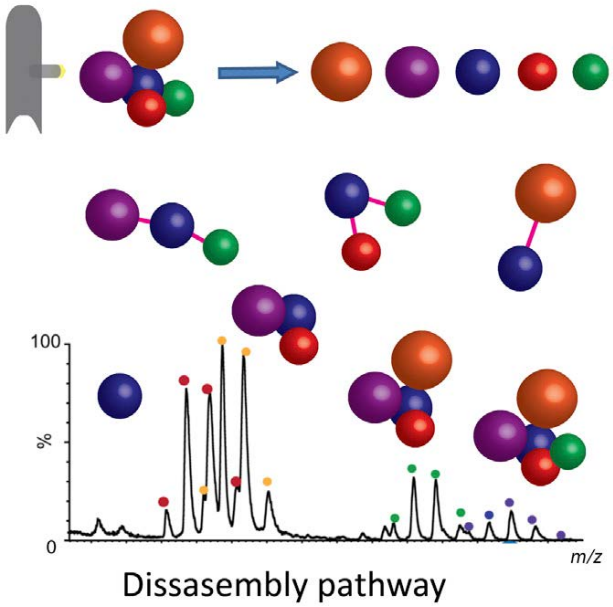
Equilibrium dialysis

- dialysis or mini-dialysis units :
  - Slide-A-Lyzer mini-dialysis (Pierce)



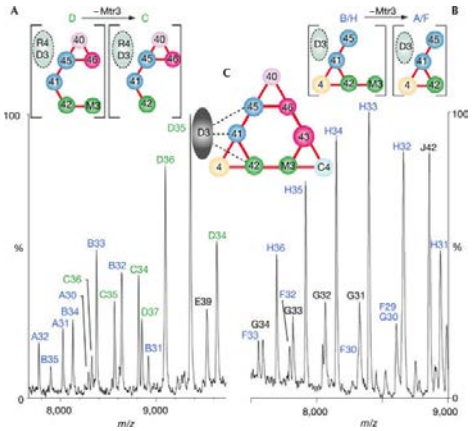
- dilutes the samples
- very easy to perform (overnight)

Insights into disassembly pathways and composition of sub-complexes

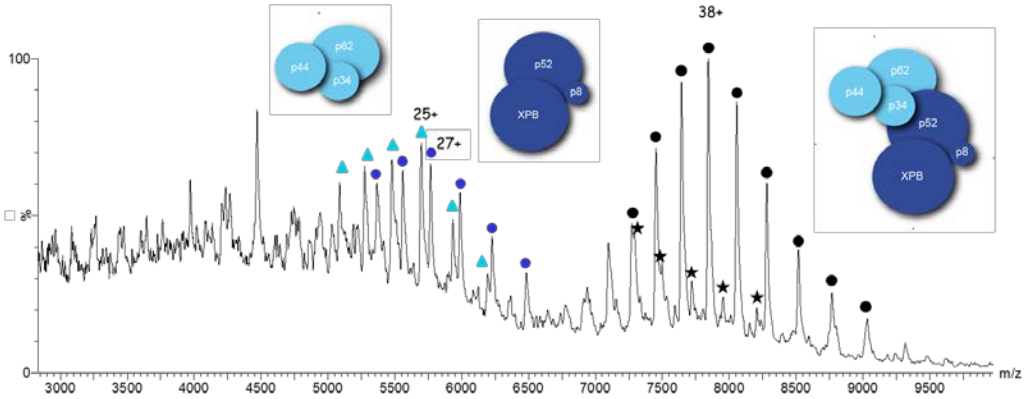


Dissasembly pathway

Subunit composition and stoichiometry
Dissociation pathways, sub-complexes.
-acceleration voltage, pressure-



Hernandez, EMBO Rep 2008



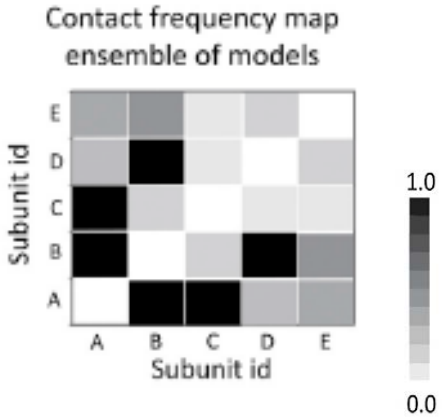
A topological model from MS data

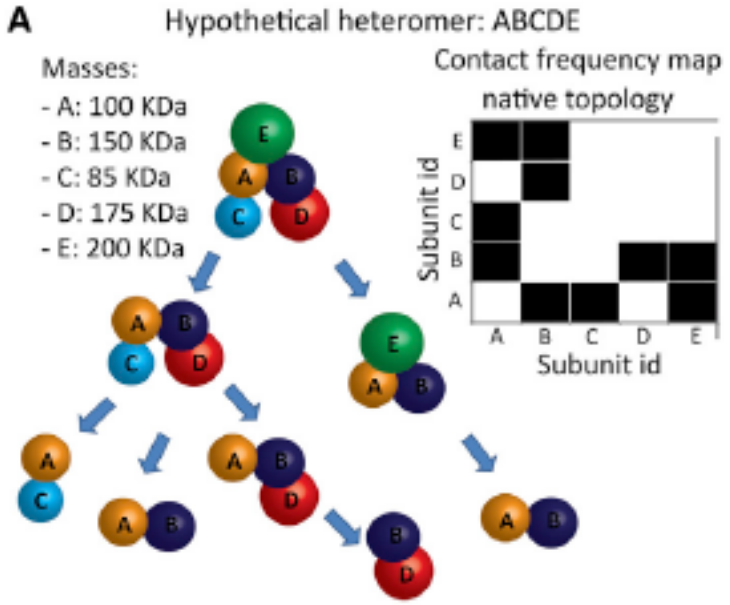
Native-MS data and excluded volume restraints (r = f((MM)) are not sufficient

Additional constrains are needed

Scoring function

$$S_F = S_{MS} + S_{CX\text{-}MS} + S_{EV}$$





Subcomplex	ABCD	ABCDE
Number of subunits	**4**	**5**
Subcomplex connectivity restraint and excluded volume restraint		
Fraction of topologies that match the native topology (%)	8.95	7.52
***Subcomplex connectivity restraint, binary interactions and excluded v**		
Fraction of topologies that match the native topology (%)	35.6	43.51

Politis et al. 2015 (Robinson and Sali's labs)

Integrated representation of the eIF3 yeast initiation translation factor

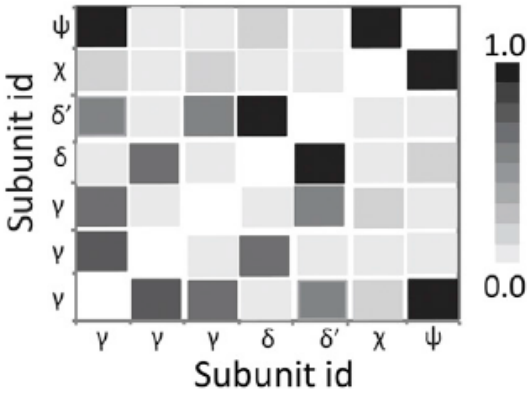
Excluded volume restraints

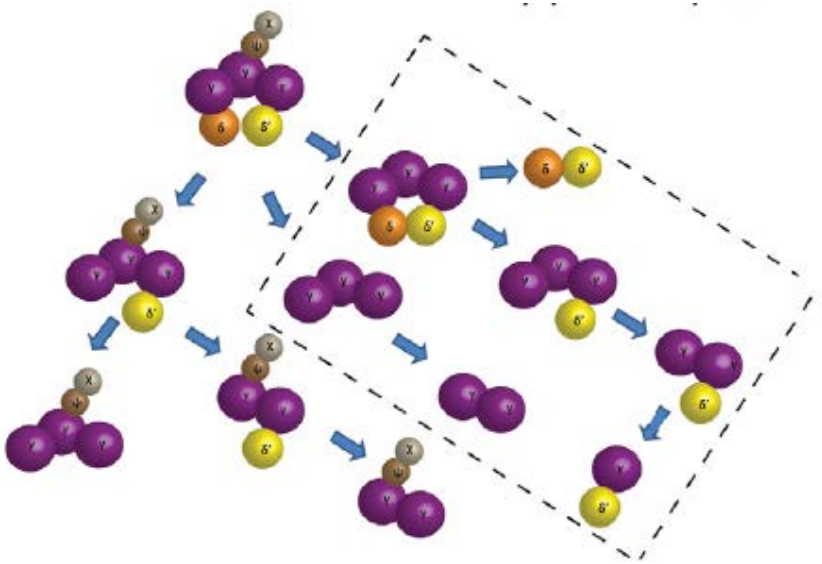
Native-MS data

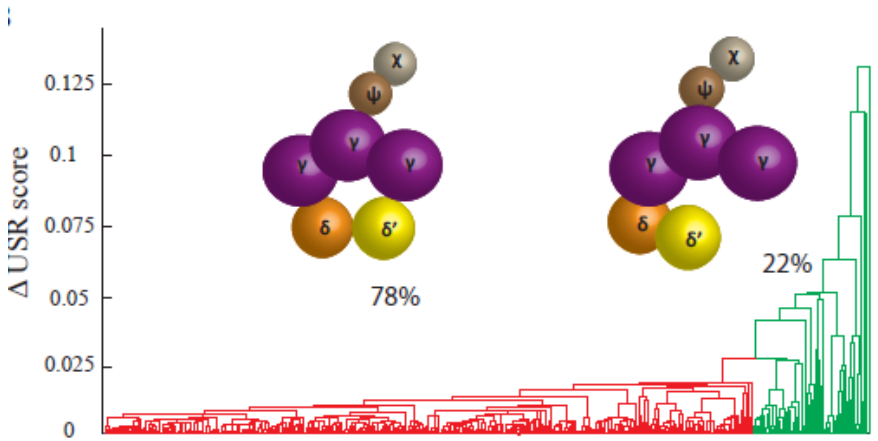
X-link-MS constrains

Scoring function

$$S_F = S_{MS} + S_{CX\text{-}MS} + S_{EV}$$



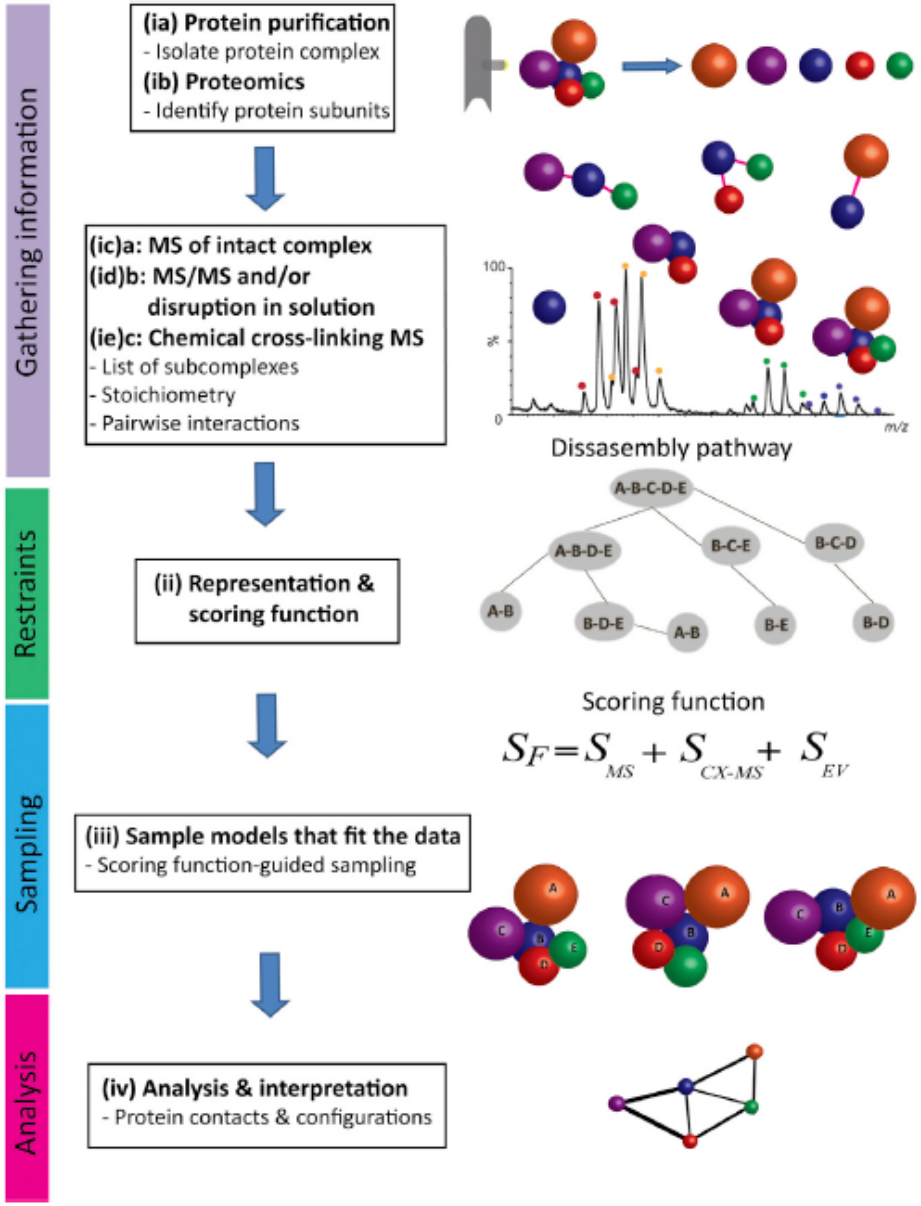




Politis et al. 2015 (Robinson and Sali's labs)

Topological models from MS-based hybrid analysis

Politis et al. 2015 (Robinson and Sali's labs)



Topological Models of Heteromeric Protein Assemblies from Mass Spectrometry: Application to the Yeast eIF3:eIF5 Complex

Argyris Politis,[1,2,5] Carla Schmidt,[1,5] Elina Tjioe,[3] Alan M. Sandercock,[4,6] Keren Lasker,[3,7] Yuliya Gordiyenko,[1,8] Daniel Russel,[3] Andrej Sali,[3,*] and Carol V. Robinson[1,*]

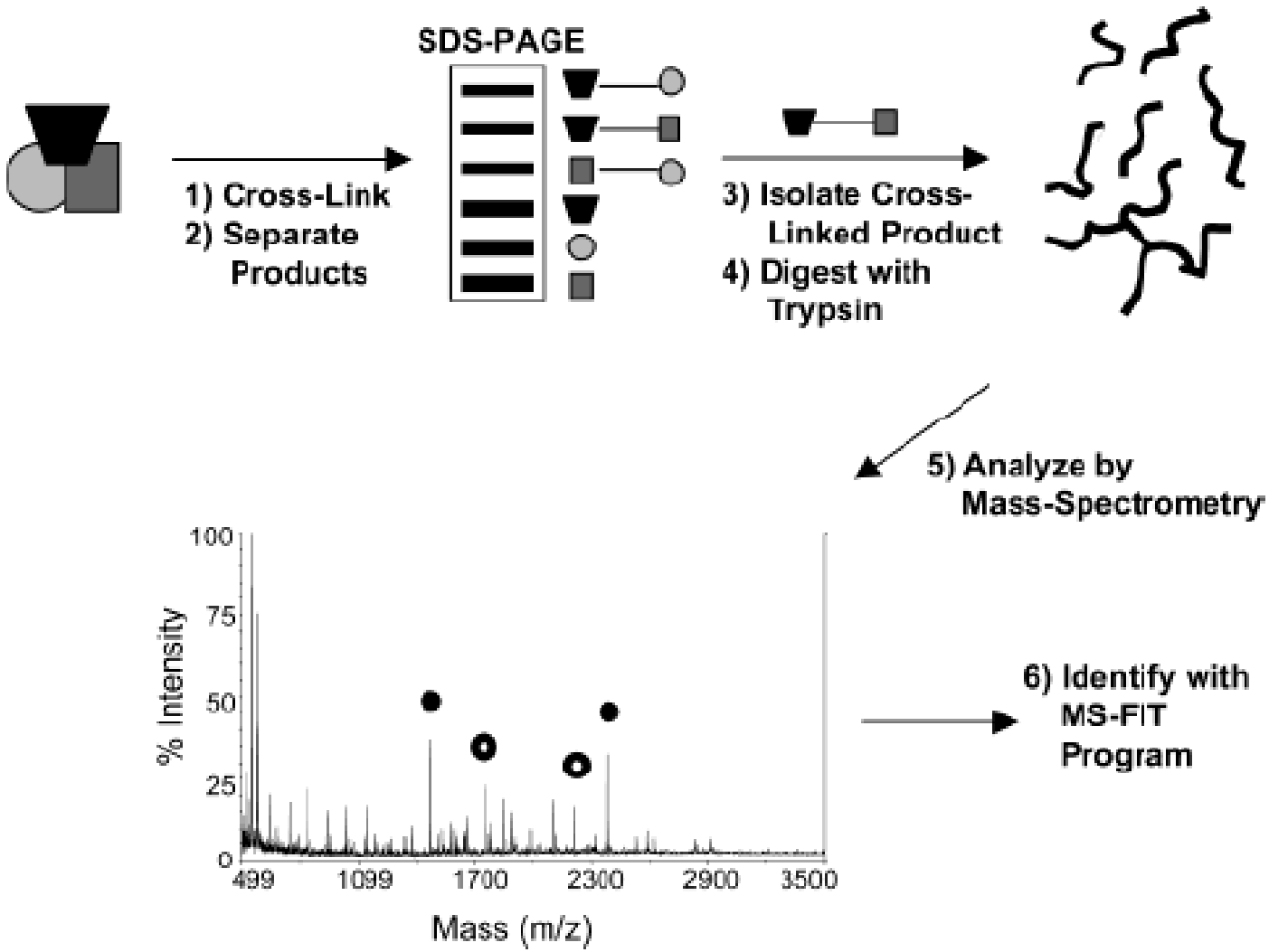
Native MS

Cross-link MS

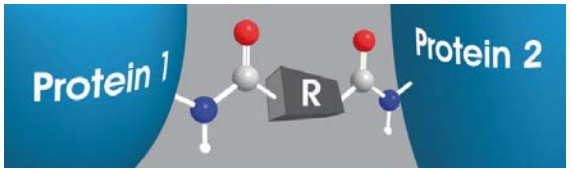
H-DX MS

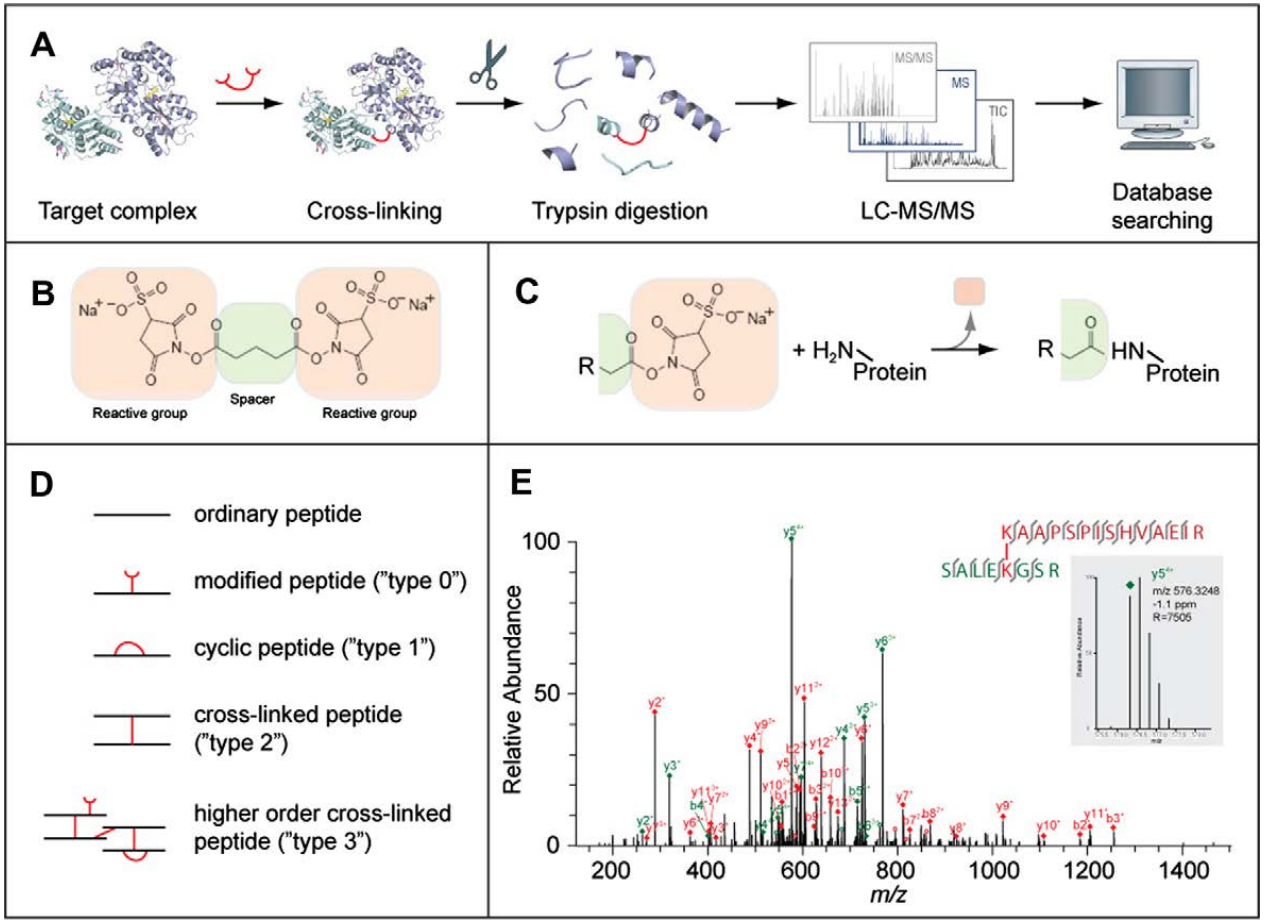
Ion Mobility MS

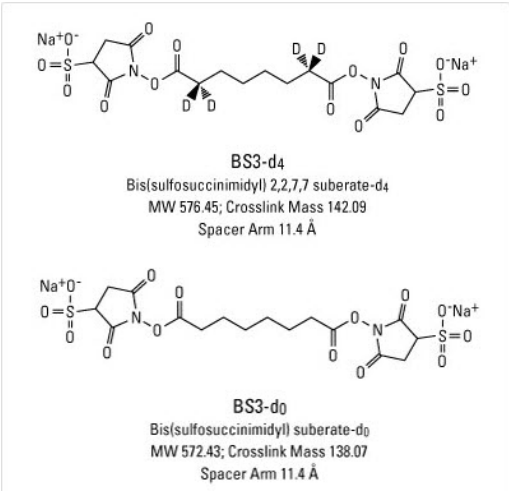
Quantitative MS

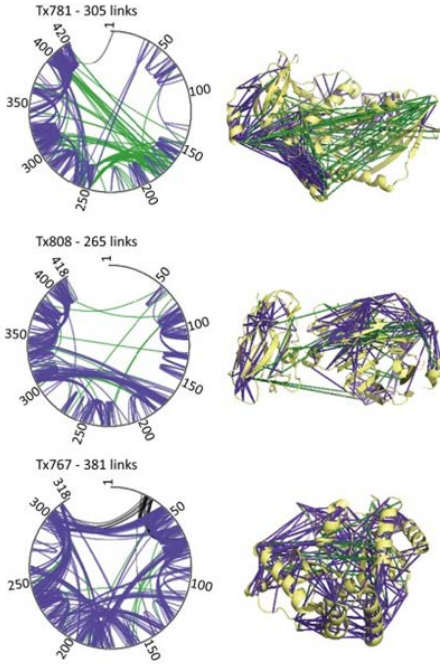

SDS-PAGE
1) Cross-Link
2) Separate
Products
3) Isolate Cross-
Linked Product
4) Digest with
Trypsin
5) Analyze by
Mass-Spectrometry
6) Identify with
MS-FIT
Program
% Intensity
100
75
50
25
0
499
1099
1700
2300
2900
3500
Mass (m/z)

