IGBMC
Institut de génétique et de biologie moléculaire et cellulaire

PREPARATION AND CHARACTERIZATION OF BIOLOGICAL SAMPLES FOR STRUCTURAL STUDIES

SOLUBILITY, STABILITY AND AGGREGATION

Marc Ruff

Integrated structural biology department
Chromatin stability and DNA mobility team

IGBMC, Illkirch, France

ruff@igbmc.fr

Preparation of biological sample

- **Extract a single, pure, homogenous, soluble and functional protein / protein complex from a high complex mixture**

“Cellular crowding”

Intracellular medium: complex mixture

- **Peptides**
  - Amino acids chains (< 6 KD)
- **Proteins**
  - Biological macromolecules (> 6 KD)
  - Nucleic acids
  - Chains of nucleotide triphosphate
- **Polysaccharides**
- **Lipids**
- **Small molecules**
- **.....**

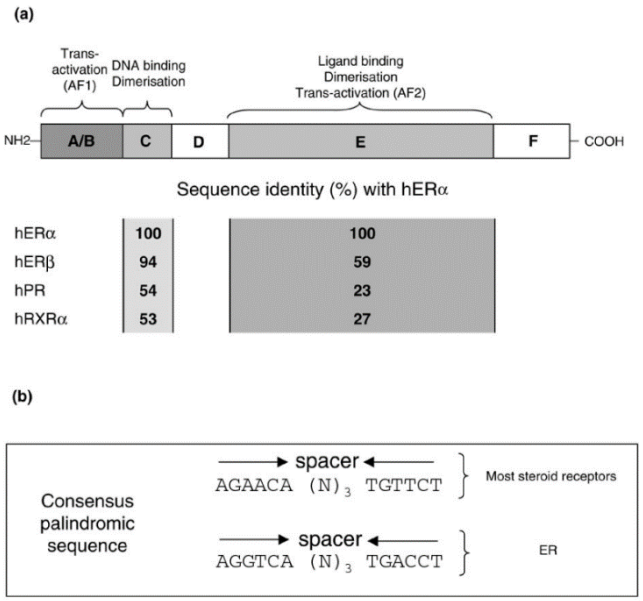
Protein families

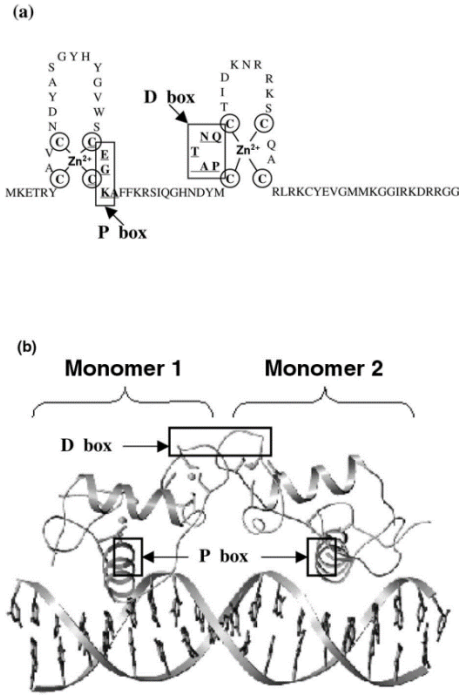
- Globular proteins
  - Generally soluble spherical proteins in water (they form a colloid)
- Fibrous proteins
  - Extended structure, Insoluble in aqueous solvents or in Lipid Bilayers, Structural function in the body or cell (tendons, bone, muscle, ligaments, hair, skin).
  - Examples: Collagen (the most abundant proteins in vertebrates, connective tissues: cartilage), Keratin (hair, nails, feathers...), Elastin (ligaments, blood vessels), Fibroid (silk, spider webs)
- Membrane proteins
  - Membrane proteins are attached or associated with cell membranes or organelles
  - They are the target of more than 50% of the modern medical drugs
  - 20 to 30% of the genes of most genomes encode membrane proteins.
- Intrinsically disordered proteins, regions
  - Intrinsically disordered proteins / regions occupy a continuum of conformational space, ranging from highly unstructured, through molten globule, to local disorder within an otherwise folded domain.
- Multi-protein complexes

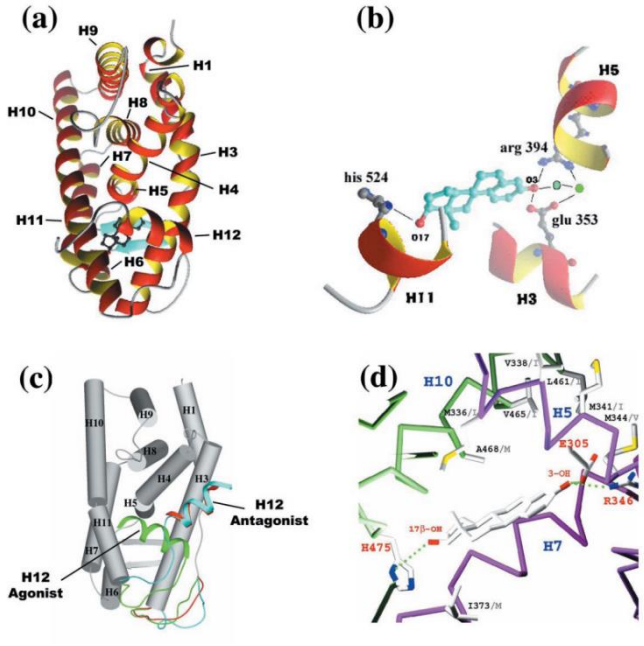
Proteins organization into domains

- **Protein organization in structural and functional domains**
  - Conserved part of a protein sequence forming an independent structure that can evolve, operate and exist independently of the rest of the protein chain.
  - Each domain forms a compact three-dimensional structure and is often independently stable and folded.
  - Many proteins consist of several structural domains.
  - A single domain can appear in a variety of different proteins.
  - Molecular evolution uses domains as building blocks, and these can be recombined in different arrangements to create proteins with different functions.
  - The domains length varies between about 25 to 500 amino acids. The shorter domains like zinc fingers are stabilized by metal ions or disulfide bridges.
  - Molecular evolution uses domains as building blocks, and these can be recombined in different arrangements to create proteins with different functions.
  - Domains often form functional units.

