Other methods and integrative approaches

Arnaud.Poterszman@igbmc.fr

The challenge of myriads of complexes

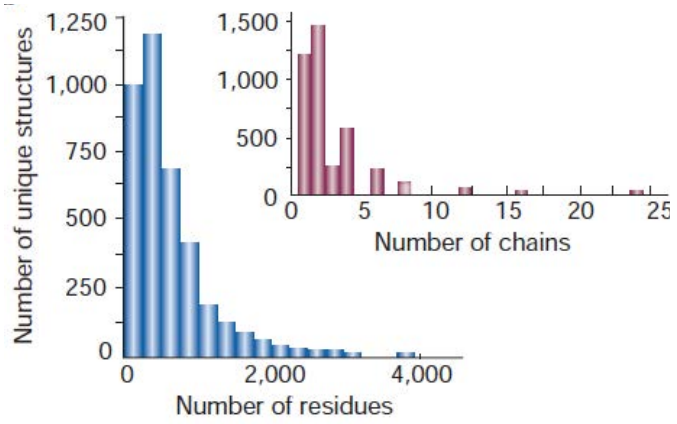
Protein act as complexes. From analysis of ~ 6200 yeast proteins

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



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 evelopes 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)

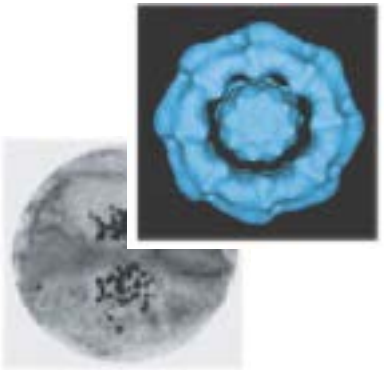
System complexity

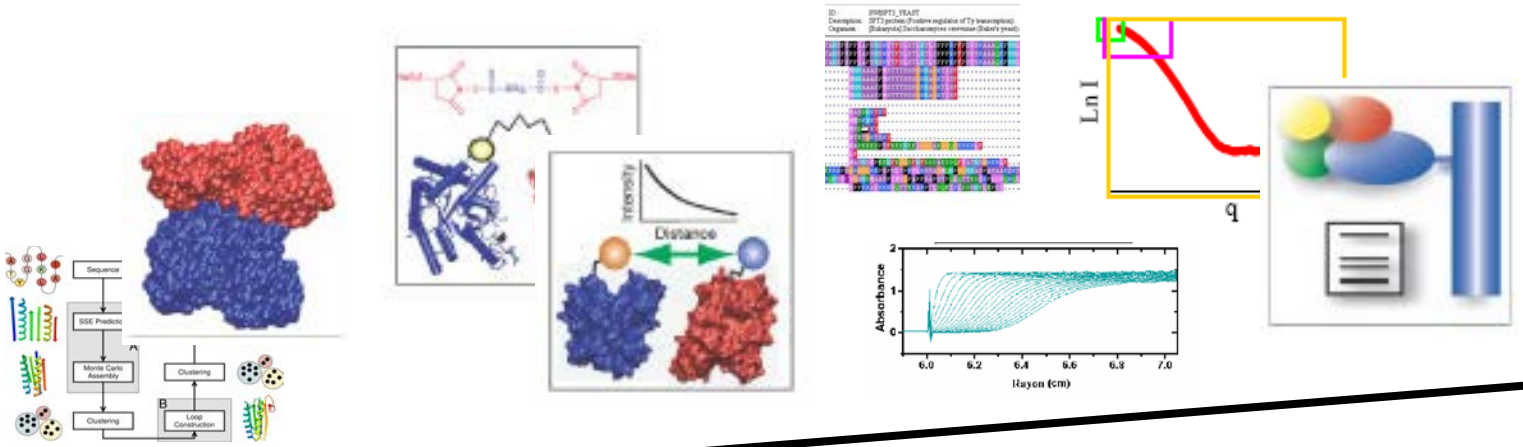
Integrative determination of macromolecular structures

Use structural information from any source

Measurements, physical principles and statistical inferences
Resolution: low or high resolution

to obtain a set of models consistent available data





Atomic positions | Residue positions | Member orientations | Member positions

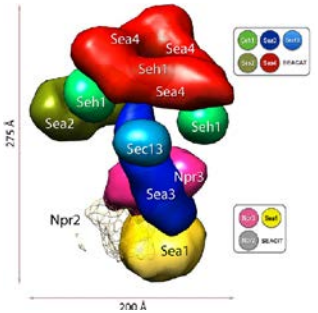
Integrative determination of macromolecular structures

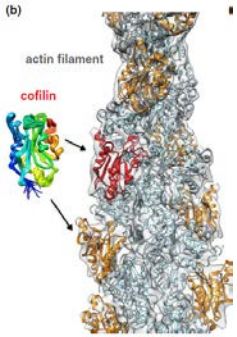
Modeling from experimental structures, comparative modeling and distance constraints

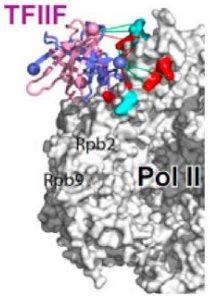
Modeling configurations from connectivity information (native, X-link-MS..)

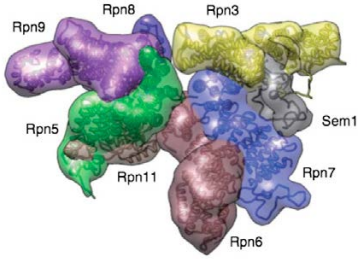
Propose multi-scale models

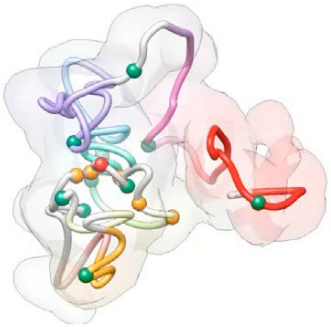
Modeling genomic region from 3C data











Atomic positions

Residue positions

Member orientations

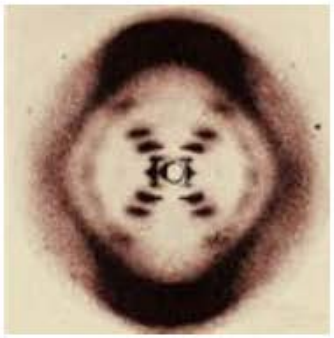
Member positions

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

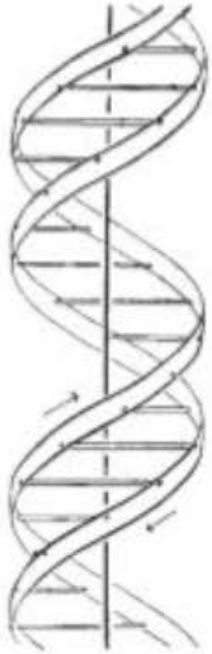
A Structure for Deoxyribose Nucleic Acid

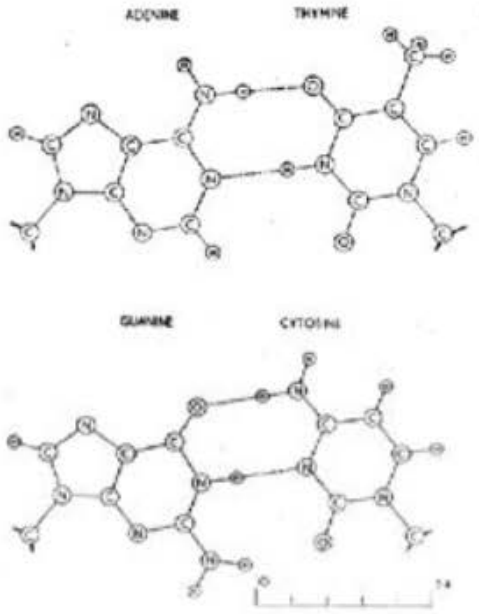
J. D. WATSON
F. H. C. CRICK

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.

Other methods and integrative approaches

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

Introduce integrative approaches which allow to combine heterogenous data and propose hybrid models to provide the best possible description of the system.

Summarize mainstream complementary experimental and in silico methods to provide structural information on a macromolecular complex and discuss their pros/cons

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.

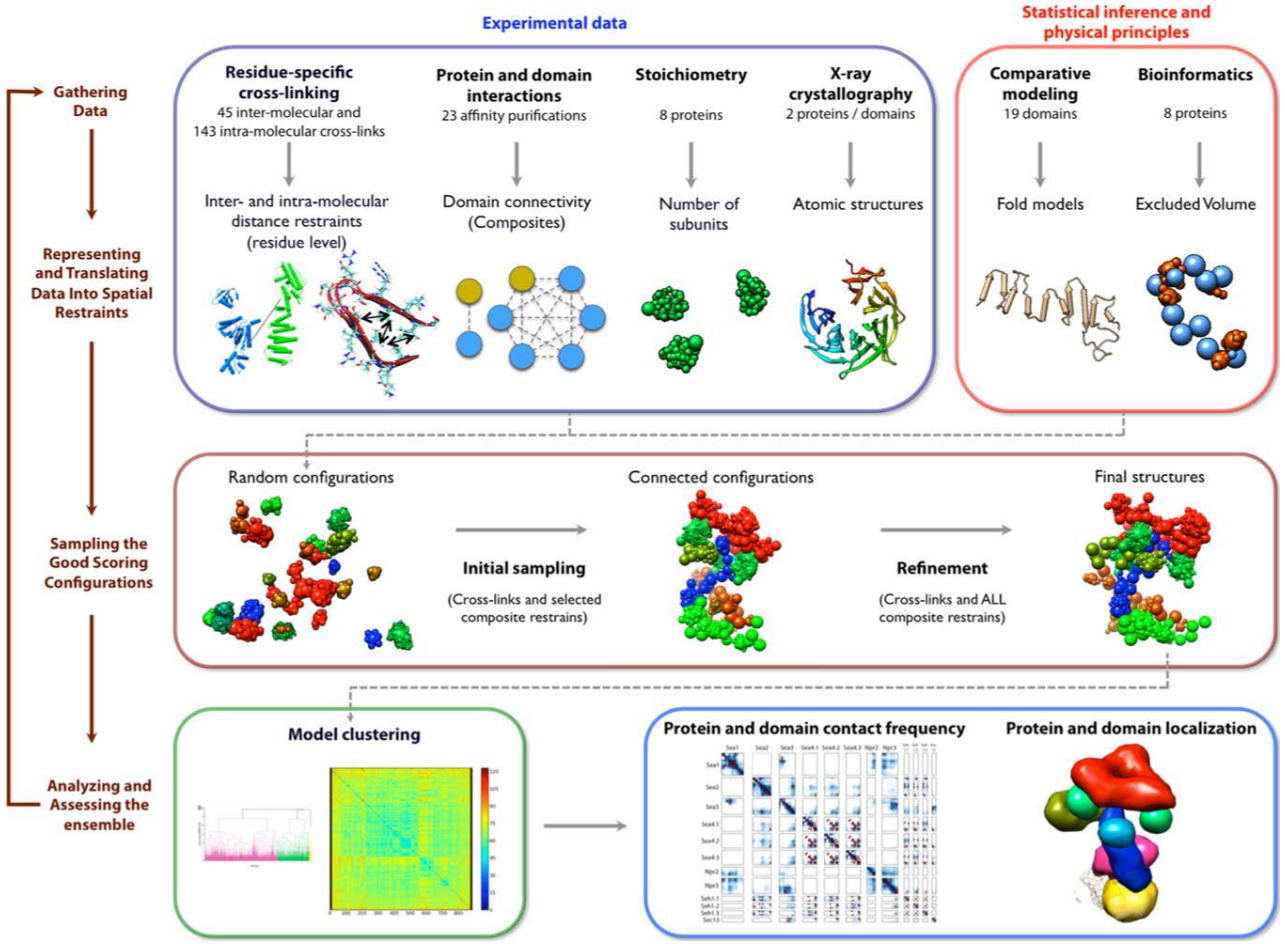
Inferential Structure Determination (ISD) framework
Rieping W, Nilges M, Habeck M. ISD: a software package for Bayesian NMR structure calculation. Bioinformatics 2008;24:1104–5.

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



A Sali

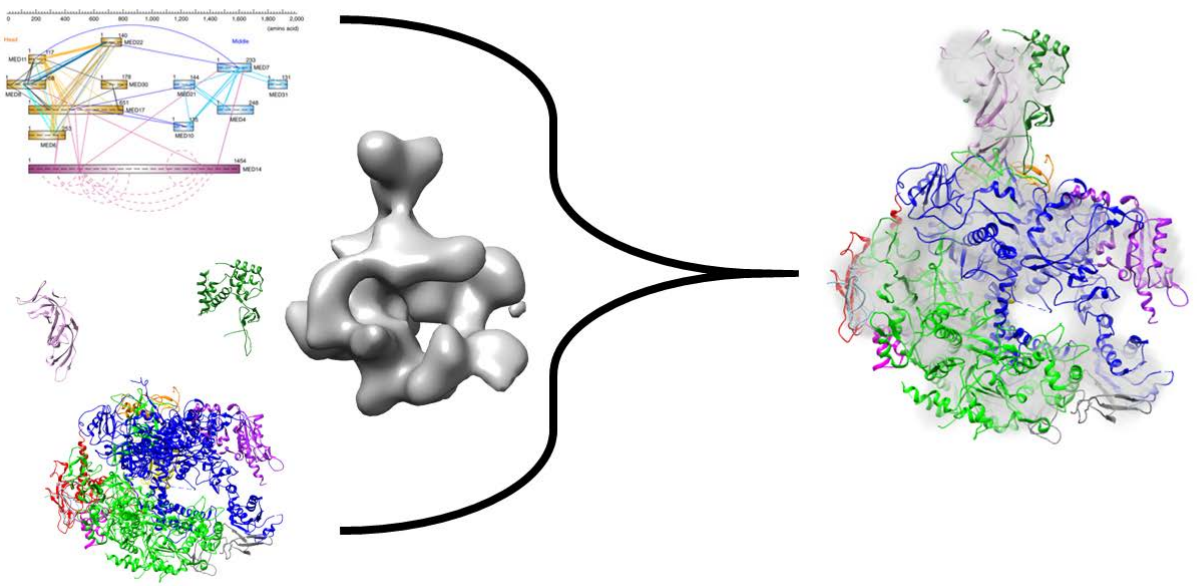
Integrative structure modeling of RNA Polymerase II stalk

1/ Gathering data

2/ Representing and translating data into restraints

3/ Sampling good scoring configurations

4/ Analysis and assesment



https://integrativemodeling.org/

Integrative structure modeling of RNA Polymerase II stalk

1/ Gathering data

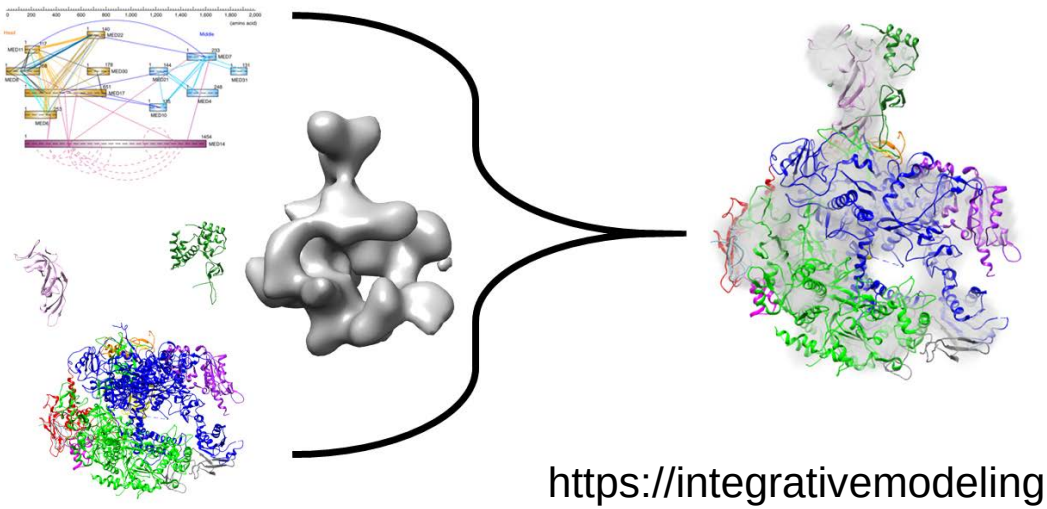
2/ Representing and translating data into restraints

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

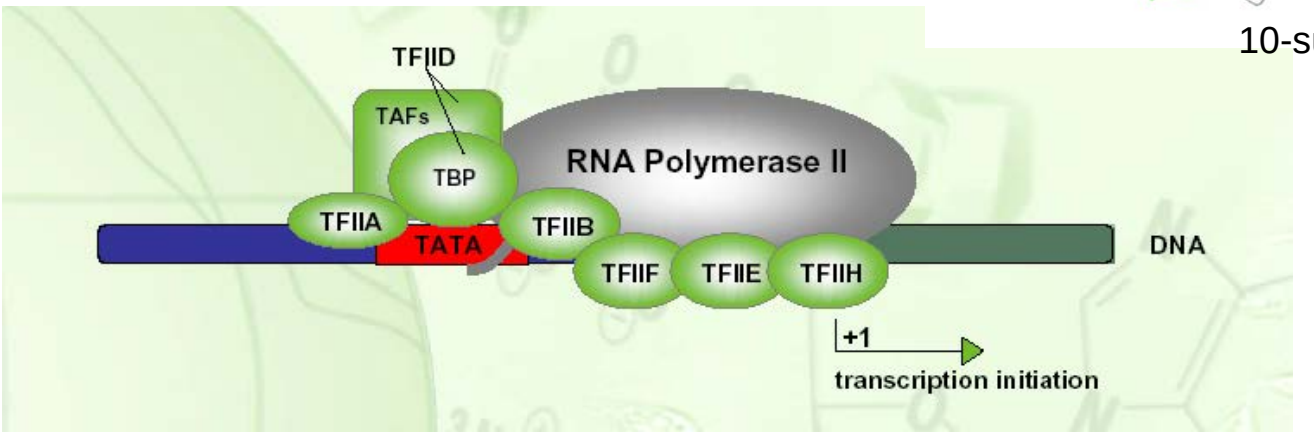
RNA Pol II is a eukaryotic complex that catalyzes DNA transcription to synthesize mRNA strands Eukaryotic RNA polymerase II contains 12 subunits, Rpb1 to Rpb12

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 of RNA Polymerase II stalk

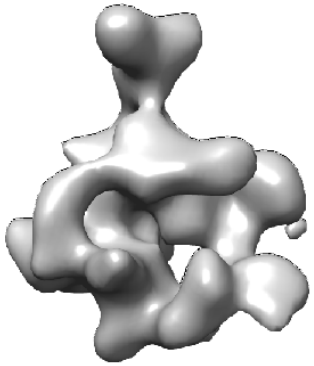
Gathering data

Represent and translate data into restraints

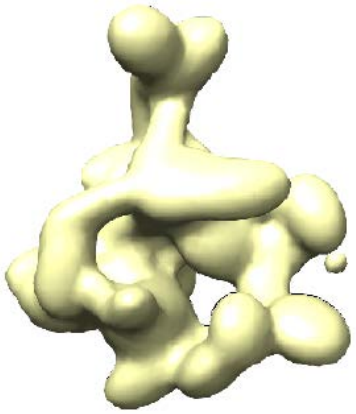
Sampling good scoring configurations

Analysis and assesment

Experimental map of entire complex at 20.9Å resolution (represented with Gaussan mixture models (GMMs))



Experimental map



50 Gaussian mixture model

https://integrativemodeling.org/

Integrative structure modeling of RNA Polymerase II stalk

Gathering data

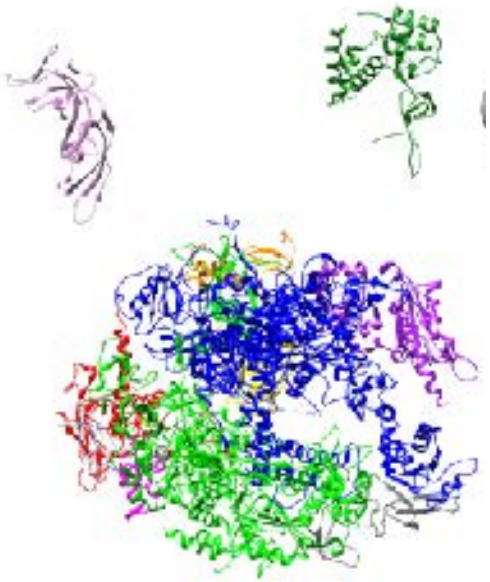
Represent and translate data into restraints

Sampling good scoring configurations

Analysis and assesment

Experimental map of entire complex at 20.9Å resolution (represented with Gaussan mixture models (GMMs))

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



https://integrativemodeling.org/

Integrative structure modeling of RNA Polymerase II stalk

Gathering data

Represent and translate data into restraints

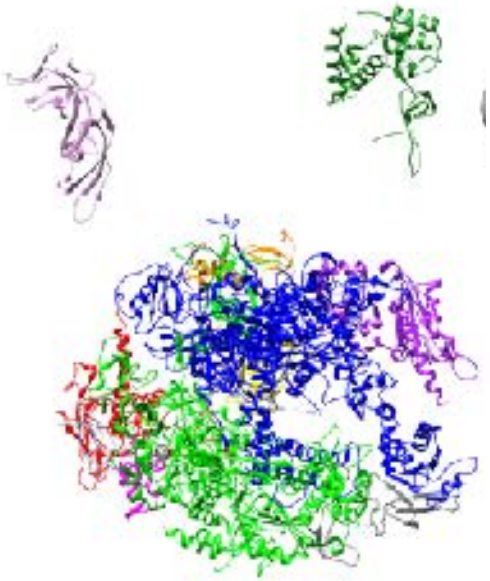
Sampling good scoring configurations

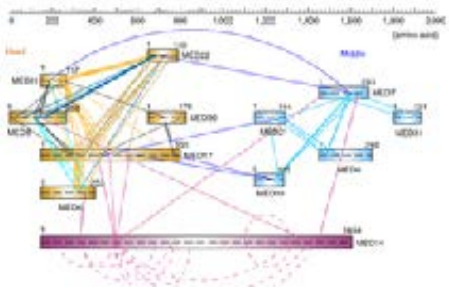
Analysis and assesment

Experimental map of entire complex at 20.9Å resolution (represented with Gaussan mixture models (GMMs))

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)





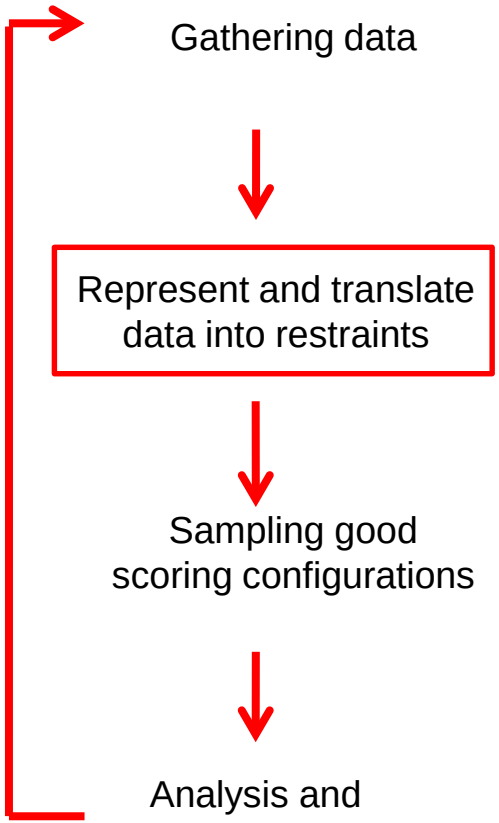
https://integrativemodeling.org/

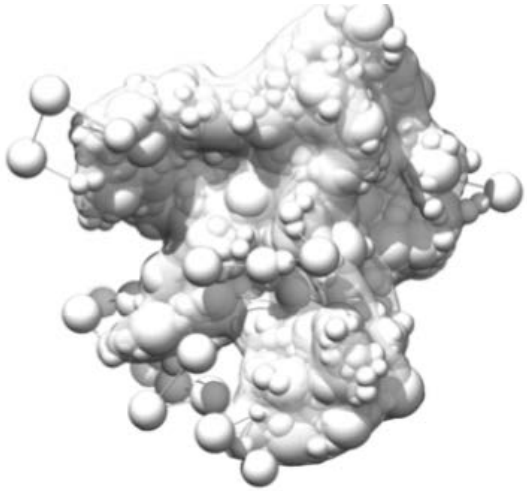
Integrative structure modeling of RNA Polymerase II stalk

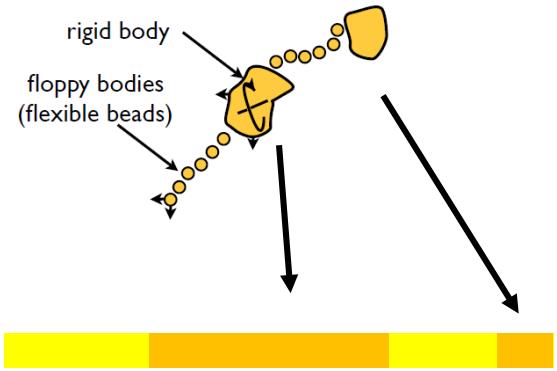
Experimental map of entire complex at 20.9Å resolution (represented with Gaussan mixture models (GMMs))