X-link/MS experiments



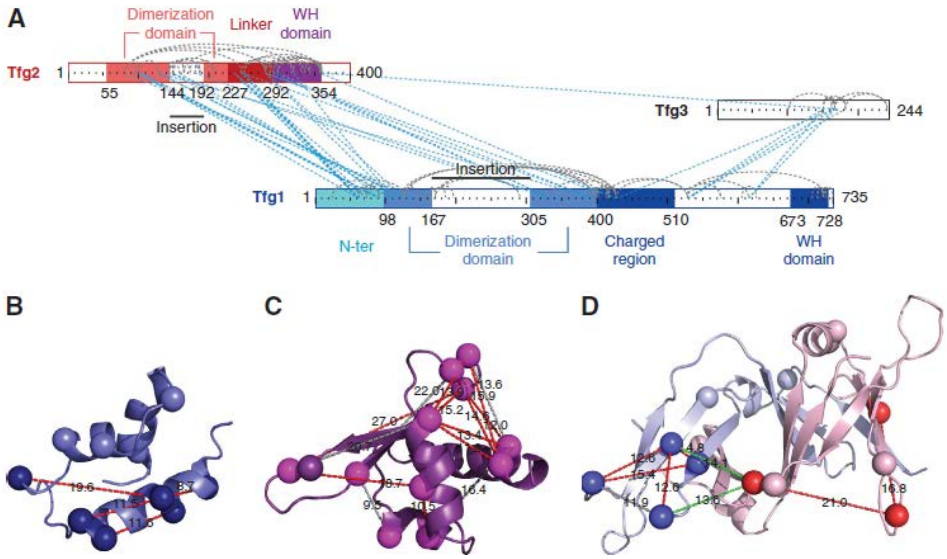




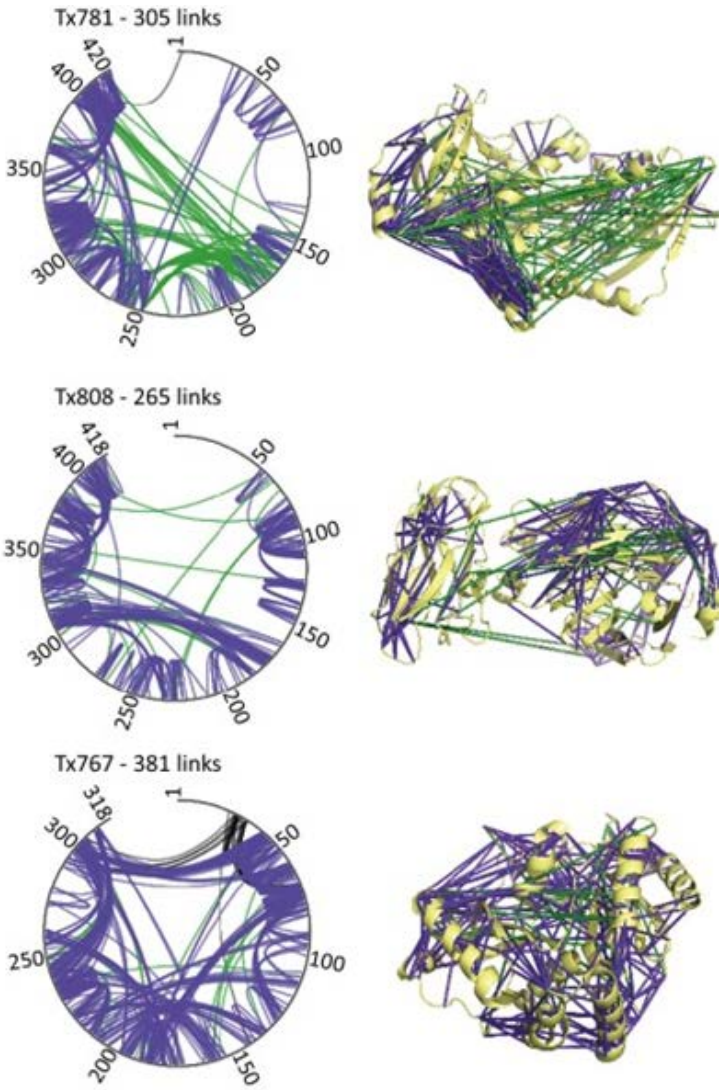


Juri Rappsilber, 2010, 2012

Linear and circular representation

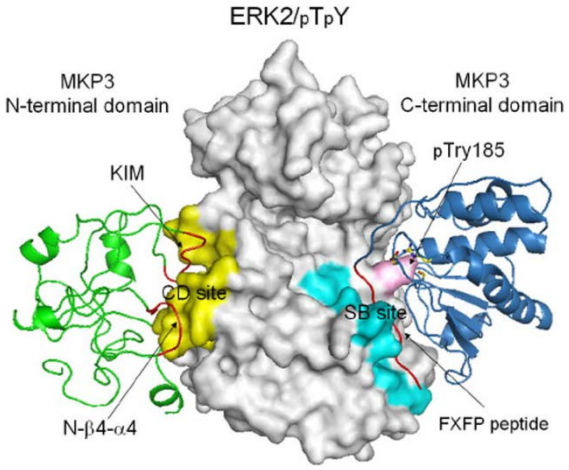


Chen, EMBO 2010

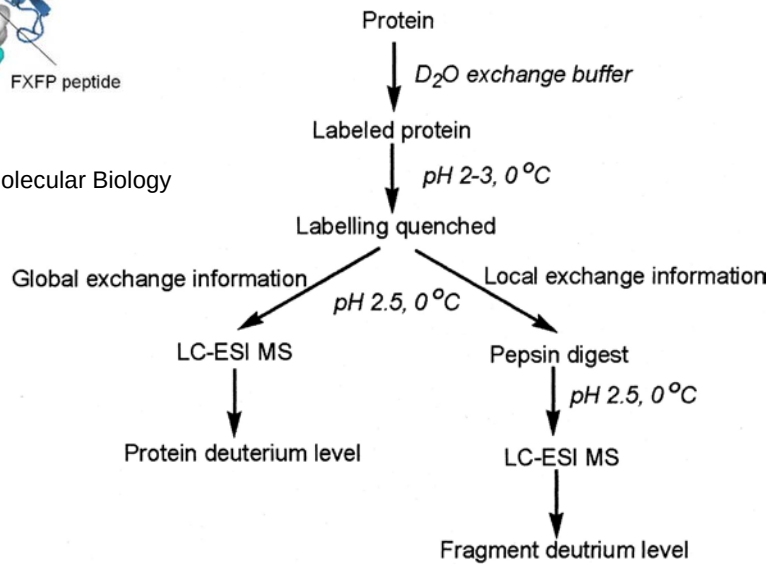


Rappsilber, 2010, 2012

H/D exchange



Protein unfolding, either natural or induced by denaturants
Measurement of folding or unfolding rates
Protein folding, on timescales from milliseconds to days
Binding, binding constants and interacting surfaces
DNA-protein interactions



Peptides that interact with a partner

Integrative approaches

1/ Principles of integrative structure determination

2/ The example of the Nuclear Pore Complex

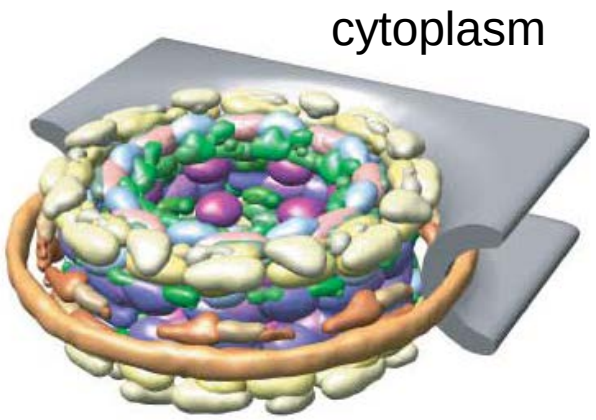
3/ A few experimental methods for complex analysis

A founding example: the nuclear pore complex (NPC)

Yeast NPCs are ~50 Mda structures built of multiple copies of some ~30 different proteins (nucleoporins), totalling at least 456 protein molecules

Each NPC is a plastic structure embedded in the nuclear envelope and is composed of eight morphologically similar ‘spokes’ surrounding a central Tube

Filling this tube and projecting into both the cytoplasmic and nuclear sides are flexible filamentous domains from proteins termed FG (phenylalanine-glycine) repeat nucleoporins; these domains form the docking sites for transport factors that carry macromolecular cargoes through the NPC



Albert et al. 2007, 2008

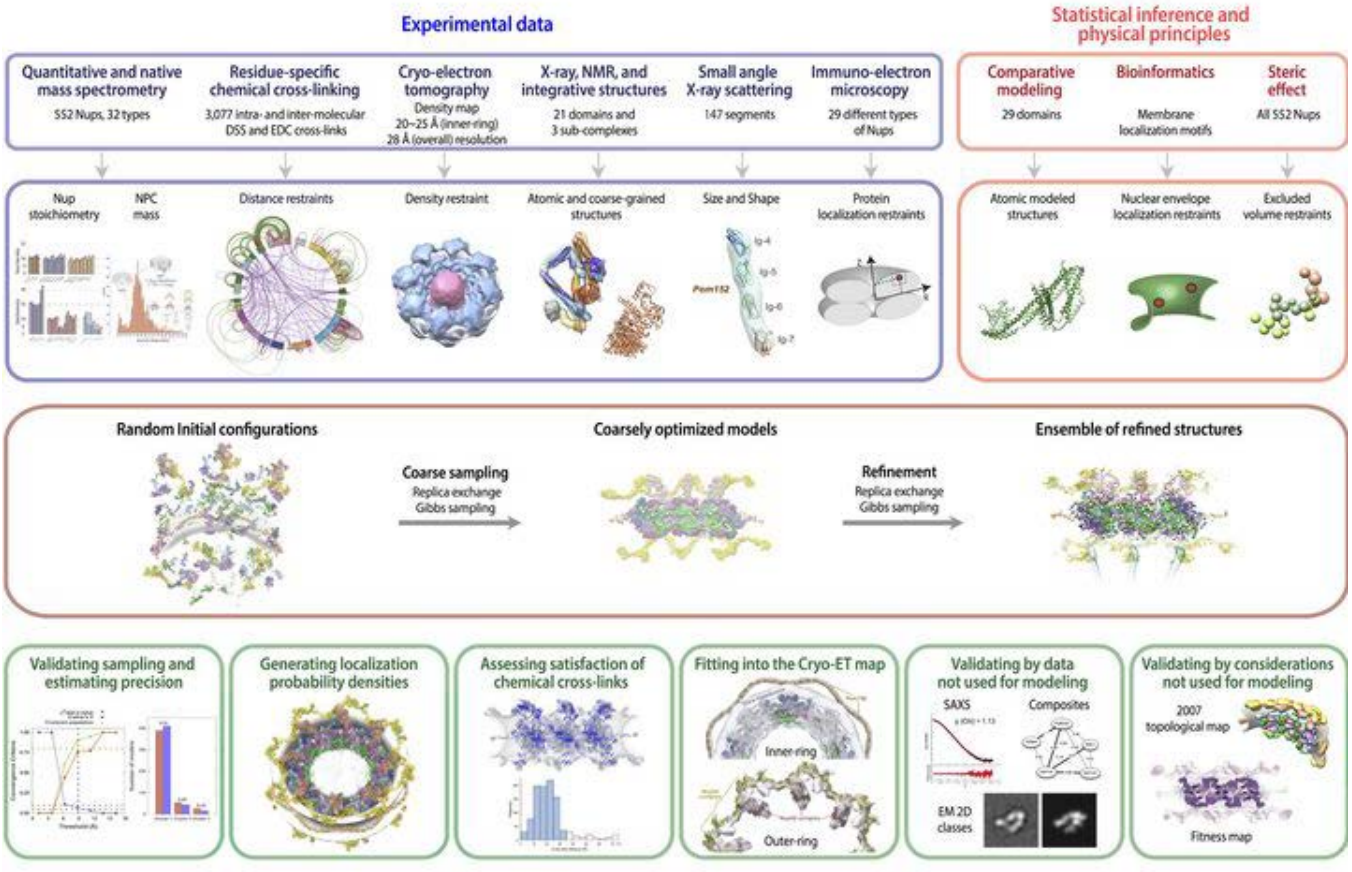
Integrating spacial restrains from proteomic data

1/ Gathering data

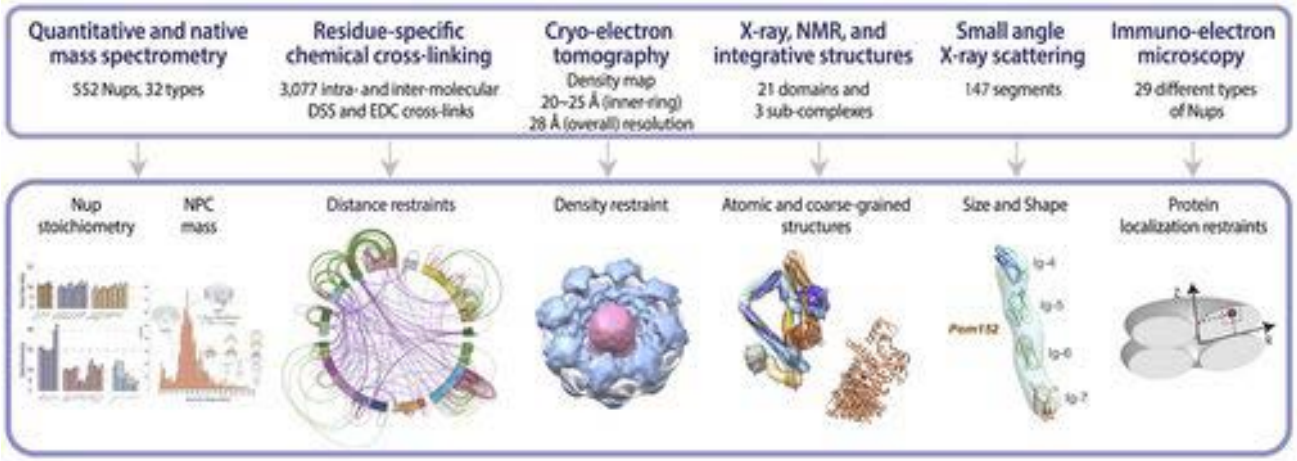
2/ Representing and translating data into restraints

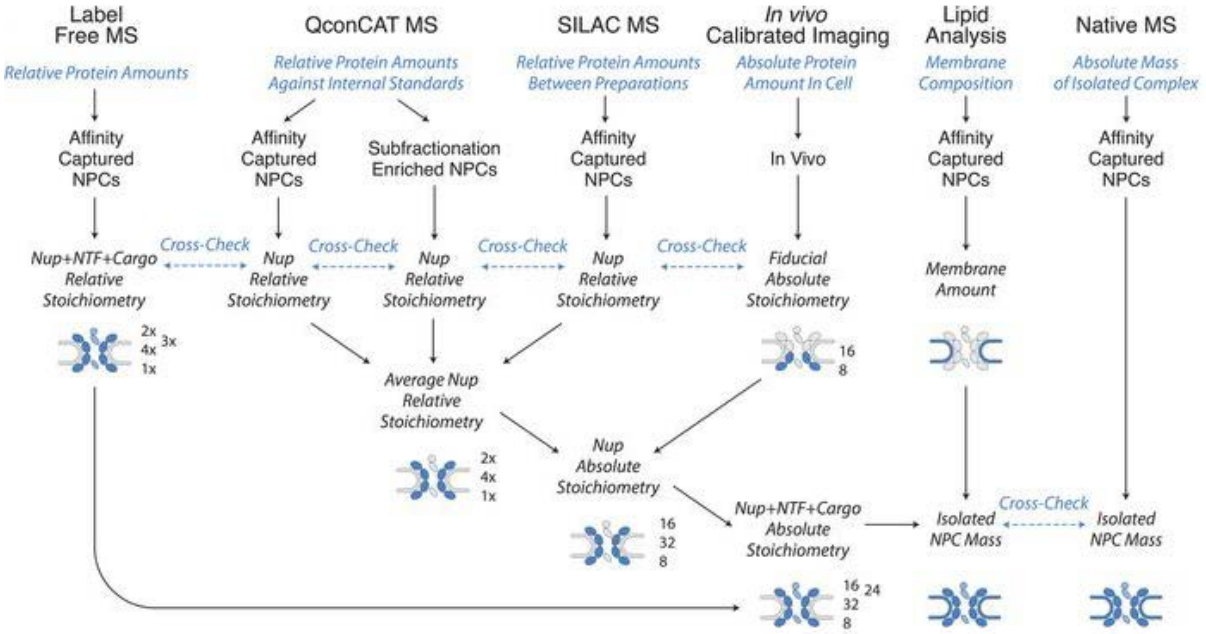
3/ Sampling good scoring configurations

4/ Analysis and assesment

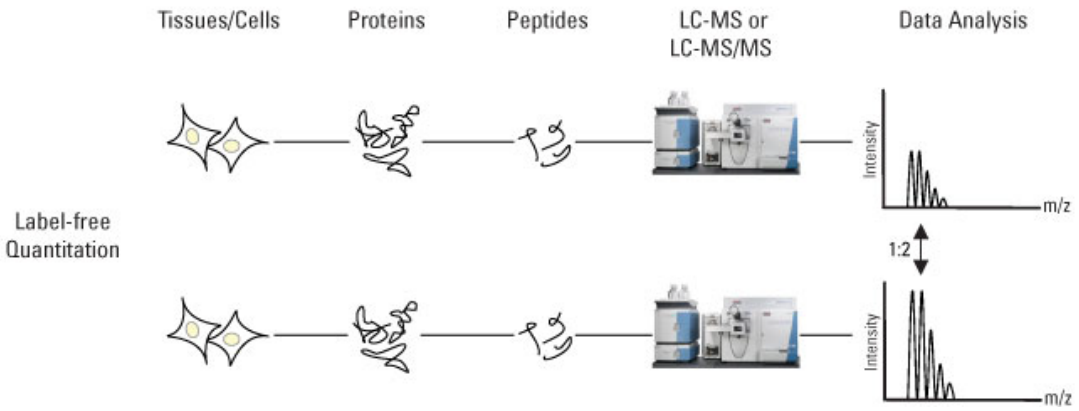


Experimental data

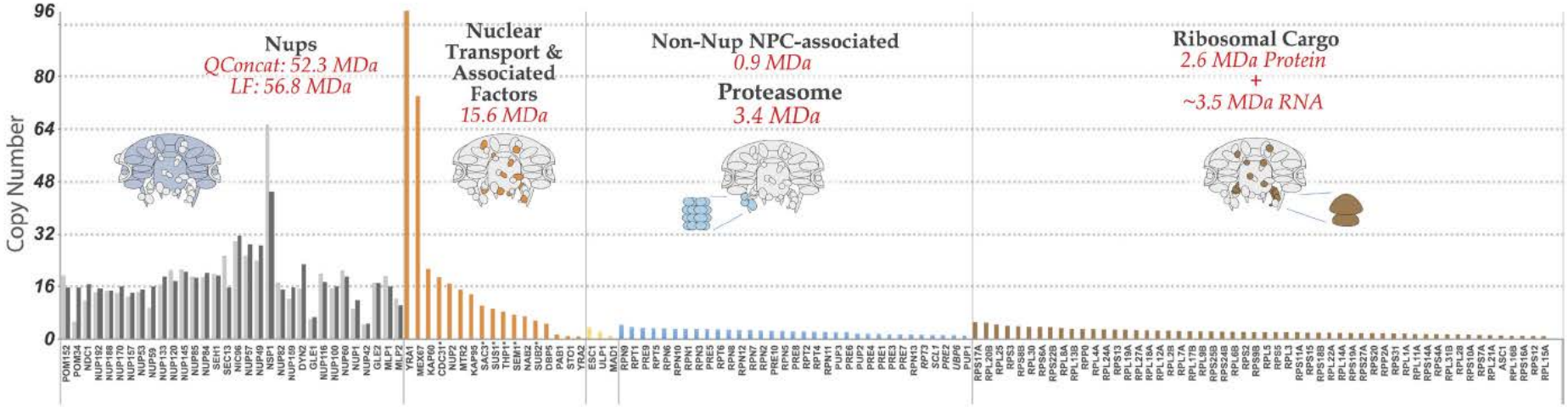




Label free quantitation by MS



MaxQuant iBAQ: Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).



Label free quantitation by MS

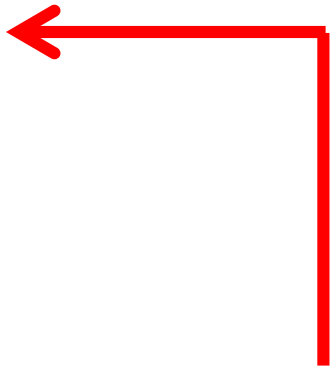
Label-free quantification approaches aim to correlate the mass spectrometric signal of intact proteolytic peptides or the number of peptide sequencing events with the relative or absolute protein quantity directly.

Relative quantitation strategies compare the levels of individual peptides in a sample to those in an identical, but experimentally modified, sample.

Absolute quantification can be obtained *estimated* from analysis of several mass spectrometric signal (TOP3 where the intensity of the selected peaks is taken into account) or the number of peptide sequencing events (emPAI == exponentially modified Protein Abundance Index).

$$\mathrm{PAI} = \frac{N_{\mathrm{obsd}}}{N_{\mathrm{obsbl}}} \qquad \mathrm{emPAI} = 10^{\mathrm{PAI}} - 1$$

$$\text{Protein content (mol \%)} = \frac{\mathrm{emPAI}}{\Sigma\,(\mathrm{emPAI})} \times 100$$



Quantitation by MS

Label free quantitation (MAxQuand iBAQ)
intensity-Based Absolute Quantification

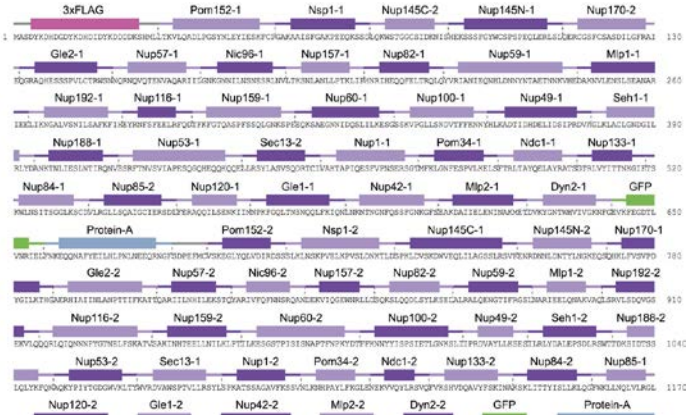
QconCAT MS.
(Internal standard = synthetic concatemer of tryptic peptides)

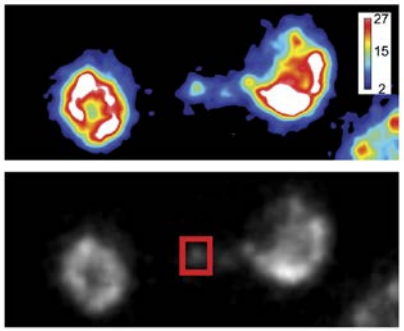
SILAC MS Relative quantification between preparations
(Stable Isotope Labelling in Culture)

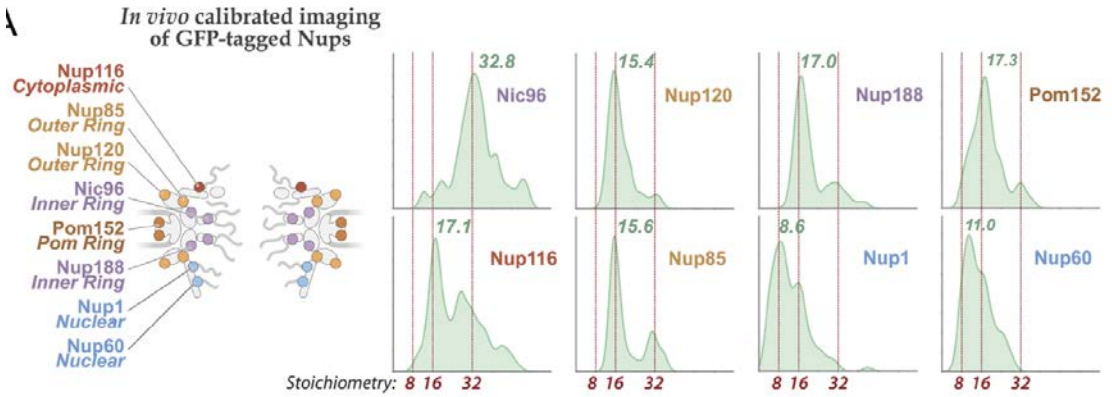
In vivo Calibrated Imaging Absolute quantification