Proteins organization into domains: nuclear receptors







Marc Ruff, Monique Gangloff, Jean Marie Wurtz and Dino Moras. Estrogen receptor transcription and transactivation. Structure–function relationship in DNA- and ligand-binding domains of estrogen receptors. *Breast Cancer Res* 2000, **2**:353–359

Intrinsically disordered proteins, regions

While many proteins need to adopt a well-defined structure to carry out their function, a large fraction of the proteome of any organism consists of polypeptide segments that are not likely to form a defined three-dimensional structure, but are nevertheless functional. These protein segments are referred to as **intrinsically disordered regions (IDRs)**.

Proteins sequences in a genome can be viewed as modular because they are made up of combinations of structured and disordered regions. **Proteins without IDRs are called structured proteins, and proteins with entirely disordered sequences that do not adopt any tertiary structure are referred to as intrinsically disordered proteins (IDPs).** The majority of eukaryotic proteins are made up of both structured and disordered regions, and both are important for the repertoire of functions that a protein can have in a variety of cellular contexts.

IDRs and IDPs are prevalent in eukaryotic genomes. For instance, 44% of human protein-coding genes contain disordered segments of >30 amino acids in length.

Johnny Habchi; Peter Tompa; Sonia Longhi; Vladimir N. Uversky; *Chem. Rev.* **2014,** 114, 6561-6588.

Robin van der Lee; Marija Buljan; Benjamin Lang; Robert J. Weatheritt; Gary W. Daughdrill; A. Keith Dunker; Monika Fuxreiter; Julian Gough; Joerg Gsponer; David T. Jones; Philip M. Kim; Richard W. Kriwacki; Christopher J. Oldfield; Rohit V. Pappu; Peter Tompa; Vladimir N. Uversky; Peter E. Wright; M. Madan Babu; *Chem. Rev.* **2014,** 114, 6589-6631. DOI: 10.1021/cr400525m.

Intrinsic disorder and protein interaction networks

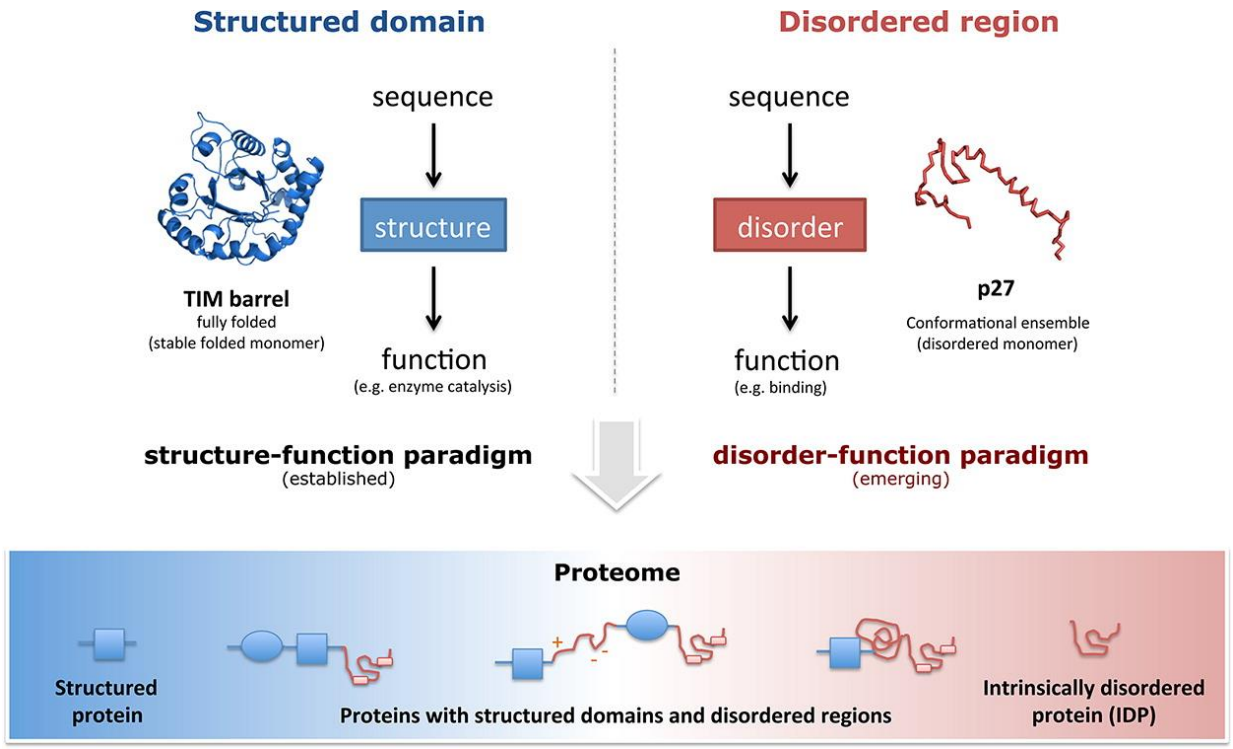
Dunker et al, Flexible nets: The roles of intrinsic disorder in protein interaction networks, FEBS J. 2005, 272, 5129-5148)

- plasticity of the disorder regions allows the binding of multiple partners
- Few proteins bind to many partners (Hub proteins) but most proteins bind to few partners
- Many disordered regions are modified post-transcriptionaly (Phosphorylation, Acetylation, ubiquitination, proteolytic cleavage)

IDPs participate in important regulatory functions in the cell, including transcription, translation, the cell cycle, and numerous signal transduction events. Disordered proteins often undergo coupled folding and binding transitions upon interaction with their cellular targets.