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

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

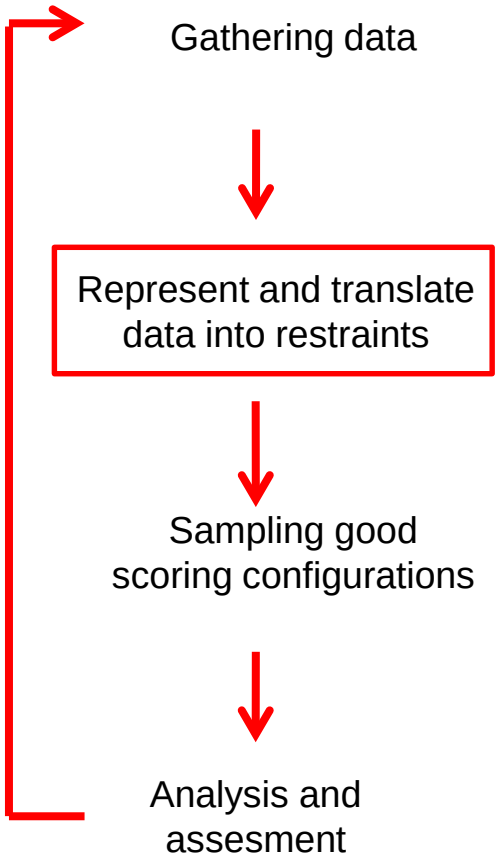






https://integrativemodeling.org/

Integrative structure modeling of RNA Polymerase II stalk



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

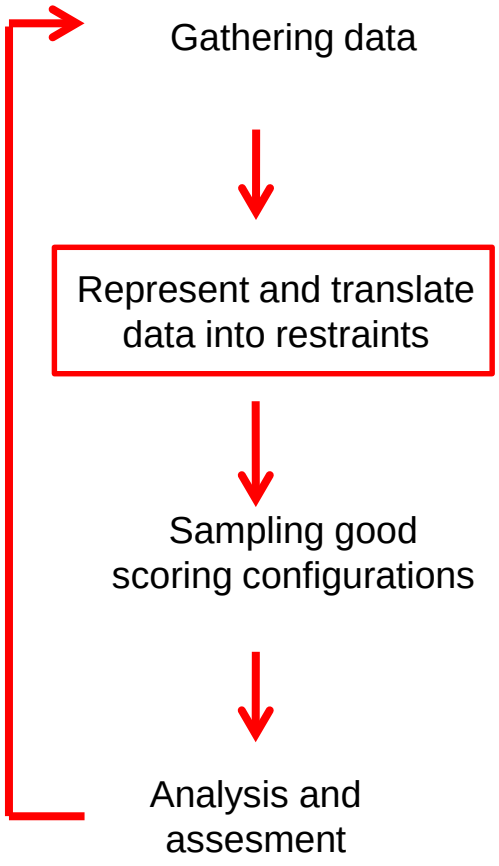
Sequence connectivity restraint: residues that are adjacent in sequence will also be close in space due to the peptide bond

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 of RNA Polymerase II stalk



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

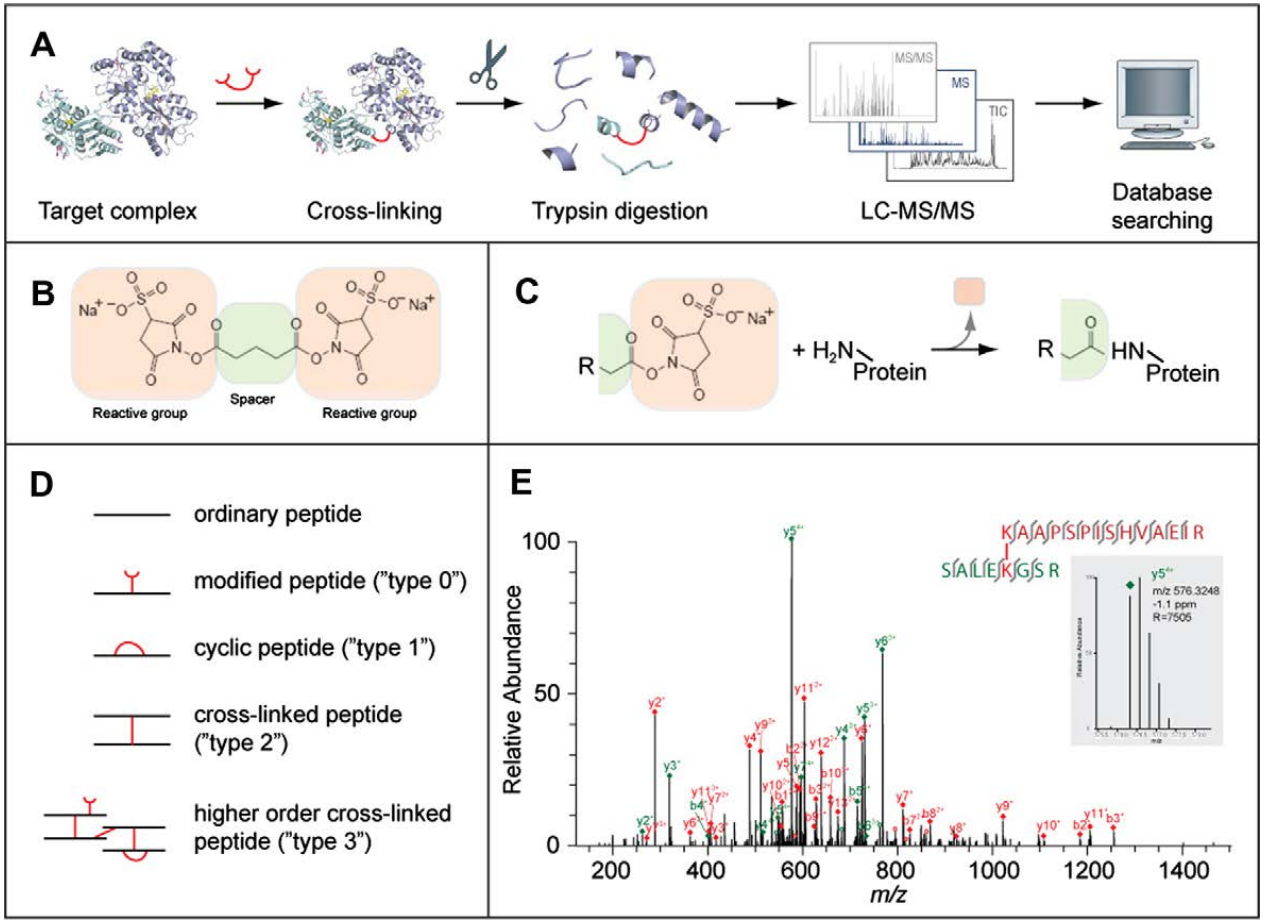
Sequence connectivity restraint: residues that are adjacent in sequence will also be close in space due to the peptide bond

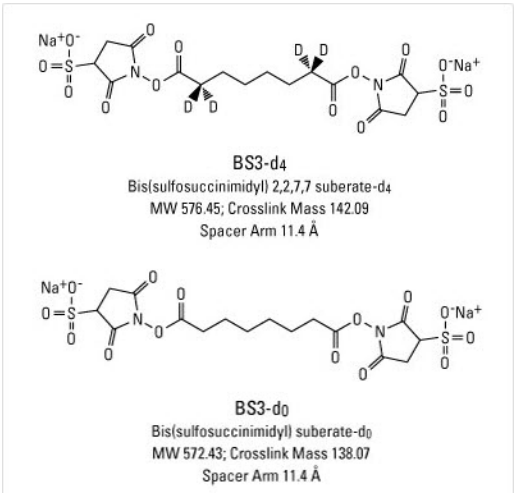
Excluded volume restraint: one protein cannot occupy the same space as another

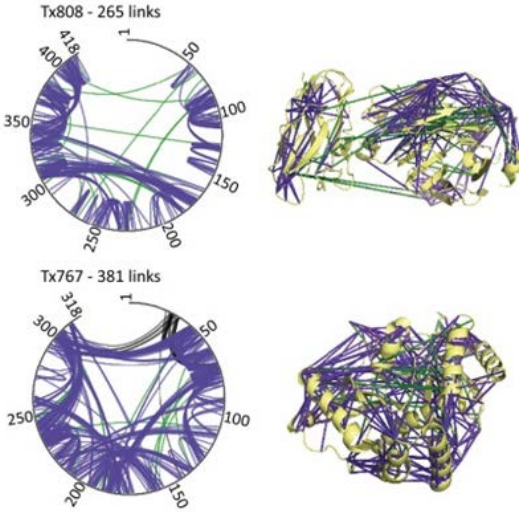
EM restraints: A density overlap function to compare the GMM approximation of our model (em_components) with that of the EM map (target_gmm_file)

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

X-link/MS experiments

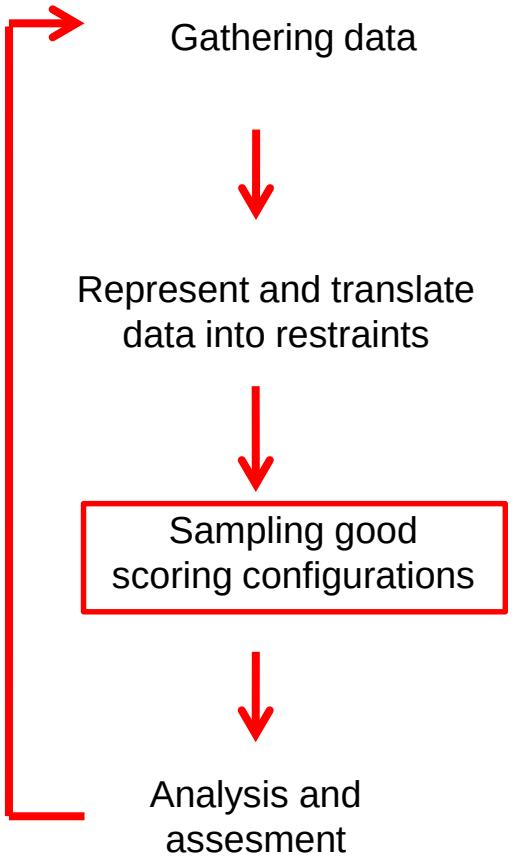






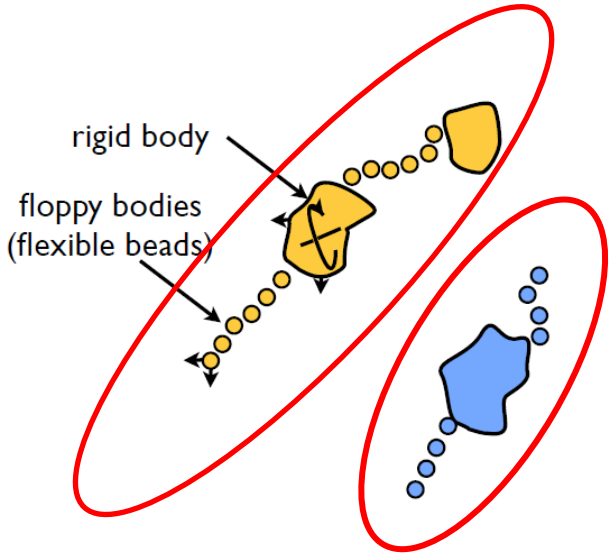
Juri Rappsilber, 2010, 2012

Integrative structure modeling of RNA Polymerase II stalk



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.



Srb (super rigid body)

https://integrativemodeling.org/

Integrative structure modeling of RNA Polymerase II stalk

Gathering data

↓

Represent and translate data into restraints

↓

Sampling good scoring configurations

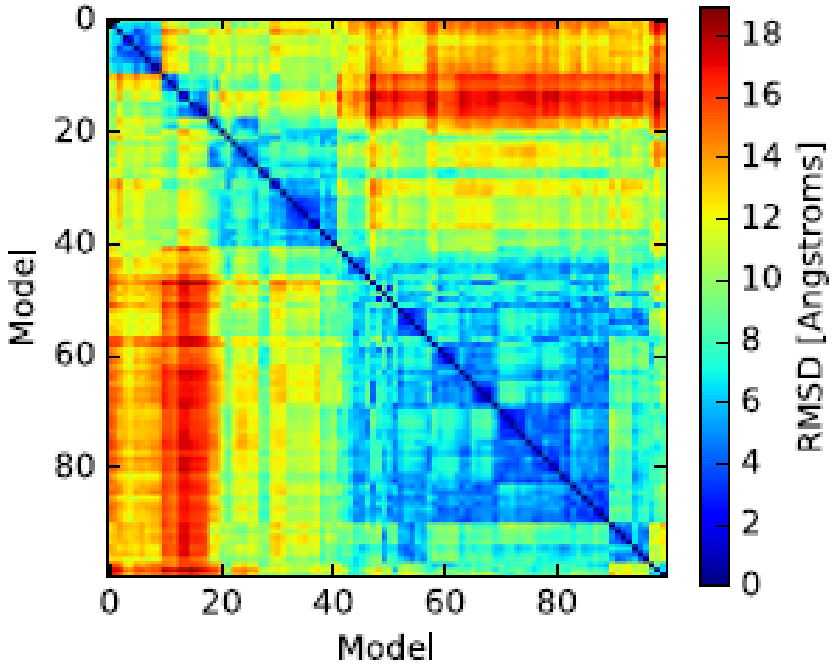
↓

Analysis and assesment

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

Rmsd_names == Rpb4, Rbp7



https://integrativemodeling.org/

Integrative structure modeling of RNA Polymerase II stalk

Gathering data

↓

Represent and translate data into restraints

↓

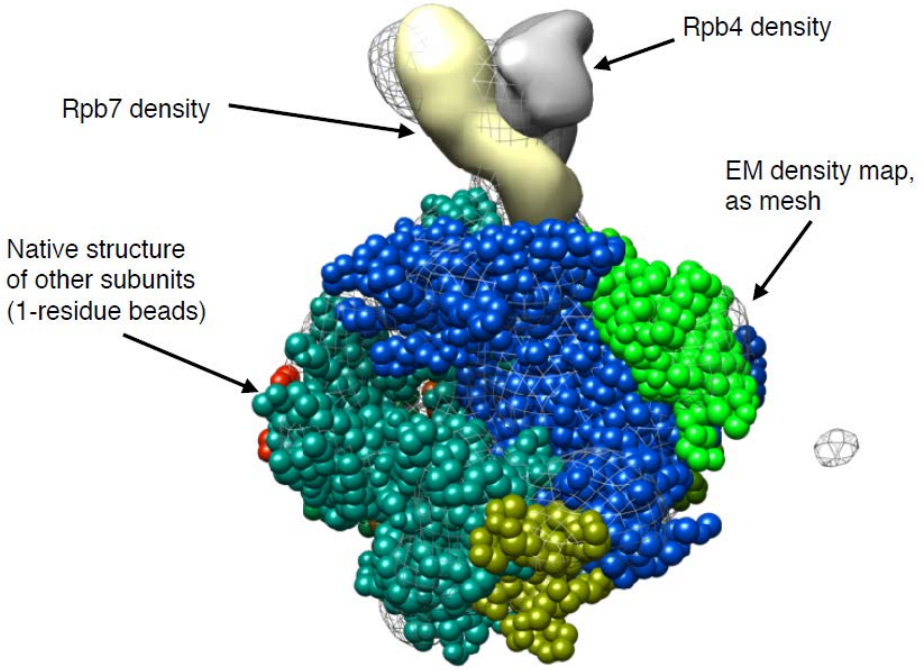
Sampling good scoring configurations

↓

Analysis and assesment

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



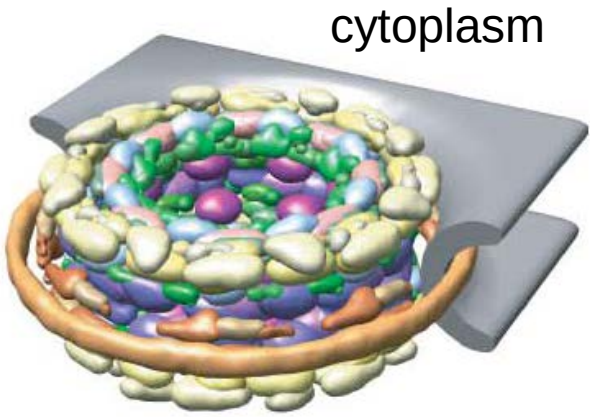
https://integrativemodeling.org/

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

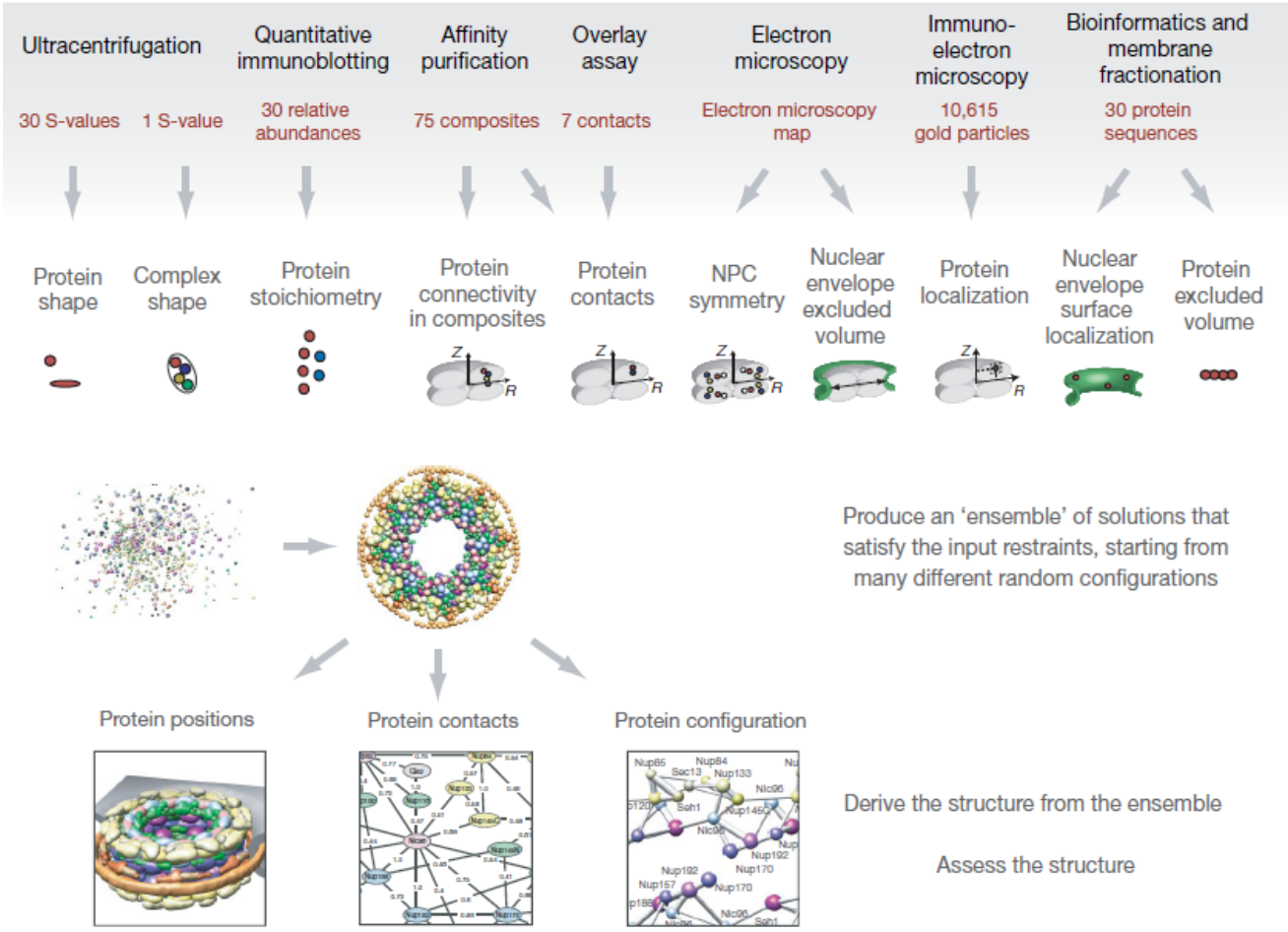
Integrating spacial restrains from proteomic data

1/ Gathering data

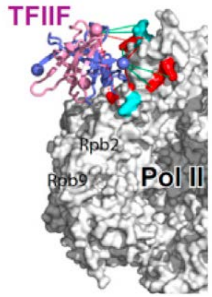
2/ Representing and translating data into restraints

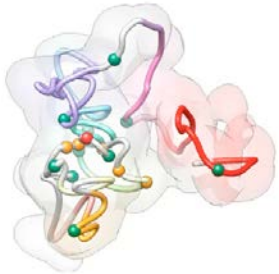
3/ Sampling good scoring configurations

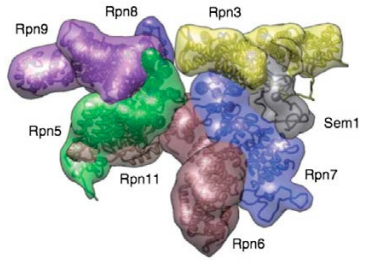
4/ Analysis and assesment

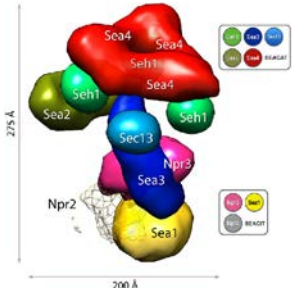


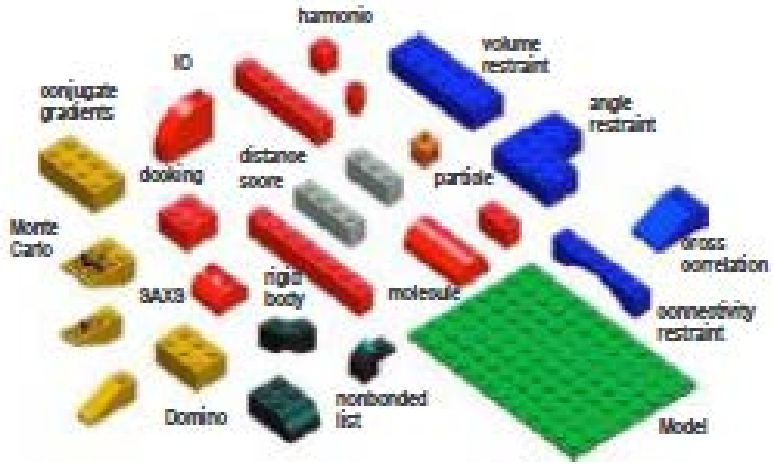
Albert et al. 2007, 2008











Representation:

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

Scoring:

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

Sampling:

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

Analysis:

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

https://integrativemodeling.org/

Sali et al. From words to literature in structural proteomics. Nature 2003, 422, pp 216-227

Alber et al. Determining the architectures of macromolecular assemblies.
Nature. 2007 Nov 29;450(7170):683-94.

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

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

A. Ward et al. Integrative structural biology. Science 339, 913-915, 2013

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.

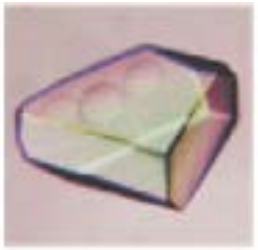
RNABuilder
Flores SC, Sherman MA, Bruns CM, Eastman P, Altman RB. Fast flexible modeling of RNA structure using internal coordinates. IEEE/ACM Trans Comput Biol Bioinform 2011;8:1247–57.

Other methods and integrative approaches

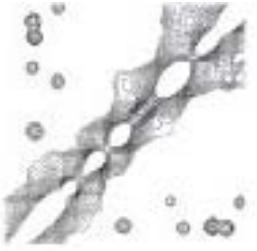
Introduce integrative approaches which allow to combine heterogenous data and propose hybrid models to provide the best possible description of the system.

Summarize mainstream complementary experimental and in silico methods to provide structural information on a macromolecular complex and discuss their pros/cons

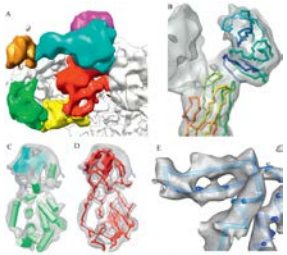
Comparative modeling and molecular docking



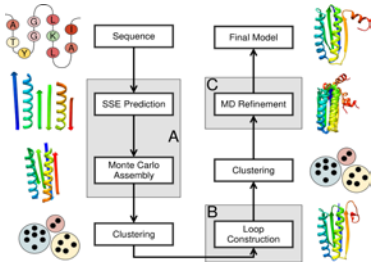
X-ray

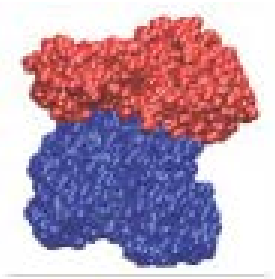


NMR



Cryo EM/ET





Comparative modelling & molecular docking

In silico models are not the product of experimental measurements of a physical sample.

They are generated computationally using various molecular modeling methods and underlying assumptions: comparative modeling, virtual docking of ligand molecules to protein targets, virtual docking of one protein to another, simulations of molecular dynamics and motions and de novo (ab initio) protein modeling.

Structures predictions

In absence of experimental 3D structure, from a sequence (or better from a multiple sequence alignment)

- Secondary structure predictions

- Prediction of 3D structure :

  - With reference to a known parental architecture parente,
    Homology modelling

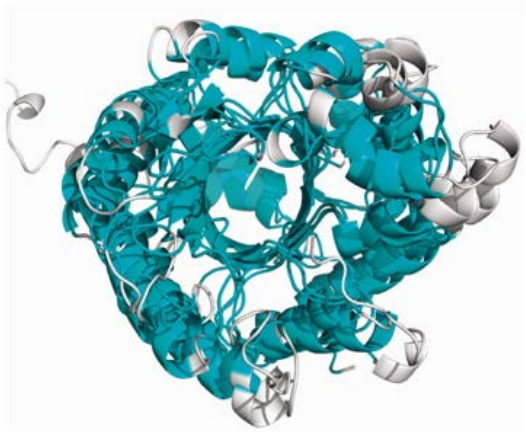
  - Without a known parental architecture (still unsolved problem)
    Fold recognition (Threading, Profile recongnition)
    Prediction of a new fold

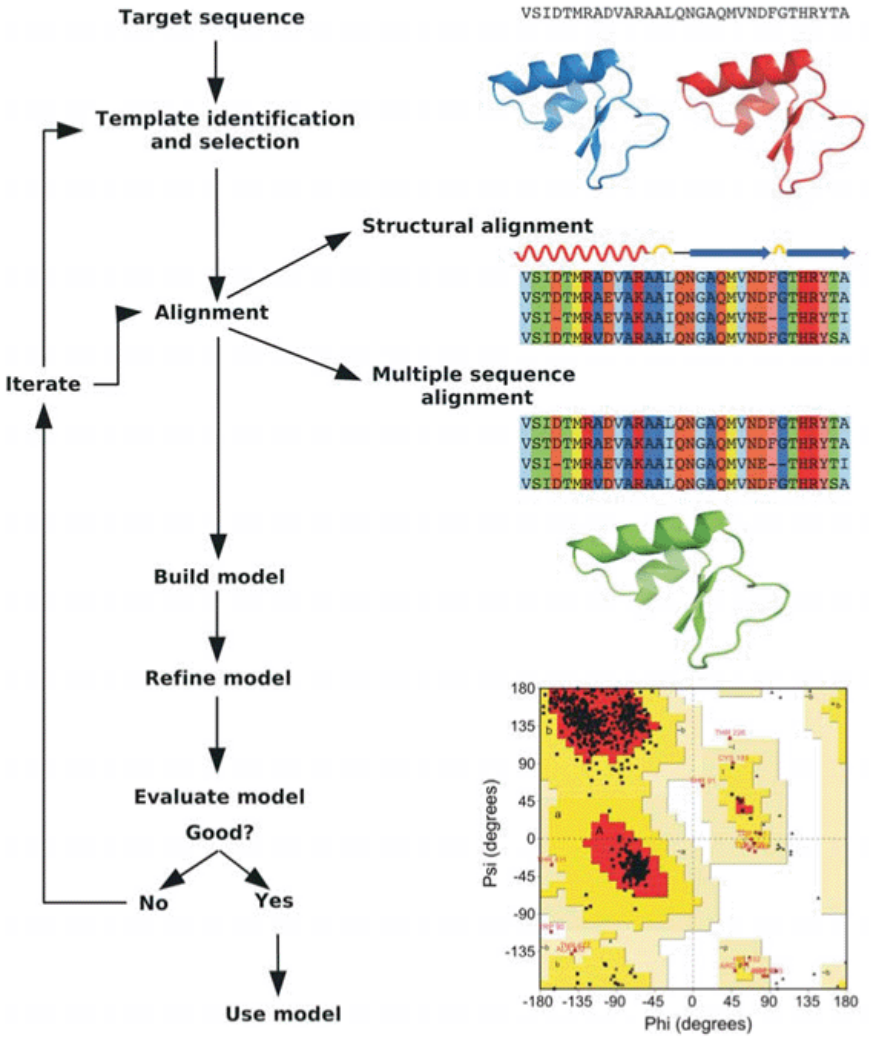
Homology modeling

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





http://swissmodel.expasy.org/

MENU

Modeling requests:

- First Approach mode
- Alignment Interface
- Project (optimise) mode

- Oligomer modeling
- GPCR mode

Model Database

- SWISS-MODEL Repository, a database for theoretical protein models.