Lipid Analysis Membrane composition

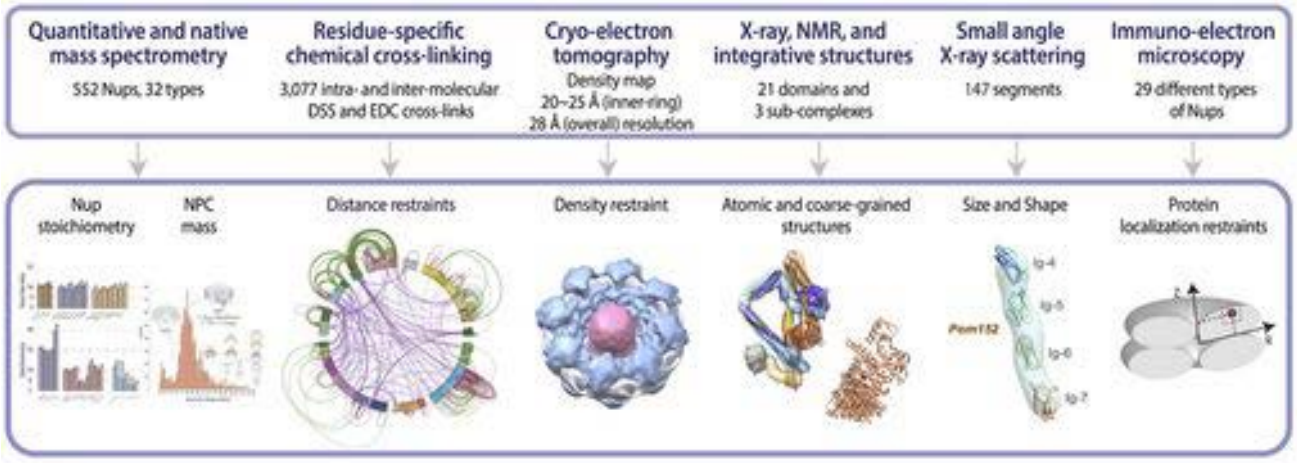
Native MS Absolute mass of isolated complexes

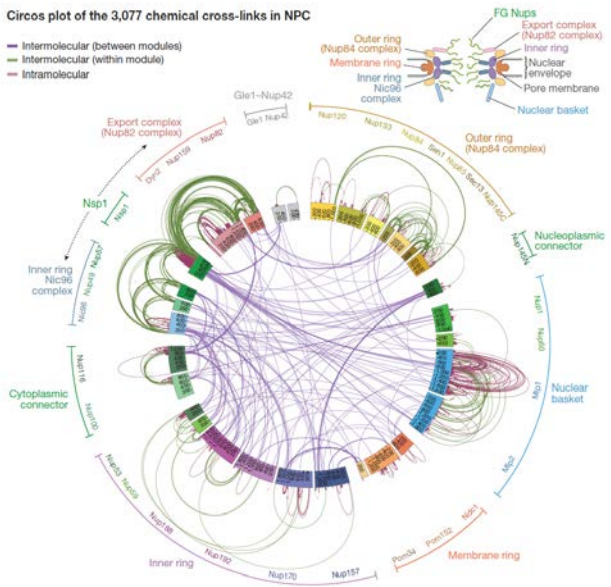


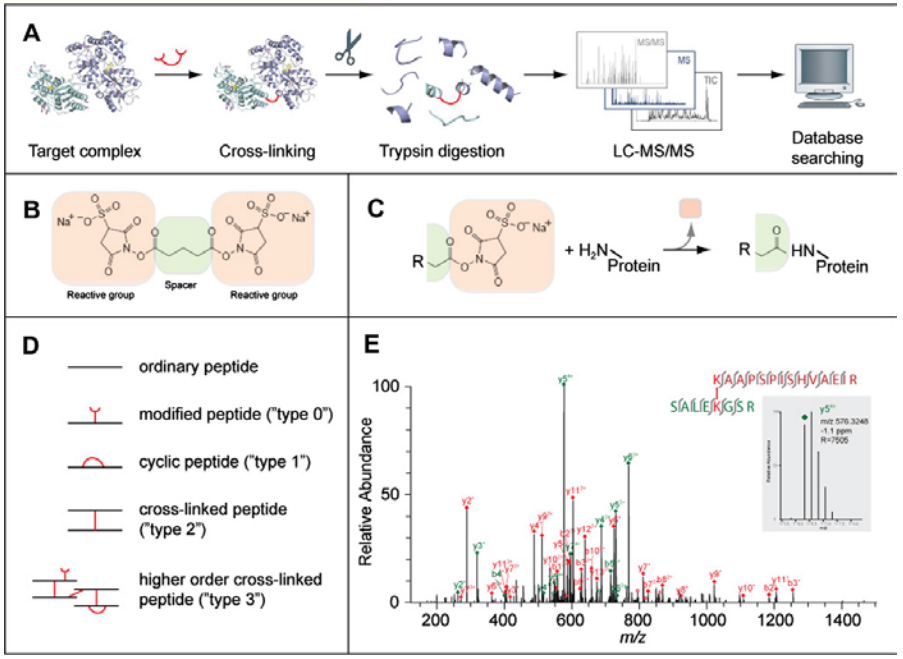




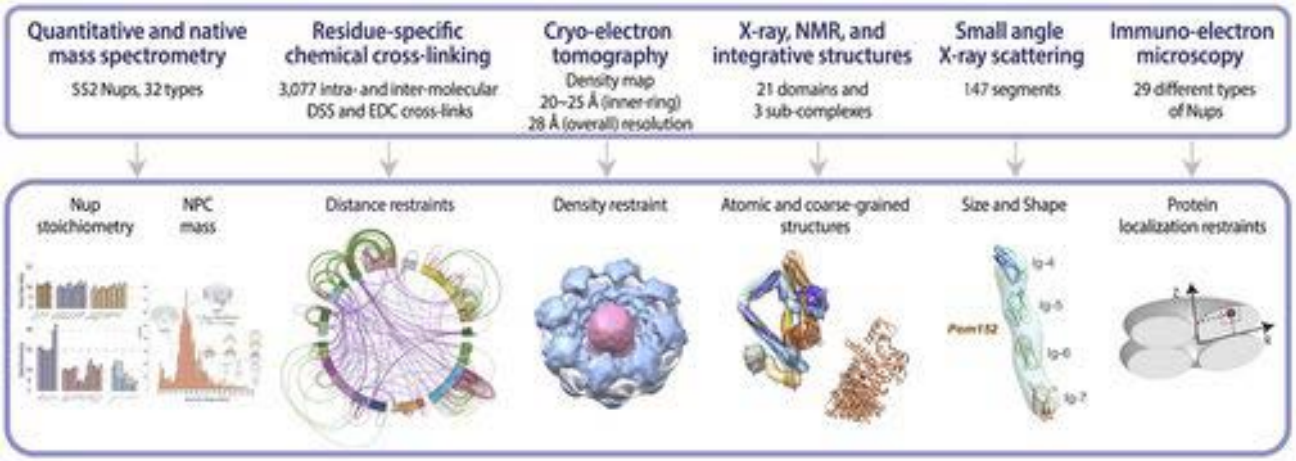
Experimental data





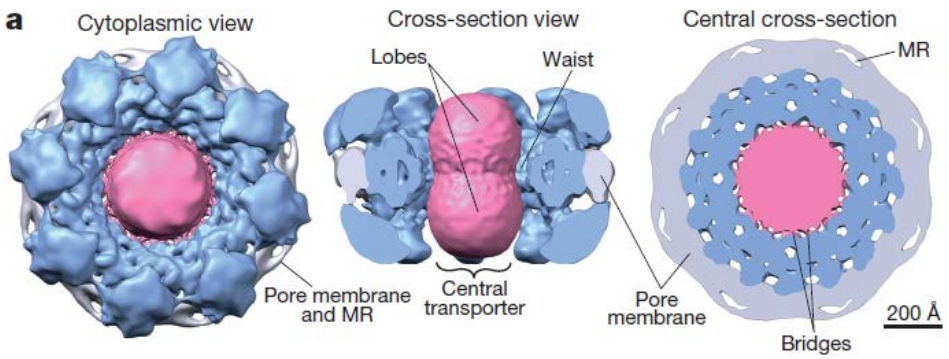


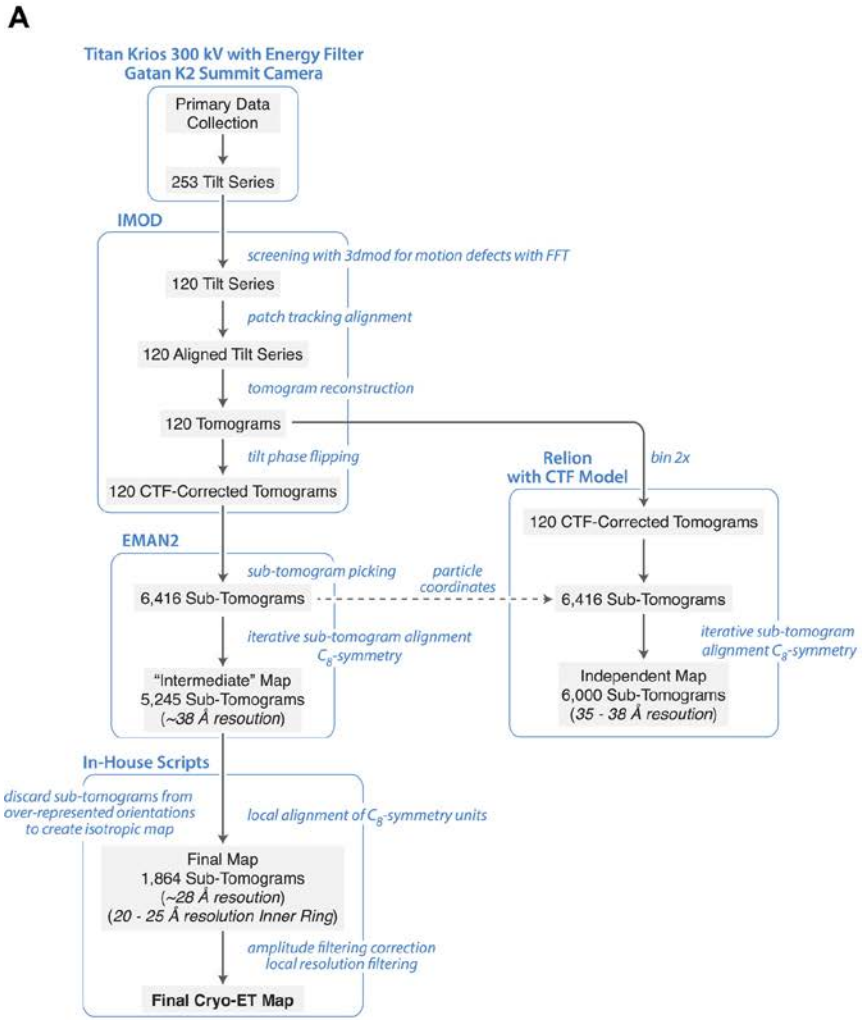
Experimental data

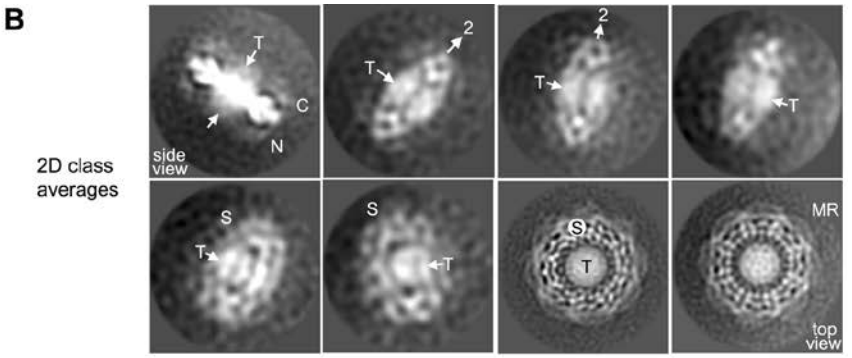


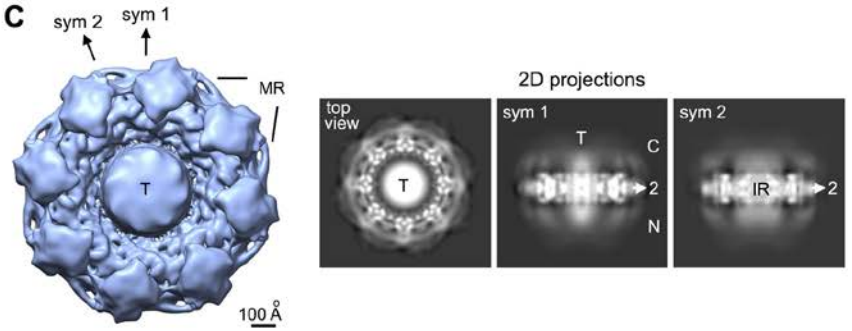
CryoET and sub-tomogram averaging

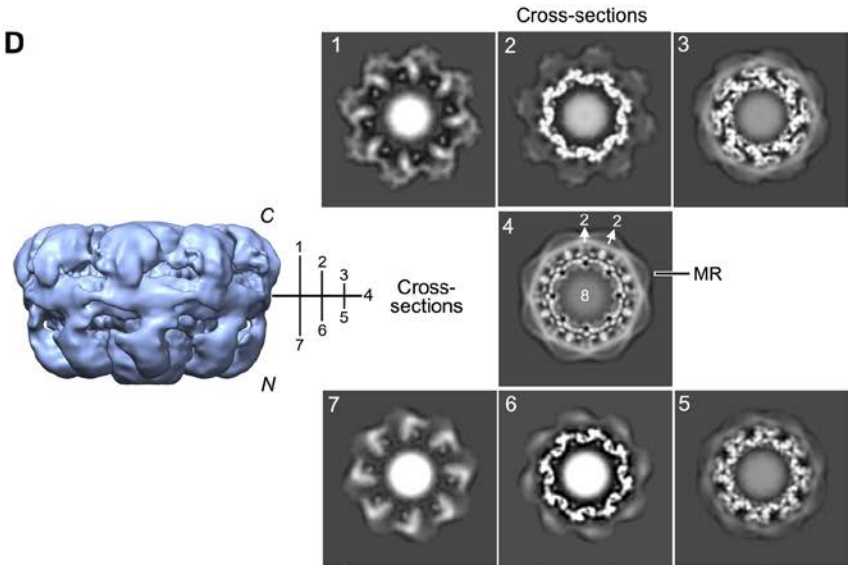
- 3D map at approximately 28 Å resolution, 1864 sub-tomograms
- nuclear envelope membrane



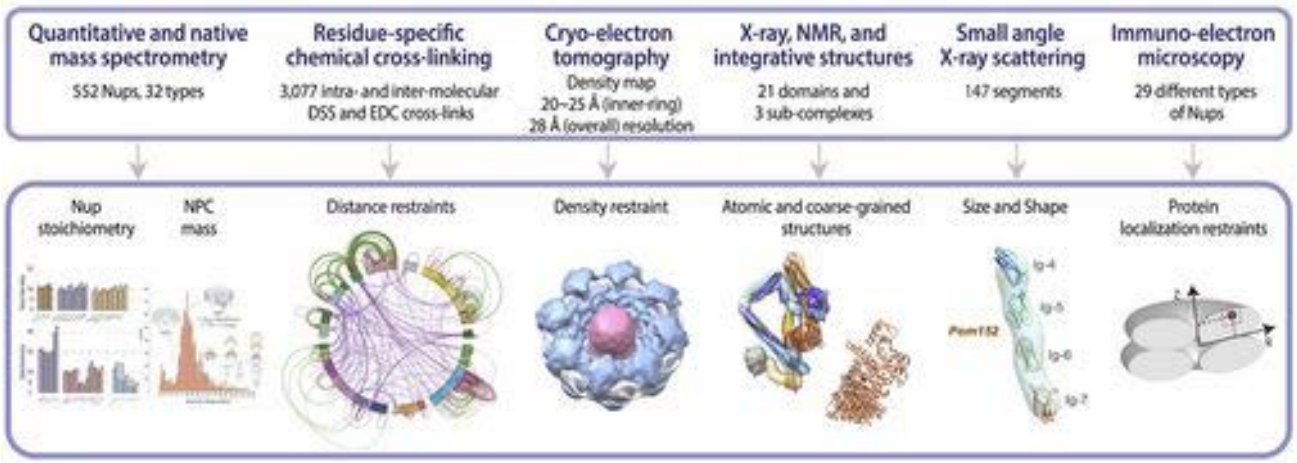

A
Titan Krios 300 kV with Energy Filter
Gatan K2 Summit Camera
Primary Data Collection
253 Tilt Series
IMOD
screening with 3dmod for motion defects with FFT
120 Tilt Series
patch tracking alignment
120 Aligned Tilt Series
tomogram reconstruction
120 Tomograms
tilt phase flipping
120 CTF-Corrected Tomograms
Relion with CTF Model
bin 2x
120 CTF-Corrected Tomograms
EMAN2
sub-tomogram picking
particle coordinates
6,416 Sub-Tomograms
6,416 Sub-Tomograms
iterative sub-tomogram alignment C8-symmetry
iterative sub-tomogram alignment C8-symmetry
"Intermediate" Map
5,245 Sub-Tomograms
(~38 Å resoution)
Independent Map
6,000 Sub-Tomograms
(35 - 38 Å resoution)
In-House Scripts
discard sub-tomograms from over-represented orientations to create isotropic map
local alignment of C8-symmetry units
Final Map
1,864 Sub-Tomograms
(~28 Å resoution)
(20 - 25 Å resolution Inner Ring)
amplitude filtering correction
local resolution filtering
Final Cryo-ET Map

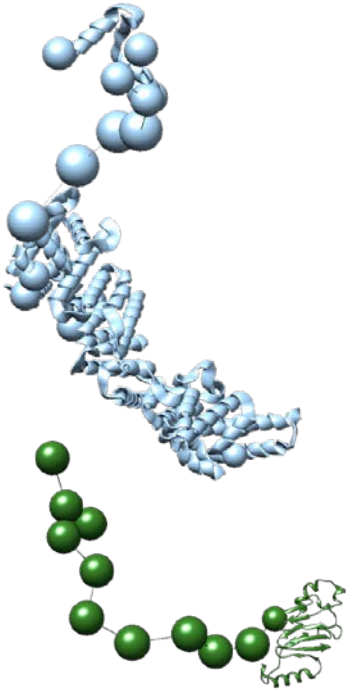

B
2D class averages
side view
top view
T
C
N
S
2
MR


C
sym 2
sym 1
MR
T
100 Å
2D projections
top view
sym 1
sym 2
T
C
N
2
IR

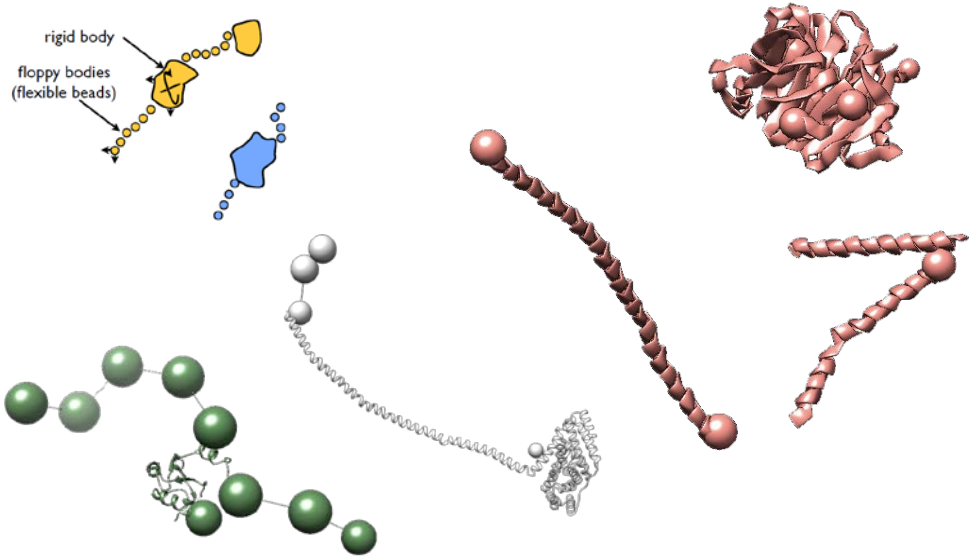

D
Cross-sections
C
N
1
2
3
4
5
6
7
Cross-sections
MR
8

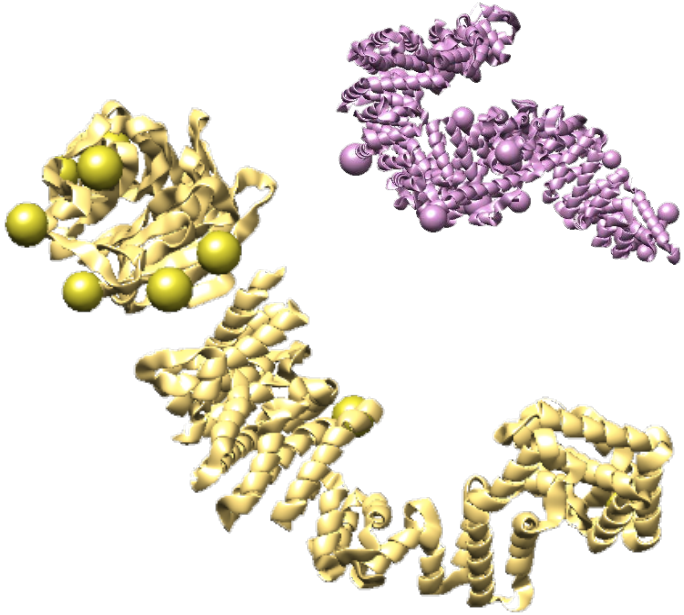
Experimental data



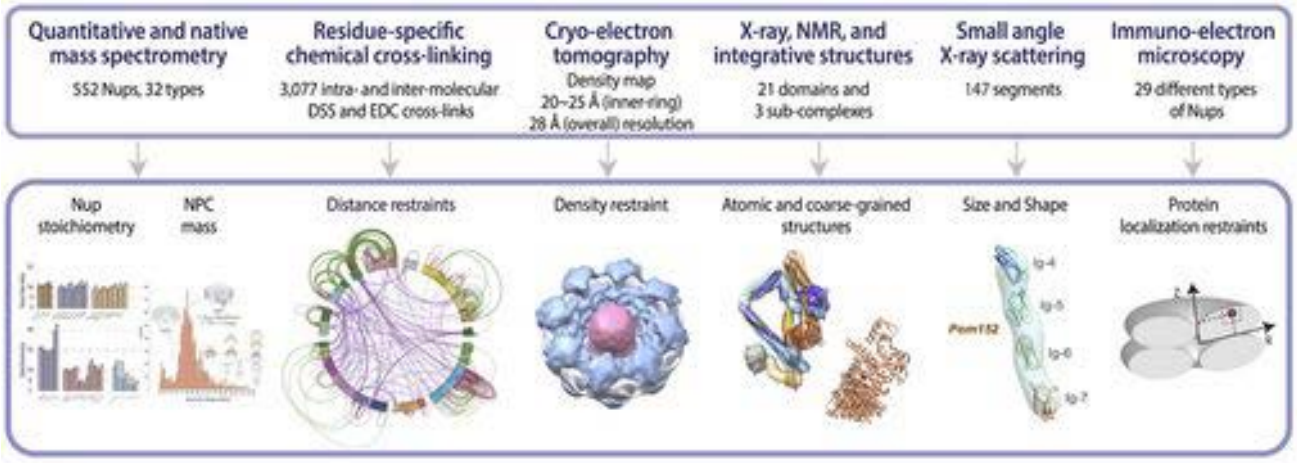


Gaussian Mixture Models (GMMs)