The lack of stable globular structure confers numerous functional advantages, but not without cost; many disordered proteins are associated with amyloid disease and with chromosomal translocations in cancer.

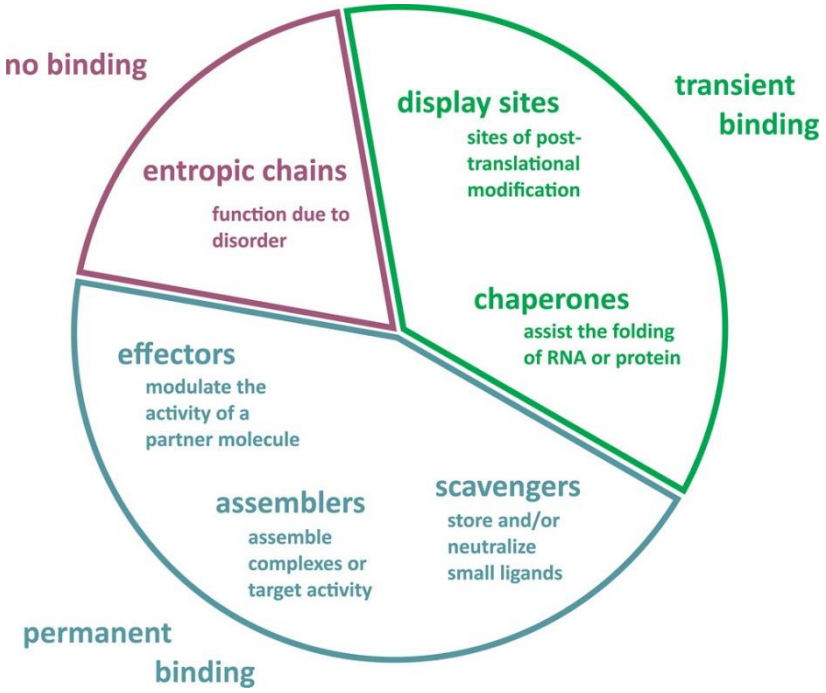
Intrinsically disordered proteins: function, folding, and flexibility



Structured domains and intrinsically disordered regions (IDRs) are two fundamental classes of functional building blocks of proteins. The synergy between disordered regions and structured domains increases the functional versatility of proteins.

Robin van der Lee; Marija Buljan; Benjamin Lang; Robert J. Weatheritt; Gary W. Daughdrill; A. Keith Dunker; Monika Fuxreiter; Julian Gough; Joerg Gsponer; David T. Jones; Philip M. Kim; Richard W. Kriwacki; Christopher J. Oldfield; Rohit V. Pappu; Peter Tompa; Vladimir N. Uversky; Peter E. Wright; M. Madan Babu; *Chem. Rev.* **2014,** 114, 6589-6631. DOI: 10.1021/cr400525m.

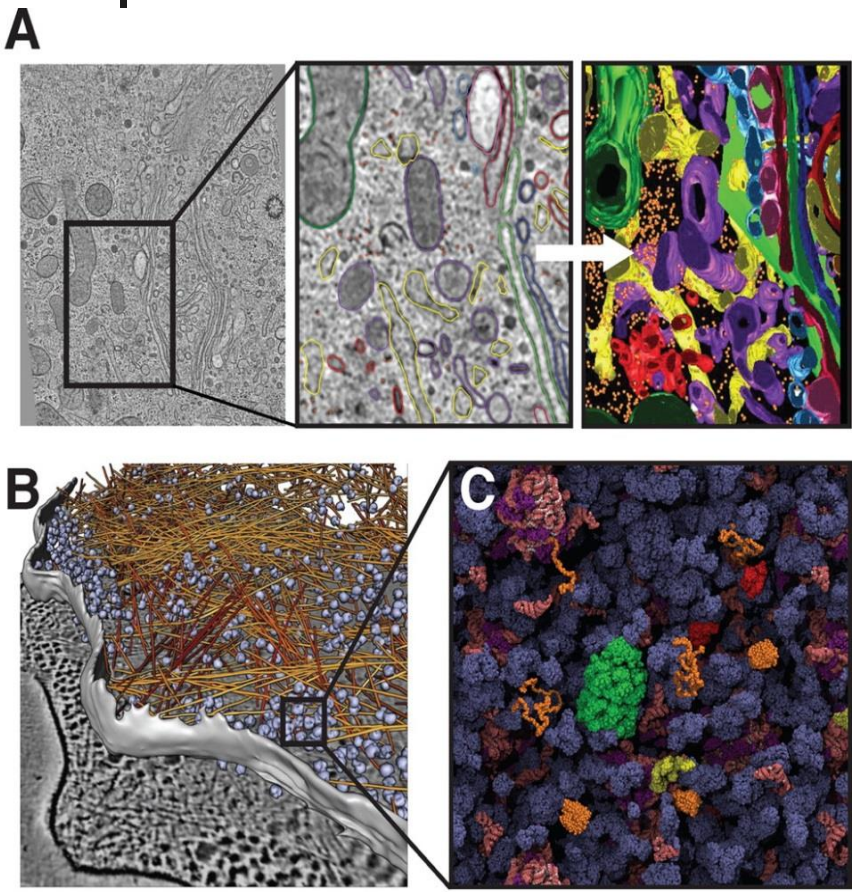
Intrinsically disordered proteins: function, folding, and flexibility



Functional classification scheme of IDRs. The function of disordered regions can stem directly from their highly flexible nature, when they fulfil entropic chain functions (such as linkers and spacers, indicated in dark-tone red), or from their ability to bind to partner molecules (proteins, other macromolecules, or small molecules). In the latter case, they bind either transiently as display sites of post-translational modifications or as chaperones (indicated in green), or they bind permanently as effectors, assemblers, or scavengers (indicated in dark-tone blue).