Interactive tools

- SWISS-MODEL Workspace, an interactive working environment for protein structure modelling and assessment.
- DeepView - Swiss-PdbViewer, a tool for viewing and manipulating protein structures and models.
- Lookup ExPDB template codes accessible to SWISS-MODEL.
- Search the SWISS-MODEL Template library.
- Examples using SWISS-MODEL and the Swiss-PdbViewer.
- ANOLEA Protein structure quality check (atomic non-local environment assessment)
- News from Swiss Model.

Other links

- Course on protein structure and comparative modeling.
- Other Web-based Comparative Protein Modeling Servers.
- PHYRE, fold recognition server at the ICRF.
- PredictProtein, Burkhard Rost's sequence analysis and structure prediction server.

HELP

- Frequently Asked Questions.
- Visualising 3D models.
- Reliability of models.
- How SWISS-MODEL works.
- How ProModII works.
- Modelling of oligomeric proteins.
- Model Confidence factors.
- About model quality.
- Methods and Programs.
- Databases used.

15 years

SWISS-MODEL

An Automated Comparative Protein Modelling Server

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists World Wide.

The present version of the server is 3.5 and is under constant improvement and debugging. In order to help us refine the sequence analysis and modelling algorithms, please report of possible bugs and problems with the modelling procedure.

SWISS-MODEL is provided by:

BIOZENTRUM SIB 10 years

History

SWISS-MODEL was initiated in 1993 by Manuel Peitsch, and further developed at Glaxo Wellcome Experimental Research in Geneva and the SIB Swiss Institute of Bioinformatics by Manuel Peitsch, Nicolas Guex and Torsten Schwede. Since 2001, SWISS-MODEL is being developped by Torsten Schwede's Structural Bioinformatics Group at the SIB & Biozentrum (University of Basel). The SWISS-MODEL Repository, a relational database of annotated three-dimensional comparative protein s tructure models, was established in 2004. In 2005, SWISS-MODEL service was extended by SWISS-MODEL Workspace, a web-based work bench for protein structure modelling and assessment. Computational resources for the SWISS-MODEL server are provided in collaboration by the Biozentrum (University Basel), the Swiss Institute of Bioinformatics and the Advanced Biomedical Computing Center (NCI Frederick, USA).

Acknowledgements

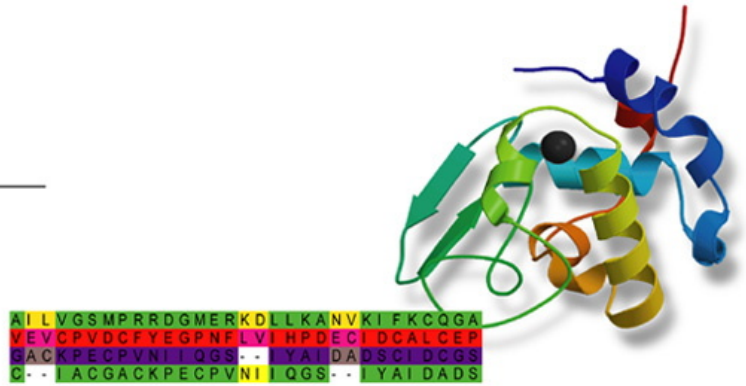
SWISS-MODEL would not have been possible without a lot of help and support. We are particularly thankful to Nicolas Guex for his many crucial contributions to the development efforts of Swiss-Model and specifically DeepView and to Gale Rhodes of the University of Southern Maine for coordinating the active DeepView user community. We also thank Alexander Diemand, Konstatin Arnold, Jürgen Kopp and Lorenza Bordoli for their many contributions to the development and operations for the modeling platform. Furthermore, we deeply indebted to Jake V. Maizel Jr, Timothy N.C. Wells, Jonathan C.K. Knowles, and Allan Baxter who have provided the necessary environment and resources during various phases of this project. Finally, we thank Stanley K. Burt, Robert W. Lebherz III, Karol Miaskiewicz and Jack R. Collins of the Advanced Biomedical Computing Center at the National Cancer Institute in Frederick Maryland for their support and operating the US mirror of the Swiss-Model server. We gratefully acknowledge the financial support by GlaxoSmithKline, Novartis, the Swiss National Science Foundation, the Biozentrum of the University of Basel and the Swiss Institute of Bioinformatics.

Disclaimer

The result of any modelling procedure is NON-EXPERIMENTAL and MUST be considered with care. This is especially true since there is no human intervention during model building. Carefully read the header section of the files to know what templates and alignments were used during the model building process.

Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



About MODELLER

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints (3,4), and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. MODELLER is available for download for most Unix/Linux systems, Windows, and Mac.

Several graphical interfaces to MODELLER are commercially available from Accelrys. Teaching licenses are also available to those institutions that acquire and maintain a research license.

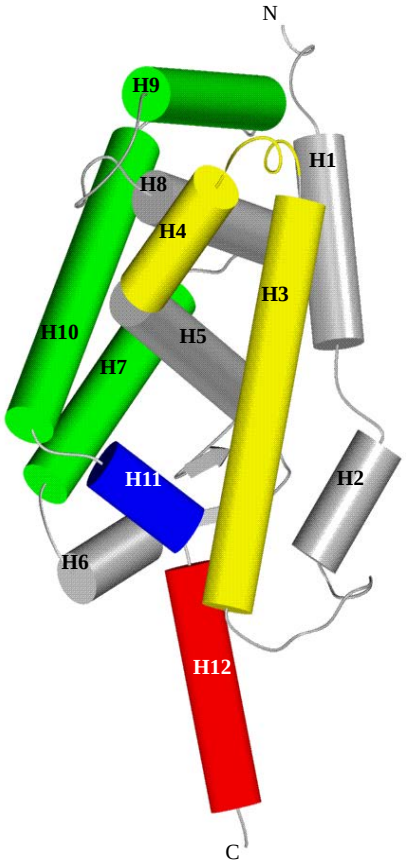
1. N. Eswar, D. Eramian, B. Webb, M. Shen, A. Sali. Protein Structure Modeling With MODELLER. in press, 2006.
2. M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.
3. A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.
4. A. Fiser, R.K. Do, & A. Sali. Modeling of loops in protein structures, Protein Science 9. 1753-1773, 2000.

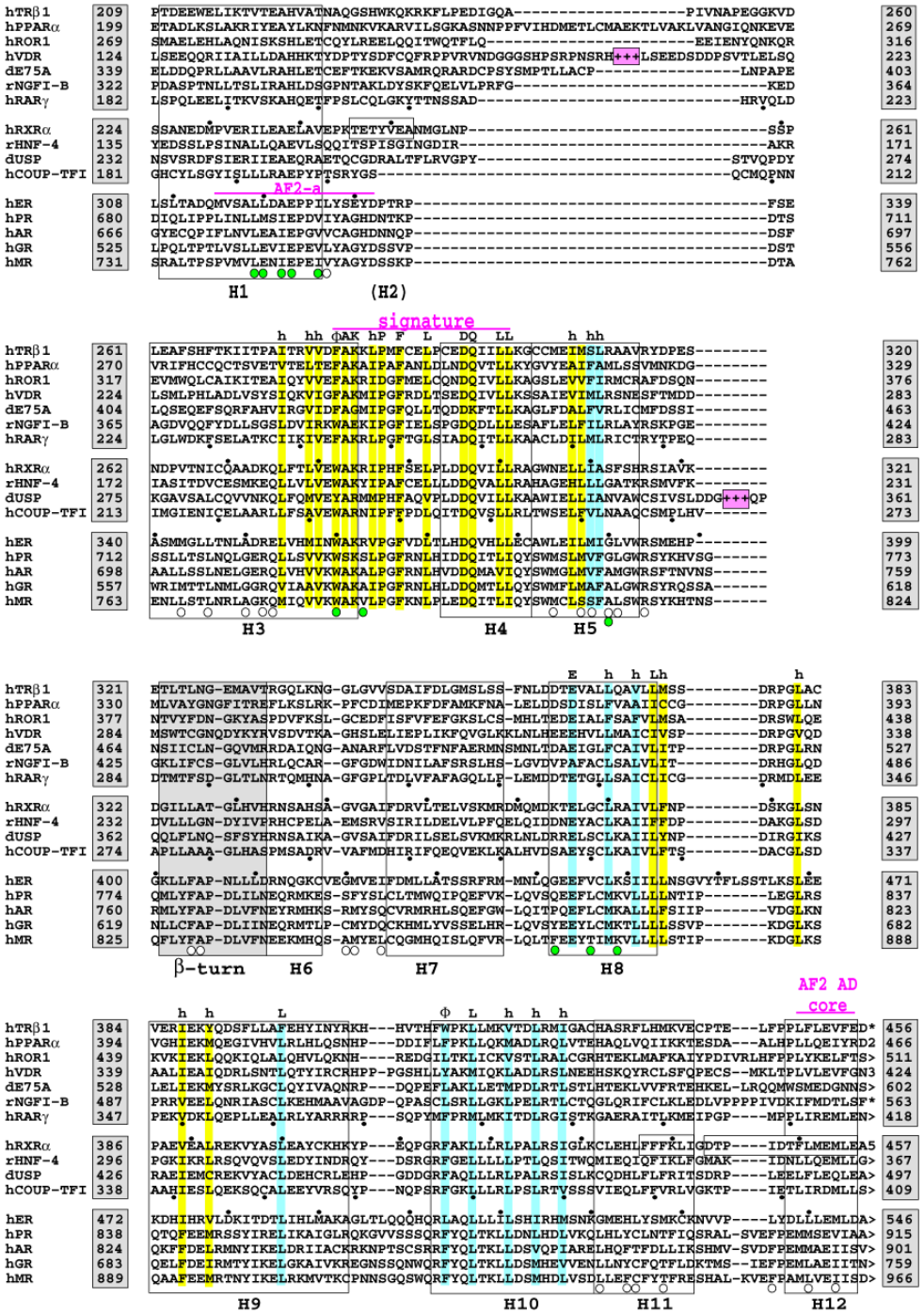
The current release of Modeller is **9v1**, which was released on January 22nd, 2007. Modeller is currently maintained by Ben Webb.

Homology modeling

1 (ou several) 3D structure(s)

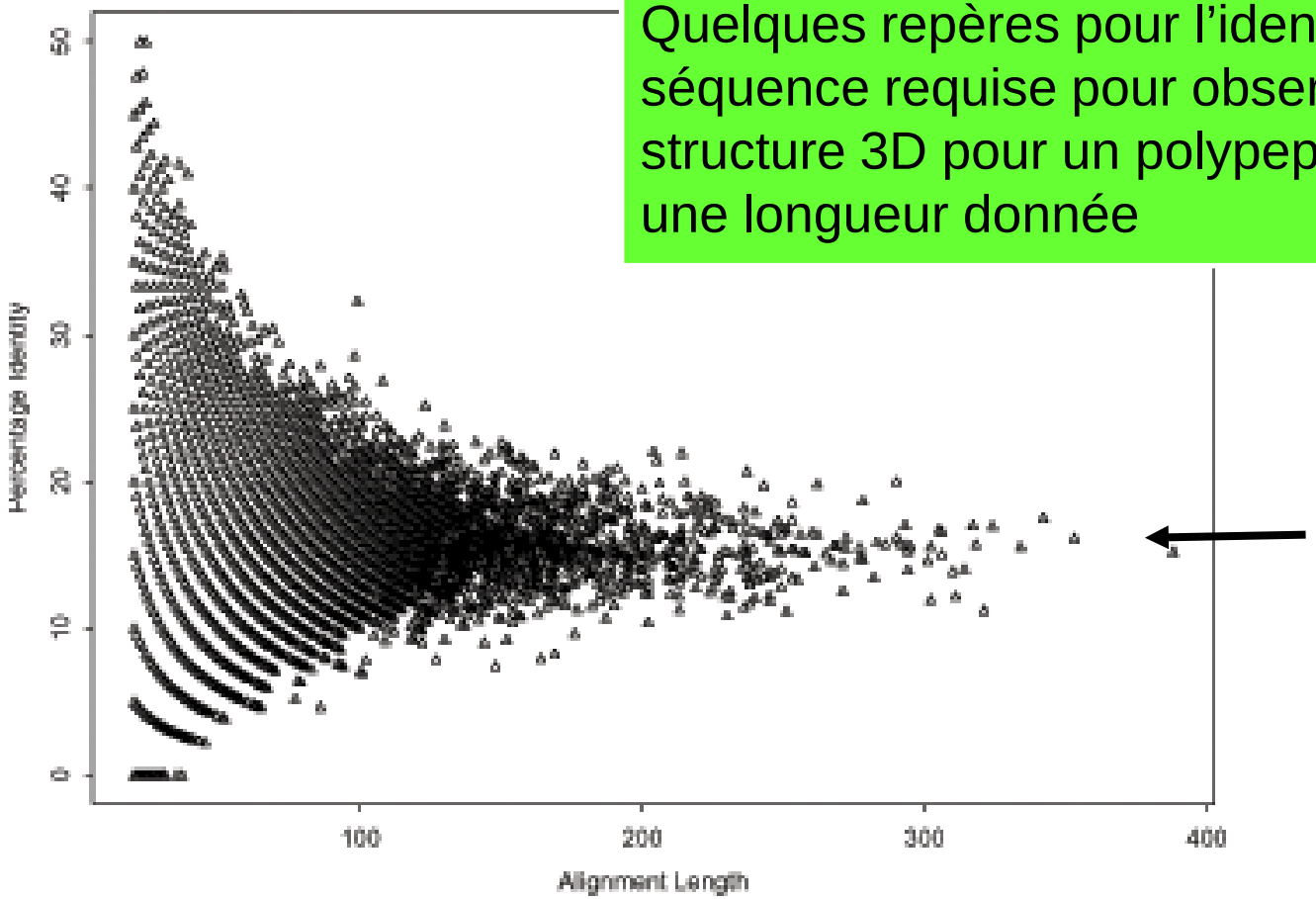
1 multiple sequence alignment





Similarités de séquences.
Signification Biologique / Signification statistique

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

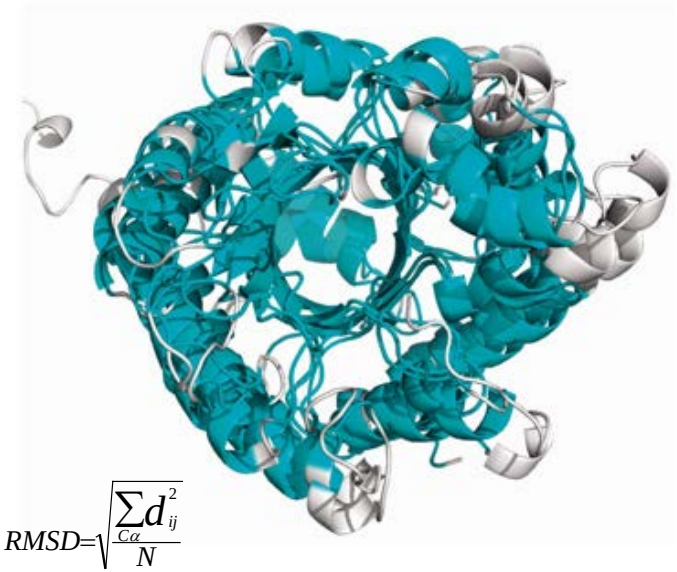


Quelques repères pour l'identité de séquence requise pour observer la même structure 3D pour un polypeptide ayant une longueur donnée

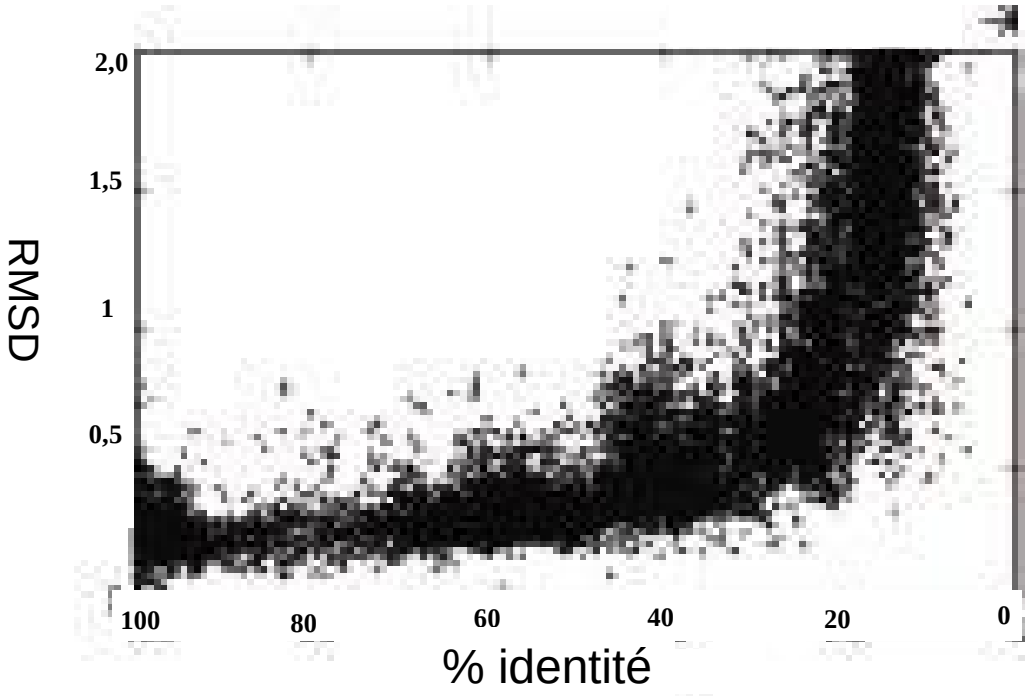
Comparaisons des structures 3D de protéines homologues

Déformations autour d'un thème commun

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



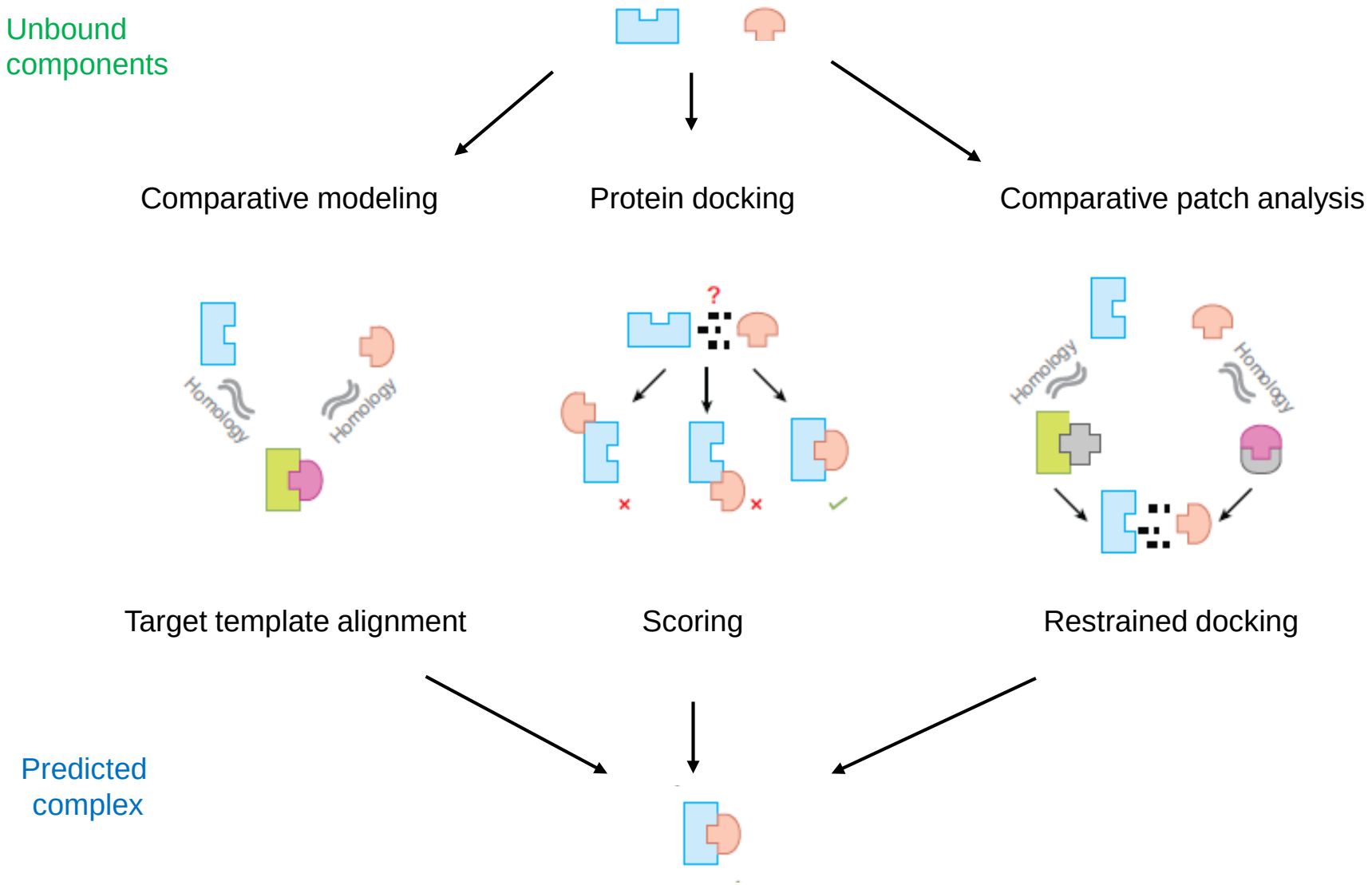
$$RMSD=\sqrt{\frac{\sum_{C\alpha} d_{ij}^2}{N}}$$



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

Relation type exponentiel % identité / déformations

Modelling structures of protein complexes



Integrative computational modeling of protein interactions

João P. G. L. M. Rodrigues and Alexandre M. J. J. Bonvin

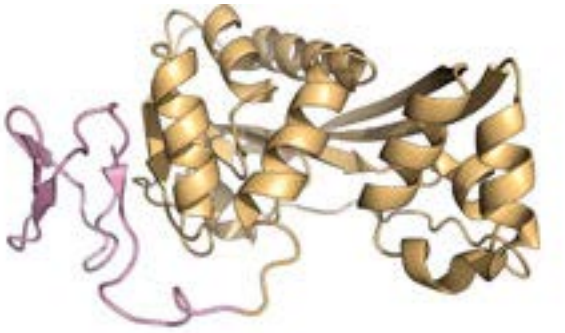
Computational Structural Biology Group, Bijvoet Center for Biomolecular Research, Utrecht University, the Netherlands

Table 1. List of top-performing docking approaches participating in CAPRI.