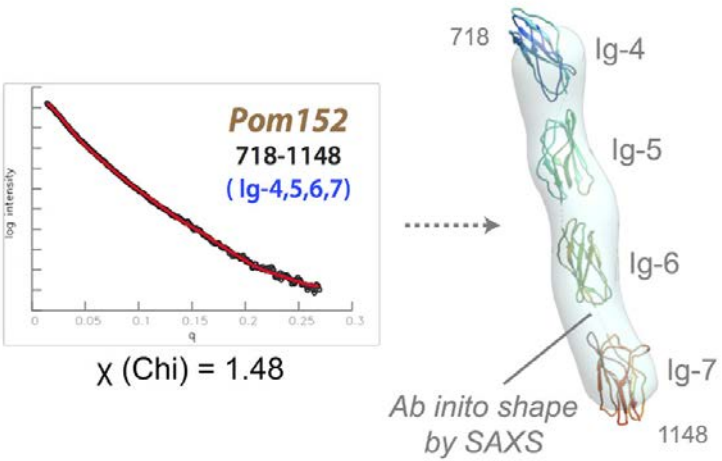




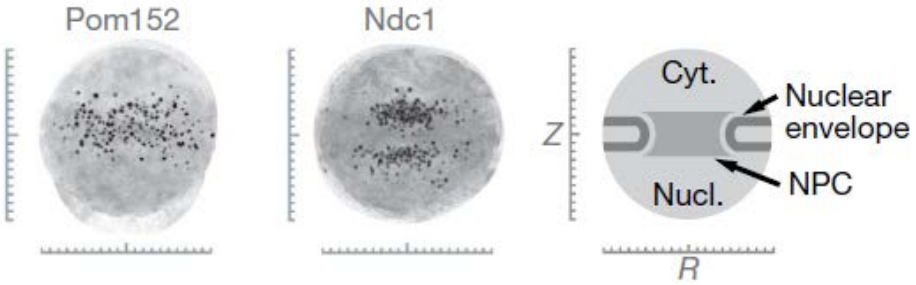
Experimental data



SAXS analysis and Ab-initio reconstruction



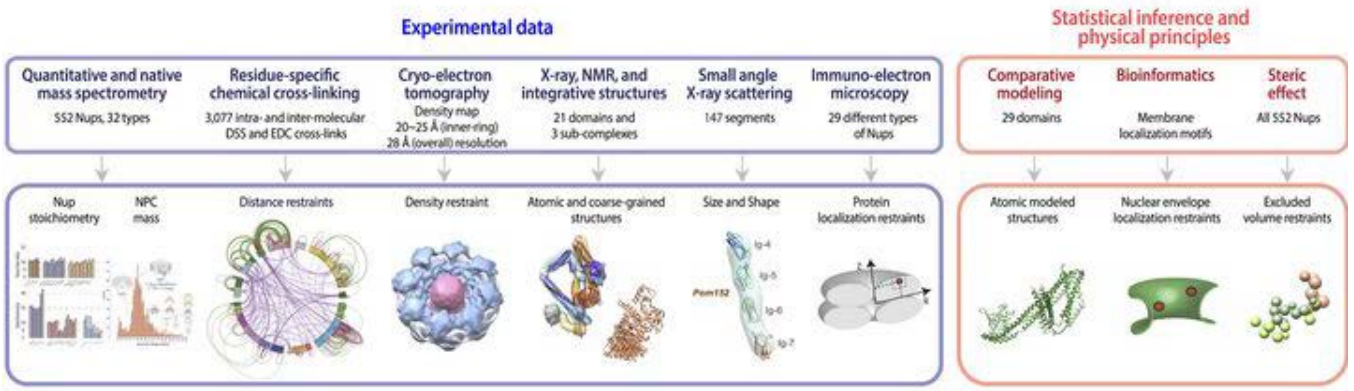
Localization of proteins by immuno-EM. Immuno-EM montages for Pom152–PrA nuclei and Ndc1–PrA nuclear envelopes with gold-labelled antibodies. (Alber, 2007)



Gathering data

↓

Representing and translating data into restraints

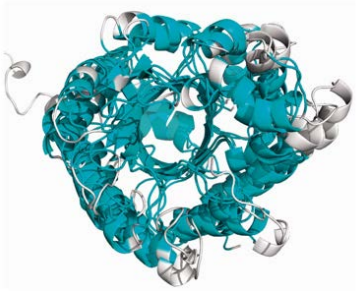


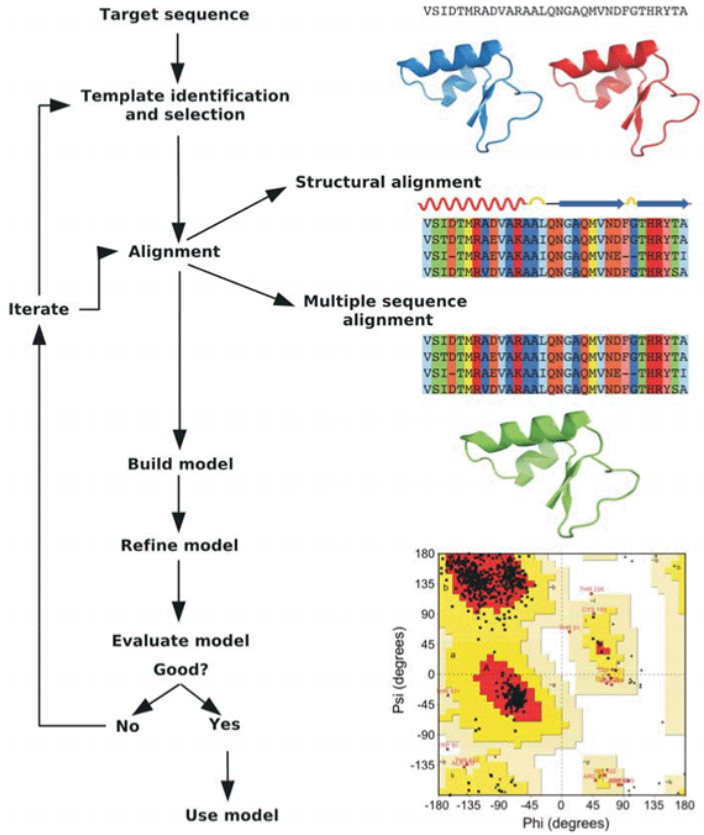
Comparative modelling

1/ Search for templates (or 'parents')

2/ Align the target sequence with the parent(s)
- structurally conserved regions
- structurally variable regions

3/ Inherit the SCRs from the parent(s)
4/ Build the SVRs
5/ Build the sidechains
6/ Refine the model
7/ Evaluate errors in the model

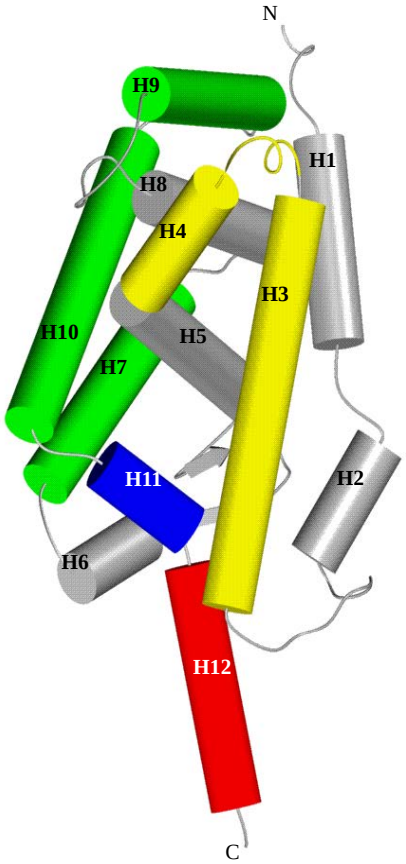


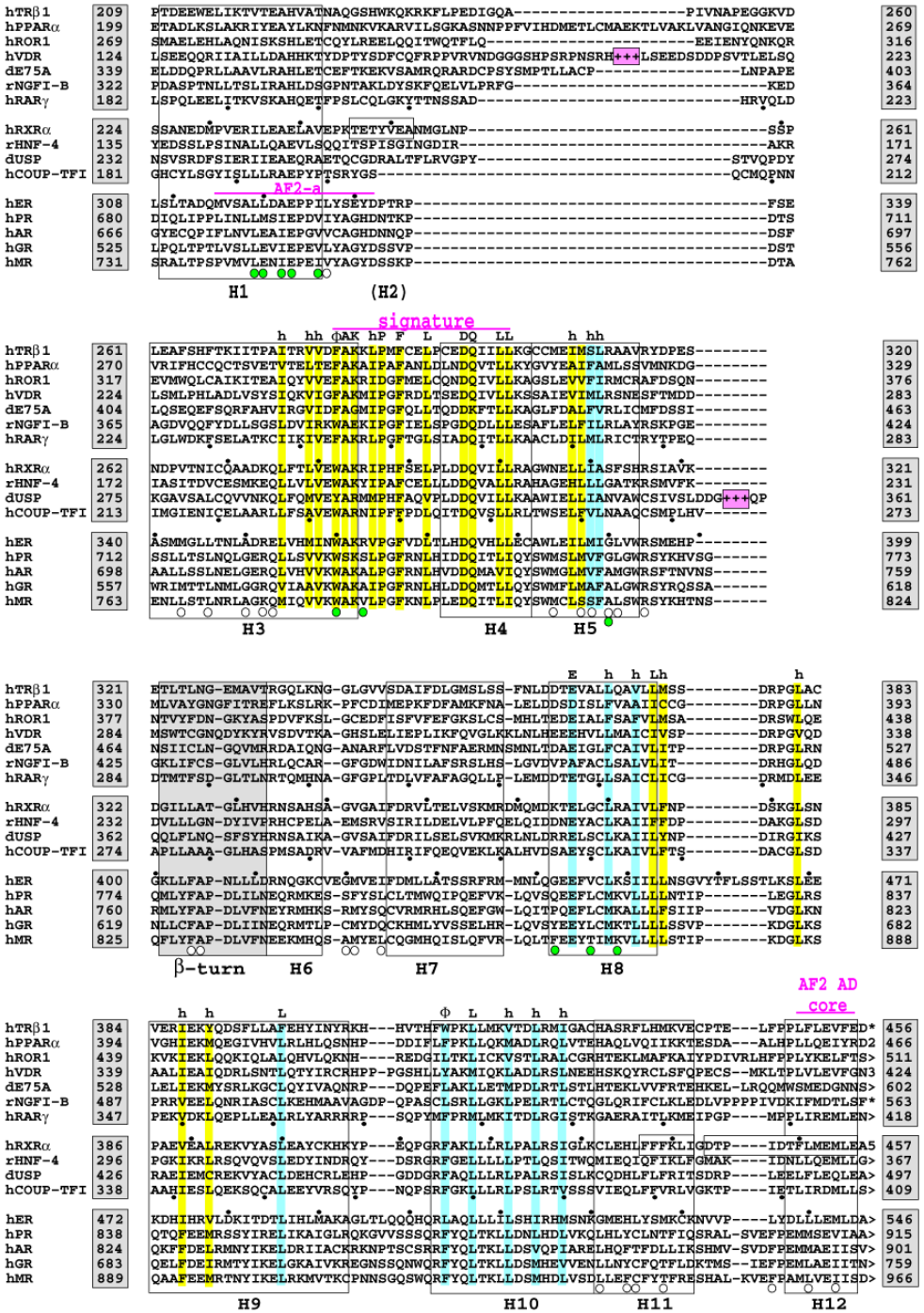


Homology modeling

1 (ou several) 3D structure(s)

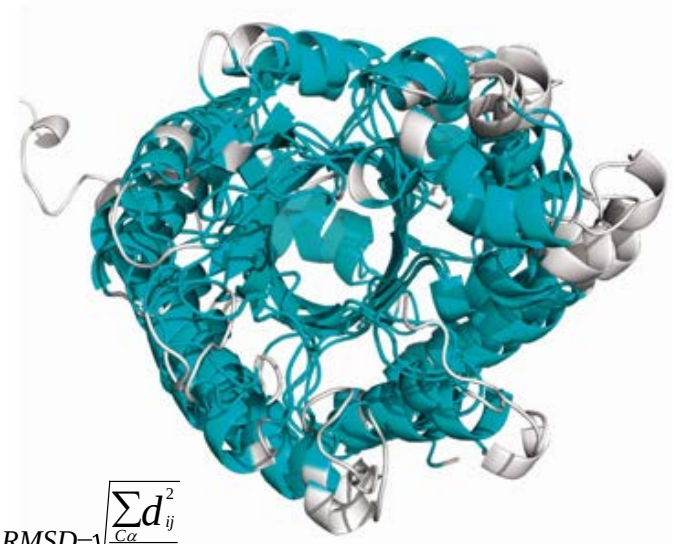
1 multiple sequence alignment



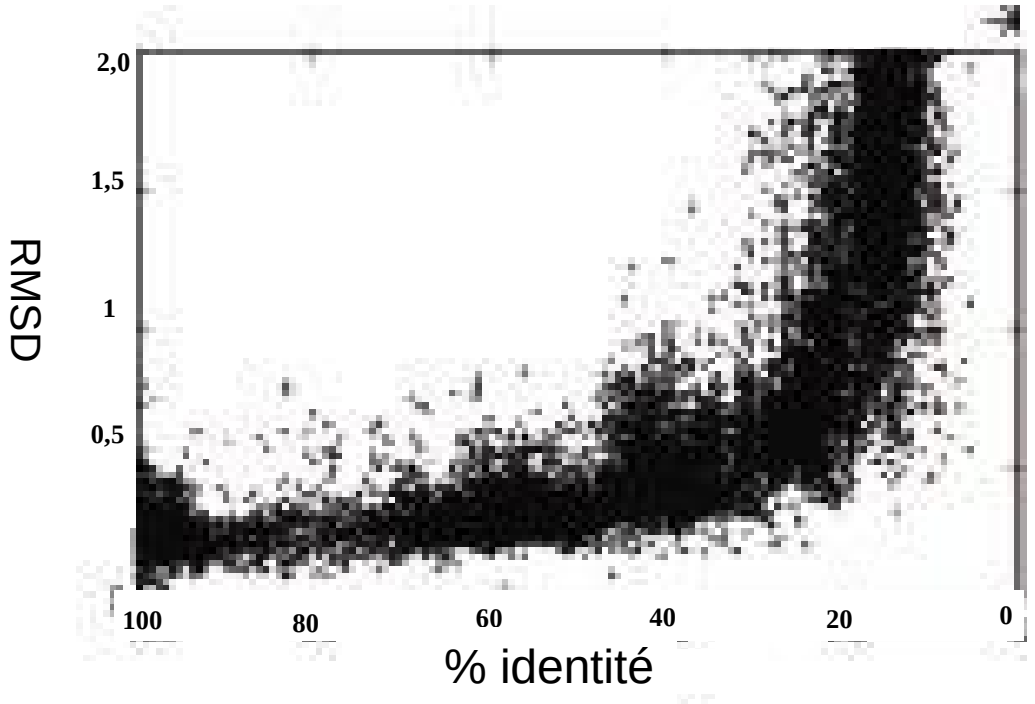


Comparaison of 3D structures from homologous proteins

Comparaison des structures 3D de protéines homologues. Relations entre la divergence en séquence et la divergence en structure 3D



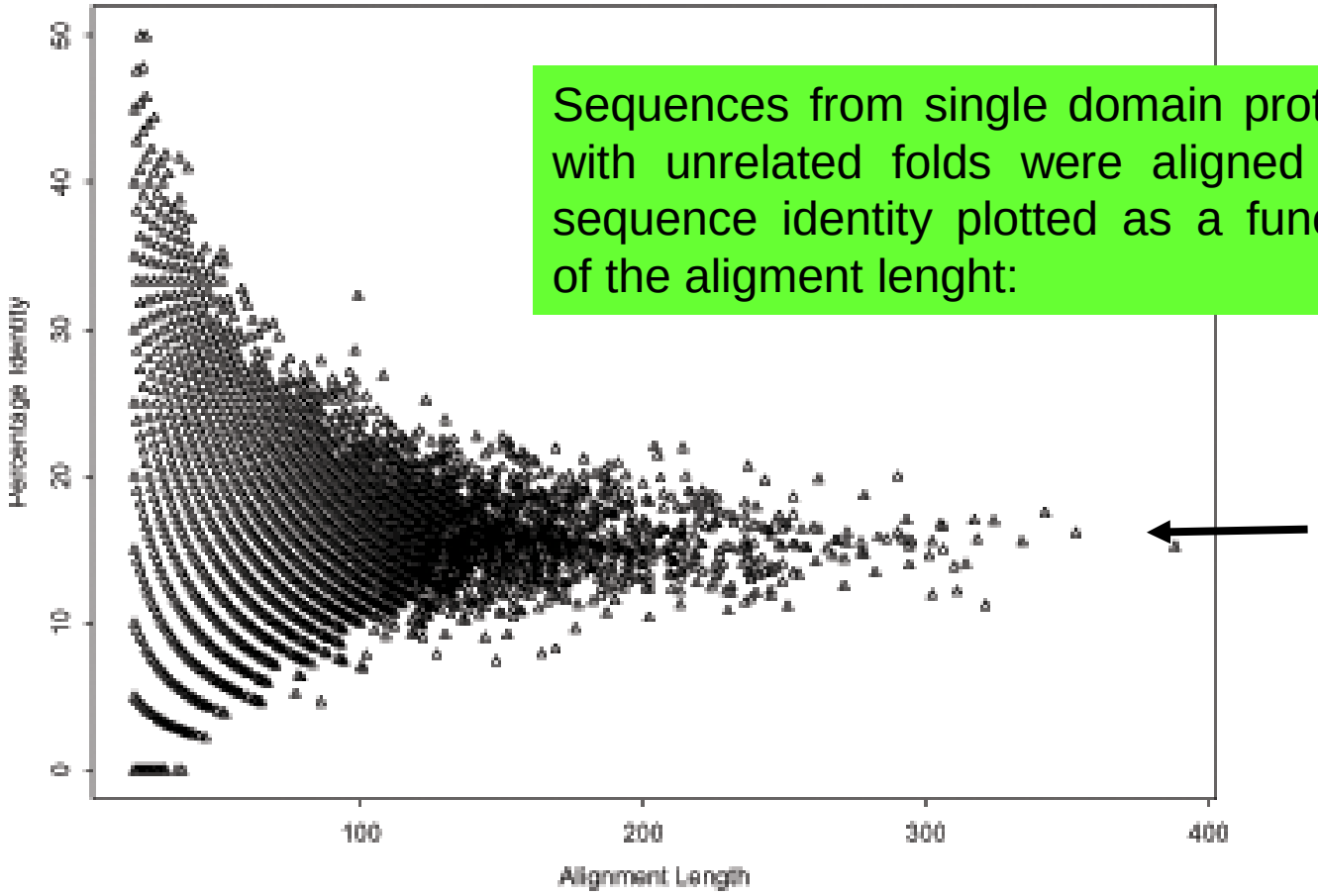
$$RMSD=\sqrt{\frac{\sum_{C\alpha} d_{ij}^2}{N}}$$



@Wilson et al., (2000), J. Mol. Biol.,297, 233-249.

Relation type exponentiel % identité / déformations

Sequence identity vs aligment lenght. Biologic / Statistical Significance

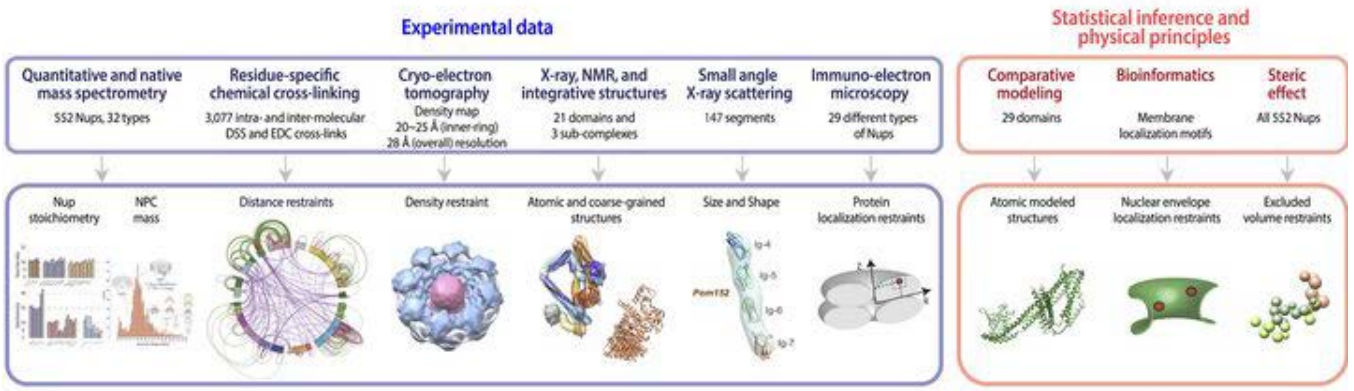


Sequences from single domain proteins with unrelated folds were aligned and sequence identity plotted as a function of the aligment lenght:

Gathering data

↓

Representing and translating data into restraints



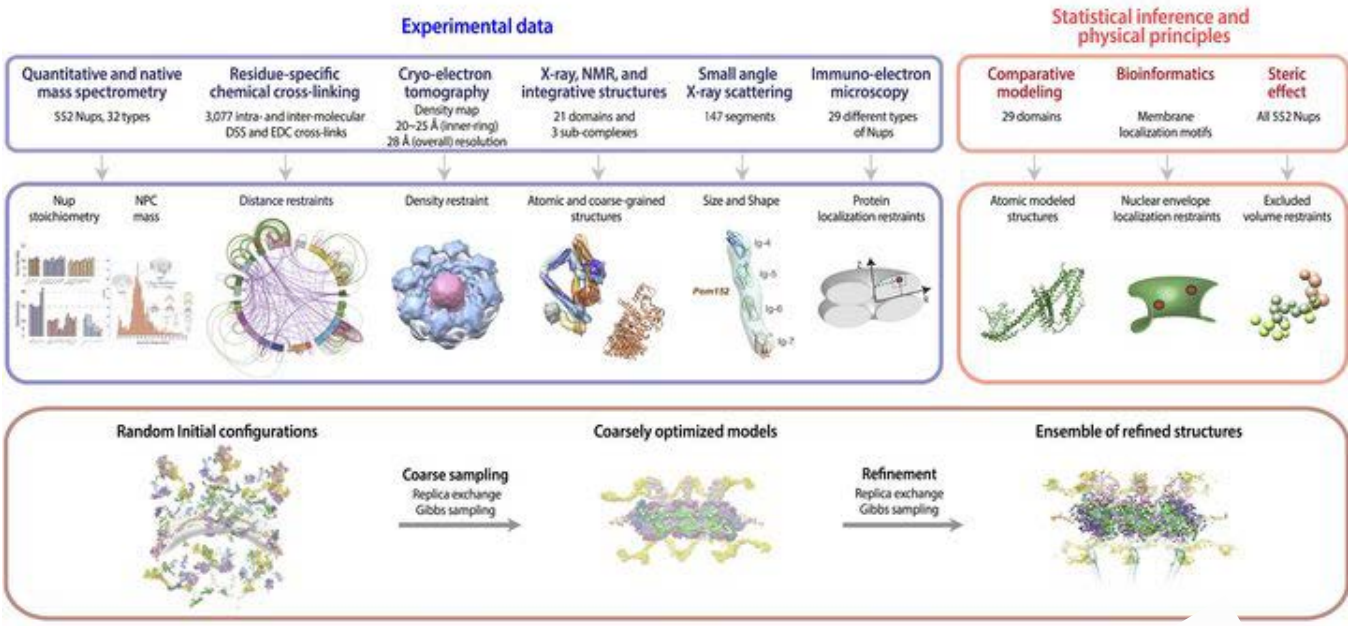
Gathering data

↓

Representing and translating data into restraints

↓

Sampling good scoring configurations



Sampling:

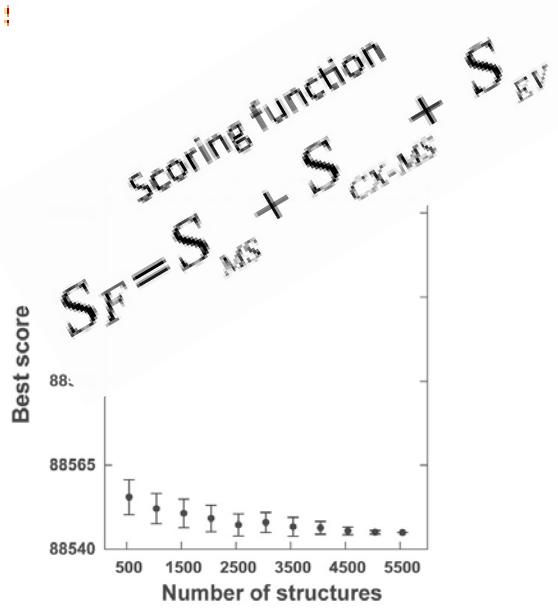
- Simplex
- Conjugate Gradients
- Monte Carlo
- Brownian Dynamics
- Molecular Dynamics
- Replica Exchange
- Divide-and-conquer enumeration

100 455 models

5 500 good scoring

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, ...)