Robin van der Lee; Marija Buljan; Benjamin Lang; Robert J. Weatheritt; Gary W. Daughdrill; A. Keith Dunker; Monika Fuxreiter; Julian Gough; Joerg Gsponer; David T. Jones; Philip M. Kim; Richard W. Kriwacki; Christopher J. Oldfield; Rohit V. Pappu; Peter Tompa; Vladimir N. Uversky; Peter E. Wright; M. Madan Babu; *Chem. Rev.* **2014,** 114, 6589-6631. DOI: 10.1021/cr400525m.

Intracellular complexity



The intracellular environment is extremely crowded. Estimates show that the concentration of biological macromolecules (proteins, nucleic acids, ribonucleoproteins, polysaccharides, etc.) inside cells is in the range of **80–400 mg/mL**. **This corresponds to a volume occupancy of 5%–40% and creates a crowded medium, with considerably restricted amounts of free water.** Such natural intracellular media, being filled with billions of protein molecules and a myriad of DNA, RNA, and polysaccharide molecules are known as "crowded" rather than "concentrated" environments, as, in general, **no individual macromolecular species may be present at high concentration.**

Francois-Xavier Theillet; Andres Binolfi; Tamara Frembgen-Kesner; Karan Hingorani; Mohona Sarkar; Ciara Kyne; Conggang Li; Peter B. Crowley; Lila Gierasch; Gary J. Pielak; Adrian H. Elcock; Anne Gershenson; Philipp Selenko; *Chem. Rev.* **2014,** 114, 6661-6714

Irina M. Kuznetsova, Konstantin K. Turoverov and Vladimir N. Uversky, What Macromolecular Crowding Can Do to a Protein, Int. J. Mol. Sci. 2014, 15(12), 23090-23140

Macromolecular crowding

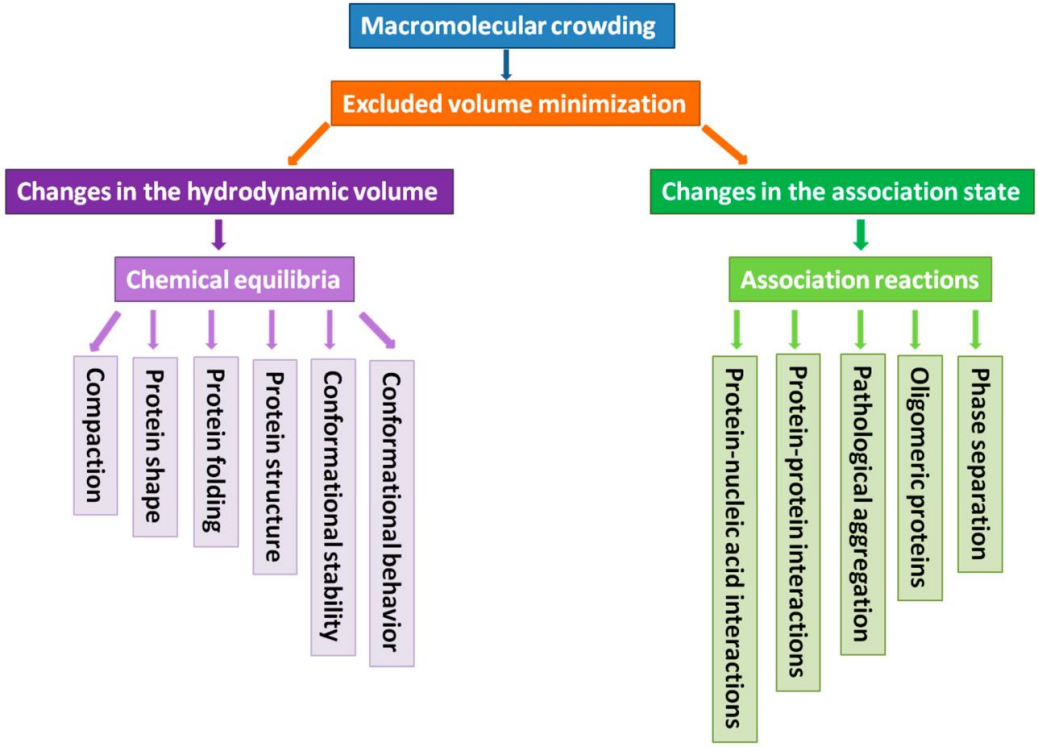
The average spacing between macromolecules in such crowded milieu can be much smaller than the size of the macromolecules themselves. Furthermore, the volume occupied by solutes is unavailable to other molecules because two molecules cannot be in the same place at the same time. As a result, any reactions that depend on available volume can be affected by macromolecular crowding effects.