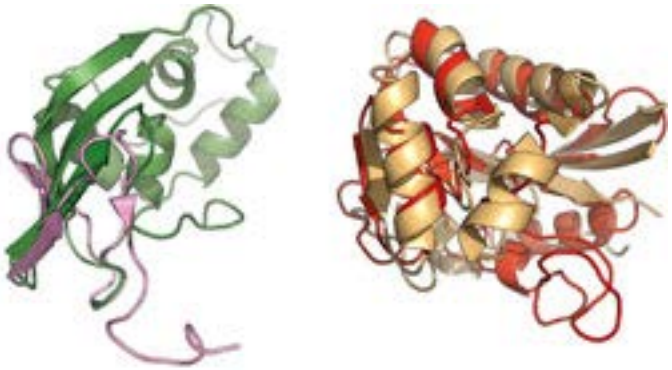
Name	Protocol	Strengths and weaknesses	Integration of data	Public Web Server
ATTRACT [39]	Energy minimization in translational and rotational degrees of freedom using a reduced protein model and normal-mode analysis to allow conformational changes upon binding.	+ Fast derivative-based search method. + Conformational changes upon binding (local and global motions). + Support for Cryo-EM density maps. – Not available via a web server	Implements interface data by adding atom/residue-specific weights, which can be negative (repulsive). Also offers the option to dock using Cryo-EM density maps.	None
ClusPro [61]	Rigid-body search via a FFT correlation approach (PIPER), followed by structural similarity (RMSD) based clustering to find the most popular interaction modes, and final refinement of selected structures using CHARMM.	+ Best automated served in the latest CAPRI evaluation [12]. + Fast and exhaustive protocol. + Several docking 'modes' depending on the biological function (antibody/antigen, multimer, others). – Cannot handle flexible complexes.	Regions/residues in the binding partners can be introduced to bias the scoring of the models. Noteworthy option of 'negative' (repulsion) contacts.	http://cluspro.bu.edu/
GRAMM-X[68]	Grid-based FFT rigid body docking approach using a softened Lennard–Jones potential function. The top predictions are minimized and re-scored using a soft Lennard–Jones potential, an evolutionary conservation term, a statistical residue–residue potential, and the volume of the local energy minima in the grid.	+ Fast. + Uses CONSURF to determine evolutionary conserved residues. – Cannot handle flexible complexes.	Regions/residues in the binding partners can be introduced to bias the scoring of the models.	http://vakser.bioinformatics.ku.edu/resources/gramm/grammx/
HADDOCK [32]	Rigid-body energy minimization followed by semi-flexible (interface) refinement and final optimization in explicit solvent. Returns clusters of models ranked by HADDOCK score.	+ Best performing team in the latest CAPRI evaluation [12]. + Restraint-based integration of ambiguous experimental/prediction data. + Explicit flexibility of the interface. + Powerful user-friendly web interface. – Slower than FFT-based methods.	Several types of restraints allow integration of different sources of data: distances, orientations, radius of gyration, symmetry type, etc. Directly integrates several NMR-derived data.	http://haddocking.org
Hex [27]	Spherical polar Fourier approach using rotational correlations generates a very large number of putative models, which are then re-scored using a shape correlations or shape plus electrostatic correlations.	+ Extremely fast approach (~ 15 s using 2 graphical processing units). – Cannot handle flexible complexes.	Can restrict the rotational search around an intermolecular axis defined by a pair of residues, one on each interacting monomer. The angular range of the search can also be defined.	http://hexserver.loria.fr/
PatchDock [25]	The surface of the molecules is divided into patches (concave, convex and flat) and only those containing 'hot-spot' residues are kept. The patches are then matched using geometric hashing and pose-clustering techniques and the candidate models are examined (to remove extreme clashes) and scored.	+ Extremely efficient and fast protocol. + Integrated suite of docking tools [24]. – Fragmentation of protocol in very specialized tools requires *a priori* knowledge of their limitations.	Regions/residues in the binding partners can be introduced to bias the scoring of the models.	http://bioinfo3d.cs.tau.ac.il/PatchDock/
pyDock [22]	Rigid-body search via a FFT search method (custom optimized FTDOCK) followed by scoring with a combined electrostatics and desolvation energy function.	+ Fast protocol. + Integration of different modules (pyDockSIPPER, pyDockRST, pyDockSAXS) to improve predictions. – Cannot handle flexible complexes.	PyDockRST module scores models based on agreement with user-defined distances. Implements SAXS scoring by generating synthetic SAXS curves of the models to user-provided data.	http://life.bsc.es/servlet/pydock/home/
RosettaDock [38]	Low-resolution, rigid-body, Monte Carlo search followed by simultaneous optimization of backbone displacement and side-chain conformations using Monte Carlo minimization.	+ Scoring based on the Rosetta energy function. + Very powerful refinement protocol. - Poor (direct) support of interface information. – Computationally demanding and slow protocol.	Molecules can be positioned manually in space (e.g. using PyMOL). Interfaces with other Rosetta tools (RosettaInterface) to validate mutagenesis data.	http://antibody.graylab.jhu.edu/docking
SwarmDock [33]	Local docking and particle swarm optimization of partner position and orientation, using normal modes to model induced fit, and final energy minimization. Uses the DComplex scoring function but the final models are re-ranked with a centroid potential prior to clustering.	+ Search method that explicitly models global flexibility upon binding. – Slow. Predictions can take days unless a local version and sufficient computational resources are available.	Residues belonging to the binding site can be selected to bias the starting positions of the binding partners.	http://bmm.cancerresearchuk.org/~SwarmDock
ZDOCK [28]	FFT-based rigid-body search using a scoring function composed of desolvation energy, electrostatics, and	+ Fast protocol + Robust performance over several CAPRI rounds.	Allows user-defined selection of 'blocked' and 'binding' residues that influence the	http://zdock.umassmed.edu/

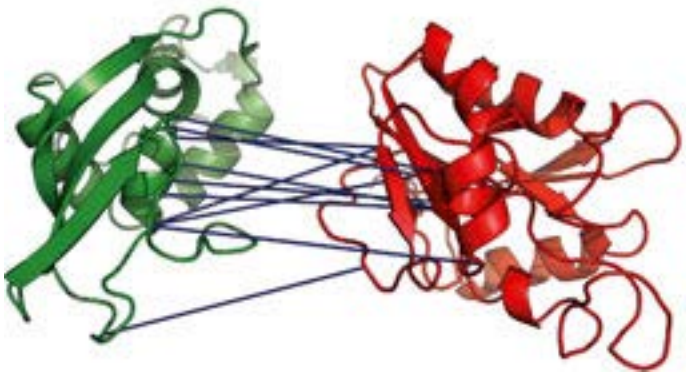
Comparative modeling of the Mtq2/Trm112 complex

zinc-binding (pink) and methyltransferase (orange) domains of RlmA(I) (PDB 1P91).

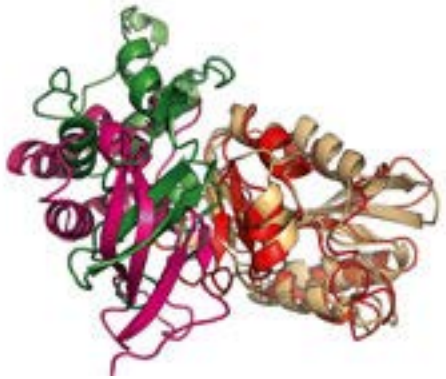


Superimposition of of Mtq2 (red) and Trm112 (green) models on their respective structural homologous domains of RlmA(I).





Distance restraints derived from Ca– Ca contacts between the domains of RlmA(I) and mapped onto the Mtq2 and Trm112 models).



Superimposition of Mtq2 (red) /Trm112 (green) models on the experimental .structure (rmsd 3.3 Å

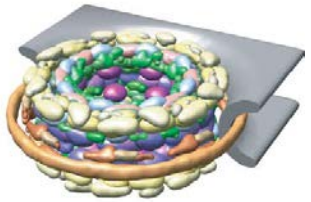
Rodrigues and Bonvin. 2013

Structural data used in integrative modeling

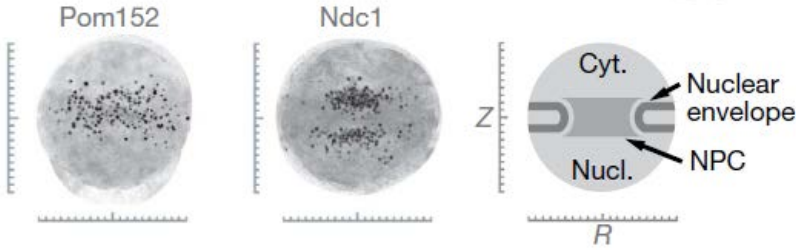
Atomic structures of parts of the system	X-ray and neutron crystallography, NMR, Cryo-EM/ET, Comparative modelling and molecular docking
Composition and components positions	Purification from source with gel analysis or MS, Electron microscopy and tomography, gold labelling, Super resolution microscopy,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

Observations of complexes in their native environment

Electron microscopy and tomography, Super resolution microscopy, FRET imaging

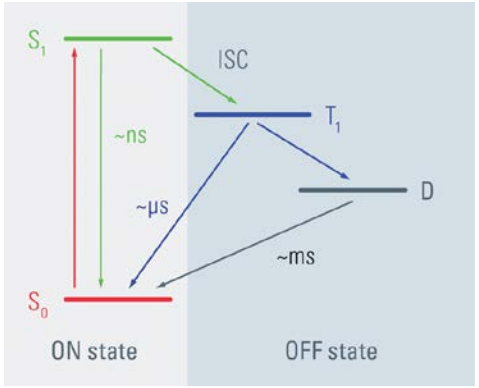


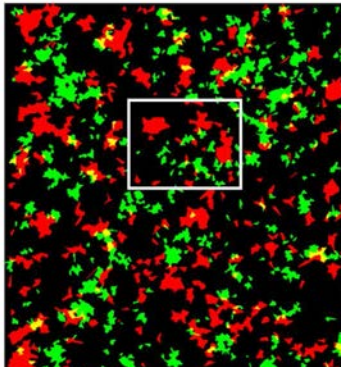
Localization of proteins by immuno-EM.
Immuno-EM montages for Pom152–PrA nuclei and Ndc1–PrA nuclear envelopes with gold-labelled antibodies. (Alber, 2007)

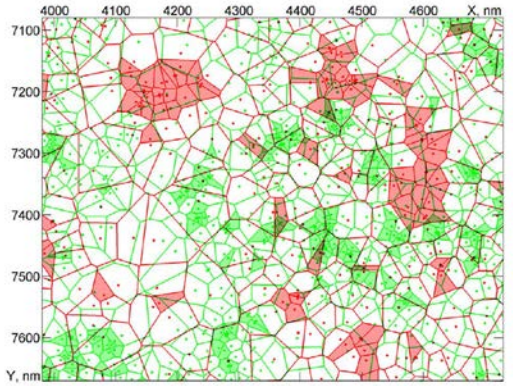


Super resolution microscopy: PALM (PhotoActivated Localization Microscopy), GSDim (Ground State Depletion imaging followed by Individual Molecule return)

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.







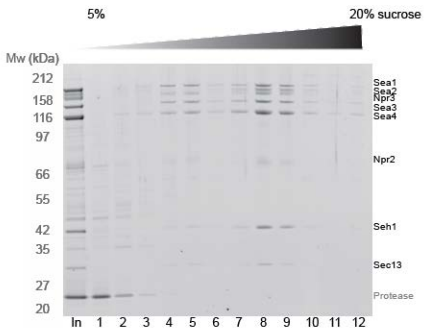
Purification from cells and mapping

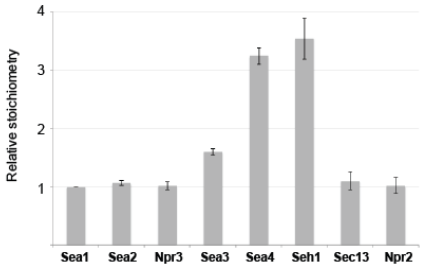
Information: subunit composition and stoickiometry

Method: purification (from endogenous source and analysis)

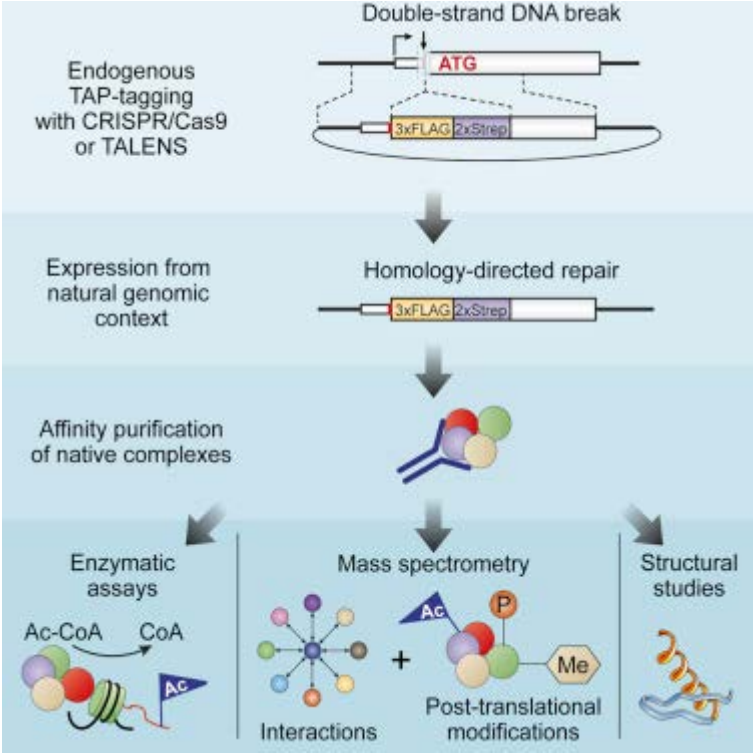
Genome editing (yeast, mammalian cells)

Quantitative densitometry
Quantitative mass spectrometry



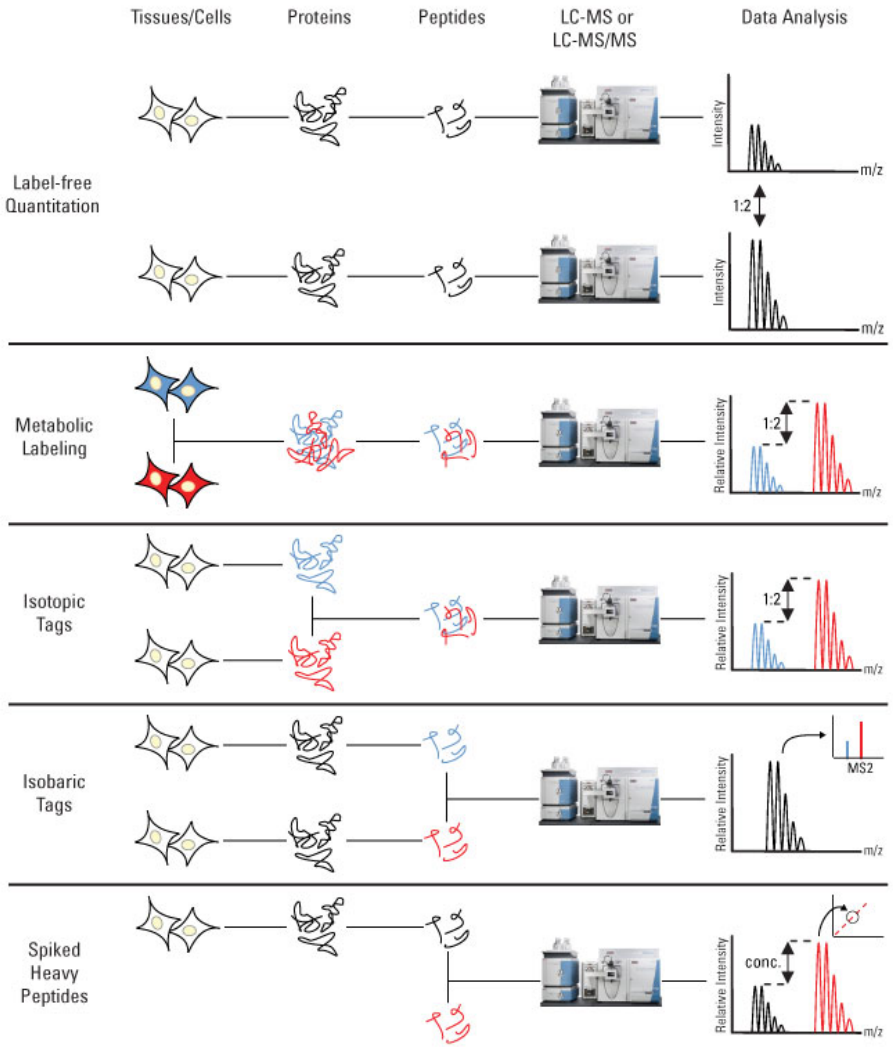


Analysis of the SEA complex relative stoichiometry by SYPRO Ruby staining (Algret et al. 2014)



Dalvail et al. 2015

Quantitation by MS using ion peak intensity



Measure peak intensities from Different samples

Relative and *absolute quantification*

Compare isotopically labelled samples ($^{14}N/^{15}N$)

Isobaric Tags for relative and Absolute Quantification (iTRAC)

Isotope-Coded Affinity Tags (iCAT)

Spiking sample with isotope-labeled reference peptides

Thermofisher web site

Label free quantitation by MS

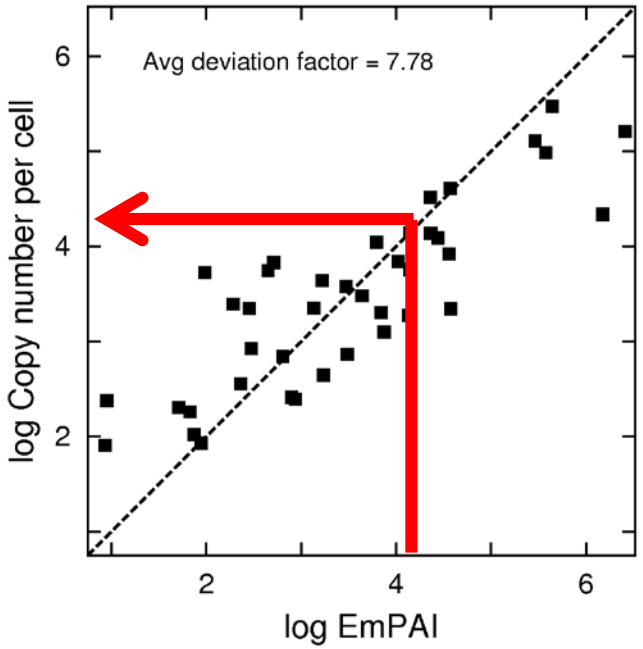
Label-free quantification approaches aim to correlate the mass spectrometric signal of intact proteolytic peptides or the number of peptide sequencing events with the relative or absolute protein quantity directly.

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).

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

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

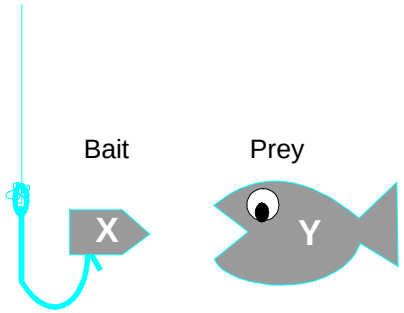


Physical proximity of components

✓*In vitro*

- Co-immunoprecipitation
- GFP, GST, His, Strep-pull down assays
- ChIPs Protein arrays
- TAP-MS

Bait – Prey model



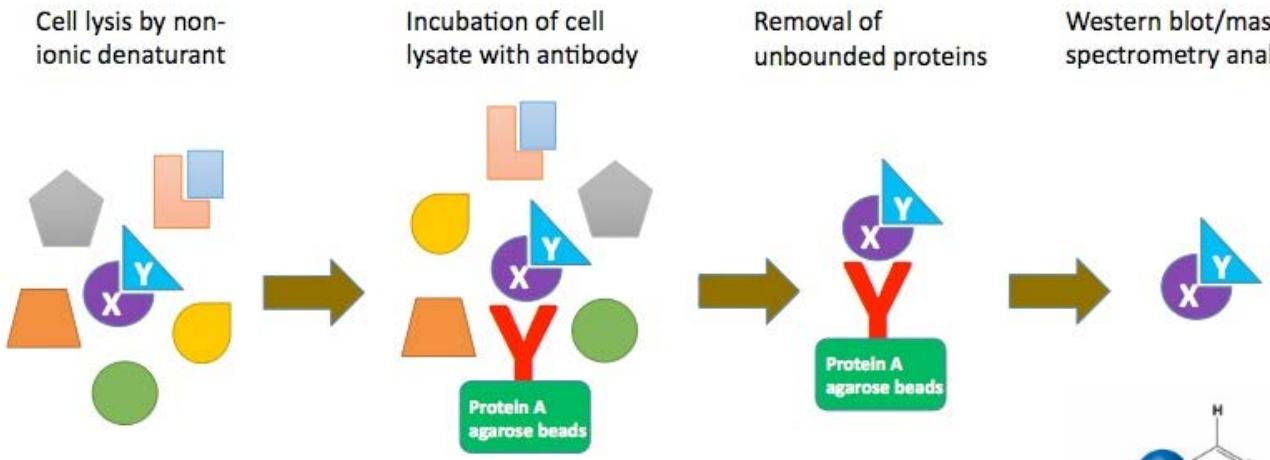
does X bind with a protein?

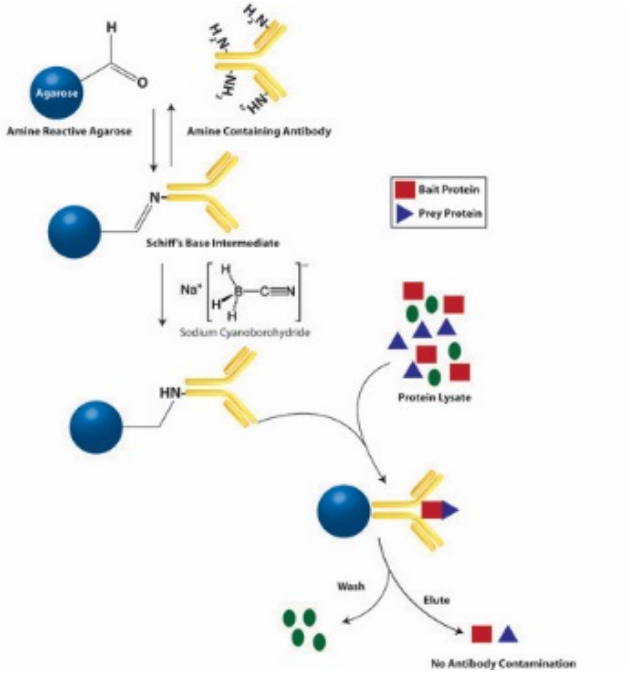
✓*In vivo*

- Yeast two-hybrid system
- Phage display

Physical interaction between protein binding domains

Pair wise analysis and purification of sub-complexes



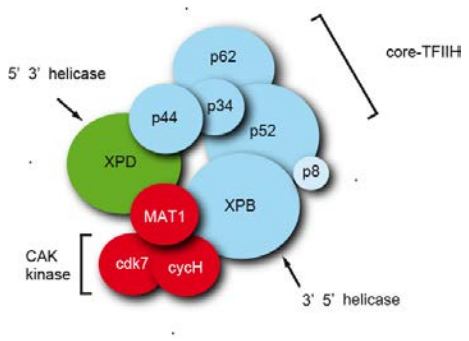


Pull down/immuno-precipitations

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

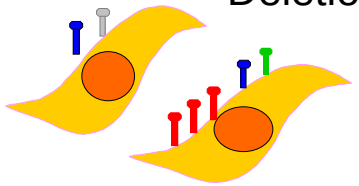
Conventional resin or magnetic beads

Systematic dissection of protein-protein interactions from recombinant proteins co-expressed in insect cells

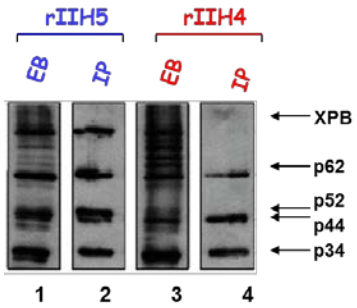


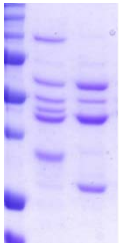
Analysis of the protein interaction network
Identification of key regulatory interactions

Deletion



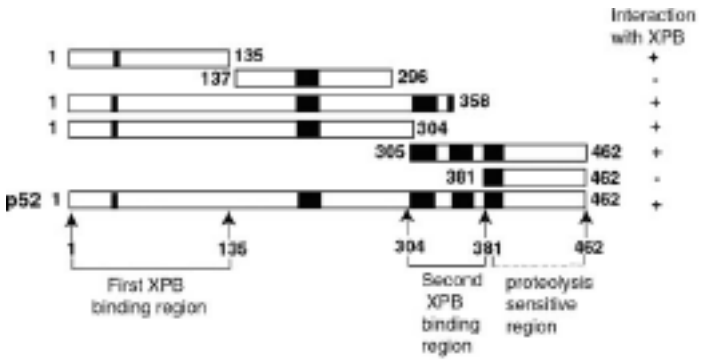
Pair-wise

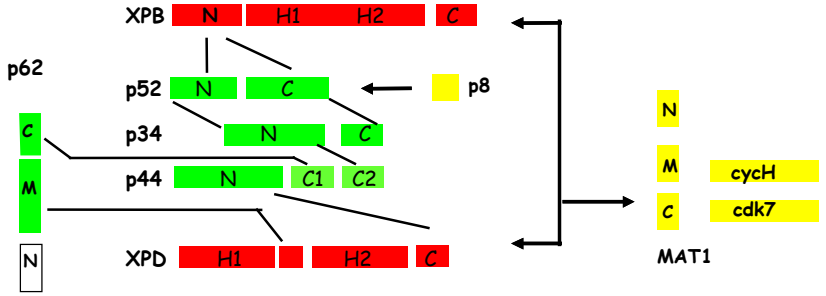




XPB
P62
*
P52,p44
P34
P34(1-233)

	p34	p44	p52	p62	XPB	XPD	MAT1
p34							
p44							
p52							
p62							
XPB							
XPD							