Gathering data

↓

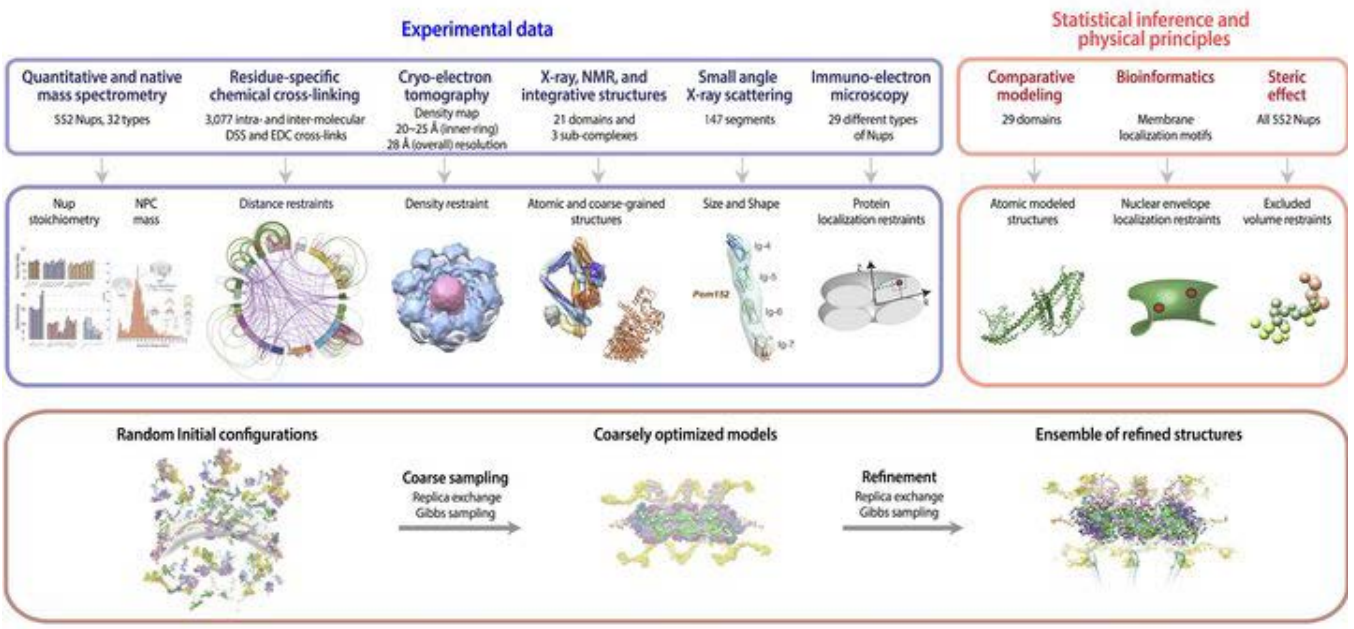
Representing and translating data into restraints

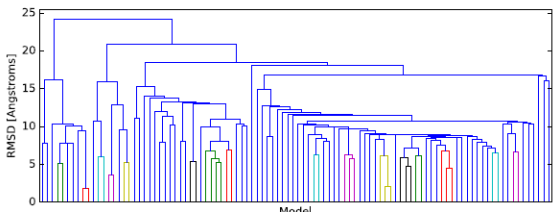
↓

Sampling good scoring configurations

↓

Analysis and assessment





Clustering solutions and generating probability densities
Satisfaction of data and considerations that

were used to compute structures
were not used to compute structures.

Gathering data

↓

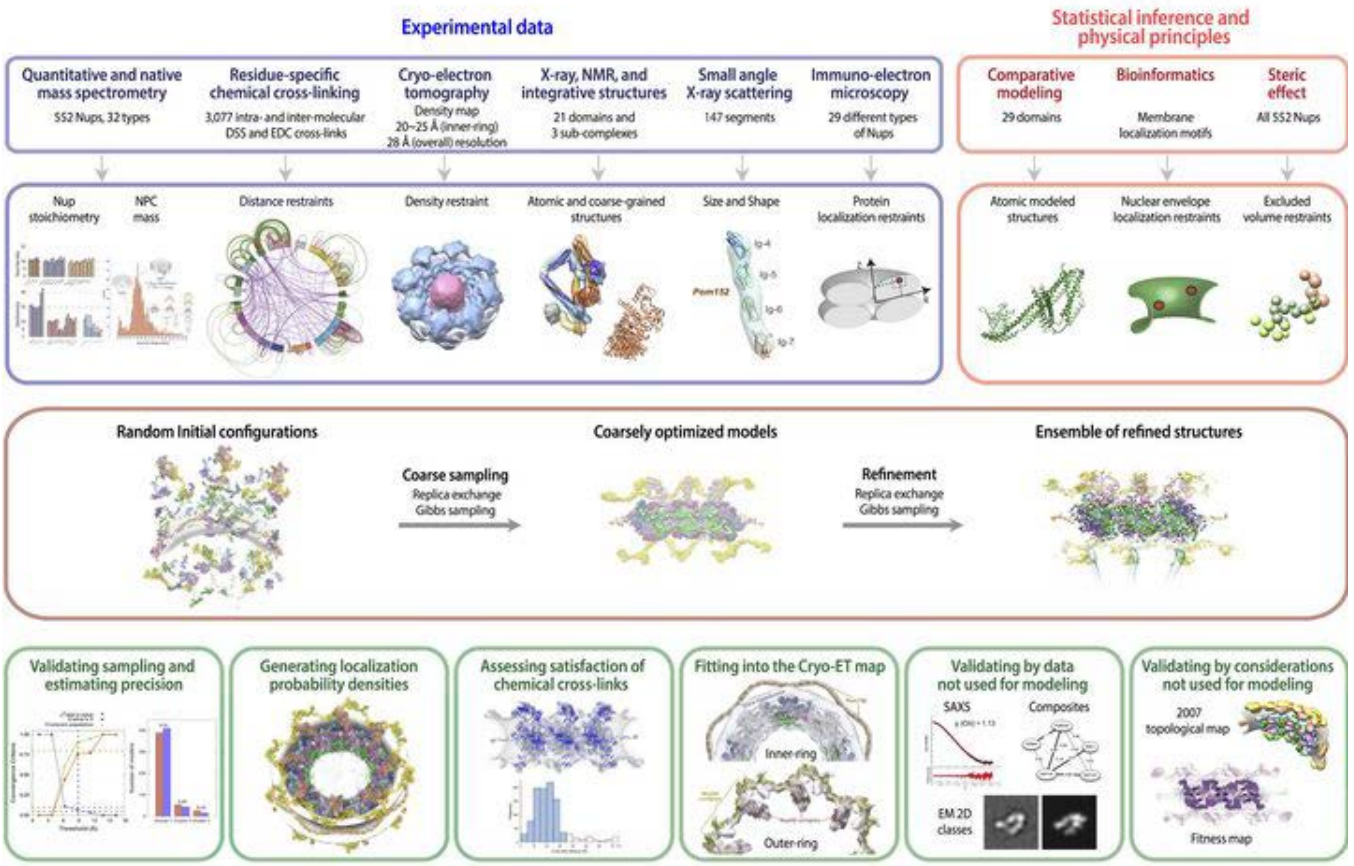
Representing and translating data into restraints

↓

Sampling good scoring configurations

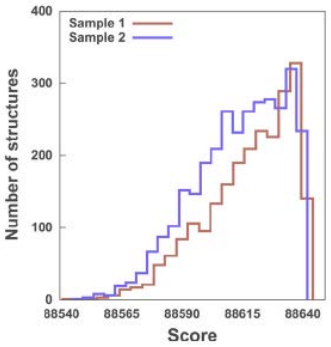
↓

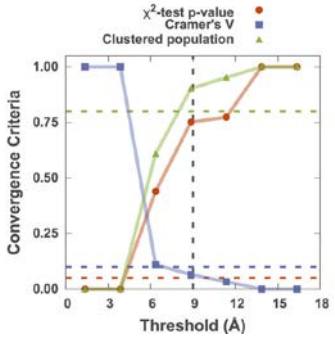
Analysis and assesment



52 (87) MDa
550 copies of 30 polypeptide
C8 symmetry
3082 intermolecular cross-links

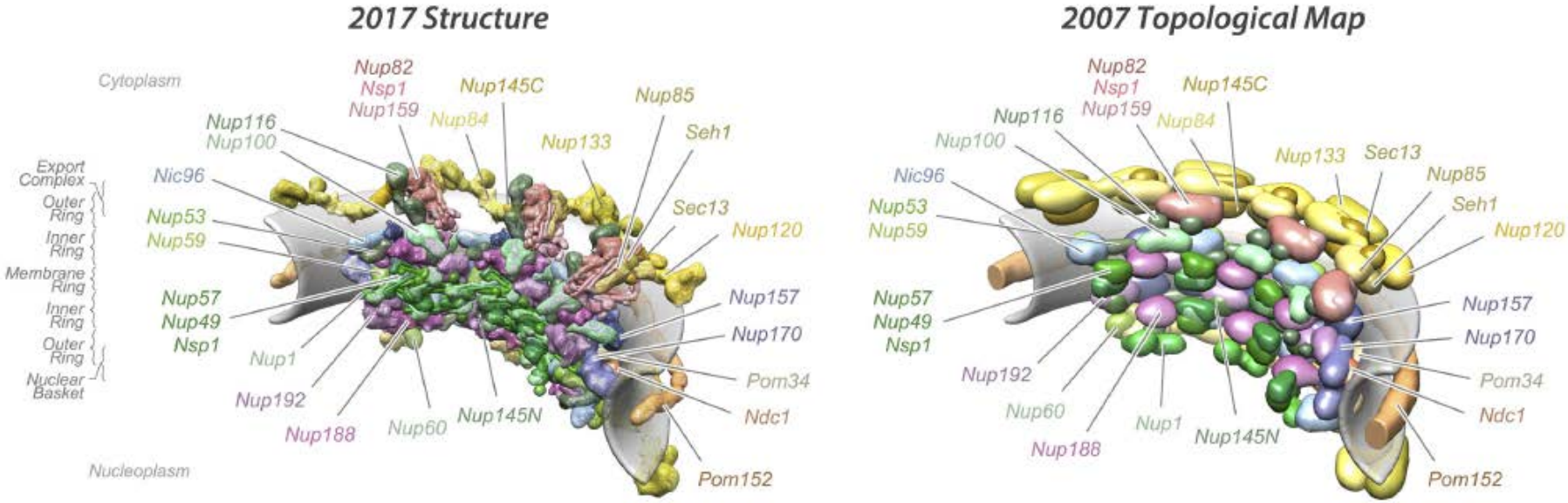
Estimated resolution: 28 Å





Integrative structure and functional anatomy of a nuclear pore complex

Seung Joong Kim[1]*, Javier Fernandez-Martinez[2]*, Ilona Nudelman[2]*, Yi Shi[3]†*, Wenzhu Zhang[3]*, Barak Raveh[1], Thurston Herricks[4], Brian D. Slaughter[5], Joanna A. Hogan[6], Paula Upla[7], Ilan E. Chemmama[1], Riccardo Pellarin[1]†, Ignacia Echeverria[1], Manjunatha Shivaraju[5], Azraa S. Chaudhury[2], Junjie Wang[3], Rosemary Williams[2], Jay R. Unruh[5], Charles H. Greenberg[1], Erica Y. Jacobs[3], Zhiheng Yu[8], M. Jason de la Cruz[8]†, Roxana Mironska[2], David L. Stokes[7], John D. Aitchison[4,9], Martin F. Jarrold[6], Jennifer L. Gerton[5], Steven J. Ludtke[10], Christopher W. Akey[11], Brian T. Chait[3], Andrej Sali[1] & Michael P. Rout[2]



Albert et al. 2007, 2008
Kim et al. 2018

1/ Principles of integrative structure determination

2/ The example of the Nuclear Pore Complex

3/ Mainstream complementary experimental methods

3.1/ Immunoprecipitation assays

3.2 Two hybrid

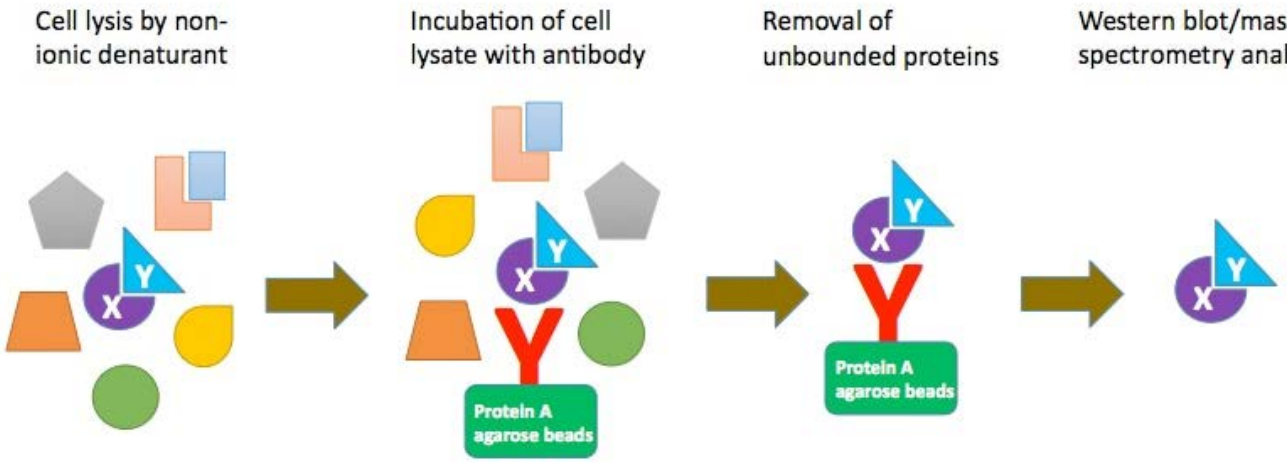
3.3/ Analytical ultracentrifugation

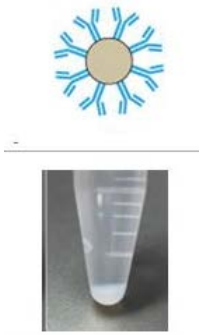
3.4/ Foerster Resonnance Energy Transfer

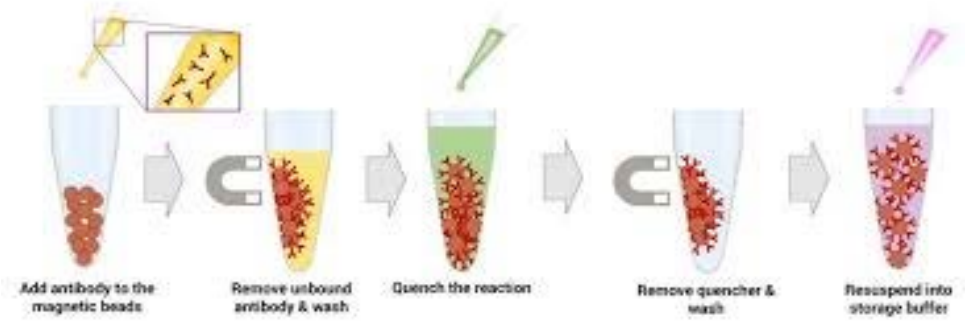
3.5/ Fluorescence anisotropy

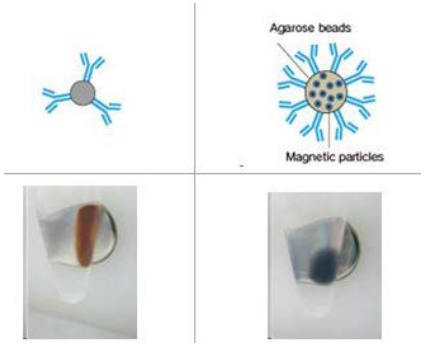
Immunoprecipitation assays

Popular (Flag, HA..) or specific epitopes
Non-Ab pull down: Affinity tags (His, Strep….)





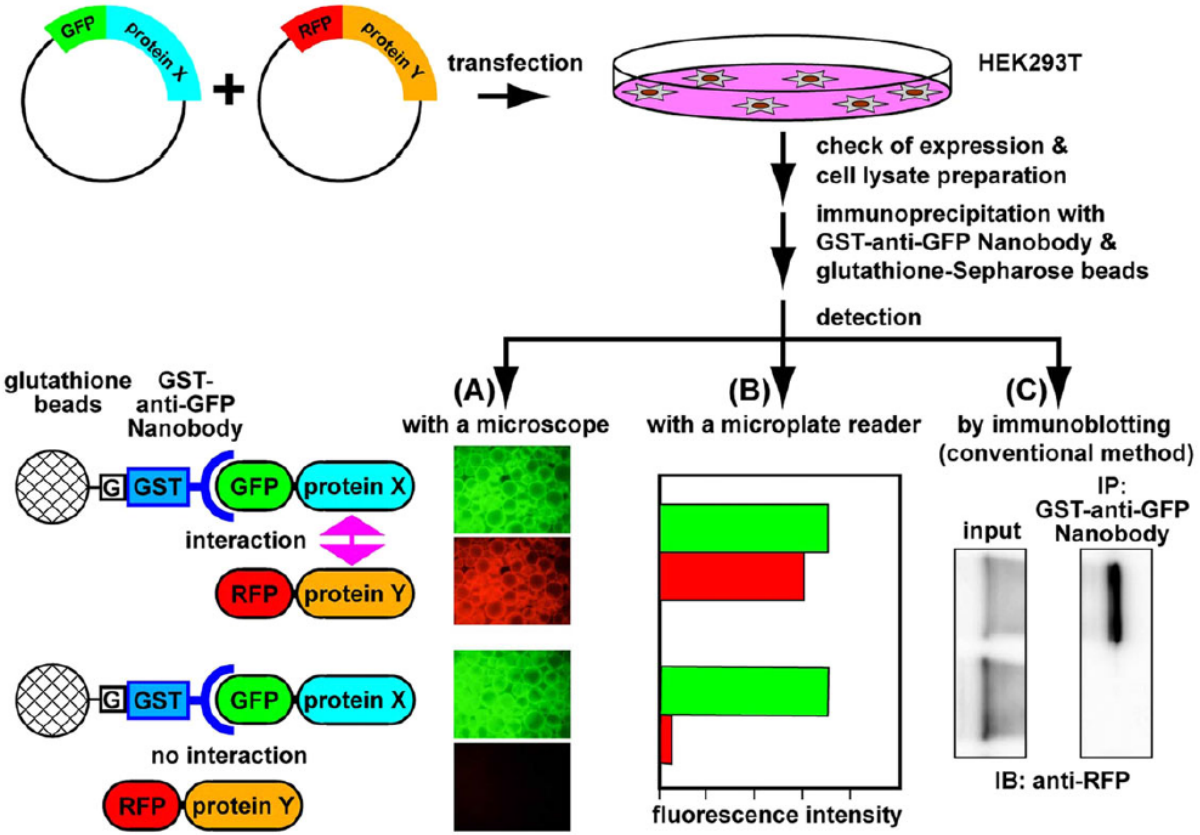


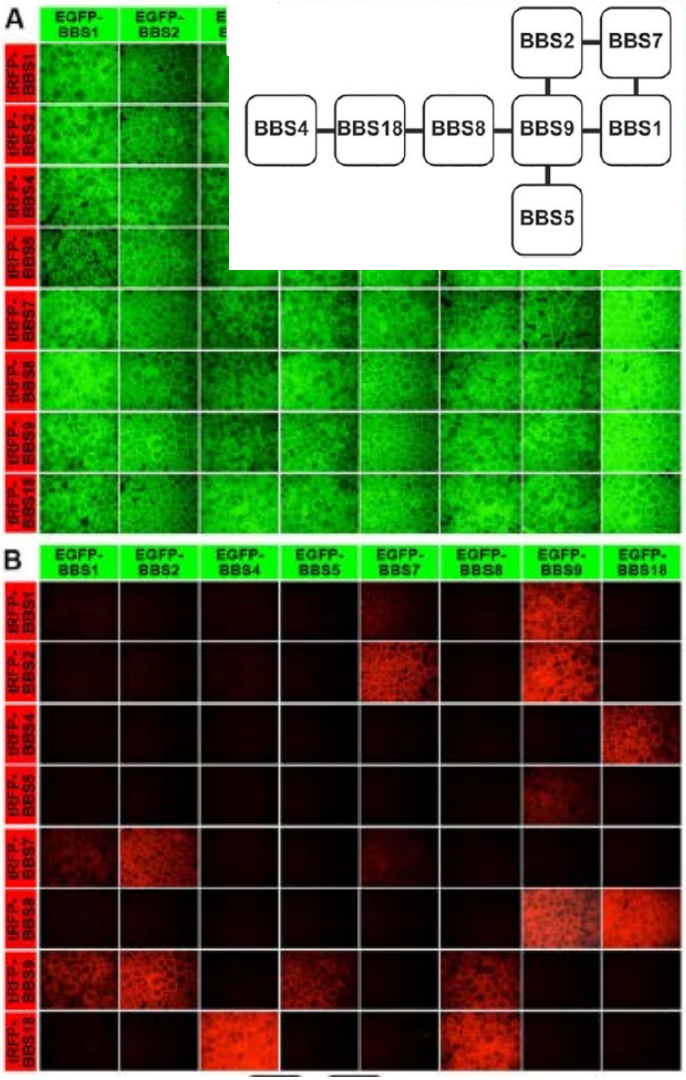


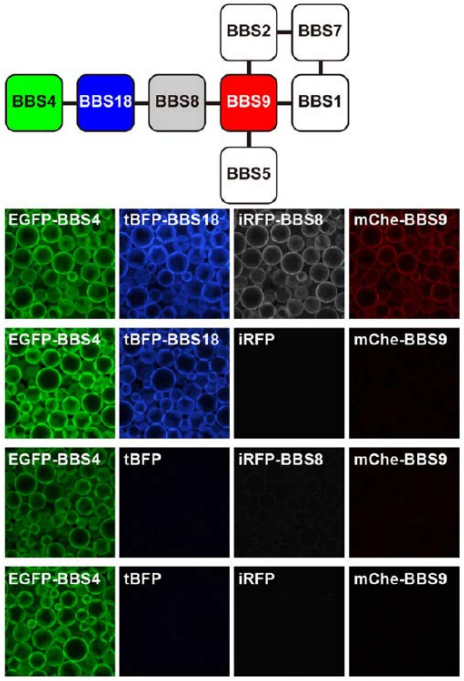
A nanobody based visible immunoprecipitation assays

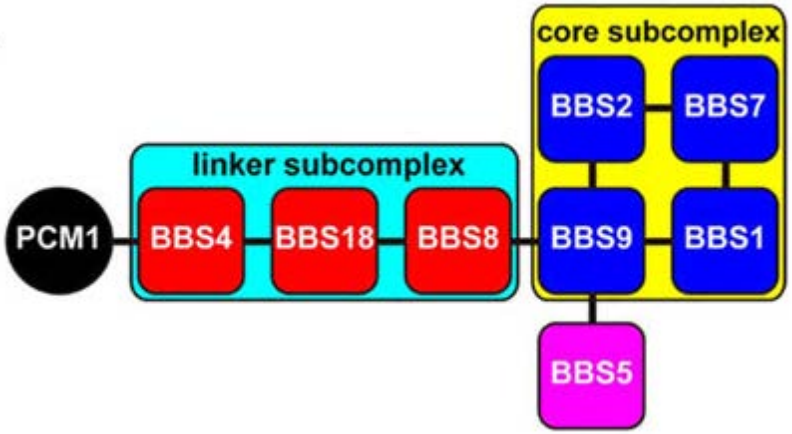
Architectures of multisubunit complexes revealed by a visible immunoprecipitation assay using fluorescent fusion proteins