The thermodynamic consequences of the unavailable volume are called excluded volume effects

Irina M. Kuznetsova, Konstantin K. Turoverov and Vladimir N. Uversky, What Macromolecular Crowding Can Do to a Protein, Int. J. Mol. Sci. 2014, 15(12), 23090-23140

Macromolecular crowding

the fact that two molecules cannot occupy the same space in solution, and that steric hindrance or impediment of a macromolecule is expected to exclude other molecules from its neighborhood give rise to the excluded volume phenomenon



Schematic representation of the potential effects of excluded volume on the behavior of proteins in crowded milieu

Irina M. Kuznetsova, Konstantin K. Turoverov and Vladimir N. Uversky, What Macromolecular Crowding Can Do to a Protein, Int. J. Mol. Sci. 2014, 15(12), 23090-23140

Sample preparation

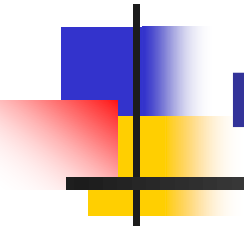
- **Extraction**
- **Purification**

- **Stability**
- **Mono-dispersity (unwanted aggregation)**
- **Solubility**

- **Functional protein**

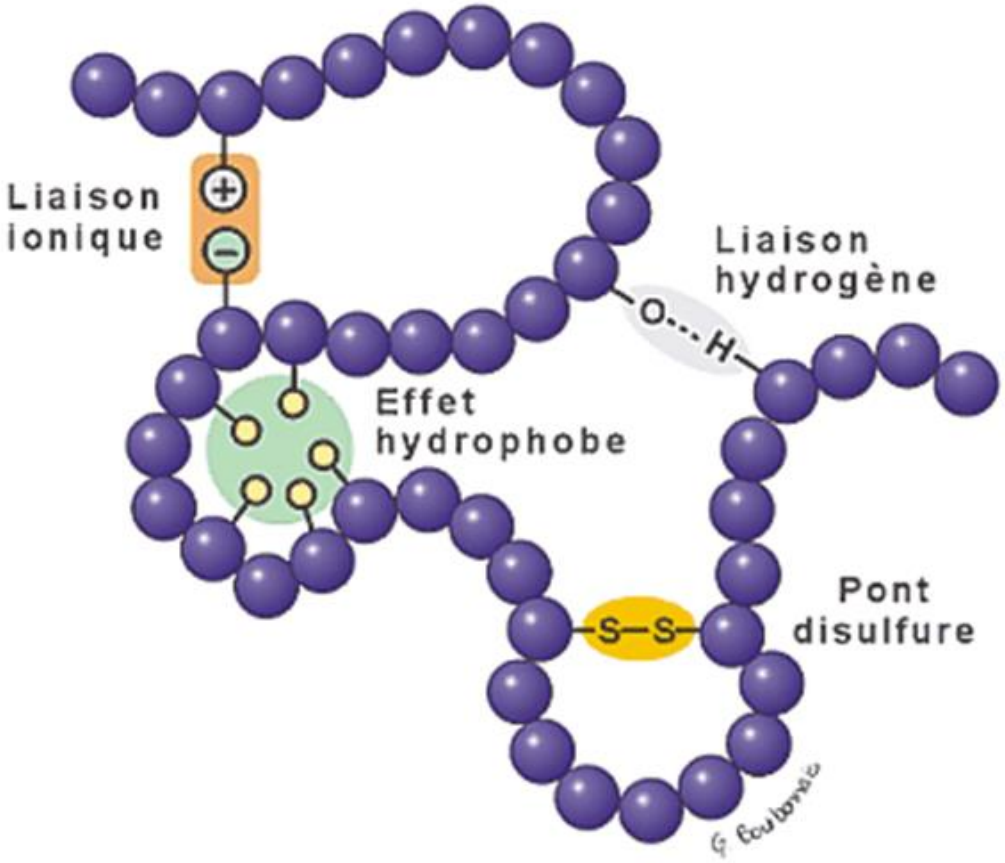
Solubility, Aggregation, Stability, Function

- Stability
  - Folded vs. unfolded state
- Aggregation
  - Protein multimerization
- Solubility
  - Protein dissolved in aqueous solvent
- Function
  - Biological function of the protein



Protein stability

Forces stabilizing proteins



Forces stabilizing proteins

(1) Based on studies of 138 hydrophobic interaction variants in 11 proteins, burying a $-CH_2-$ group on folding contributes 1.1 ± 0.5 kcal/mol to protein stability.
(2) The burial of non-polar side chains contributes to protein stability in two ways: first, a term that depends on the removal of the side chains from water and, more importantly, the enhanced London dispersion forces that result from the tight packing in the protein interior.
(3) Based on studies of 151 hydrogen bonding variants in 15 proteins, forming a hydrogen bond on folding contributes 1.1 ± 0.8 kcal/mol to protein stability.
(4) The contribution of hydrogen bonds to protein stability is strongly context dependent.
(5) Hydrogen bonds by side chains and peptide groups make similar contributions to protein stability.
(6) Polar group burial can make a favorable contribution to protein stability even if the polar group is not hydrogen bonded.
(7) Hydrophobic interactions and hydrogen bonds both make large contributions to protein stability.

C. Nick Pace, J. Martin Scholtz, Gerald R. Grimsley, Forces stabilizing proteins, FEBS Letters 588 (2014) 2177–2184

Proteins stability

- **Protein stability is the net balance of forces, which determine whether a protein will be in its native folded conformation or a denatured state.**

- **Protein stability refers to the physical (thermodynamic) stability or the chemical stability.**