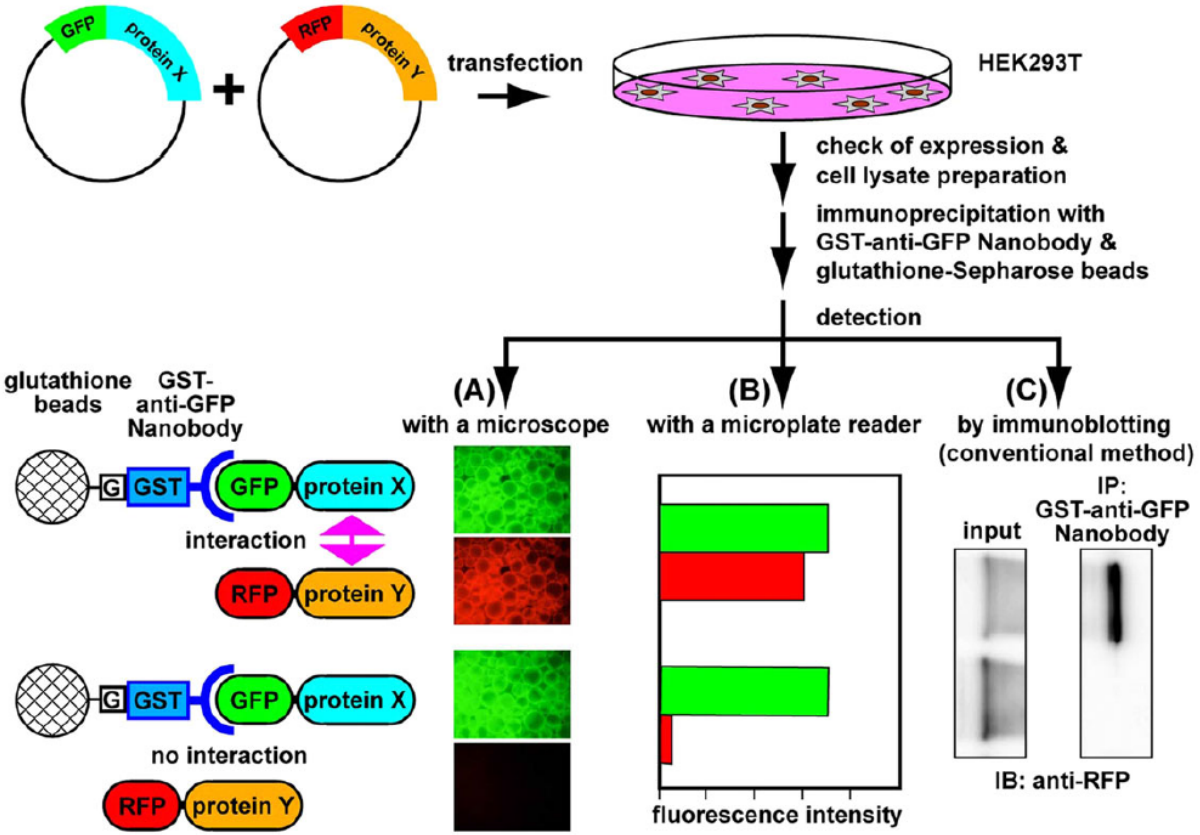


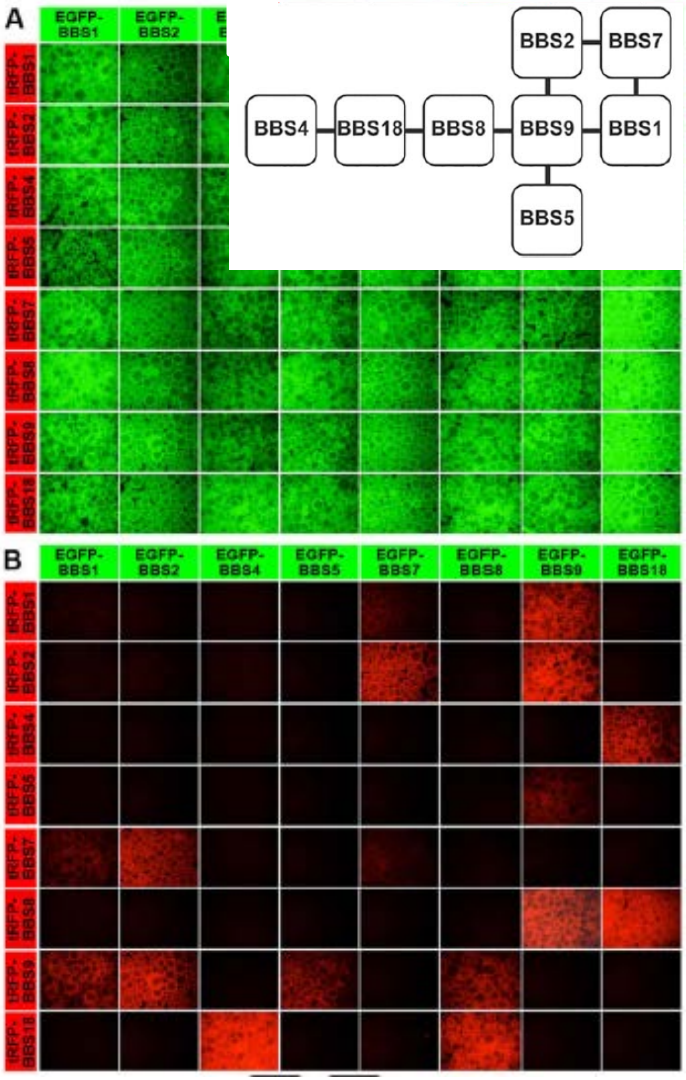
Jawahri at al. 2002, Radu et al, in prep

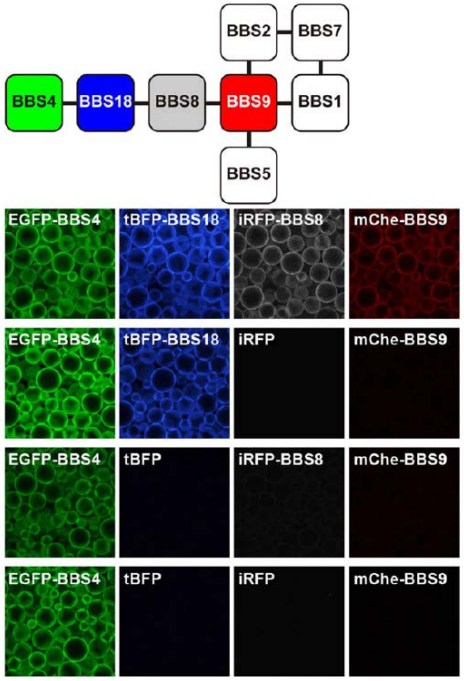
RESEARCH ARTICLE

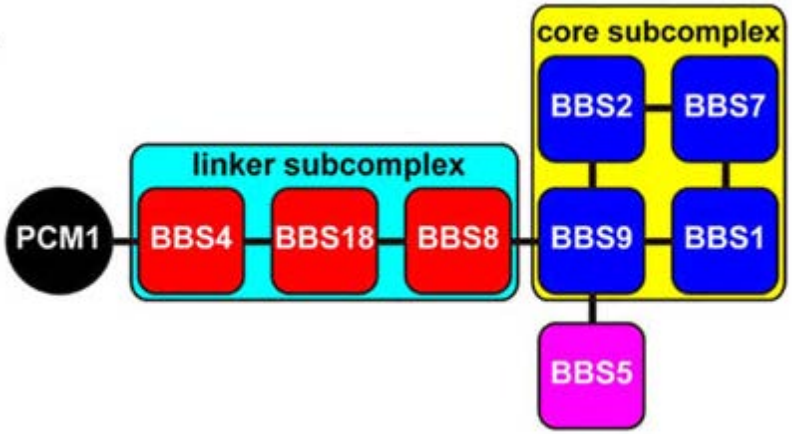
Architectures of multisubunit complexes revealed by a visible immunoprecipitation assay using fluorescent fusion proteins

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



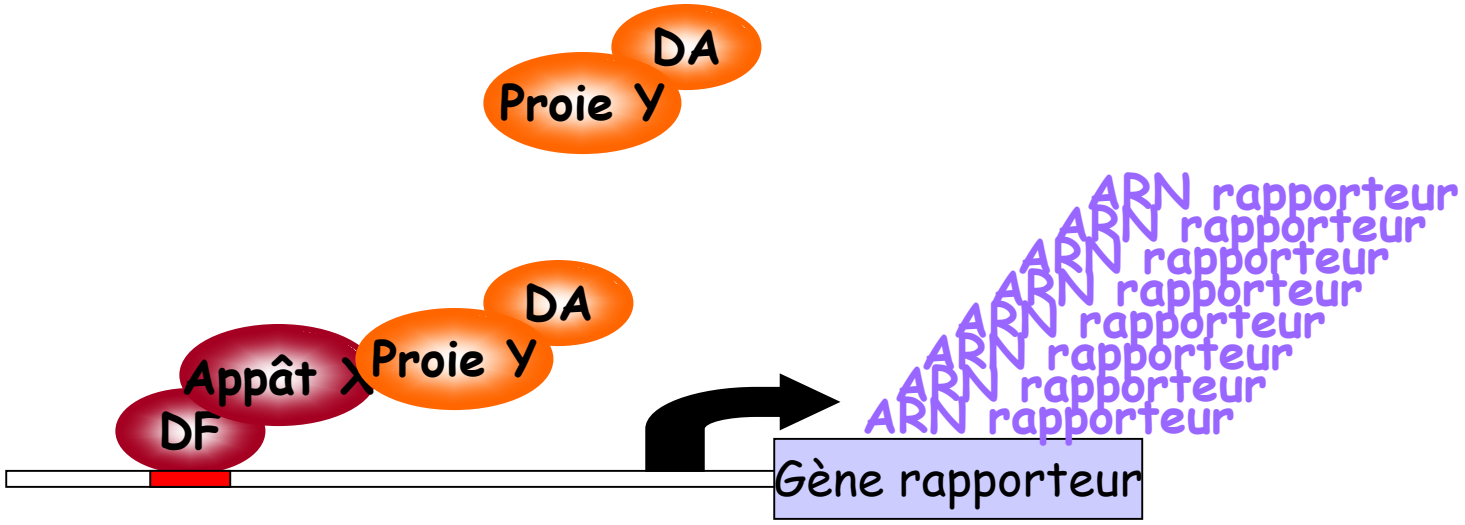

A
EGFP-BBS1
EGFP-BBS2
IRFP-BBS1
IRFP-BBS2
IRFP-BBS4
IRFP-BBS5
IRFP-BBS7
IRFP-BBS8
IRFP-BBS9
IRFP-BBS18
BBS2
BBS7
BBS4
BBS18
BBS8
BBS9
BBS1
BBS5
B
EGFP-BBS1
EGFP-BBS2
EGFP-BBS4
EGFP-BBS5
EGFP-BBS7
EGFP-BBS8
EGFP-BBS9
EGFP-BBS18
IRFP-BBS1
IRFP-BBS2
IRFP-BBS4
IRFP-BBS5
IRFP-BBS7
IRFP-BBS8
IRFP-BBS9
IRFP-BBS18


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

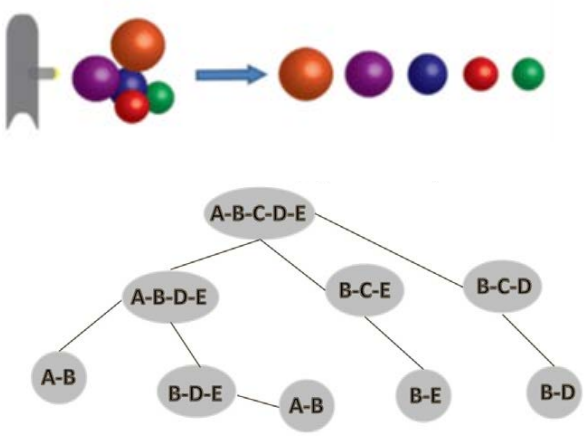
Yeast two-hybrid system

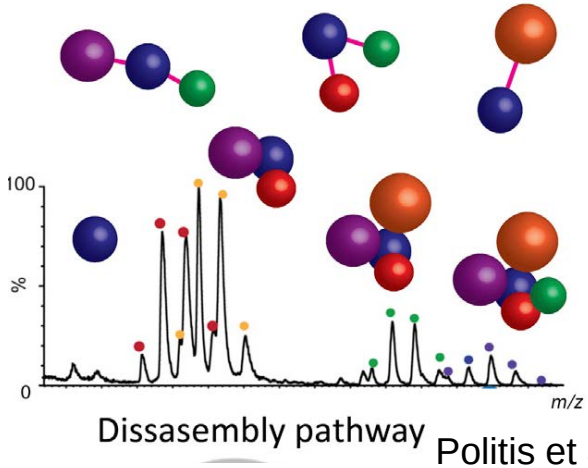
- Detecting protein-protein interactions in yeast
- Transcriptional regulator system
- “prey”-”bait” model :fusion proteins with a transcriptional activating domain (AD, prey), a DNA-binding domain (DBD, bait)
- Term “two-hybrid” derives from these two chimeric proteins.
- Most commonly used method for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



Native mass spectrometry of large entities

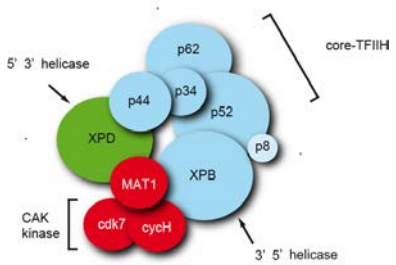
Subunit stoichiometry
Dissociation pathways and identification of sub-complexes.

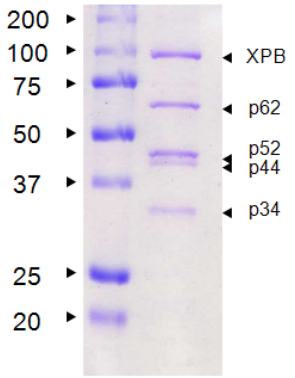


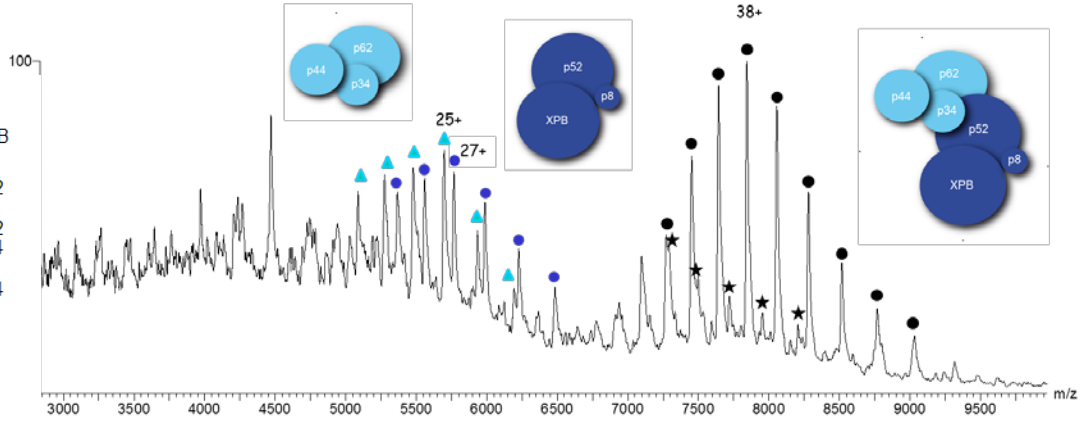


Dissasembly pathway

Politis et al. 2015

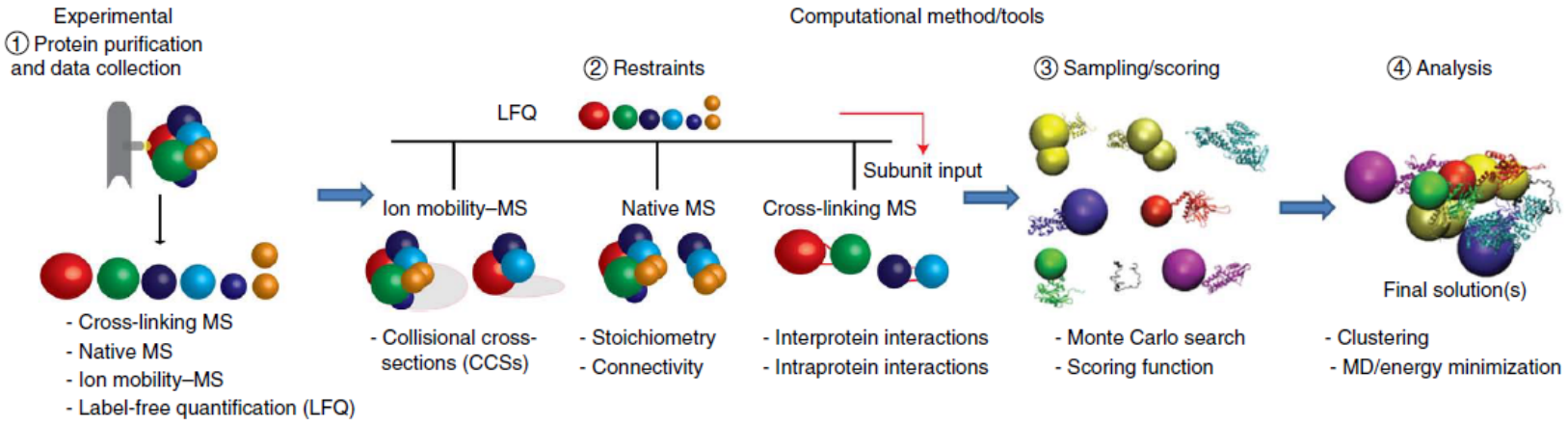




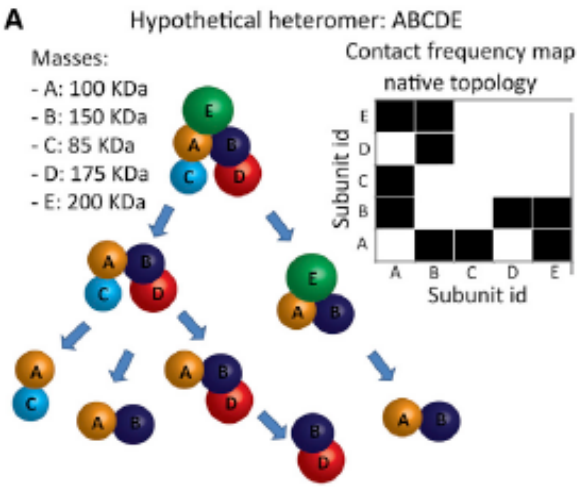


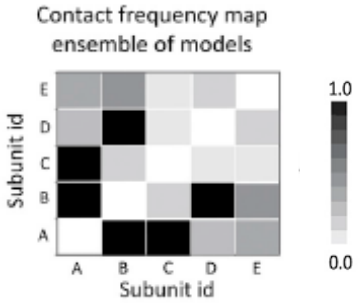
Radu et al., in prep

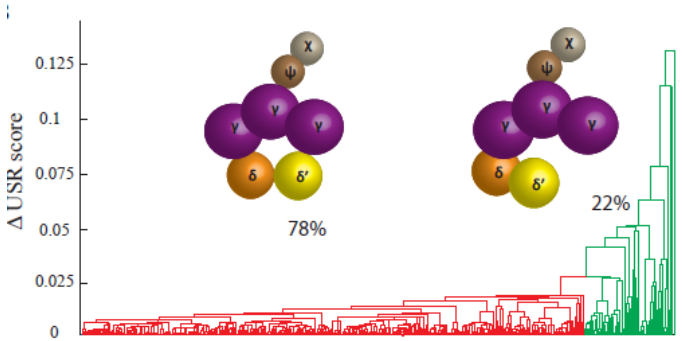
Topological models from MS-based hybrid analysis



Build an hypothetical contact map from experimental data
Sampling/scoring (Monte Carlo search and scoring)
Clustering (and minimization) to identify most the best-scoring models







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

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 !

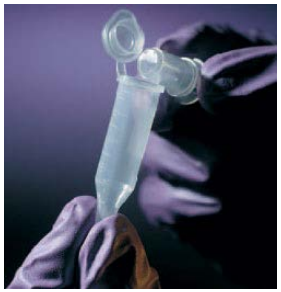
Size exclusion chromatography

- 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)

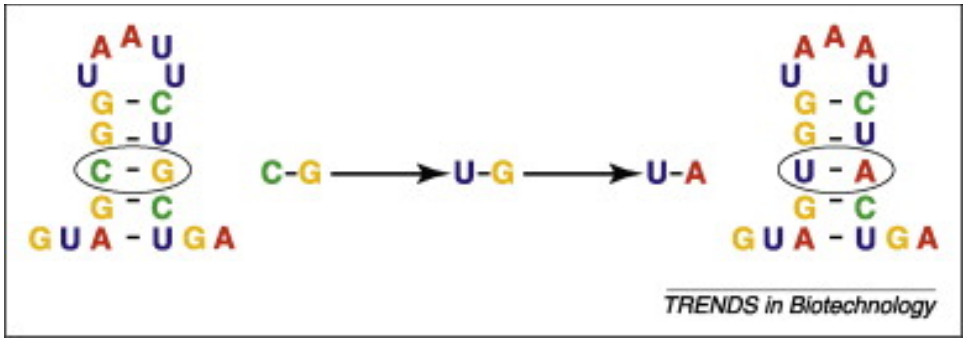


- dilutes the samples
- very easy to perform (overnight)

Covariation of RNA sequences

- RNA typically produced as a single stranded molecule (unlike DNA)
- Strand folds upon itself to form base pairs & secondary structures
- Structure conservation is important

- RNA sequence analysis is different from DNA sequence



Variations in RNA sequence maintain base-pairing patterns for secondary structures (conserved patterns of base-pairing)

When a nucleotide in one base changes, the base it pairs to must also change to maintain the same structure

Such variation, referred to as *covariation* indicates base pairing (A:U, G-C or G-U (Wobbel)).

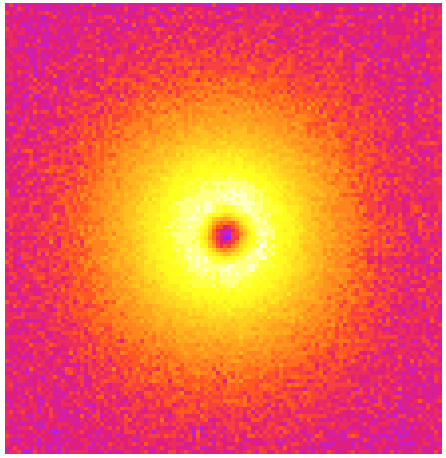
Structural data used in integrative modeling

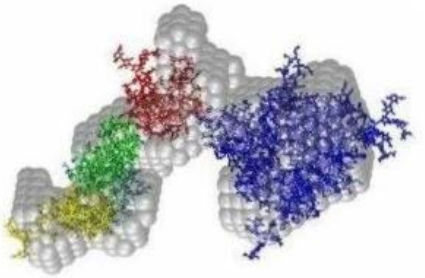
Atomic structures of parts of the system	X-ray and neutron crystallography, NMR, Cryo-EM/ET, Comparative modelling and molecular docking
Composition and components positions	Purification from source with gel analysis or MS, Electron microscopy and tomography, gold labelling, Super resolution microscopy,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

Small Angle X-ray Scattering

With a crystaline sample, diffraction pattern forms from the constructive interference of light passing through a crystal

With a non crystaline sample, there is no prefferential orientation of the molecules and light is scattered in all directions resulting in averaged diffusion patterns





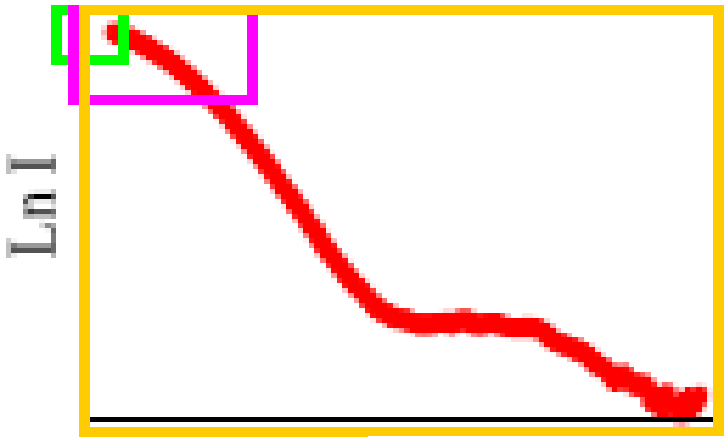
q = 0: Determination of I(0)

Small q Dimension of particules
Guinier, Krazky

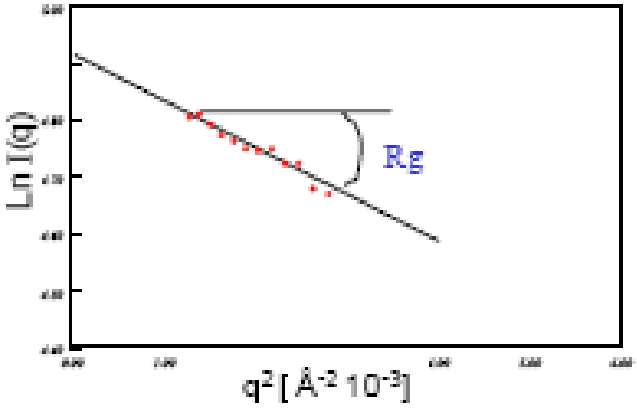
Larger q Shape of particules
(Ab-initio analysis, fitting)

$$I(q) = I(0)\exp\left(-\frac{R_g^2 q^2}{3}\right)$$

$qRg \leq 1.0$



q =4πsinθ/λ



Neutron scattering (SANS)

Photons in interact with the electronic shell
Neutrons interact with the nucleus

The intensity of the interaction depends on the diffusion lenght b:

The diffusion lenght is fonction of:

the atomic number (Z) for X-ray diffusion
the spin of the nucleus for neutron scattering

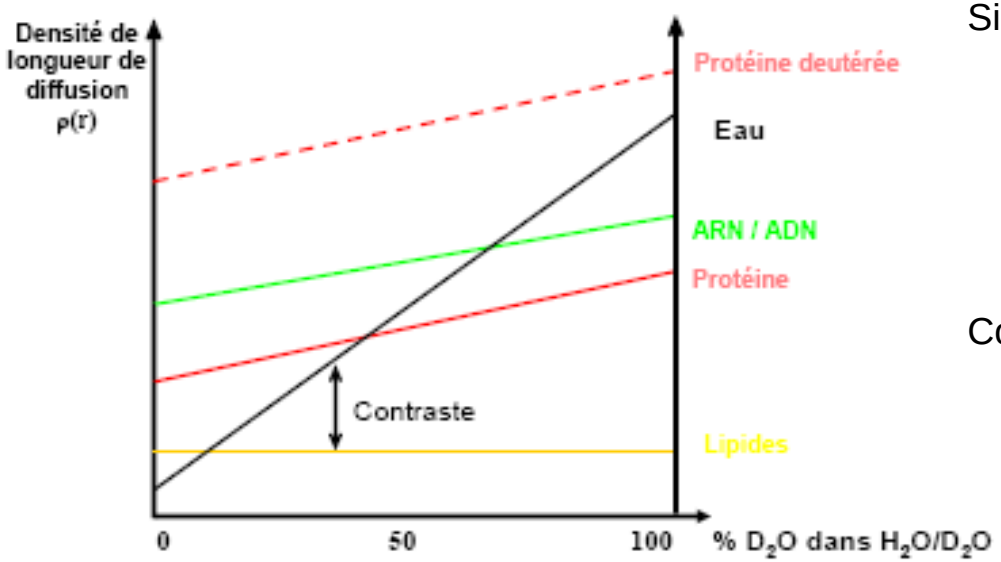
H	$-0.374 \cdot 10^{-12}$ cm
D	$0.667 \cdot 10^{-12}$ cm
C	$0.665 \cdot 10^{-12}$ cm
N	$0.940 \cdot 10^{-12}$ cm
O	$0.580 \cdot 10^{-12}$ cm