Yohei Katoh*, Shohei Nozaki*, David Hartanto, Rie Miyano and Kazuhisa Nakayama‡



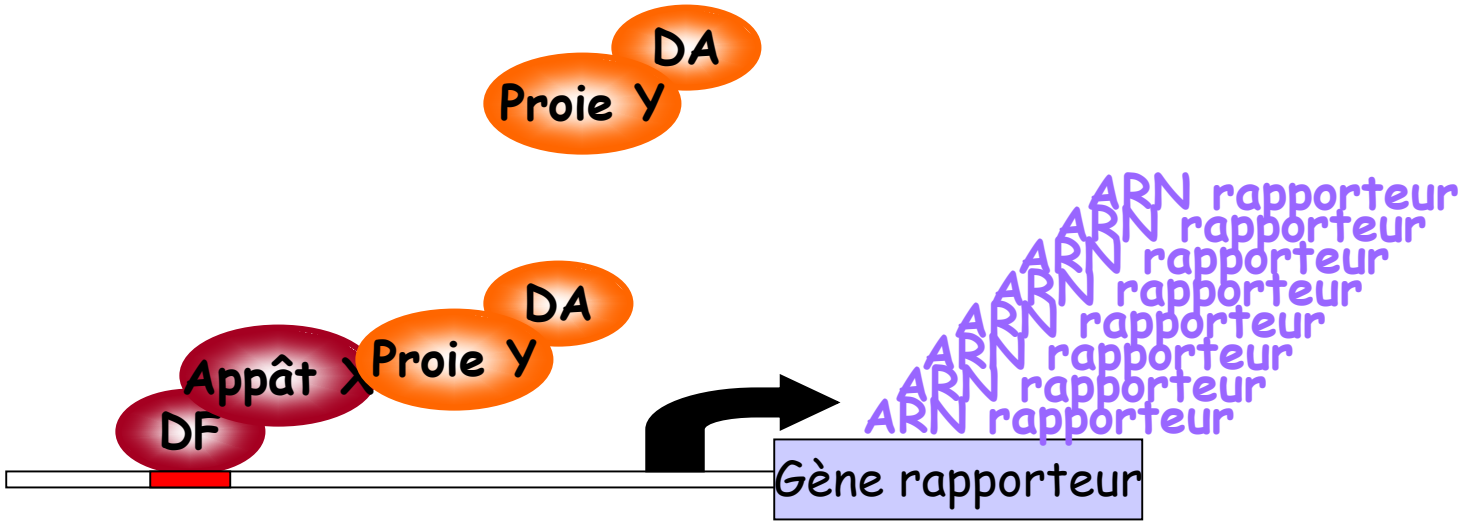

A
EGFP-BBS1
EGFP-BBS2
IRFP-BBS1
IRFP-BBS2
IRFP-BBS4
IRFP-BBS5
IRFP-BBS7
IRFP-BBS8
IRFP-BBS9
IRFP-BBS18
BBS2
BBS7
BBS4
BBS18
BBS8
BBS9
BBS1
BBS5
B
EGFP-BBS1
EGFP-BBS2
EGFP-BBS4
EGFP-BBS5
EGFP-BBS7
EGFP-BBS8
EGFP-BBS9
EGFP-BBS18
IRFP-BBS1
IRFP-BBS2
IRFP-BBS4
IRFP-BBS5
IRFP-BBS7
IRFP-BBS8
IRFP-BBS9
IRFP-BBS18


BBS2
BBS7
BBS4
BBS18
BBS8
BBS9
BBS1
BBS5
EGFP-BBS4
tBFP-BBS18
iRFP-BBS8
mChe-BBS9
EGFP-BBS4
tBFP-BBS18
iRFP
mChe-BBS9
EGFP-BBS4
tBFP
iRFP-BBS8
mChe-BBS9
EGFP-BBS4
tBFP
iRFP
mChe-BBS9


core subcomplex
BBS2
BBS7
linker subcomplex
PCM1
BBS4
BBS18
BBS8
BBS9
BBS1
BBS5

Yeast two-hybrid system

- Detecting protein-protein interactions in yeast
- Transcriptional regulator system
- “prey”-”bait” model :fusion proteins with a transcriptional activating domain (AD, prey), a DNA-binding domain (DBD, bait)
- Term “two-hybrid” derives from these two chimeric proteins.
- Most commonly used method for large scale, high-throughput identification of potential protein-protein interactions



1/ Integrative determination of macromolecular structures

2/ The example of the Nuclear Pore Complex

3/ Mainstream complementary experimental methods

3.1/ Immunoprecipitation assays

3.2/ Two hybrid

3.3/ Analytical ultracentrifugation

3.4/ Förster Resonnance Energy Transfer

3.5/ Fluorescence anisotropy

Analytical Ultracentrifugation

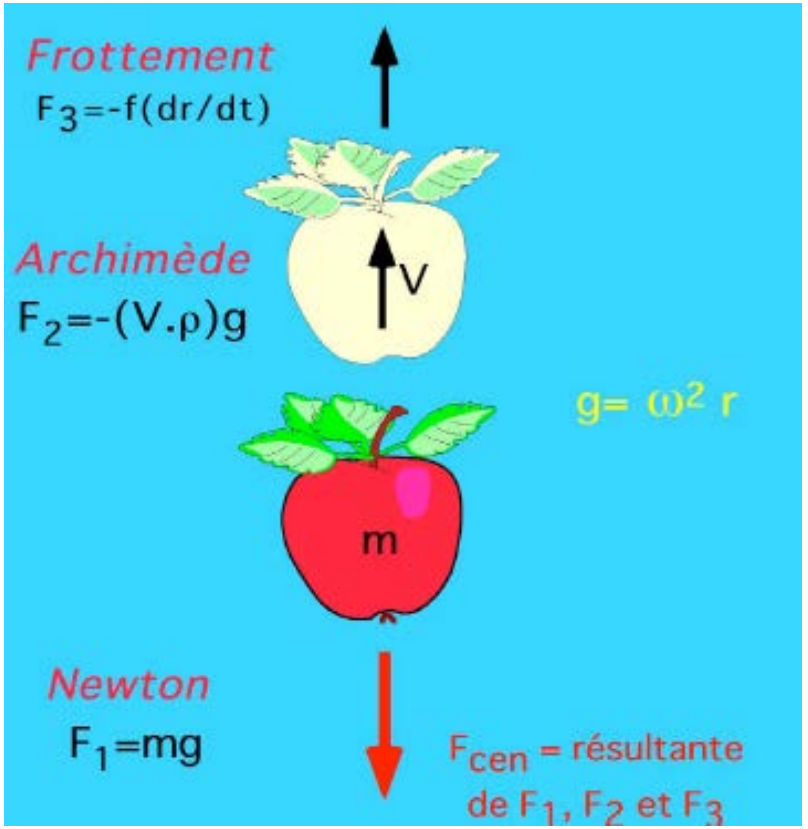
There are 3 forces acting on a sedimenting particle, buoyancy, viscous drag and centrifugal force. As soon as the rotor accelerates to a constant speed, the particle reaches terminal velocity and an equilibrium between these 3 forces is established. There are several experimental conditions and sample properties that influence the sedimentation behavior:

Experimental conditions:

1. rotor speed
2. distance from the rotor center
3. density of the solution
4. viscosity of the solution
5. temperature

Sample properties:

1. molecular weight
2. shape
3. partial specific volume

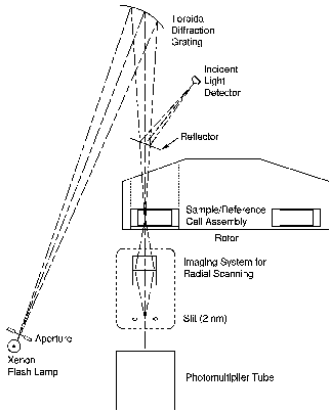


Analytical ultracentrifuge

Ultracentrifuge that posseses a detection system allowing the measure of the solute concentration as a function of the distance to the rotation axis (optical density, interferrométry).

Absorbance measurement: choice of the wavelenght:

- NA and protein 260 et 280 nm
- ligand eg 380
- peptidic bond 220: low concentra



Avantage: large choice in the experimental conditions: buffer and ionic strength

Sedimentation velocity

Principle: the protein migrates towards the bottom of the tube;
The speed of the particule is measured

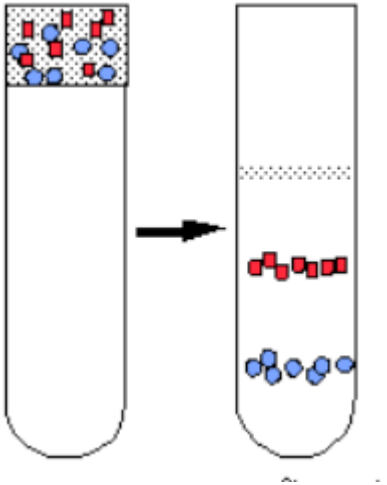
Determine sedimentation coefficient s (SVEDBERG) 1 S = 10^{-13} s

Speed of sedimentation v = dr/dt per acceleration unit

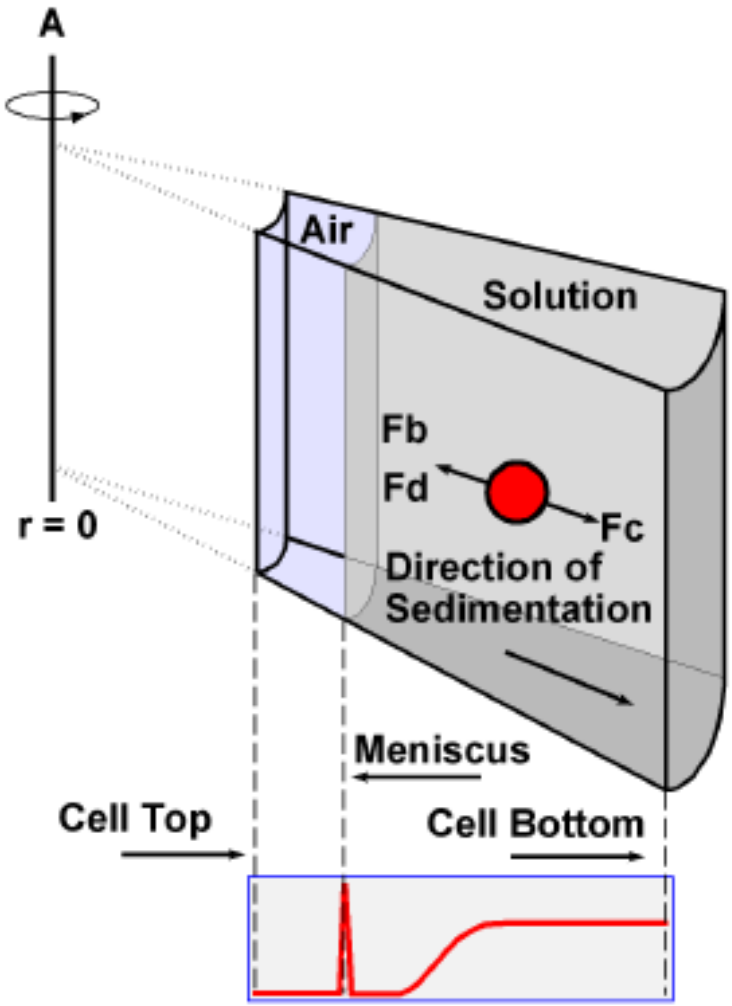
$s = dr/dt \cdot (1/w^2 r)$

$dr/dt = s \cdot w^2 r$

dr/dt = speed of the particule
r = distance of the particule to the rotation axis
w = angular speed of the rotor



$[MS^{-1}\ S^2M^{-1}]\ [S]$



Sedimentation:

Forces at Equilibrium:

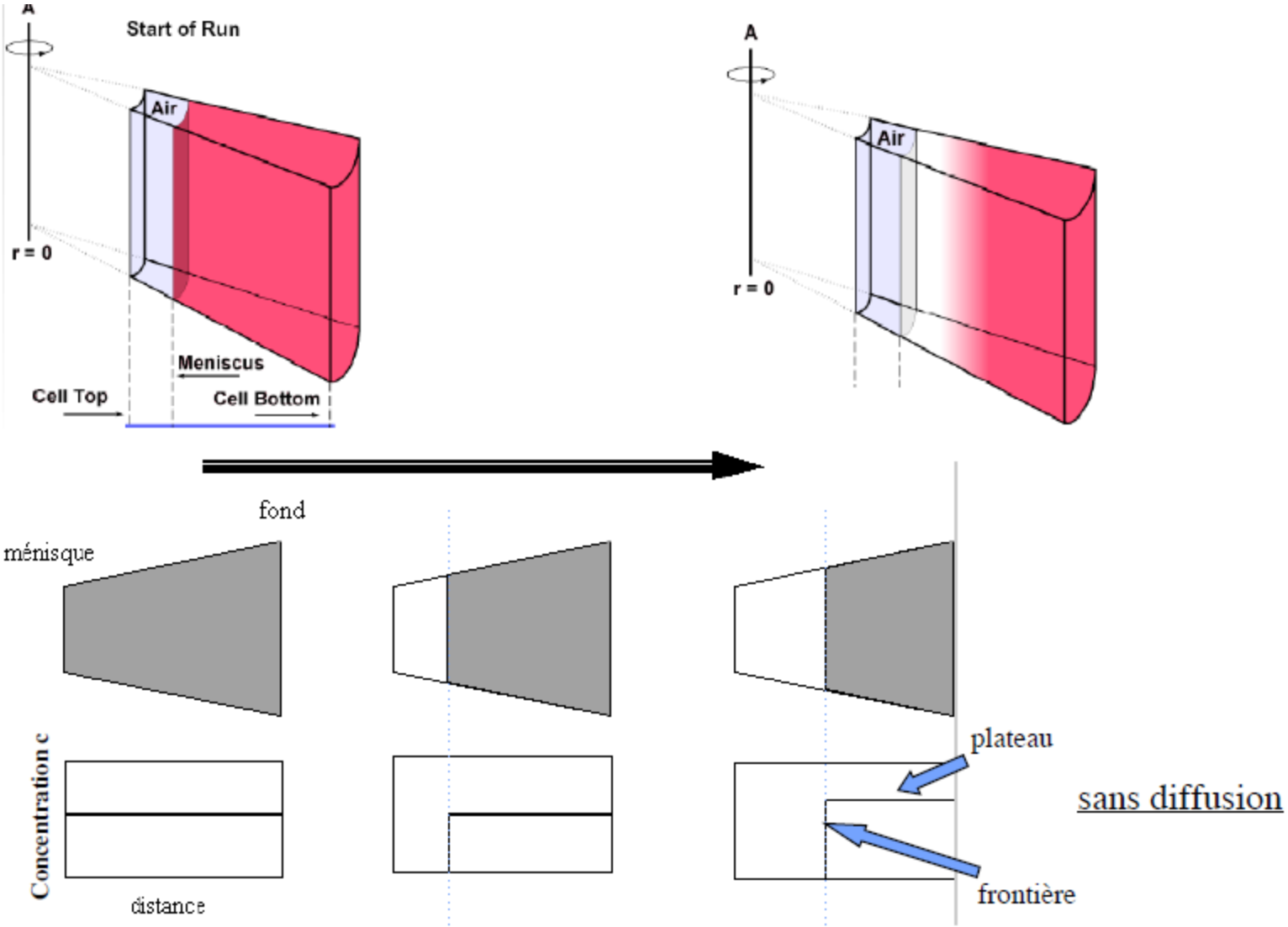
$Fc - Fb - Fd = 0$

Fb **(buoyancy)**	$= \omega^2 r m_0$
Fd **(viscous drag)**	$= fv$
Fc **(centrifugal force)**	$= \omega^2 r m$

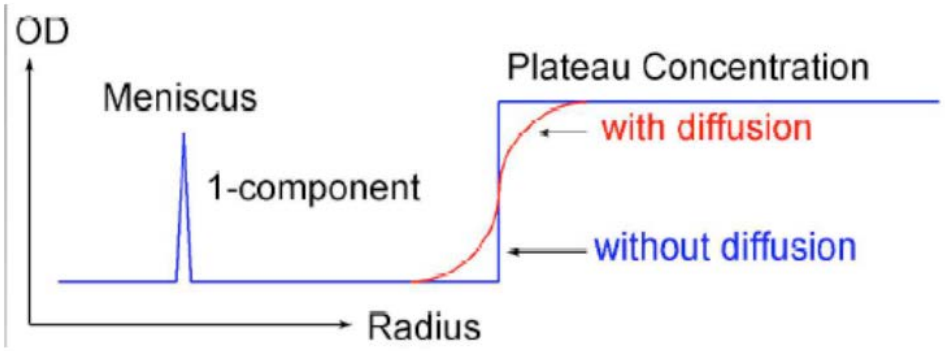
Explanation:

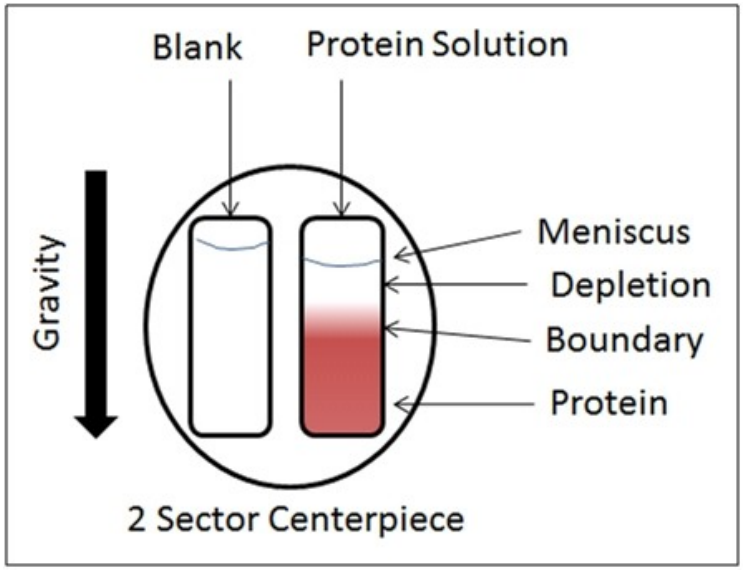
Fb is the buoyancy force - the force required to displace the buffer surrounding the solute, and m_0 is the mass of the displaced solvent.

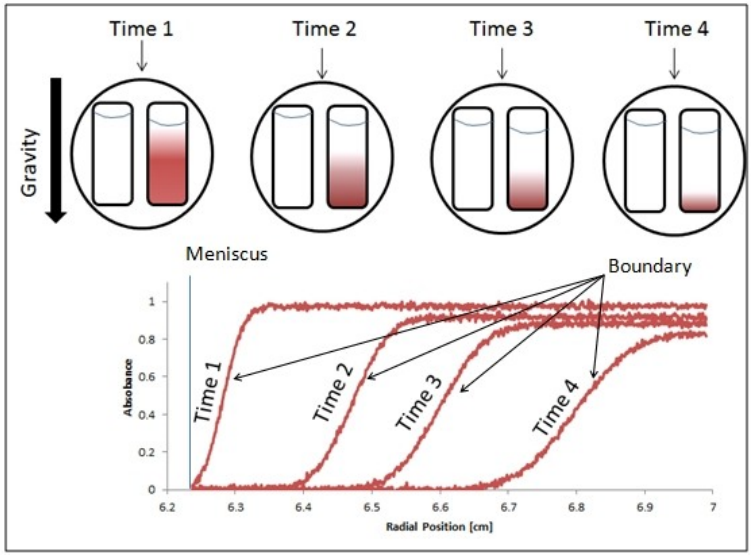
In practice, one can measure how the sedimentation boundary moves



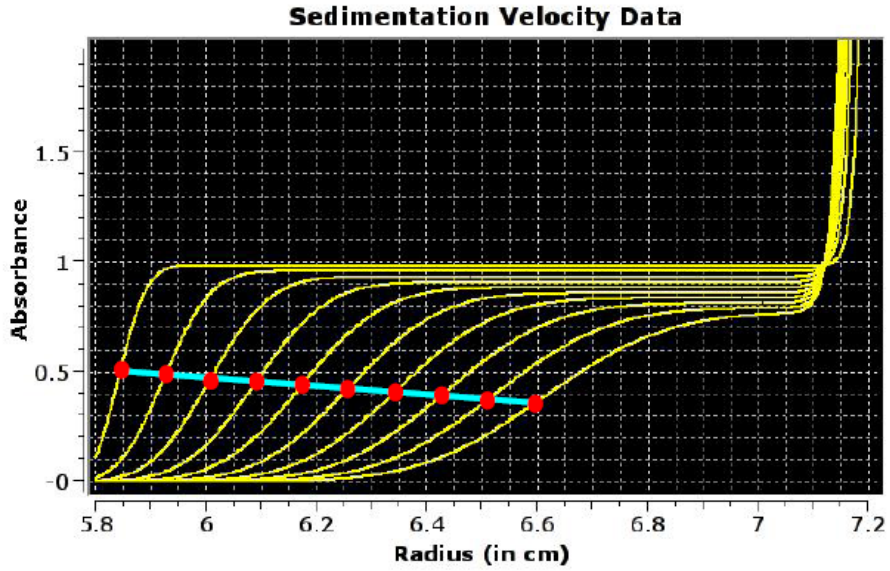
Diffusion impacts on the shape of sedimentation boundaries
They are recorded at regular intervals







How does the midpoint of the boundary move?



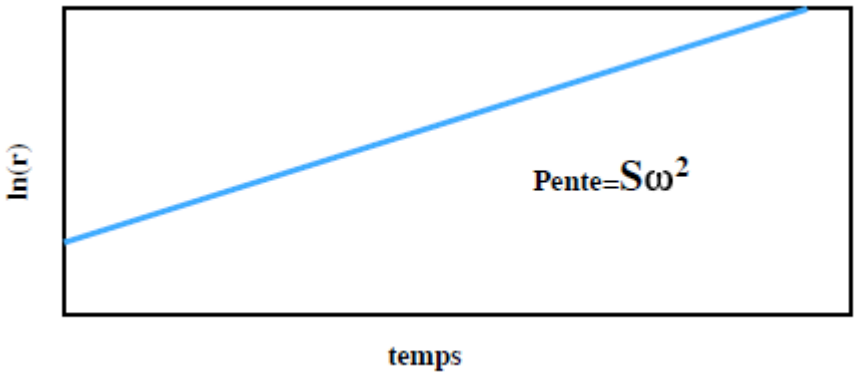
$s = dr/dt \cdot (1/\omega^2 r)$

$dr/dt = s \cdot \omega^2 r$

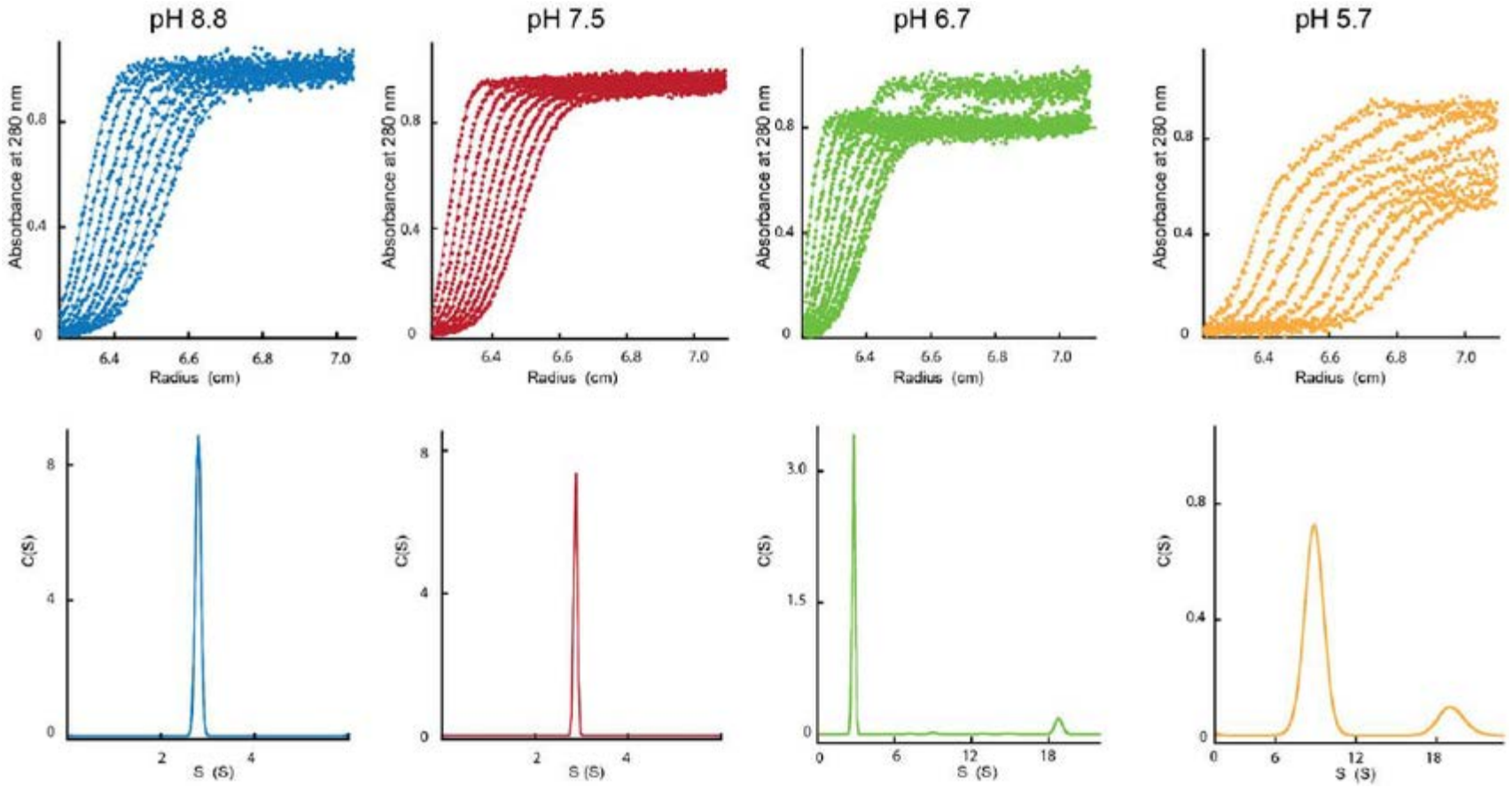
$$\frac{dr}{r} = \omega^2 s\,dt$$

$$\mathrm{Ln}\,(r(t)/r(t_\circ)) = \omega^2 s(t-t^\circ)$$

$$\frac{M\left(1-\bar{v}\rho\right)}{N f} = \frac{v}{\omega^2 r} = s$$



S depends on both M and f

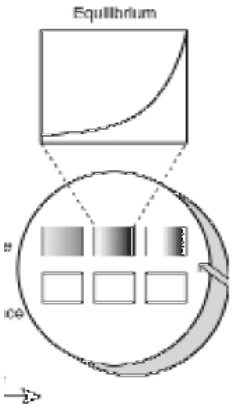


Characterization of Monomeric Intermediates during VSV Glycoprotein Structural Transition. Albertini et al. 2012 Plos Pathogens