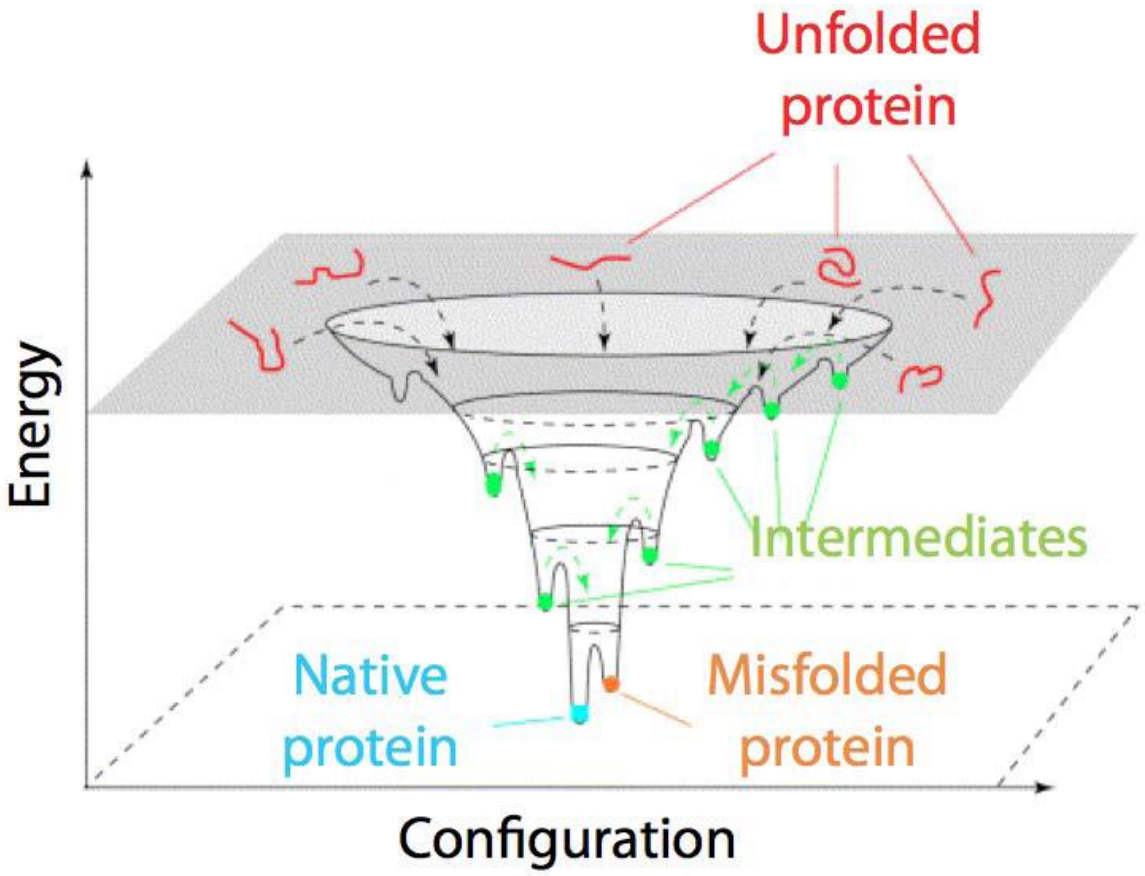
Chemical stability

- **The chemical stability implies a loss of integrity due to the breaking of chemical bonds.**
  - deamination of asparagine and/or glutamine residues,
  - hydrolysis of the peptide bond of Asp residues at low pH,
  - oxidation of the amino acid Met at high temperature,
  - Elimination of the disulfide bonds,
  - Exchange of the disulfide bonds at neutral pH.
- **Other processes include thiol-catalyzed disulfide interchange and oxidation of cysteine residues.**

Thermodynamic stability

- The native state of a protein is achieved when it reaches the state of the lowest free energy. From a thermodynamic point of view, the free energy of a protein depends on enthalpy and entropic contributions.
- The enthalpy contributions include covalent bonds, hydrogen bonds, electrostatic interactions and Van der Waals interactions.
- The entropic contributions involve the hydrophobic effect and conformational entropy

Thermodynamic stability: the folding funnel



- There is no single route for folding, but a large number of structures that follow a multidimensional funnel up to the native structure
- Progress in the funnel is accompanied by an increase of the native structures during the process of folding

Protein denaturation

Denaturation is a process in which proteins or nucleic acids lose the tertiary structure and secondary structure which is present in their native state, by application of some external stress or compound such as a strong acid or base, a concentrated inorganic salt, an organic solvent (e.g., alcohol or chloroform), or heat.

If proteins in a living cell are denatured, this results in disruption of cell activity and possibly cell death. Denatured proteins can exhibit a wide range of characteristics, from loss of solubility to communal aggregation.

Protein denaturation: denaturing agents

The denaturing agents are numerous and can be either of physical nature (temperature, pressure) or be chemical agents.

- Chemical agents
  - Chaotropic agents as urea or guanidine chloride. At high concentrations, these compounds greatly weaken hydrogen bonds (main links of low energies responsible for the maintenance of secondary, tertiary and quaternary protein structures).
  - Thiols reducing agents like 2-mercaptoéthanol or DTT (Dithiothreitol). They allow the reduction (rupture) of the disulfide bonds and can thus contribute to destabilize the tertiary or quaternary protein structure.
  - Bases and acids, by alteration of pH.
  - Detergents, by modification of the interaction with the aqueous solvent.
  - Alcohol Disrupts Hydrogen Bonding
  - Heavy metal salts act to denature proteins in much the same manner as acids and bases. Heavy metal salts usually contain Hg^{+2}, Pb^{+2}, Ag^{+1} Tl^{+1}, Cd^{+2} and other metals with high atomic weights. Since salts are ionic they disrupt salt bridges in proteins. The reaction of a heavy metal salt with a protein usually leads to an insoluble metal protein salt.

Protein denaturation: denaturing agents

Physical agents: Thermal denaturation

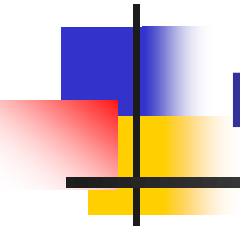
Overall, protein stability depends on the balance between enthalpy and entropic changes.

For globular proteins, the free energy of unfolding is commonly found to be positive between about 0 °C and 45 °C. It decreases through zero when the temperature becomes either hotter or colder, with the thermodynamic consequences of both cold and heat denaturation.

Heat denaturation is primarily due to the increased entropic effects of the non-polar residues in the unfolded state. The increase in temperature causes a thermal agitation of the atoms in the molecule which causes the break of weak interactions such as hydrogen bonds, that stabilize the spatial structure.