Weak negative signal from H, positive for D

Varies as a function of the number of H/D atoms

No sample dammage

Contrast variation



Signal = contrast between

macromolecule

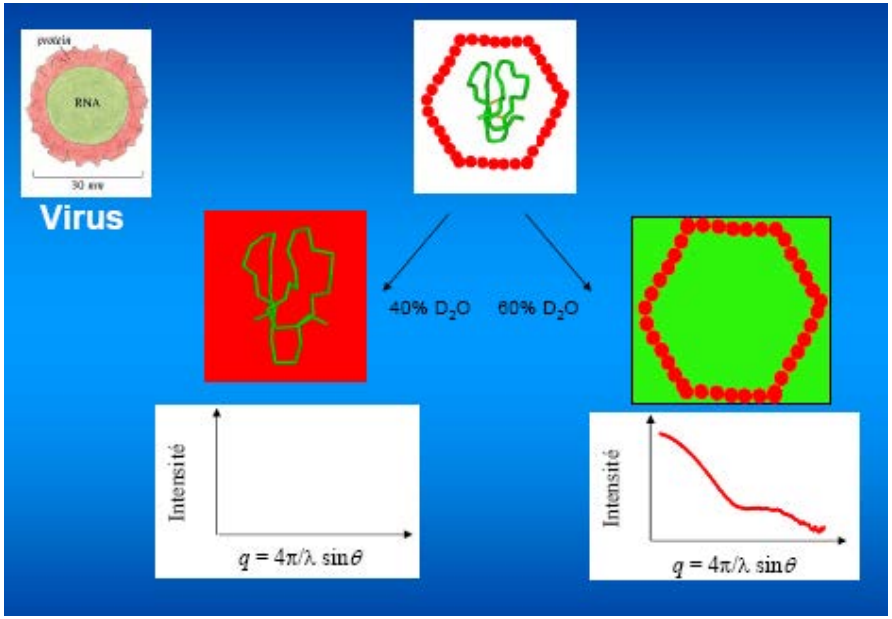
solvent

Contrast variation

Protein/NA interactions

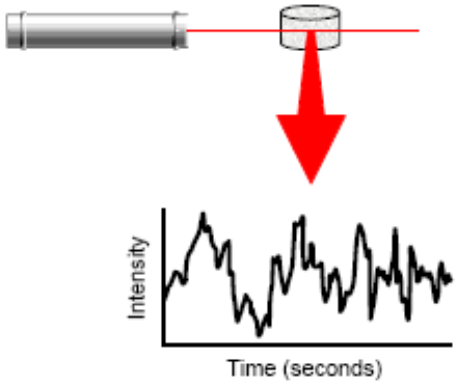
Protein lipids interactions

$$I(q) = \frac{1}{V} \left| \int_V (\rho(\vec{r}) - \rho_s)\, e^{i\vec{q}.\vec{r}}\, d^3r \right|^2$$

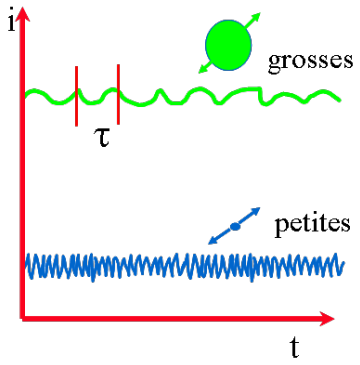


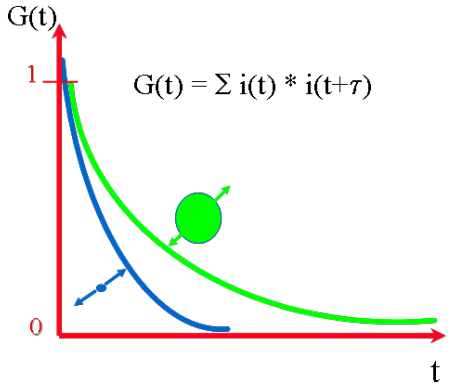
	X-rays	Neutrons	UV/visible
λ	# 1.5 A	# 1-20 A	# 5000 A
Energy	# keV	# meV	# 10eV
Source	Synchrotron	Reactor	Laser
	10^{12} to 10^{15} ph/cm^2/s	10^4 to 10^8 n/cm^2/s	10^{22} ph/cm^2/s
	Structure	Structure Internal dynamics	(Structure) SLS Dynamics DLS

Dynamic Light scattering



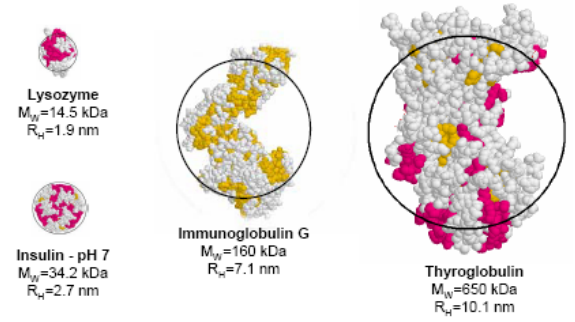
- The rate of intensity fluctuation is dependent upon the size of the particle/molecule





$$D = kT/f$$

$$f = 6p\eta r$$

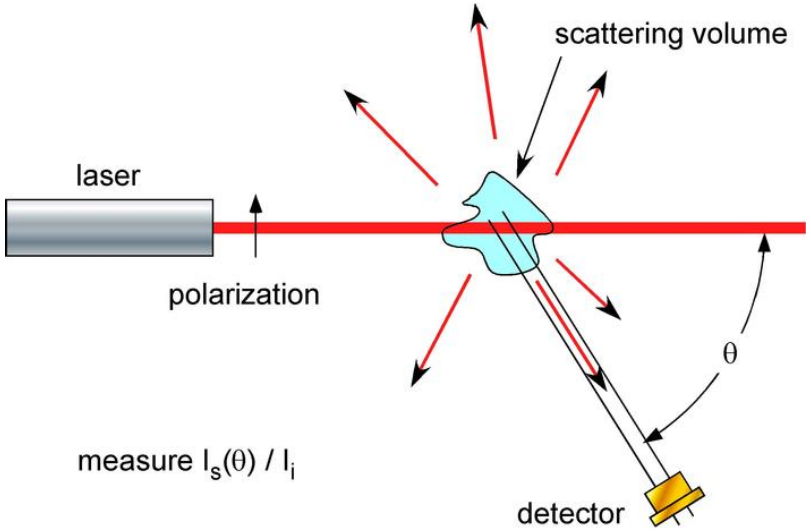


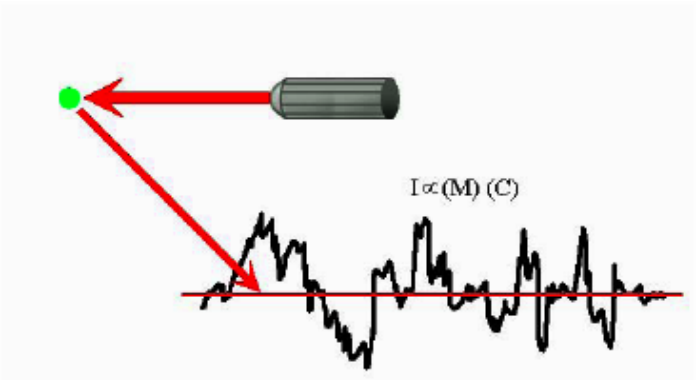
$$Vs = (M/N).[\underline{\upsilon 2} + \upsilon 1°.\ \delta 1\ ]$$

Hydrodynamic radius

Static Light Scattering

Measurement of the average scattering intensity



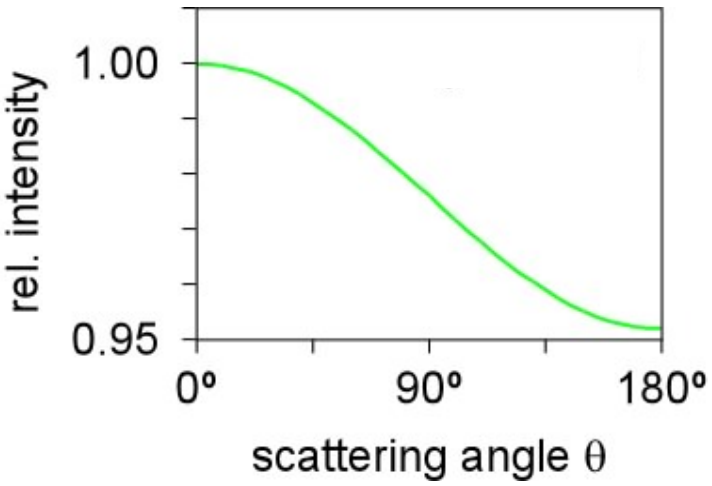


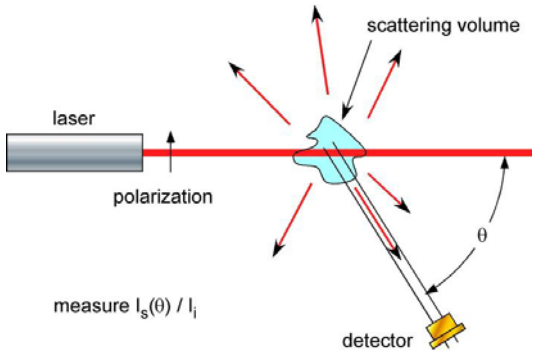
$$I(\theta)_{\text{scattered}} \propto Mc\left(\frac{dn}{dc}\right)^2 P(\theta)$$

The intensity is a function of the particule's molecular weight, its concentration, shape (form factor) and of the refractive index increment of the solution.

Form factor

In general, the scattered intensity varies with Θ, the angle between the incident beam and the detector.





$$I(\theta)_{\text{scattered}} \propto Mc\left(\frac{dn}{dc}\right)^2 P(\theta)$$

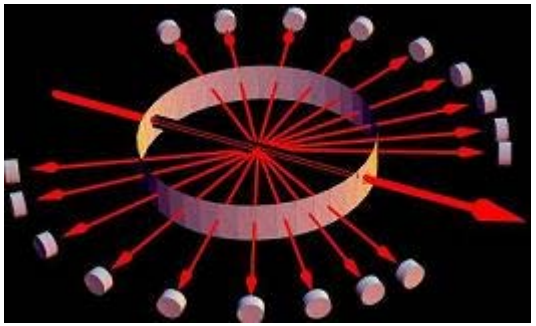
The form factor, which accounts for this dependence is also a function of the wavelenght and of the dimensions of the particule.

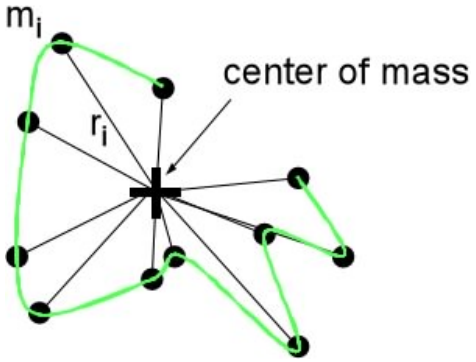
$$P(\theta) = 1 - \frac{16\pi^2 n_0^2}{3\lambda_0^2}\sin^2\left(\frac{\theta}{2}\right)\left\langle r_g^2 \right\rangle + \ldots$$

Objects smaller than the wavelength of light

The form factor does not dependent on Θ if $r_g << \lambda$ and diffusion is isotropic.

$$P(\theta) = 1 - \frac{16\pi^2 n_0^2}{3\lambda_0^2} \sin^2\left(\frac{\theta}{2}\right)\left\langle r_g^2 \right\rangle + \ldots \approx 1$$



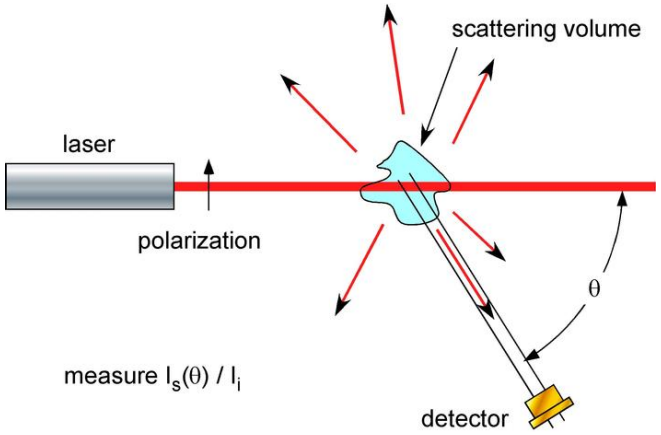


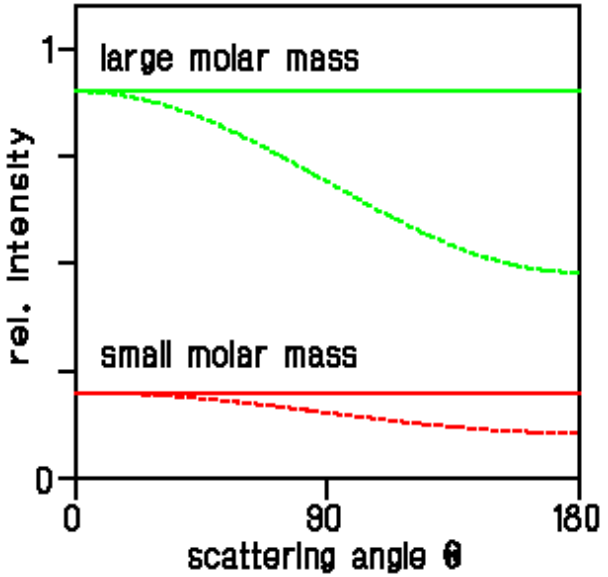
$$\frac{KC}{\Delta R_\theta} = \left(\frac{1}{M} + 2A_2C\right)P_\theta = \left(\frac{1}{M} + 2A_2C\right)$$

When the size of objects is not negligeable

The scattered intensity not only depends on M, A2 but also varies as function of θ

$$I_{\text{scattered}}(\theta) \propto R(\theta) = K^* McP(\theta)\left[1 - 2A_2 McP(\theta)\right]$$

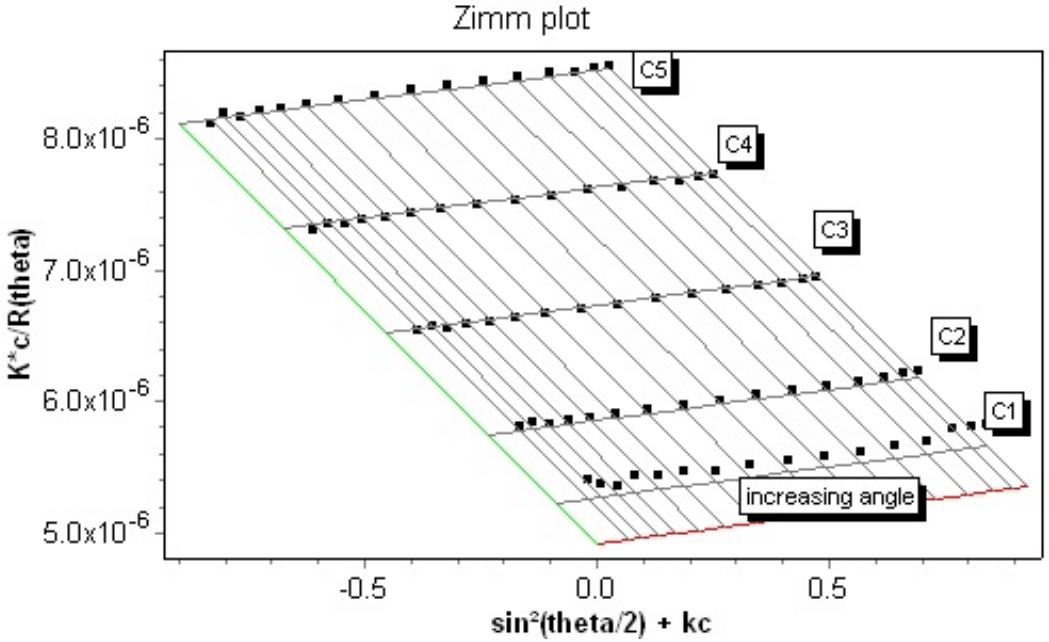


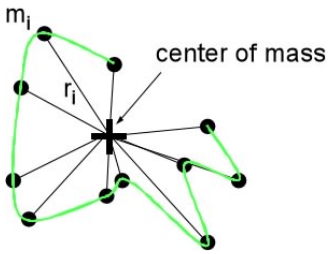


Fit of data to light scattering equation provides M, A_2 and Rg.

$$\frac{Kc}{\Delta R(\theta,c)} = \frac{1}{M_w P(\theta)} + 2A_2 c = \frac{1}{M_w}\left(1 + \frac{q^2 R_g^2}{3}\right) + 2A_2 c$$

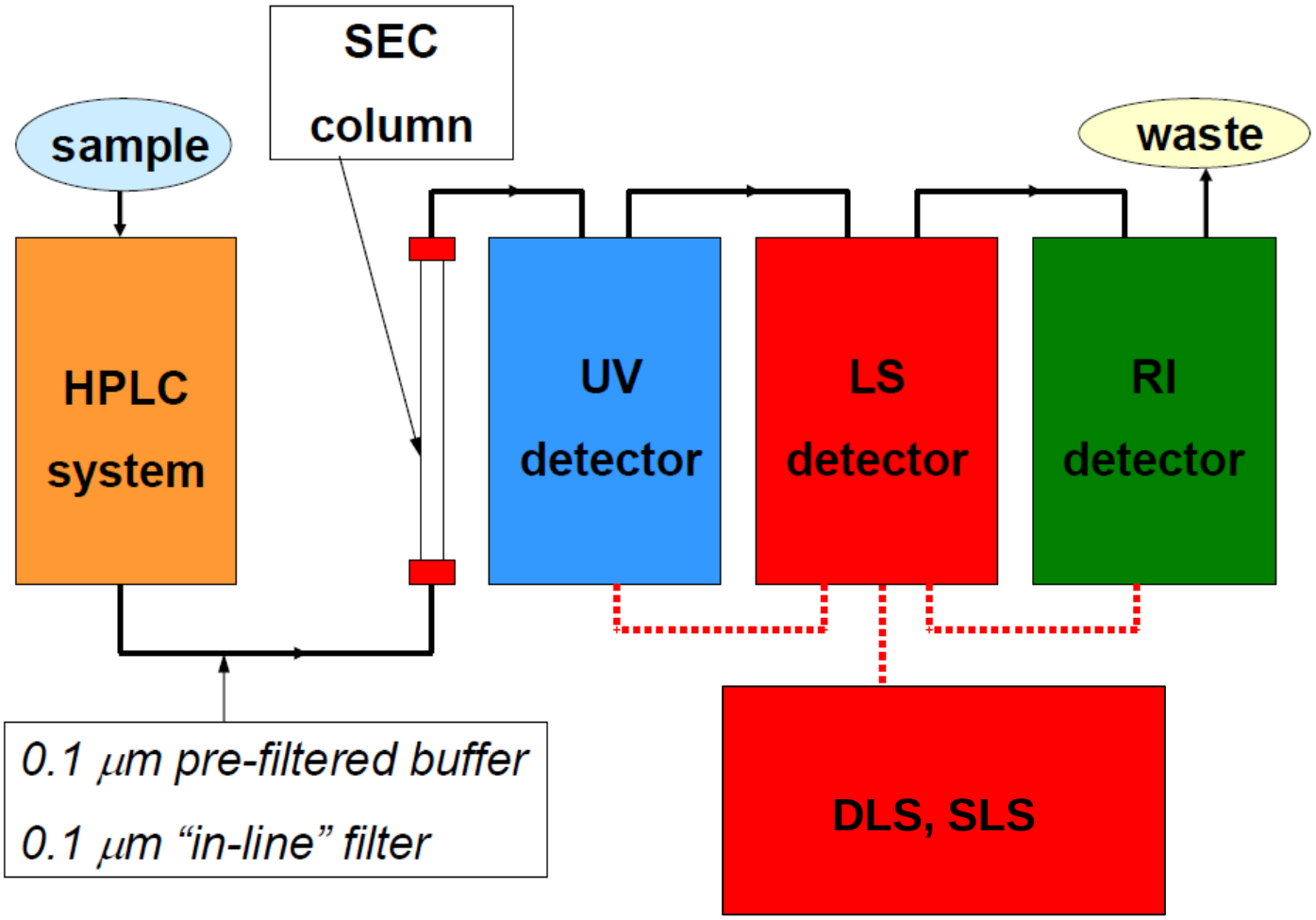
q =4πsinθ/λ





Mass	(2.032 ± 0.008) e+5 g/mol
rms radius	18.1 ± 0.7 nm
A2	(5.110 ± 0.041) e-4 mol mL/g²

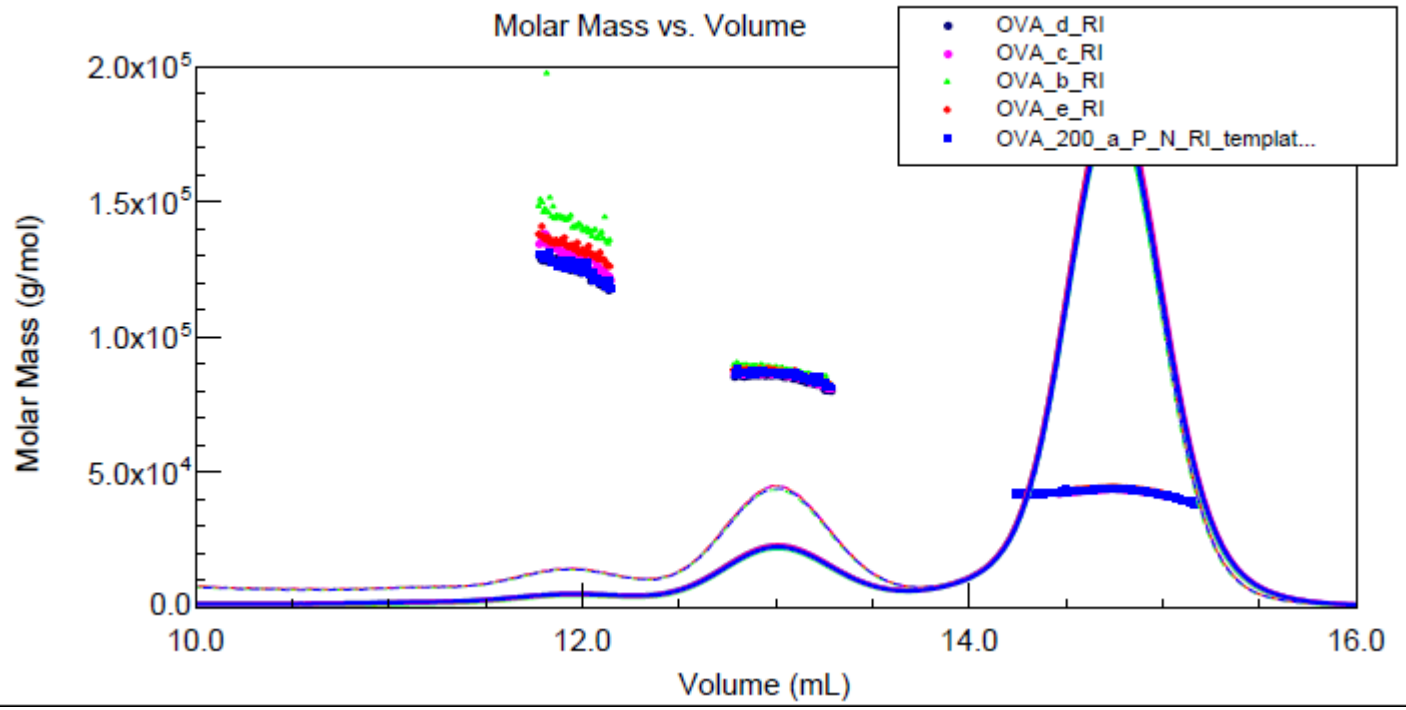
SEC-MALLS combines Light Scattering with fractionation



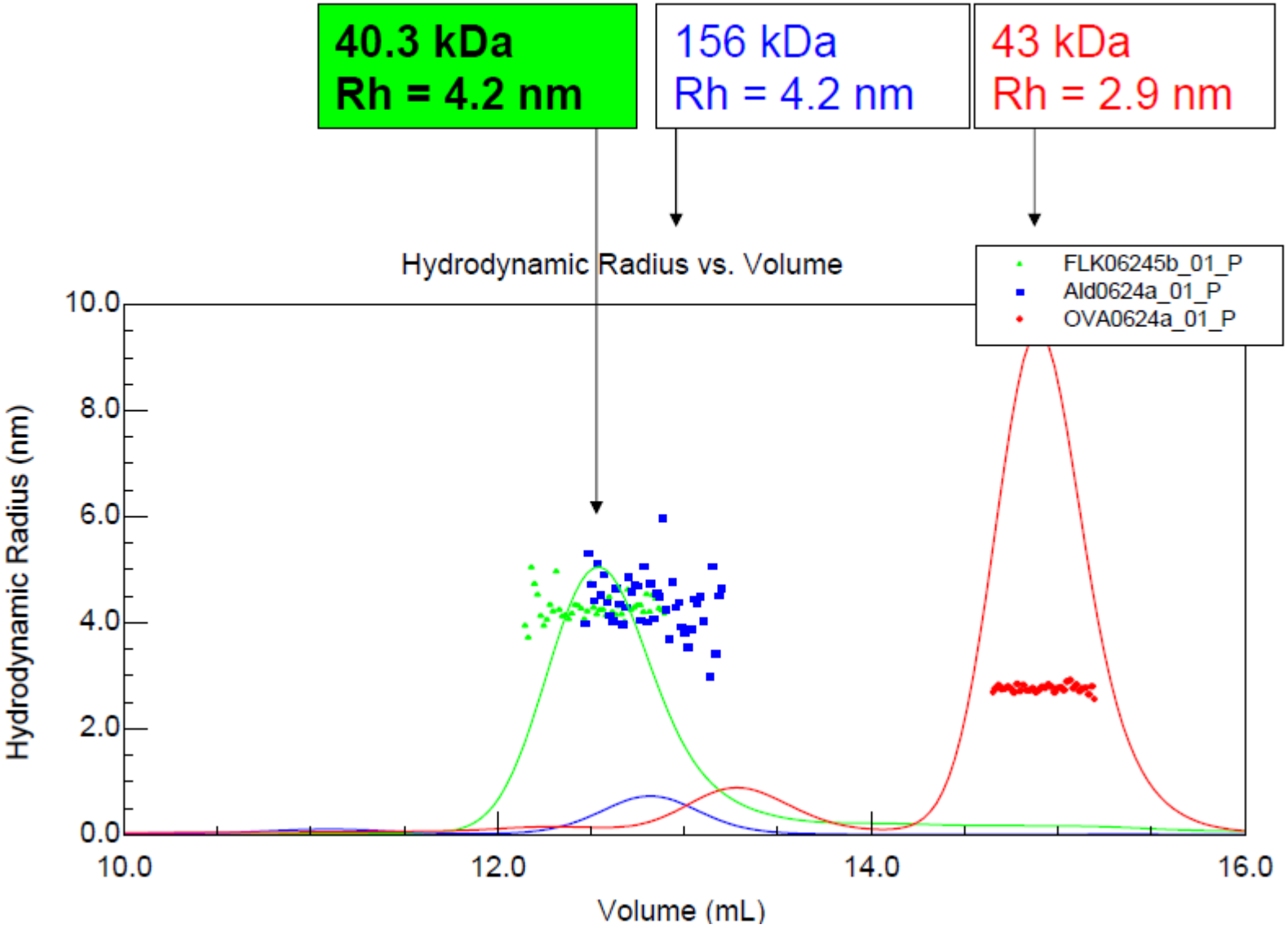
MALLS: Multi-Angle Laser Light scattering