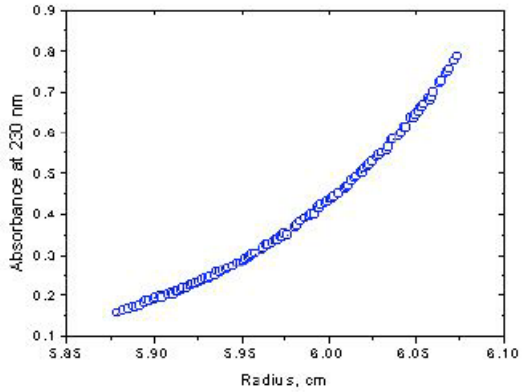
Equilibrium Sedimentation

At low speed, diffusion is not negligeable

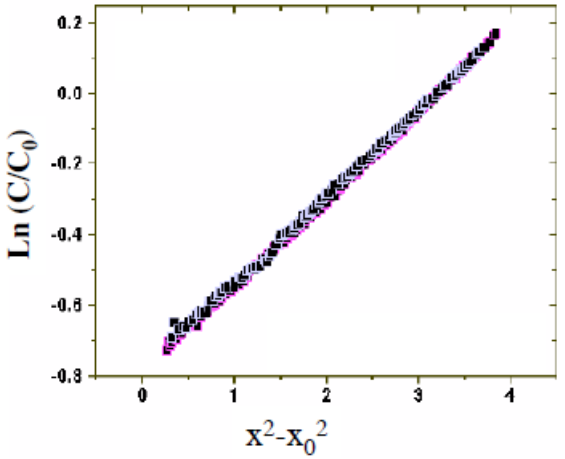
The system reaches an equilibrium:



centrigulation force = diffusion force



$$Ln(C_x/C_0)=M\omega^2(1-v\rho)(x^2-x_0^2)/2RT$$



C_x macromolecule concentration

x distance to the axis

$\underline{v}$ (cm^3/g) volume specific partial (hydrated) of the macromolecule

ρ volumic mass of the solvant (g/cm^3)

A real case

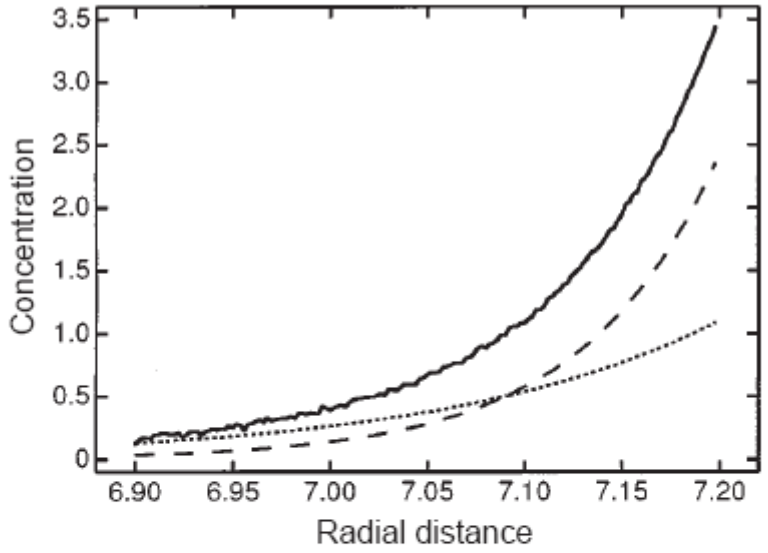


FIGURE 3

Sedimentation equilibrium data. Simulated data for a reversible monomer-dimer equilibrium: (—) total, (···) monomer, (---) dimer. The concentration distribution of the dimer is steeper than that of the monomer, and the relative amounts of monomer and dimer at each radial point are determined by mass-action equilibrium.

Mixture of noninteracting solutes

$$a(r) = \sum_n c_{n,0}\varepsilon_n d \exp\left[\frac{M_n(1-\bar{v}_n\rho)\omega^2}{2RT}(r^2 - r_0^2)\right] + \delta$$

Self-association

$$a(r) = \sum_n n\varepsilon_1 d K_n (c_{1,0})^n \exp\left[\frac{nM_1(1-\bar{v}\rho)\omega^2}{2RT}(r^2 - r_0^2)\right] + \delta \quad with\ K_1 = 1 \qquad (10)$$

Hetero-association

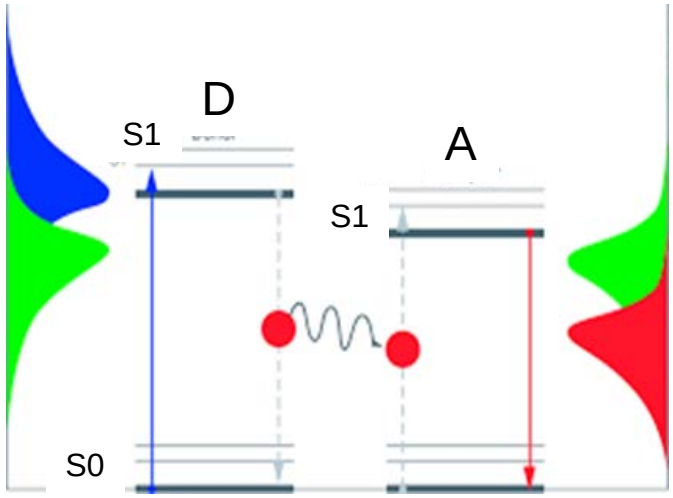
$$a(r) = c_{A,o}\varepsilon_A d \exp\left[\frac{M_A^*\omega^2}{2RT}(r^2 - r_o^2)\right] + c_{B,o}\varepsilon_B d \exp\left[\frac{M_B^*\omega^2}{2RT}(r^2 - r_o^2)\right] + c_{A,o}c_{B,o}K_{AB}(\varepsilon_A + \varepsilon_B)d \exp\left[\frac{(M_A^* + M_B^*)\omega^2}{2RT}(r^2 - r_o^2)\right] + \delta$$

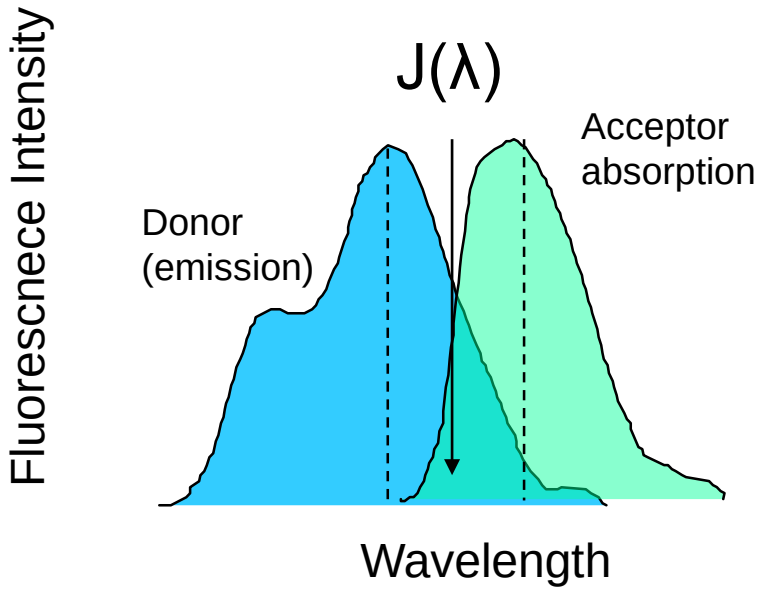
Förster Resonance Energy Transfer (FRET)

The mechanism of FRET involves a donor fluorophore (D) in an excited electronic state, which may transfer its excitation energy to a nearby acceptor chromophore (A)

Non-radiative process through long-range dipole-dipole interactions that results in the emission of light by the acceptor

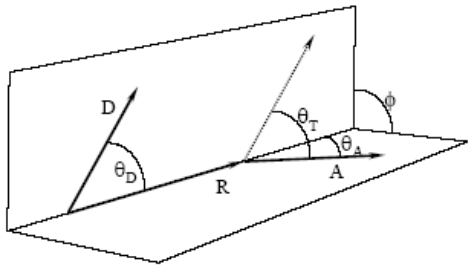
The absorption spectrum of the acceptor must overlap fluorescence emission spectrum of the donor





FRET strongly depends on:

- The relative orientation of the transition dipole moments of the Donor and the Acceptor
- The distance between the fluorophores

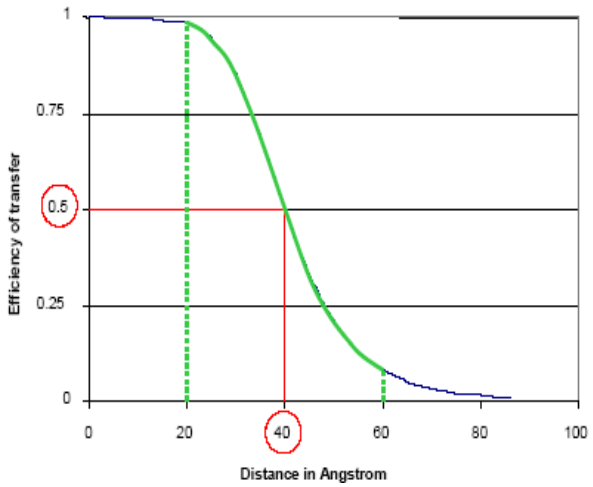


Energy transfer studies give information about

- distance between groups
- orientation of two groups and
- the refractive index of the donor-acceptor intervening medium

$$\kappa^2 = (\cos\theta_T - 3\cos\theta_D \cos\theta_A)^2$$

The efficiency of transfer varies with the inverse sixth power of the distance.



R_0 in this example was set to 40 Å.
When the E is 50%, $R=R_0$

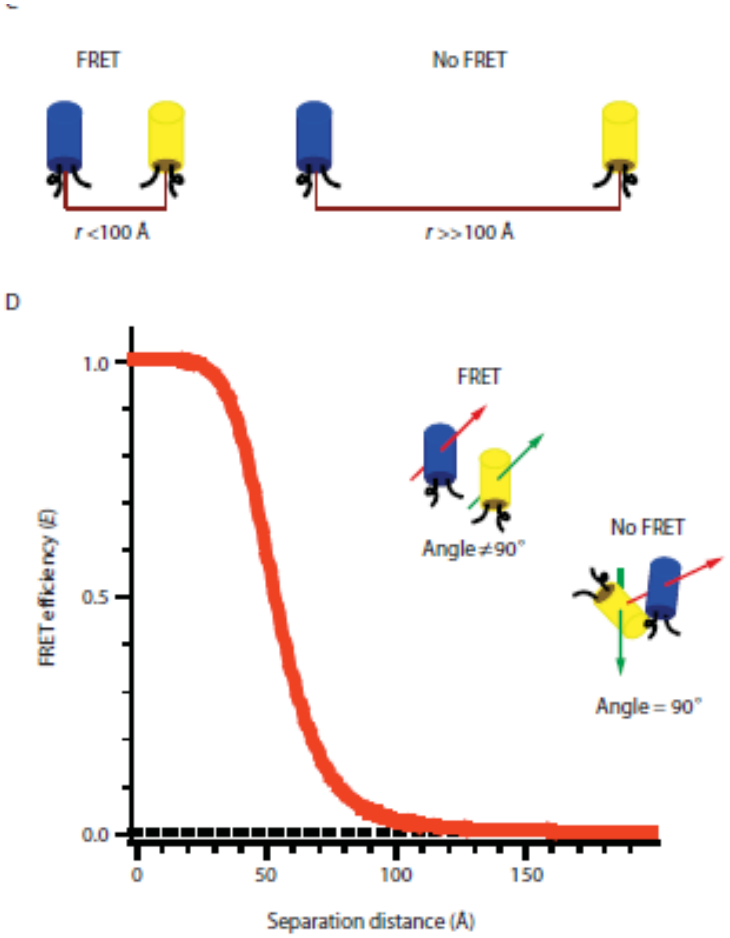
Distances can generally be measured between ~0.5 R_0 and ~1.5R_0

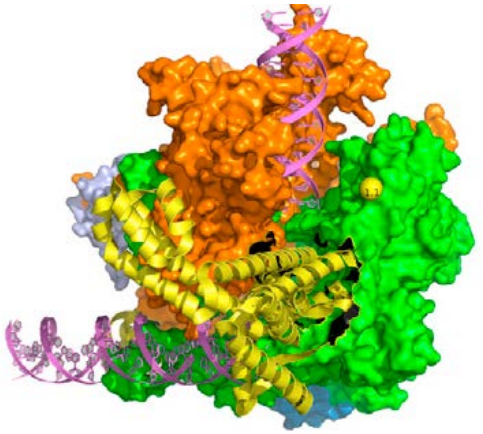
$$E = \frac{k_{DA}}{k_{DA} + k_{fD}} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}$$

In vivo and in vitro FRET analysis

Screening for compounds that inhibit or modulate A/B interactions

Use or fluorescent proteins fused to the proteins of interest or of fluorescent probes that are chemically coupled to the donnor and to the acceptor molecules





Structural organization of the bacterial (Thermus aquaticus) RNA polymerase-promoter open complex obtained by FRET (Mekler et al., 2002) was subsequently validated by a crystal structure (Zhang et al., 2012).

Fluorescence Anisotropy

Fluorescence is measured with a linearly polarized beam

Vertical emission

Horizontal emission

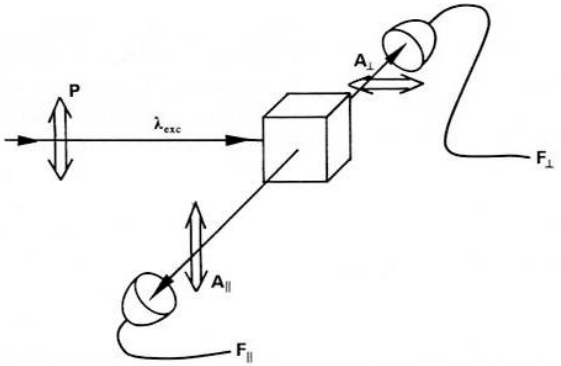
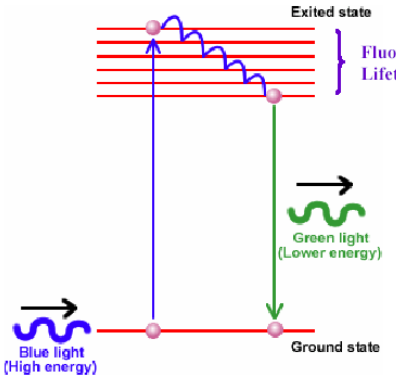


Figure 8.13. **Mesure de la polarisation de fluorescence. Le polariseur P polarise le faisceau incident dans le plan vertical. On observe les composantes parallèles et perpendiculaires du faisceau émis à travers deux analyseurs $A_\parallel$ et $A_\perp$.**



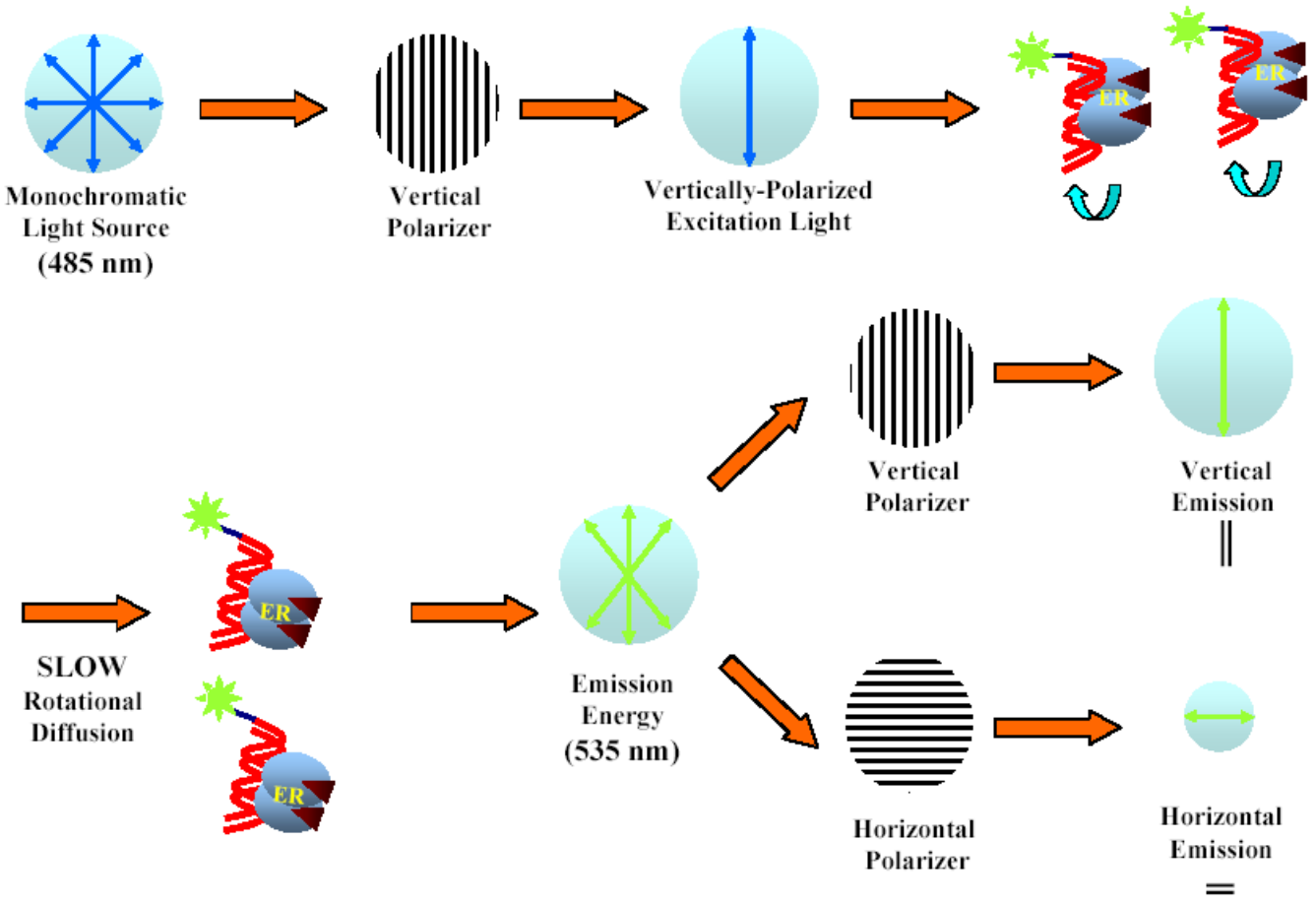
≈ 4 nanoseconds for Fluorescein

- Rotational diffusion before emission
- Diffusion dependent on size of molecule & viscosity of solvent & temperature

The basic idea is that a **fluorophore excited by polarized light** will also **emit polarized light**. However, if a molecule is moving, it will tend to "**scramble"** the polarization of the light by radiating at a different direction from the incident light.

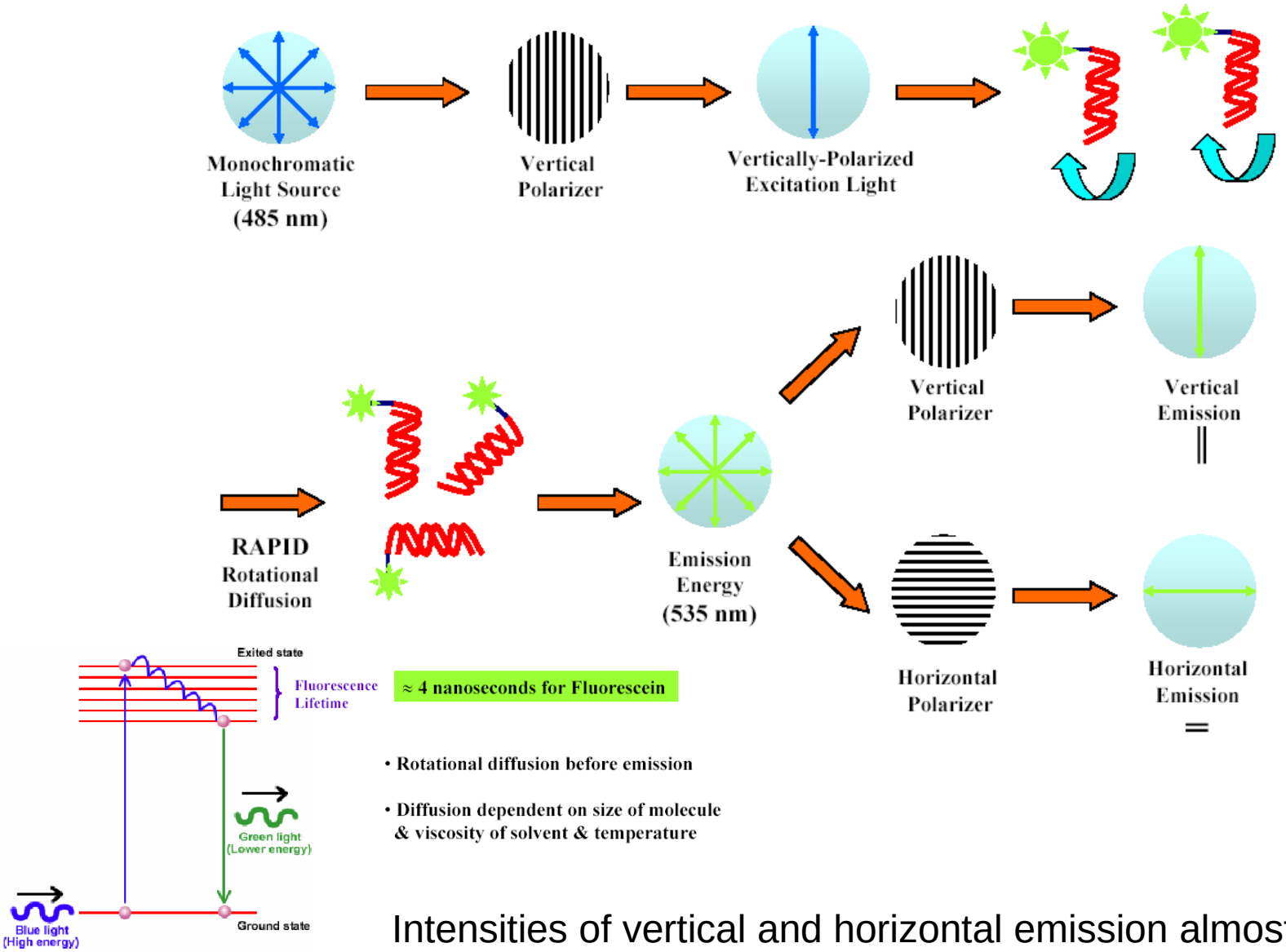
The "scrambling" effect is greatest with fluorophores tumbling in solution and **decreases** with **decreased rates of tumbling ie with increased MW**.