- Ordered molecular systems are usually more stable at low temperatures, where thermal fluctuations are suppressed, but native proteins tend to be most stable at a temperature, T*, near room temperature.
- As a consequence, the denaturated state population can be increased either by heating (for $T > T^*$) or by cooling (for $T < T^*$). At sufficiently high or low temperatures, the native protein is thus denatured.

The hydration of the internal non-polar groups is mainly responsible for cold denaturation as their energy of hydration is greatest when cold. Thus, it is the increased natural structuring of water at lower temperatures that causes cold destabilization of proteins in solution.



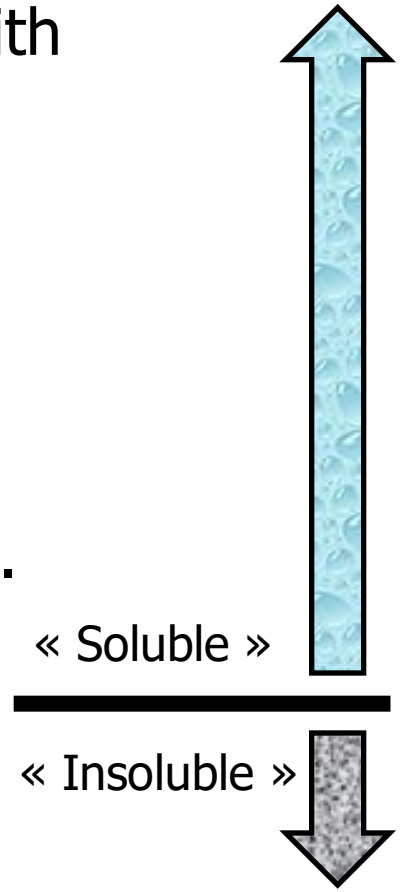
Protein aggregation

Protein aggregation

- Aggregation is a general term that encompasses several types of interactions or characteristics. Aggregates of proteins may arise from several mechanisms and may be classified in numerous ways, including soluble/insoluble, covalent/non-covalent, reversible/irreversible, and native/denatured.
- For protein therapeutics, the presence of aggregates of any type is typically considered to be undesirable because of the concern that the aggregates may lead to an immunogenic reaction (small aggregates) or may cause adverse events on administration (particulates).

Protein aggregation

- **The term aggregate covers a broad spectrum of different types and sizes of associated protein States.**
  - 1. Small reversible non-covalent oligomers with fast kinetics (dimers, trimers, tetramers...).
  - 2. Irreversible non-covalent oligomers.
  - 3. Covalent Oligomers (eg. disulfides).
  - 4. « Large » aggregates (> 10-mer).
    - May be reversible if non-covalent.
  - 5. « very large » aggregates (~50nm à 3µm).
    - May be reversible if non-covalent.
  - 6. Visible particles.



Protein aggregation

- **3 types of protein aggregation.**
  - Natural and productive aggregation as the reaction n(G-actin) →(F-actin)n controlling mobility and shape of cells.
  - Aggregation unwanted in biology including α synuclein, amyloid β, polyglutamine and Prion proteins that play an important role in neuro-degenerative disease as Parkinson's, Alzheimer's, Huntington and the prion-related diseases (mad cow).
  - Aggregation unwanted in vitro important for the conservation of proteins for various applications in research and industry.