SEC and SLS

Ovalbumin (expected MM) total mass in eluting peak	MM ± SD (5 analyses)	Precision SD (%)	Accuracy
Monomer (42.8 kDa) 178ug	43.0 ± 0.7	0.2%	0.4%
Dimer (85.6 kDa) 25ug	82.7 ± 0.4	0.5%	3%
Trimer (128.4 kDa) ? 5ug	114 ± 4	3%	11%



SEC, SLS and DLS



Analytical Ultracentrifugation

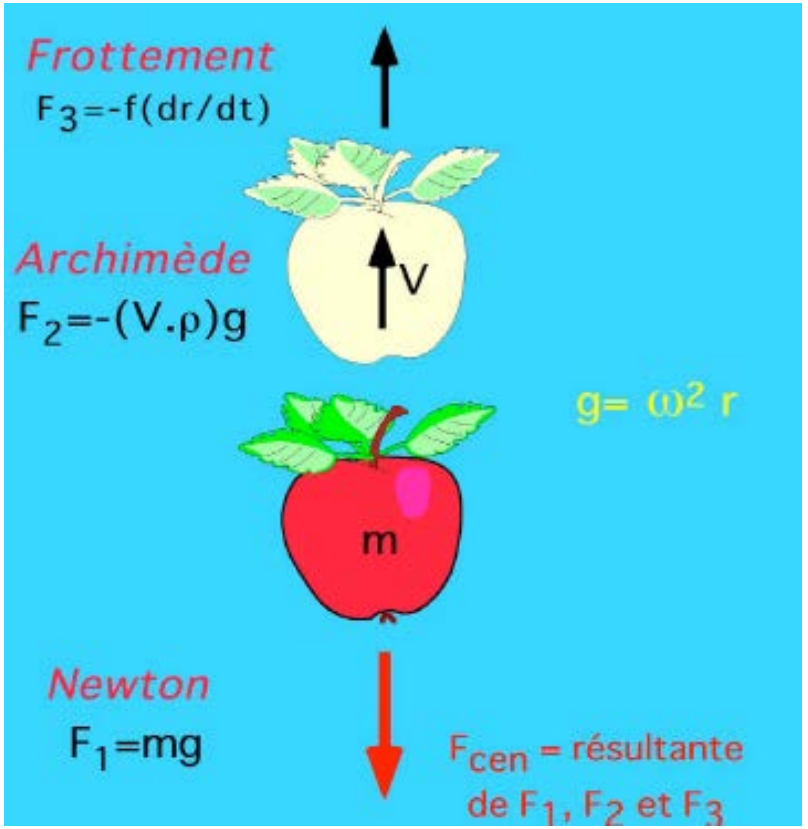
There are 3 forces acting on a sedimenting particle, buoyancy, viscous drag and centrifugal force. As soon as the rotor accelerates to a constant speed, the particle reaches terminal velocity and an equilibrium between these 3 forces is established. There are several experimental conditions and sample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

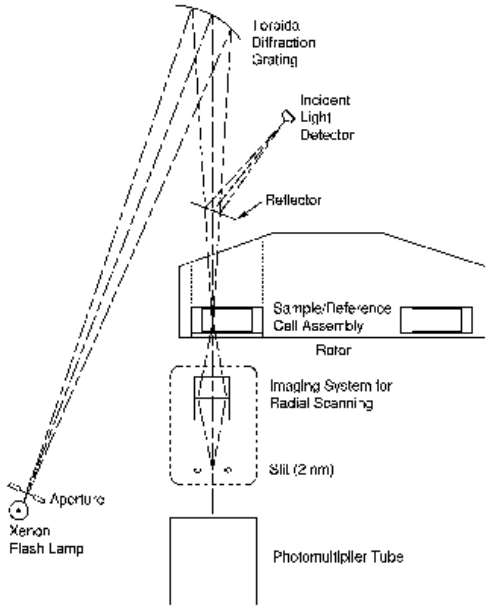
Ultracentrifuge that posseses a detection system allowing the measure of the solute concentration as a function of the distance to the rotation axis (optical density, interferrométry).

Absorbance measurement: choice of the wavelenght:

NA and protein 260 et 280 nm

ligand eg 380

peptidic bond 220: low concentrations



Limitations: detector response: inf à 0.5 OD

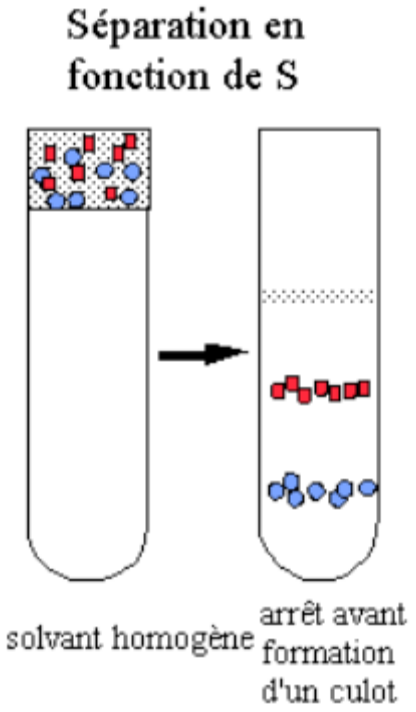
Avantage: large choice in the experimental conditions: buffer and ionic strength

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

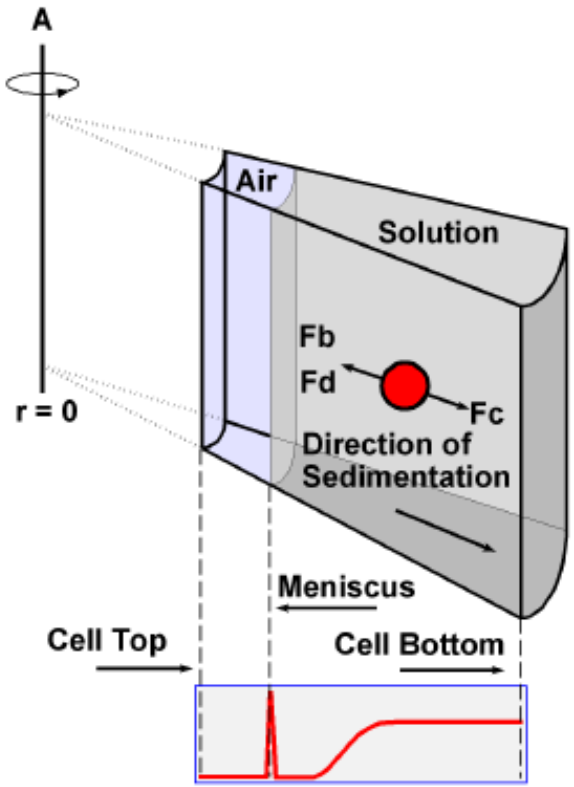


$s = dr/dt \cdot (1/w^2r)$

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

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:

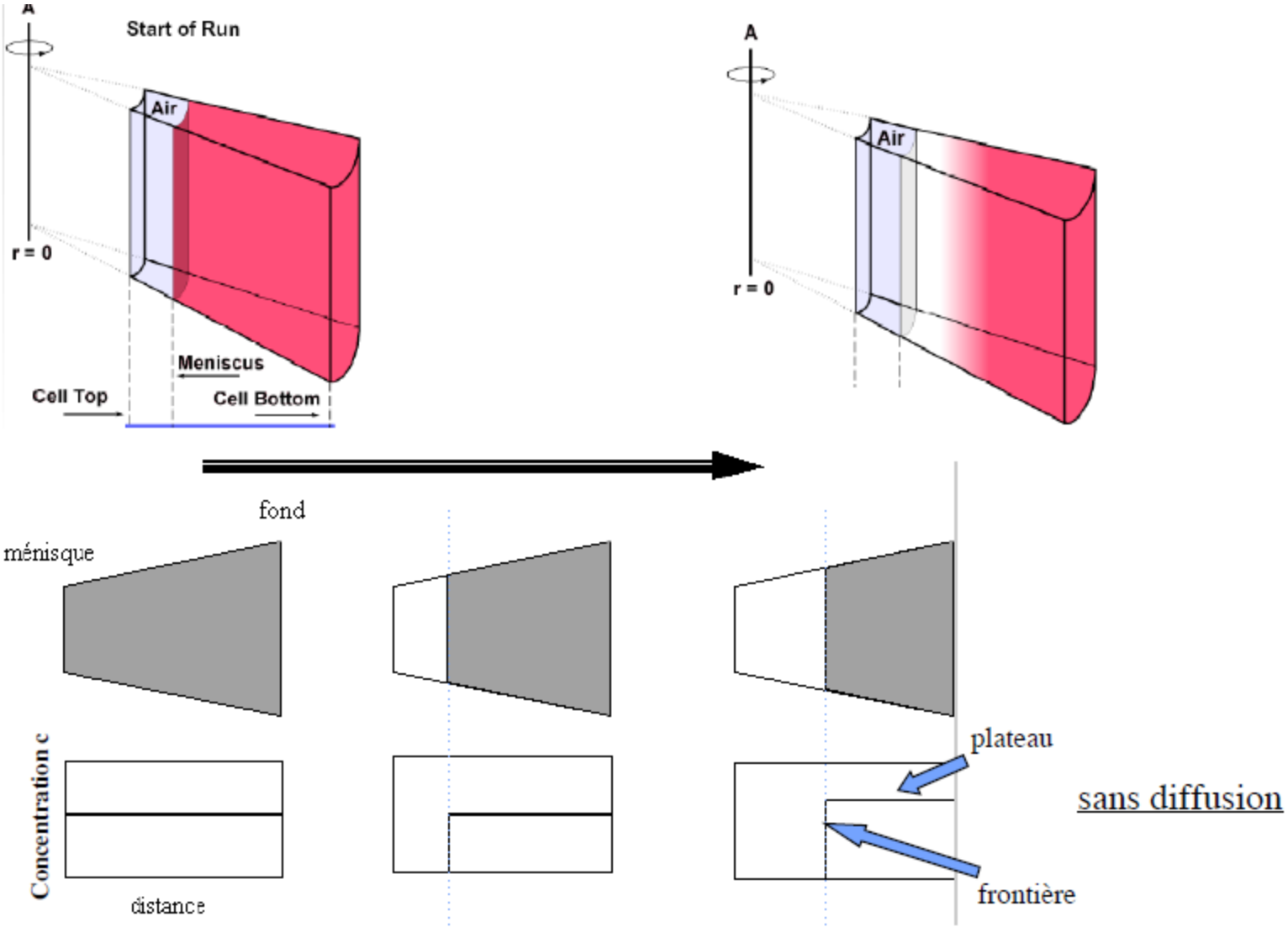
$Fc - Fb - Fd = 0$

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

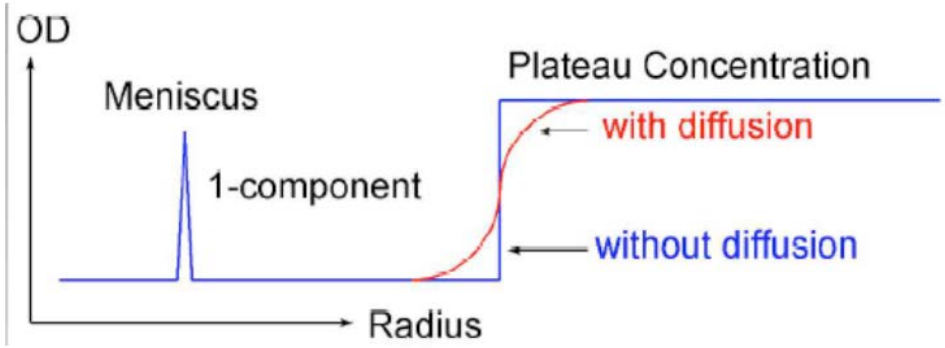
Explanation:

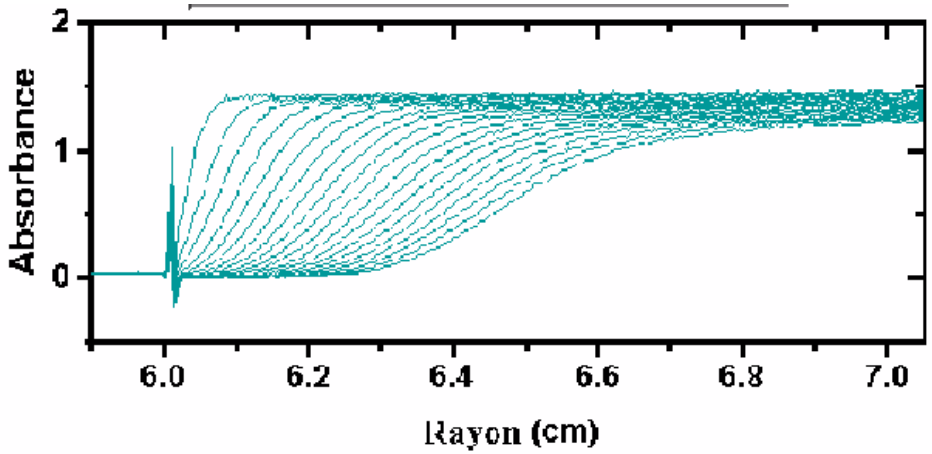
Fb is the buoyancy force - the force required to displace the buffer surrounding the solute, and m_0 is the mass of the displaced solvent.

In practice, one can measure how the sedimentation boundary moves

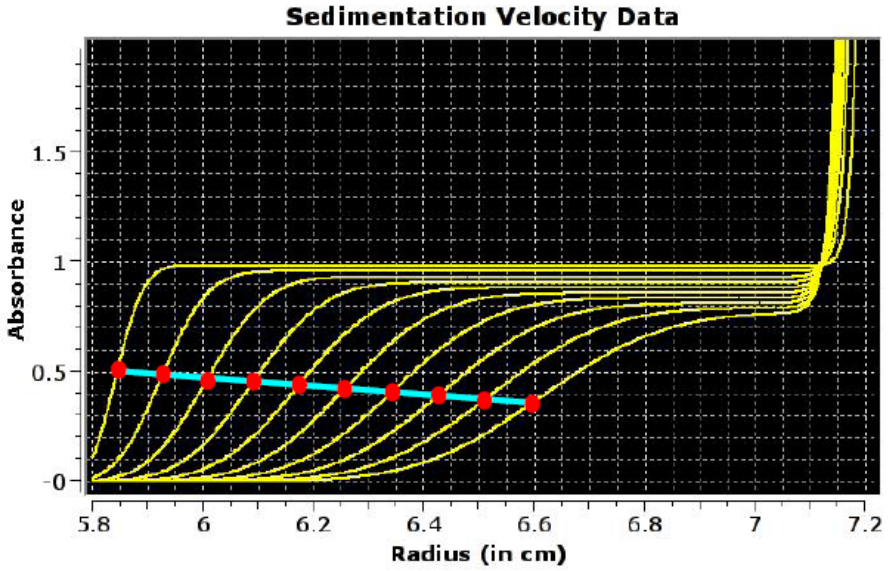


Diffusion impacts on the shape of sedimentation boundaries
They are recorded at regular,intervals





How does the midpoint of the boundary move?

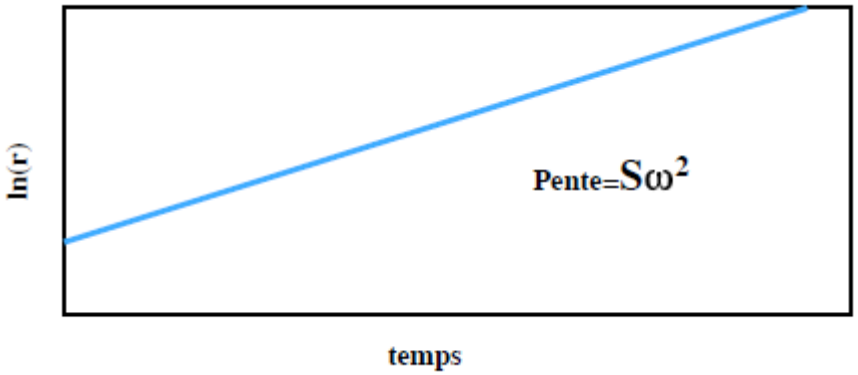


$s = dr/dt \cdot (1/\omega^2 r)$

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

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

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



Volume (occupied by the macromolecule) = Mass * Specific Volume
Mass of displaced solvant = Volume * Volumic Mass

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

Substitute the mass of the solvent, m_0,
with the mass of the solute, *m*

$$m_0 = m\bar{v}\rho, \quad Fb = \omega^2 r m \bar{v}\rho$$

Rearrange the force equation:
Fc - Fb - Fd = 0 **and substitute**

$$\omega^2 r m - \omega^2 r m \bar{v}\rho = fv$$

Place molecular parameter on one side
and experimental parameters on the other

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

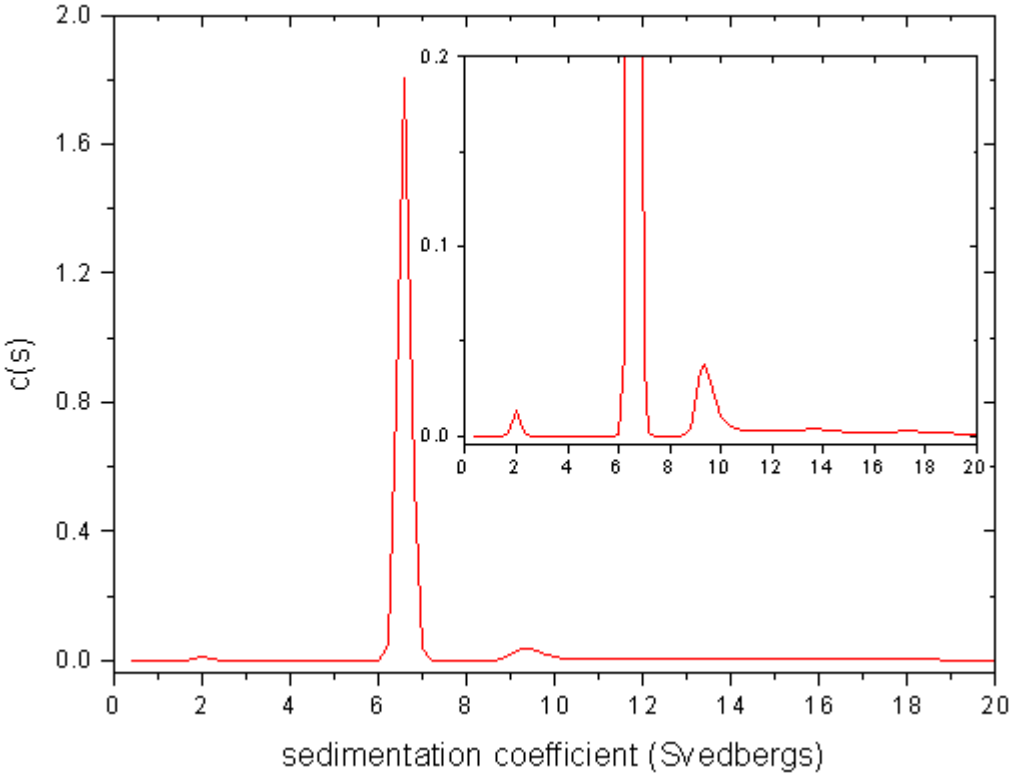
Put into molar units by multiplying with
Avogadro's number, *N*

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

Important: The sedimentation coefficient depends on both ***M*** and ***f***

The sedimentation coefficient is directly proportional to the molecular weight of the solute, and inversely proportional to the frictional coefficient of the solute. A large molecular weight increases the sedimentation speed, while an irregular shape will slow it down.

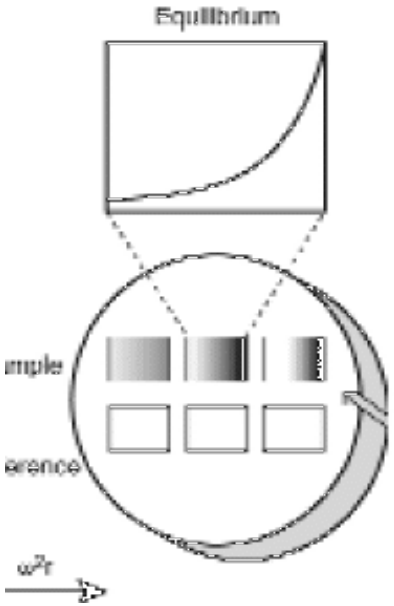
A real case

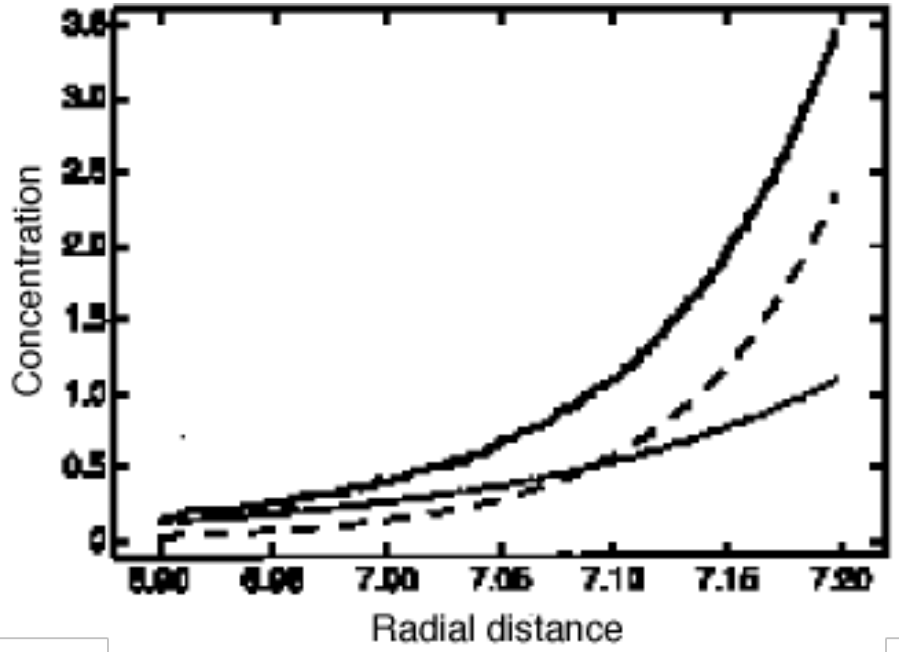

c(s)
sedimentation coefficient (Svedbergs)
2.0
1.6
1.2
0.8
0.4
0.0
0
2
4
6
8
10
12
14
16
18
20
0.2
0.1
0.0

Equilibrium Sedimentation

At low speed, diffusion is not negligeable

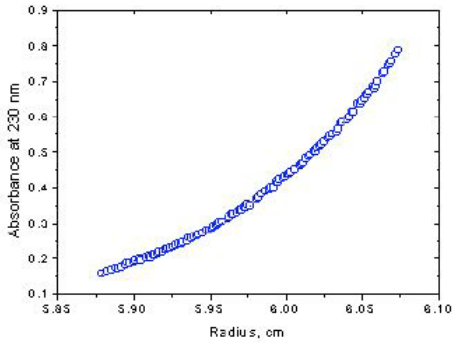
The system reaches an equilibrium: centrigulation force = diffusion force





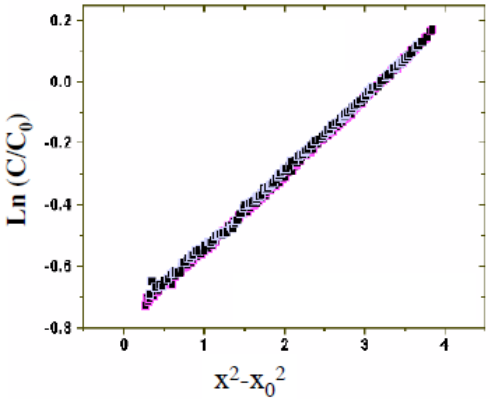
$$\frac{d\ln(C_r)}{dr^2} = \frac{M(1-r\bar{\rho})\omega^2}{2RT}$$