Slow rotational diffusion



Intensities of vertical and horizontal emission differ = Anisotropy

Rapid rotational diffusion



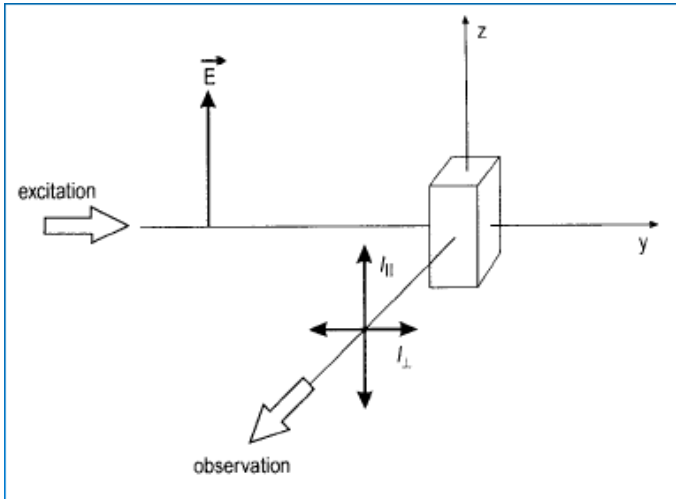
- Rotational diffusion before emission
- Diffusion dependent on size of molecule & viscosity of solvent & temperature

Intensities of vertical and horizontal emission almost equal

Fluorescence Polarization: Equations

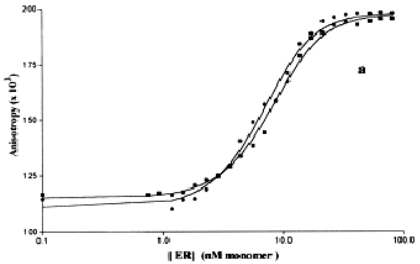
$$\text{Polarization Value (P)} = \frac{(\text{Intensity}_{\parallel}) - (\text{Intensity}_{=})}{(\text{Intensity}_{\parallel}) + (\text{Intensity}_{=})}$$

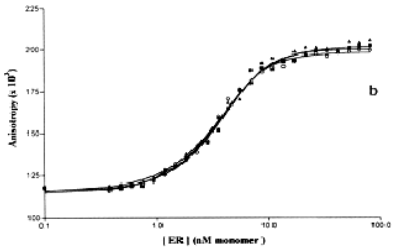
$$\text{Anisotropy Value (A)} = \frac{(\text{Intensity}_{\parallel}) - (\text{Intensity}_{=})}{(\text{Intensity}_{\parallel}) + 2(\text{Intensity}_{=})}$$



(High Rotational Diffusion) $0 < \text{Anisotropy} < 1$ (Low Rotational Diffusion)

• Boyer M, Poujol N, Margeat E, Royer CA. Nucleic Acids Res. 2000 Jul 1;28(13):2494-502.





$$A = \frac{A_0}{1 + \frac{\tau_F}{\tau_c}}$$

τ_F temps de vie de fluorescence

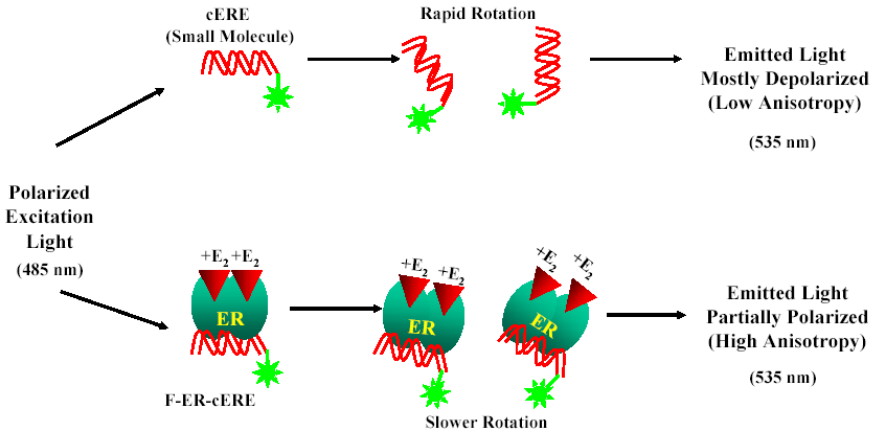
τ_c temps de correlation de rotation

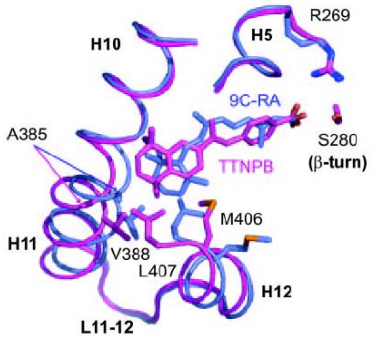
Titration experiments for Kd measurements

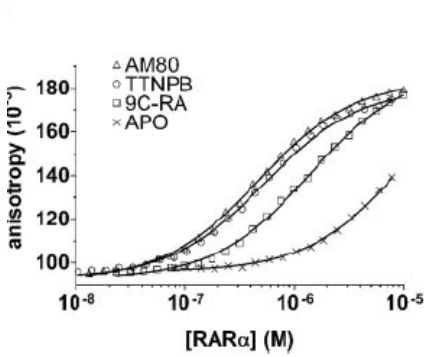
one molecule (DNA, peptide or protein) is fused to a fluorophore

measure static fluorescence anisotropy

association results in an entity which rotates more slowly than the fluorophore and results in less scrambling of polarized light





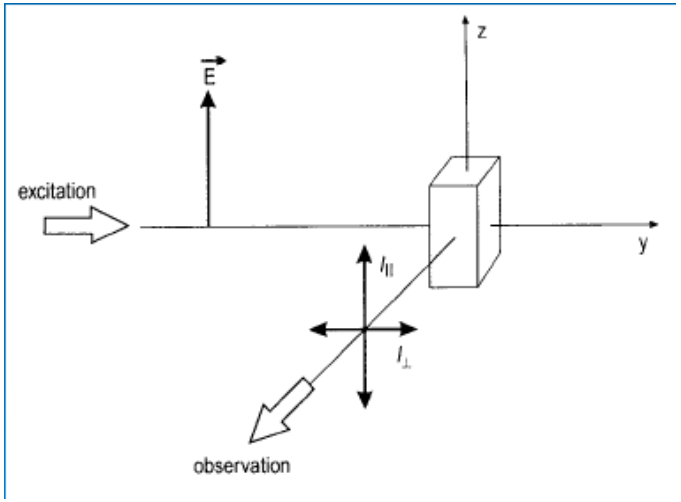


Boyer et al, 2009 NAR

Fluorescence Polarization: Equations

$$\text{Polarization Value (P)} = \frac{(\text{Intensity}_{\parallel}) - (\text{Intensity}_{=})}{(\text{Intensity}_{\parallel}) + (\text{Intensity}_{=})}$$

$$\text{Anisotropy Value (A)} = \frac{(\text{Intensity}_{\parallel}) - (\text{Intensity}_{=})}{(\text{Intensity}_{\parallel}) + 2(\text{Intensity}_{=})}$$



(High Rotational Diffusion) $0 < \text{Anisotropy} < 1$ (Low Rotational Diffusion)

Estimates of the apparent molecular volume (Perrin Equation)

$$< A >^{-1} = A_0^{-1}\left(1 + \frac{\tau_F}{\tau_r^{(2)}}\right) = A_0^{-1}\left(1 + \tau_F \frac{k_B T}{\eta V}\right)$$

$$\left(\frac{1}{P} - \frac{1}{3}\right) = \left(\frac{1}{P_0} - \frac{1}{3}\right)\left(1 + \frac{\tau_F}{\tau_r^{(2)}}\right) = \left(\frac{1}{P_0} - \frac{1}{3}\right)\left(1 + \tau_F \frac{k_B T}{\eta V}\right)$$

$$\tau_r^{(2)} = \frac{\eta V}{k_B T} \qquad \tau_r^{(2)} = \frac{1}{6 D_r} = \frac{f_r}{6 k_B T}.$$

τ_F Fluorescence life time

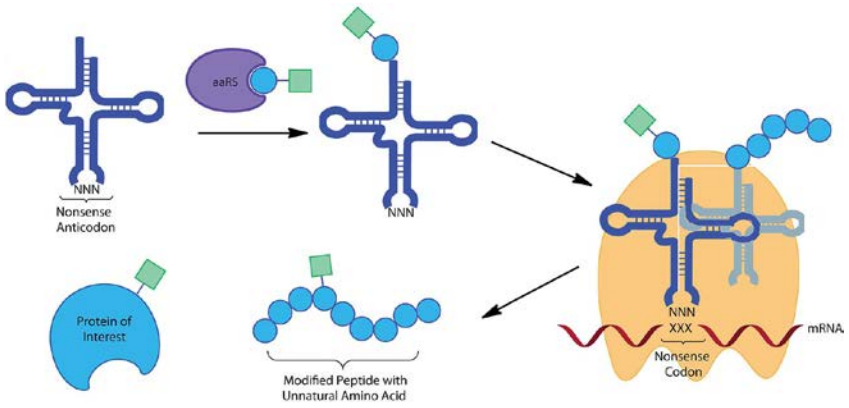
τ_c Rotation correlation time

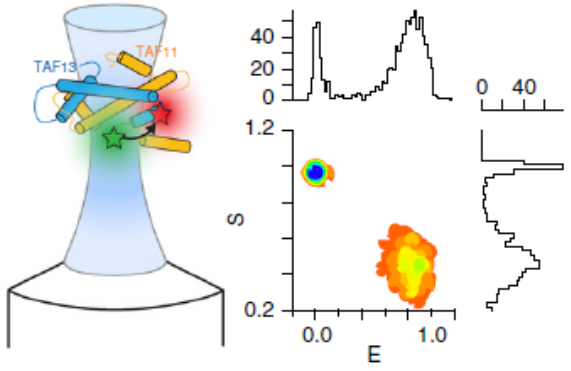
Fluorescence labelling

- Chemical synthesis of labelled peptide or oligonucleotides

- Genetically encoded tags (EGFP, mCherry,…)

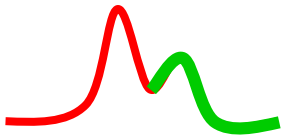
- Cys-maleimide chemistry,
- Incorporation of non-natural reactive amino acids
- Protein tags such as Halo(haloalkane dehalogenase), SNAP/CLIP(O6-alkylguanine-DNA alkyltransferase),
- Peptidic tags Avi(biotin ligase recognition peptide), Sfp phosphopantetheinyl transferase(CoA), Sortase,….





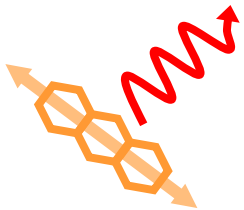
Fluorescence properties that can be measured

- spectra (environmental effects)

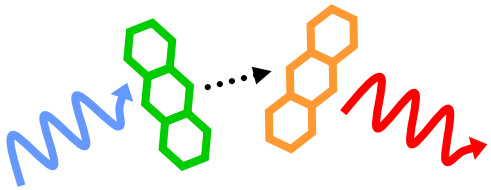


- fluorescence life times

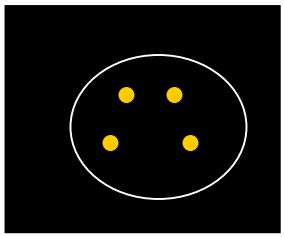
- polarization (orientation and dynamics)



- excitation transfer (distances -> dynamics)



- location of fluorescence



Widely used for binding analysis

Super resolution microscopy

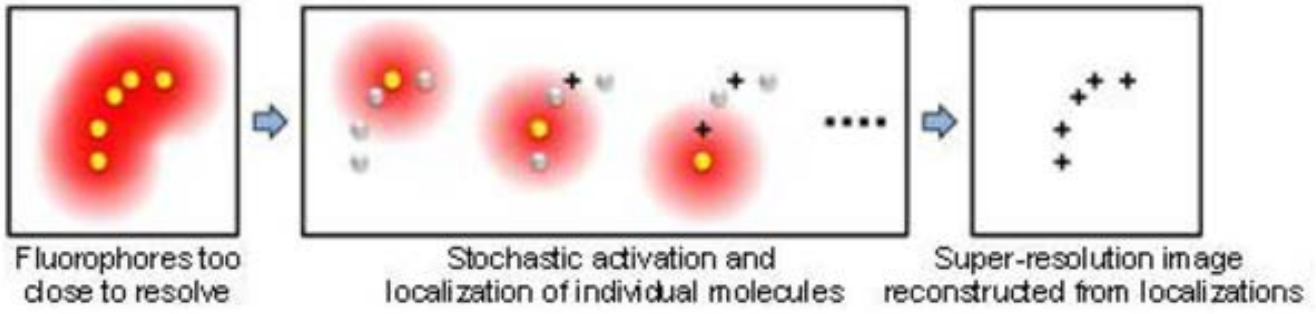
Nobel Price 2014 (Eric Betzig, Stefan W. Hell and William E. Moerner "for the development of super-resolved fluorescence microscopy".)

Different fluorescent probes marking the sample structure are activated at different time points, allowing subsets of fluorophores to be imaged without spatial overlap and to be localized to high precision.

Details cannot be separated by focusing light (diffraction limited)

But by using 2 molecular states of the fluorophore to limit the spacial overlap between airy circles

$$d = \frac{\lambda}{2n\sin\alpha\sqrt{1+I/I_s}} \longrightarrow 0$$



Super resolution microscopy

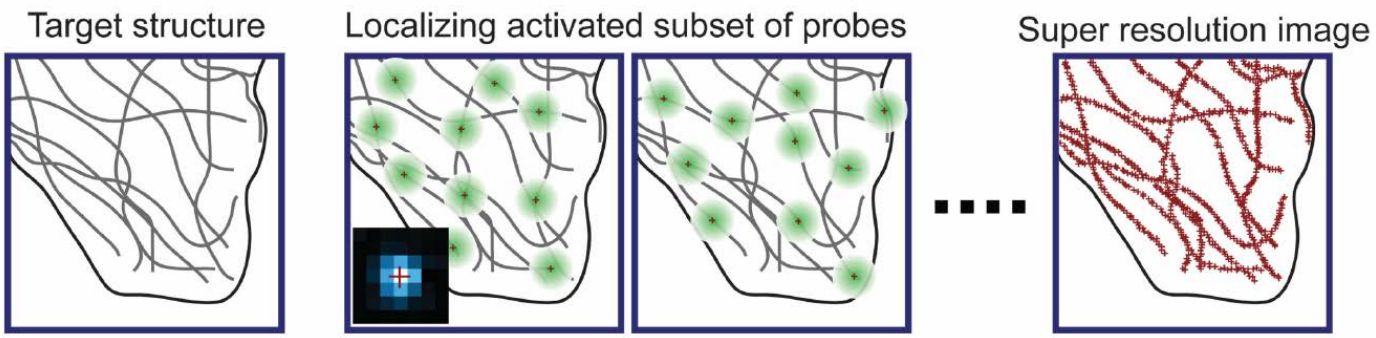
STED (STimulated Emission Depletion microscopy,

PALM (PhotoActivated Localization Microscopy),

STORM (stochastic optical reconstruction microscopy,

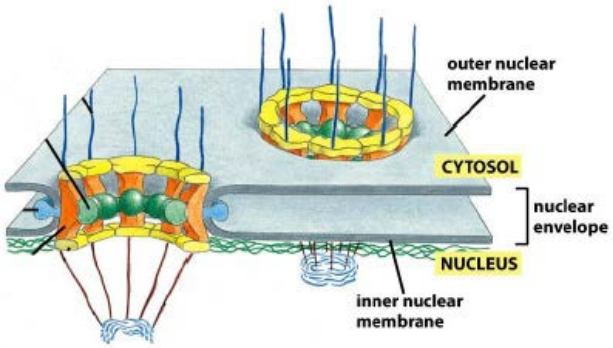
(FPALM) fluorescence photoactivation localization microscopy,

GSDim (Ground State Depletion imaging followed by Individual Molecule return).

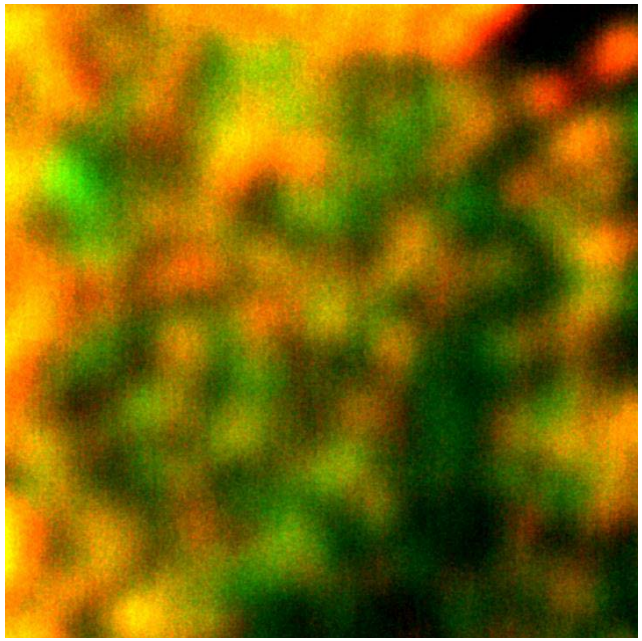


Observe individual proteins with a resolution down to 20 nm in intact cells, and second-order statistics to study the spatial interactions of the proteins.

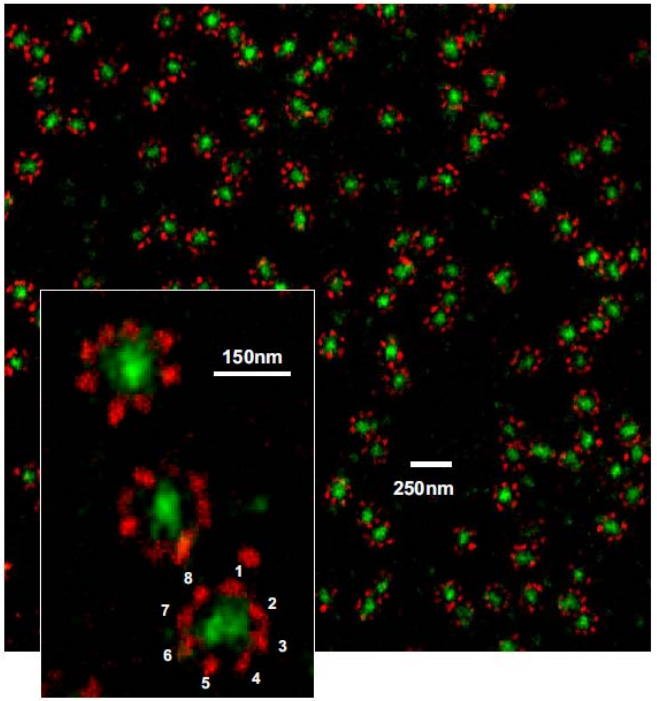
Huang, 2009



Standard confocal



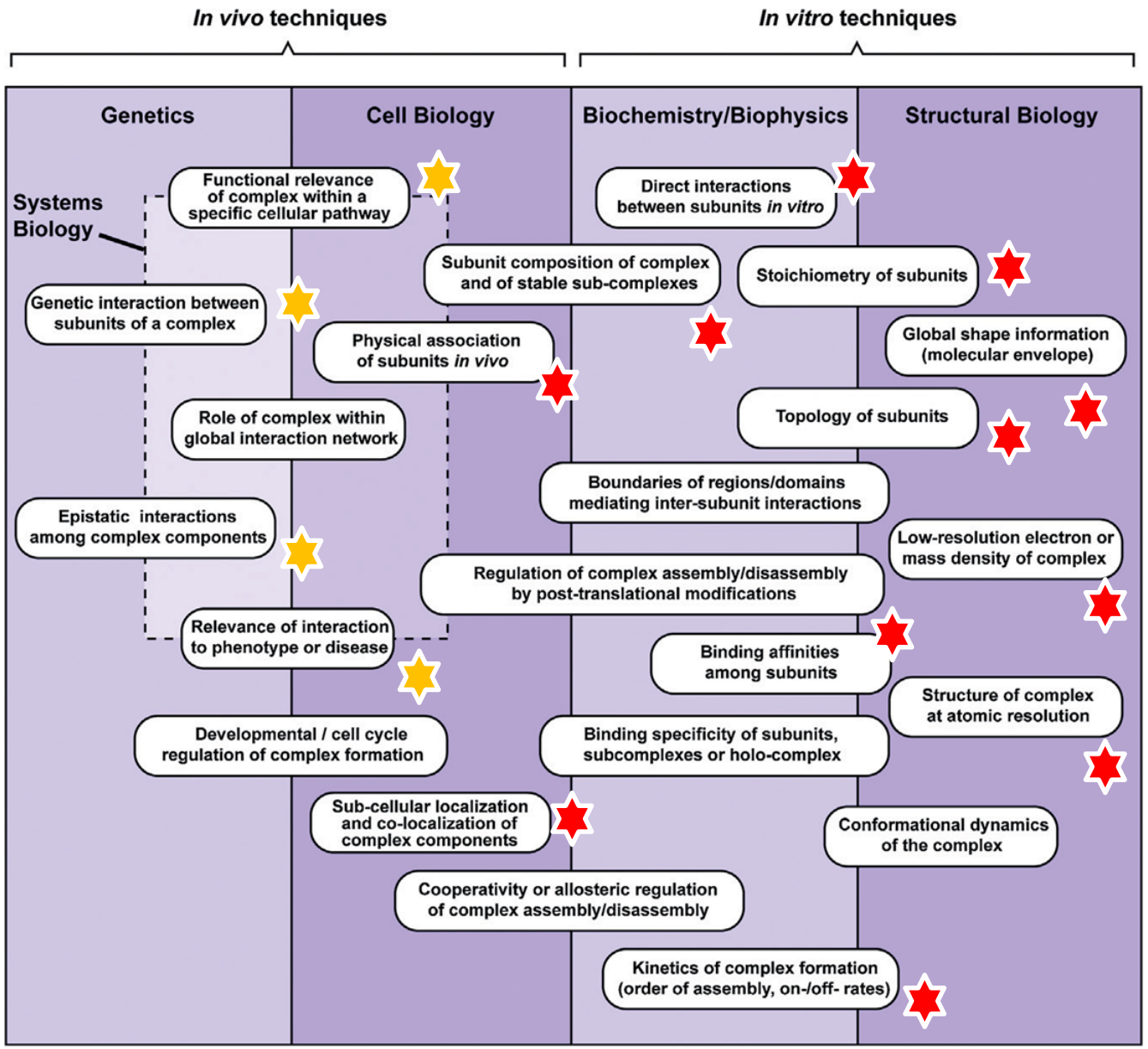
STED



Göttfert, Wurm *et al* Biophys J (2013)

Structural data used in integrative modeling

Composition	Purification from source with gel analysis or MS
Atomic structures of parts of the system	X-ray and neutron crystallography, NMR, Cryo-EM/ET, Comparative modelling and molecular docking
3D maps, 2D images, components positions	Electron microscopy and tomography, gold labelling, Super resolution microscopy, FRET imaging
Atomic and protein distances	NMR, FRET, EPR, X-link/ w/o MS……
Binding site mapping	NMR, FRET, H-D/MS, mutagenesis
Size and shape	AUC, SAS, atomic force microscopy, ion mobility Fluorescence correlation spectroscopy or anisotropy
Physical proximity	Co-purification, native MS, genetic methods, sequence convariance, Chromosome conformation Capture and other data, Y2H
Solvent accessiblity	Footprinting methods including H-DX/MS, NMR and chemical modifications



Perrakis et al., 2011

November 19th to 23rd 2018 - IGBMC, Strasbourg-Illkirch, France

Preparation and characterization of macromolecular complexes

instruct ERIC

INSTRUCT-FRISBI course

FRISBI Integrated Structural Biology

Hand-on workshop:

1/ Purification of complexes from engineered cell lines

2/ Production of recombinant multiprotein complexes using baculoviruses

3/ Biophysical characterization and sample optimization

Participants will have the opportunity to work with their own samples .

END