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



Representation of Ln(Cx/Co) as a function of x^2-xo^2 yields a linear function

The slope is only a function of M, ω, v and ρ



r (cm^3/g) volume specific partial (hydrated) of the macromolecule

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

cr macromolecule concentration

r distance to the axis

A real case

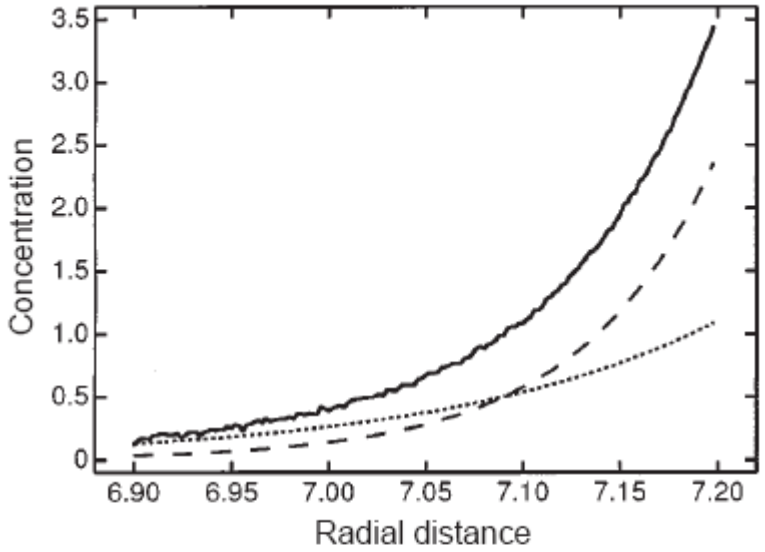


FIGURE 3

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

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

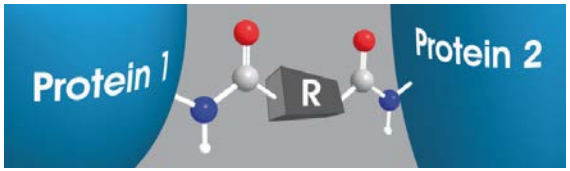
Hetero-association

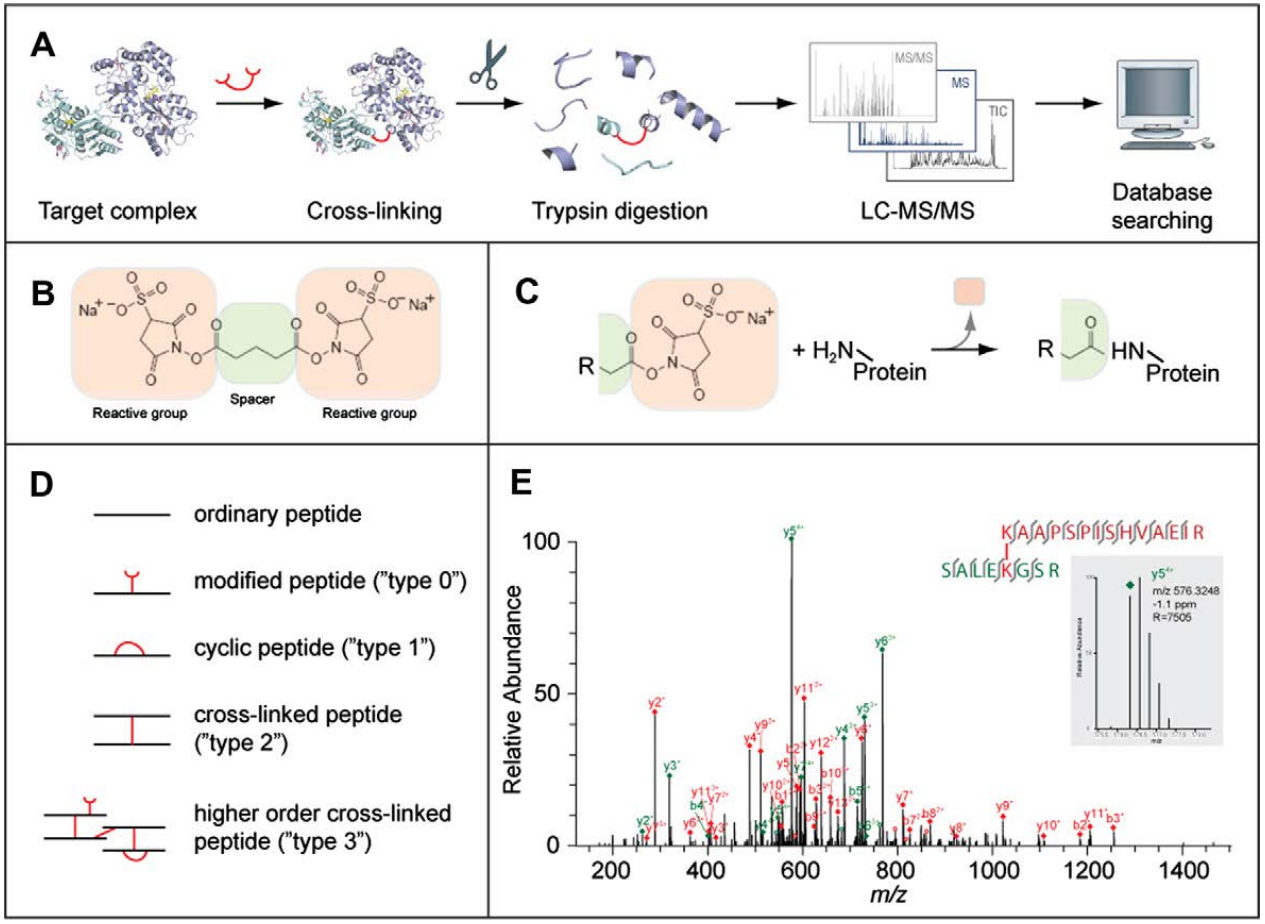
$$a(r) = c_{A,o}\varepsilon_A d \exp\left[\frac{M_A^*\omega^2}{2RT}(r^2 - r_o{}^2)\right] + c_{B,o}\varepsilon_B d \exp\left[\frac{M_B^*\omega^2}{2RT}(r^2 - r_o{}^2)\right] + c_{A,o}c_{B,o}K_{AB}(\varepsilon_A + \varepsilon_B)d \exp\left[\frac{(M_A^* + M_B^*)\omega^2}{2RT}(r^2 - r_o{}^2)\right] + \delta$$

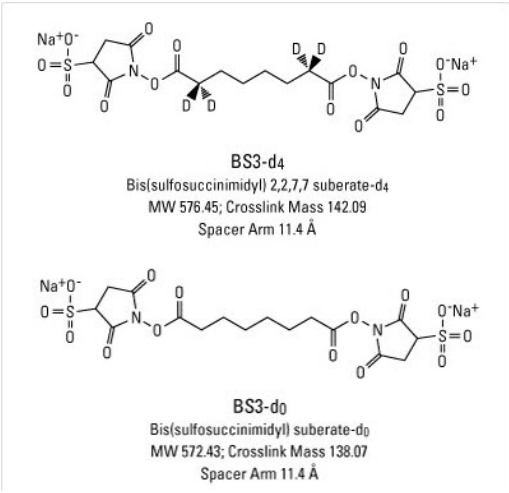
Structural data used in integrative modeling

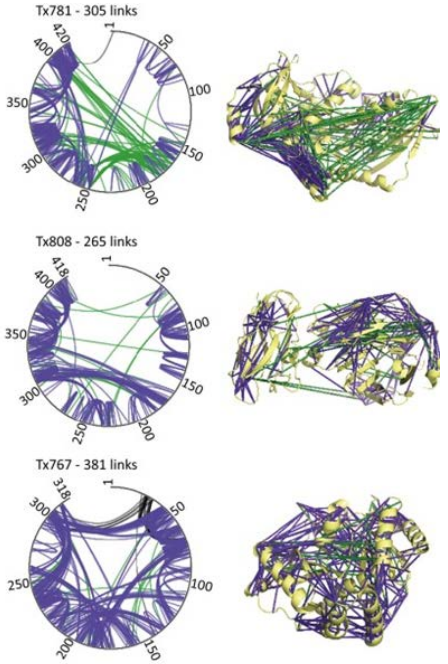
Atomic structures of parts of the system	X-ray and neutron crystallography, NMR, Cryo-EM/ET, Comparative modelling and molecular docking
Composition and components positions	Purification from source with gel analysis or MS, Electron microscopy and tomography, gold labelling, Super resolution microscopy,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

X-link/MS experiments





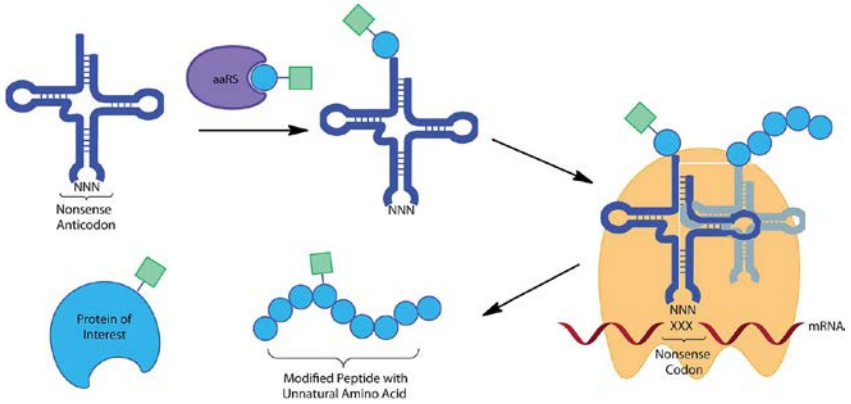


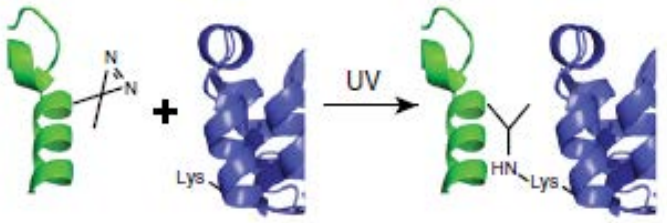


Juri Rappsilber, 2010, 2012

Expansion of the genetic code and incorporation of non-natural amino acids

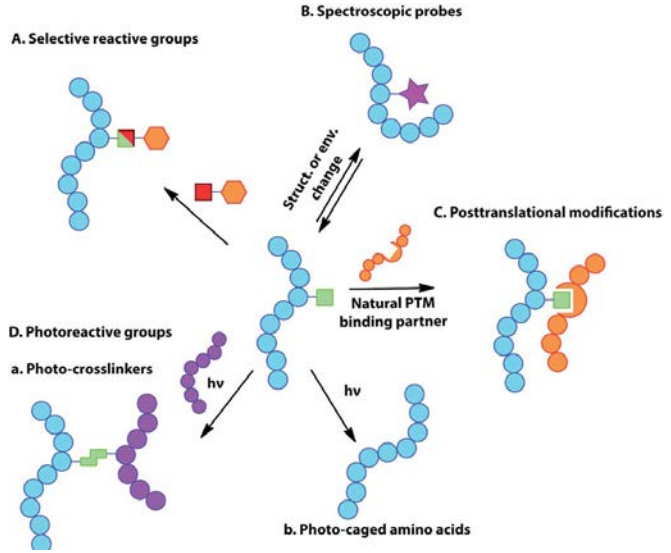
Incorporate photo-activable amino acids in proteins:



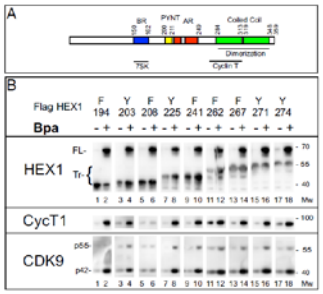


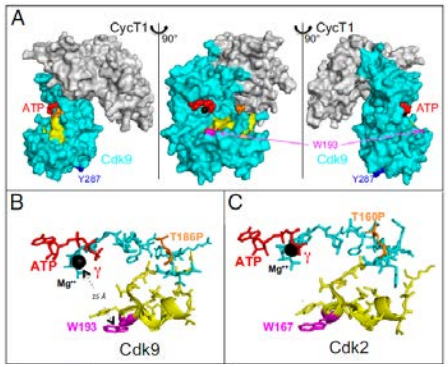
pTefb = cdk9/cyclinT/Hexim + 7SK RNA

Map hexim peptides that bind and inhibit cdk9



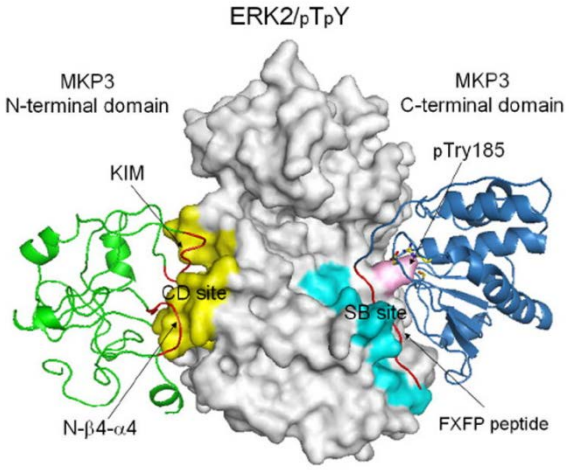
An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb



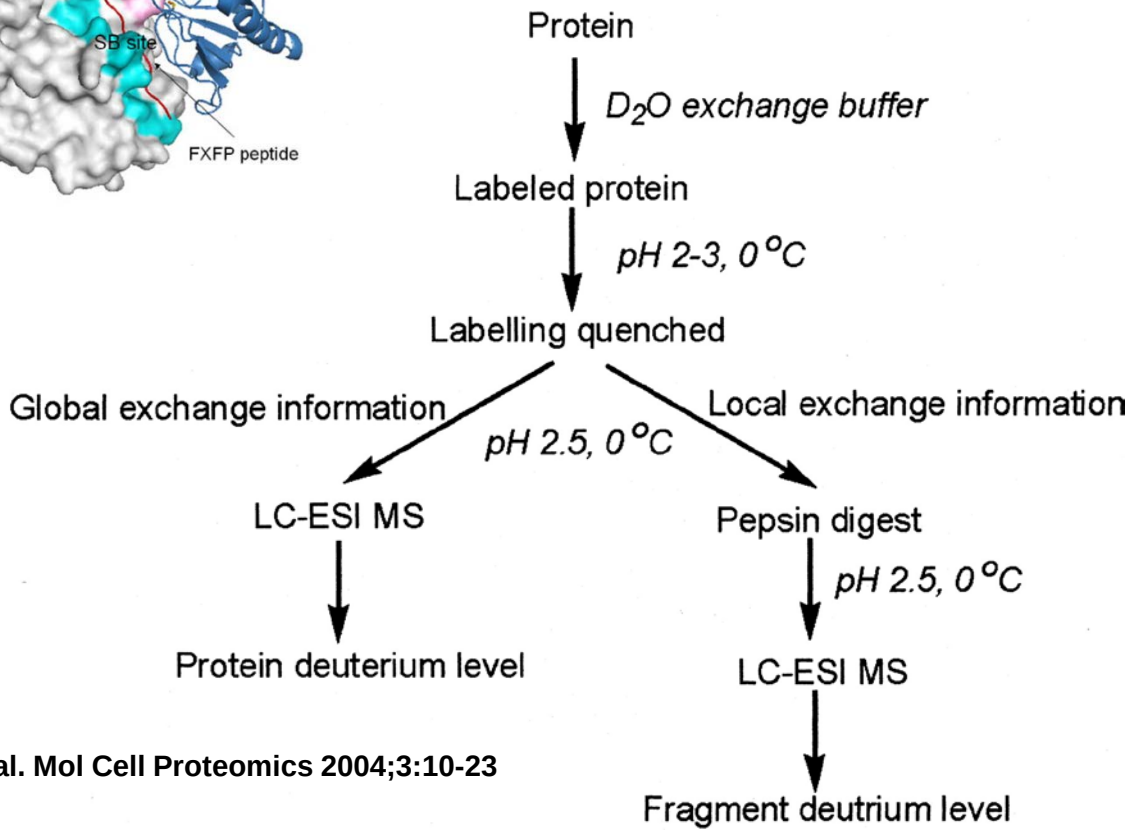


Kobbi et al. 2016

H/D exchange MS experiments



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



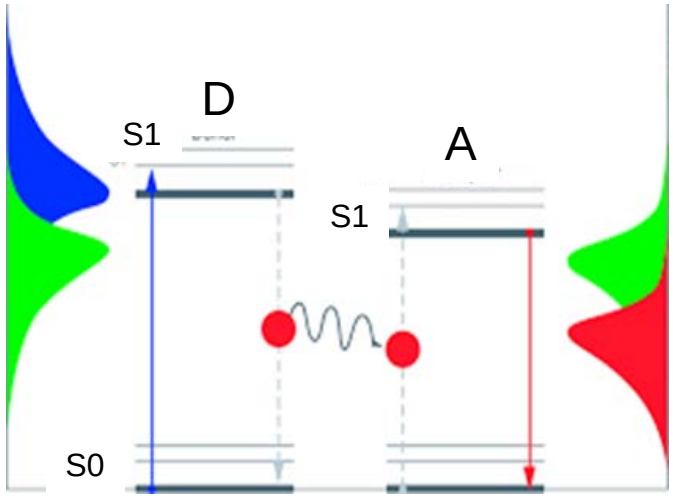
Yan X et al. Mol Cell Proteomics 2004;3:10-23

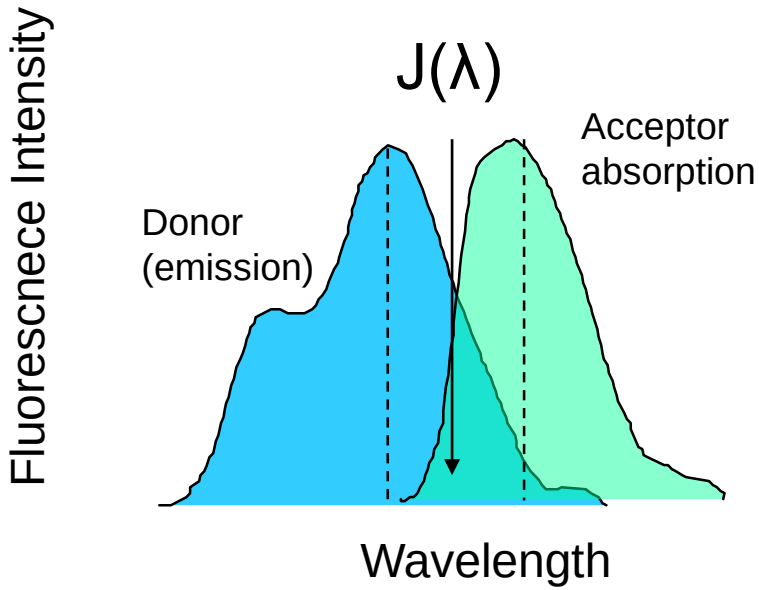
Förster Resonance Energy Transfer (FRET)

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

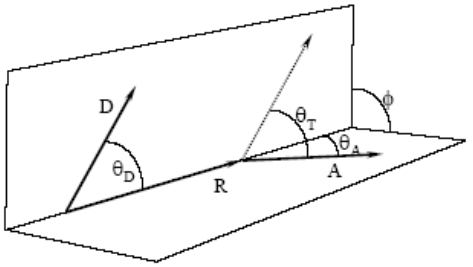
The absorption spectrum of the acceptor must overlap fluorescence emission spectrum of the donor





FRET strongly depends on:

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

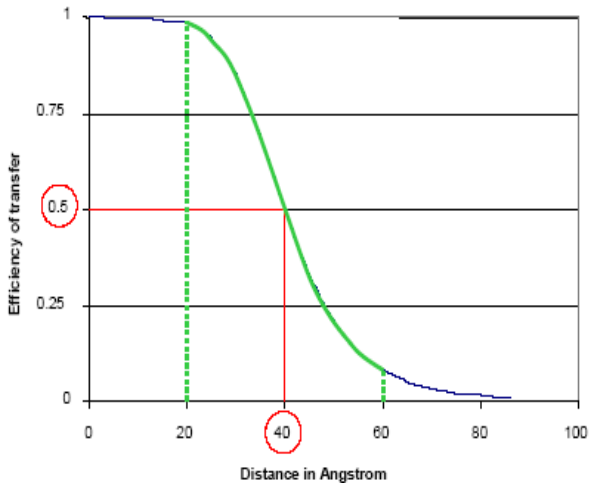


Energy transfer studies give information about

- distance between groups
- orientation of two groups and
- the refractive index between two groups

$$\kappa = \vec{D}\cdot\vec{A} - 3\left(\vec{D}\cdot\vec{R}\right)\left(\vec{A}\cdot\vec{R}\right)$$

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%, $R=R_0$

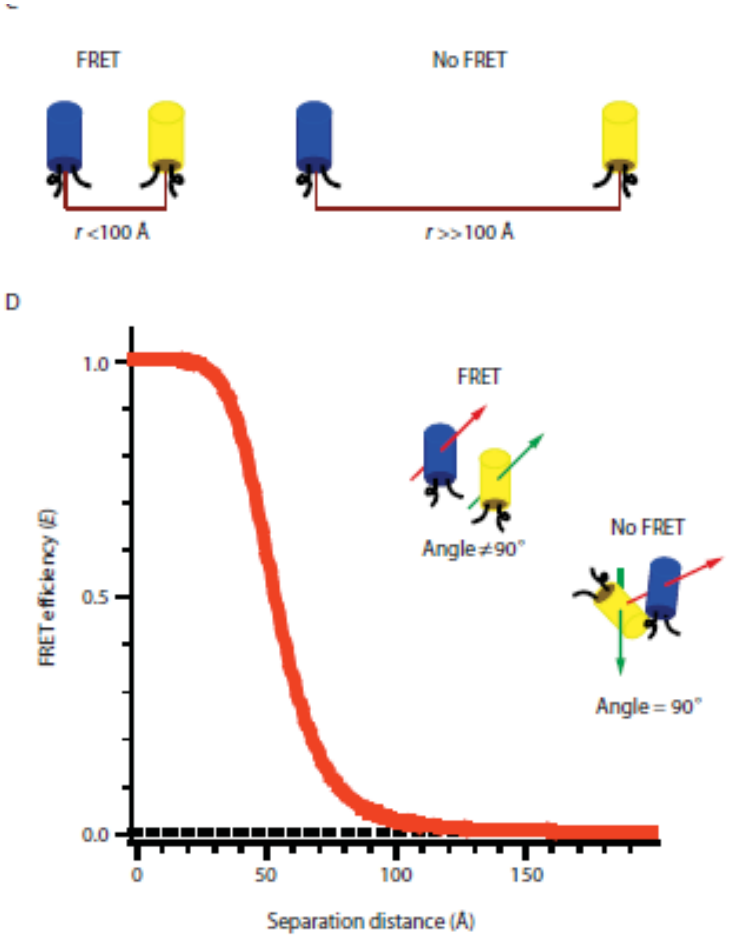
Distances can generally be measured between ~0.5 R_0 and ~1.5R_0

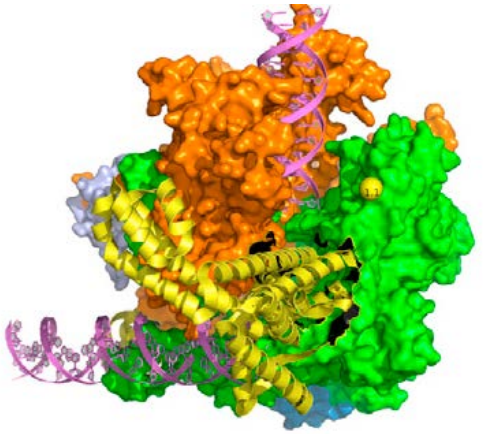
$$E = \frac{k_{DA}}{k_{DA} + k_{fD}} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}$$

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

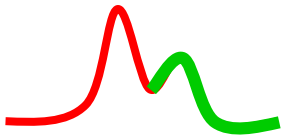




Structural organization of the bacterial (Thermus aquaticus) RNA polymerase-promoter open complex obtained by FRET (Mekler et al., 2002) was subsequently validated by a crystal structure (Zhang et al., 2012).

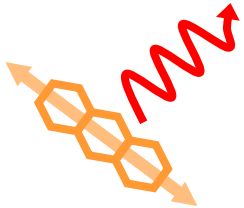
Fluorescence properties that can be measured

- spectra (environmental effects)

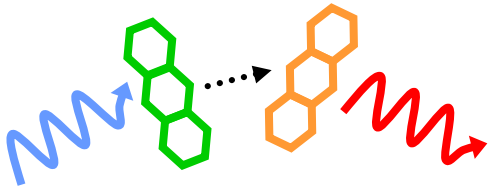


- fluorescence life times

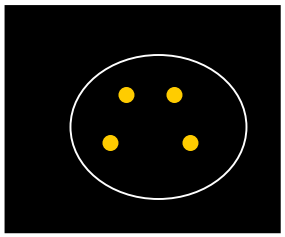
- polarization (orientation and dynamics)



- excitation transfer (distances -> dynamics)

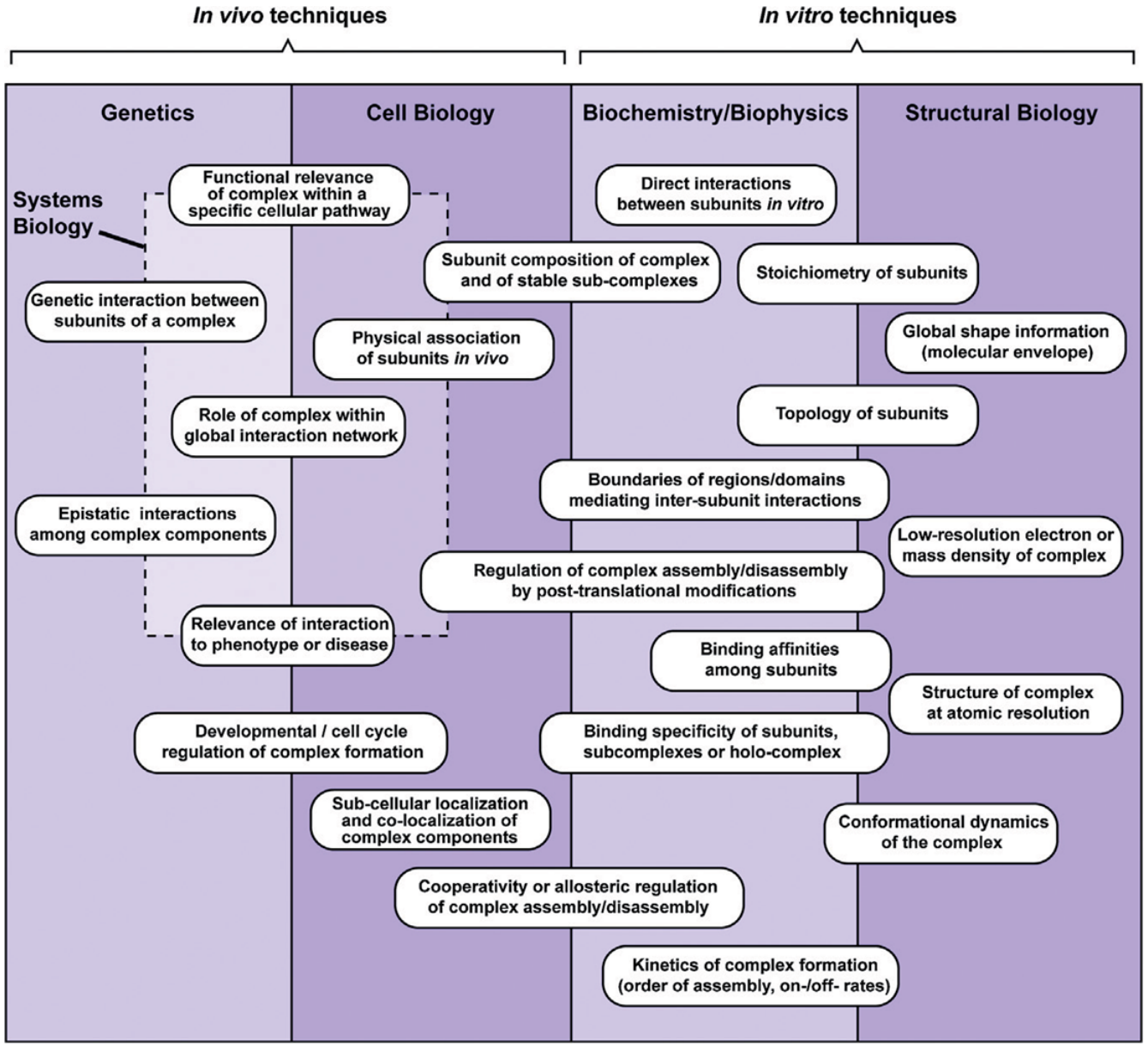


- location of fluorescence



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

END