Protein aggregation

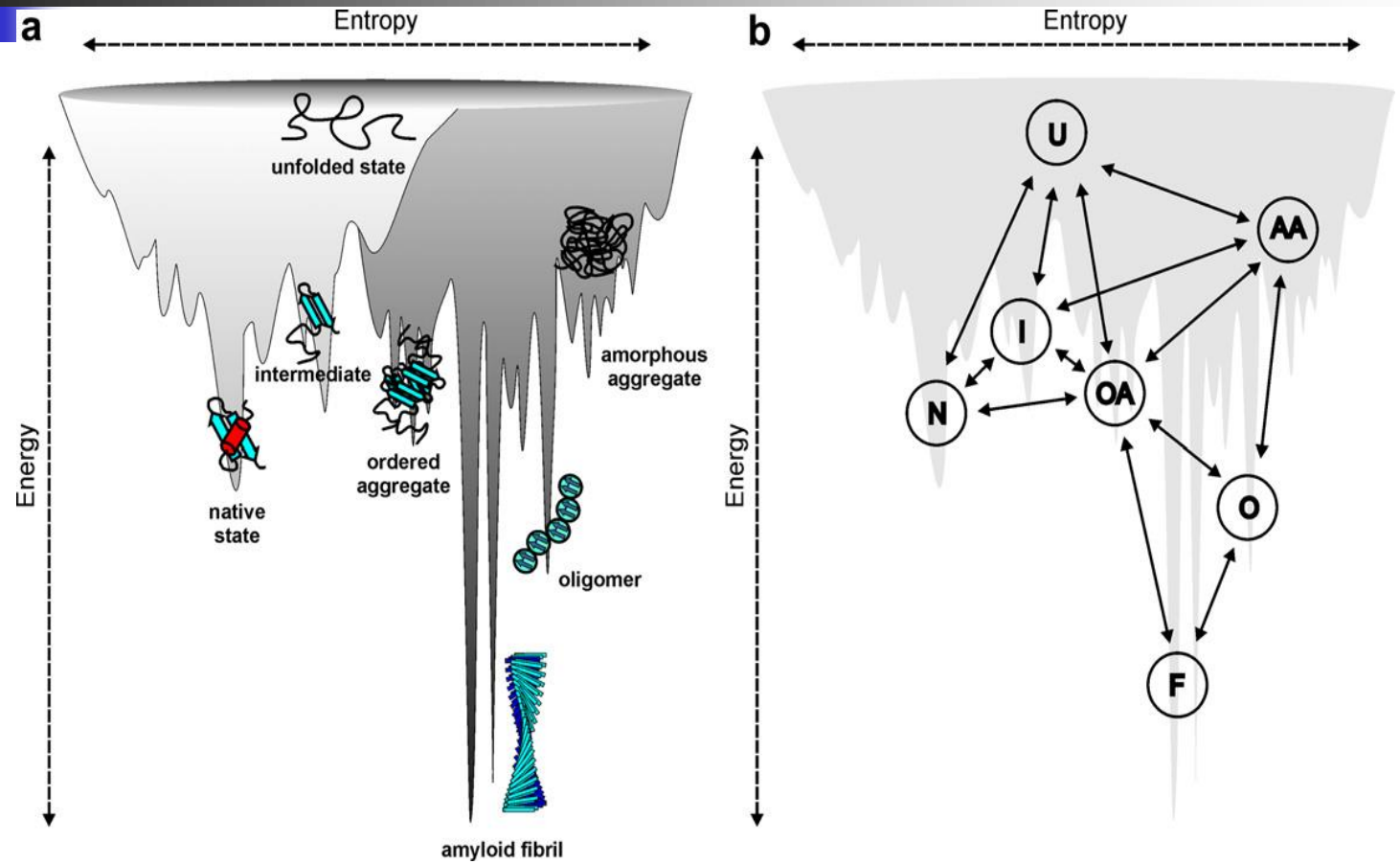
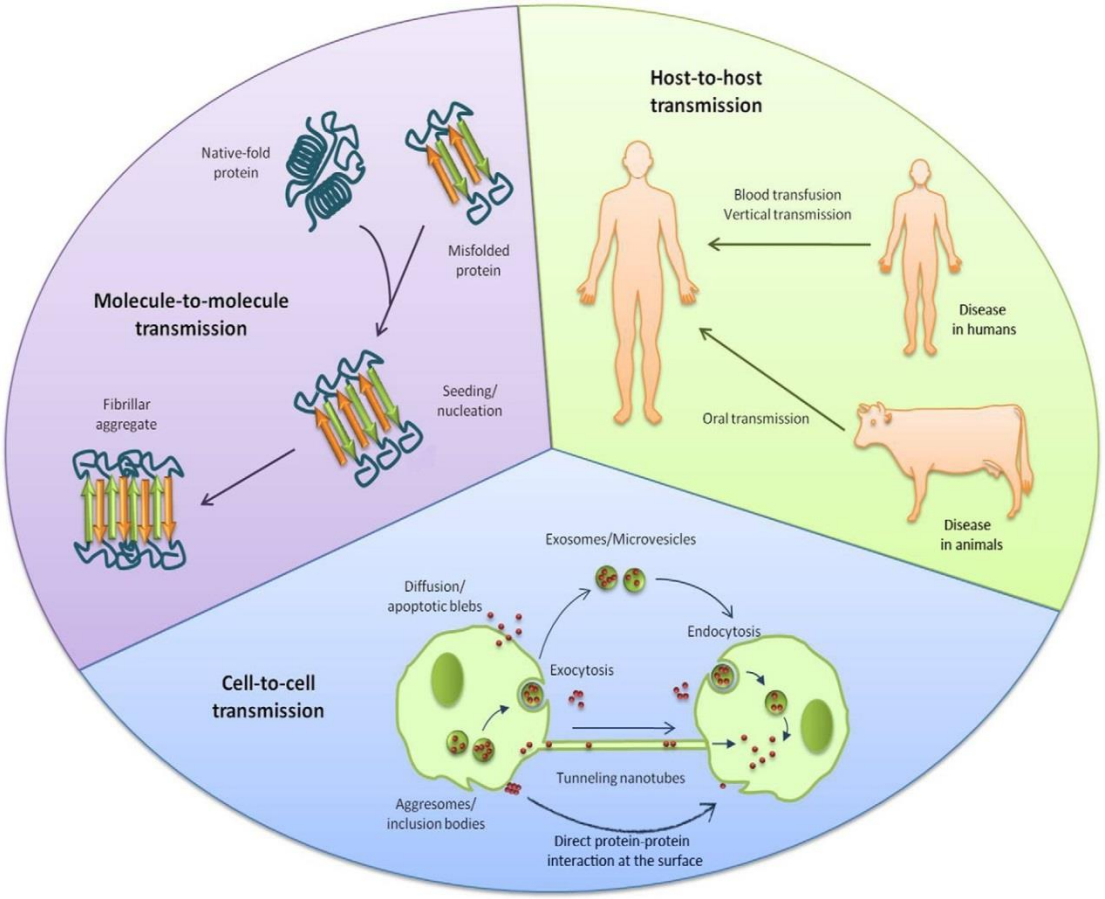


Illustration of a combined energy landscape for protein folding and aggregation. (a) The surface illustrates the roughness of the protein energy landscape, showing the multitude of conformational states available to a polypeptide chain. While rather simple folding funnels (light grey) can describe the conformational search of a single polypeptide chain to a functional monomer, intermolecular protein association dramatically increases ruggedness (dark grey). (b) Proposed pathways linking the conformational states shown in (a) populated on the combined folding and aggregation energy landscape.

T.R. Jahn, S.E. Radford / Archives of Biochemistry and Biophysics 469 (2008) 100–117

Neuro-degenerative diseases



Transmission of protein misfolding between molecules, cells and individuals. Prion-like transmission of protein misfolding may operate at various levels, including molecule-to-molecule, cell-to-cell and host-to-host. Propagation of the pathological conformational changes and downstream effects to cells, tissues and the entire individual appears to be a universal property of misfolded protein aggregates.

Ines Moreno-Gonzalez, Claudio Soto, Seminars in Cell & Developmental Biology 22 (2011) 482– 487

Protein aggregation: sequence and gatekeepers

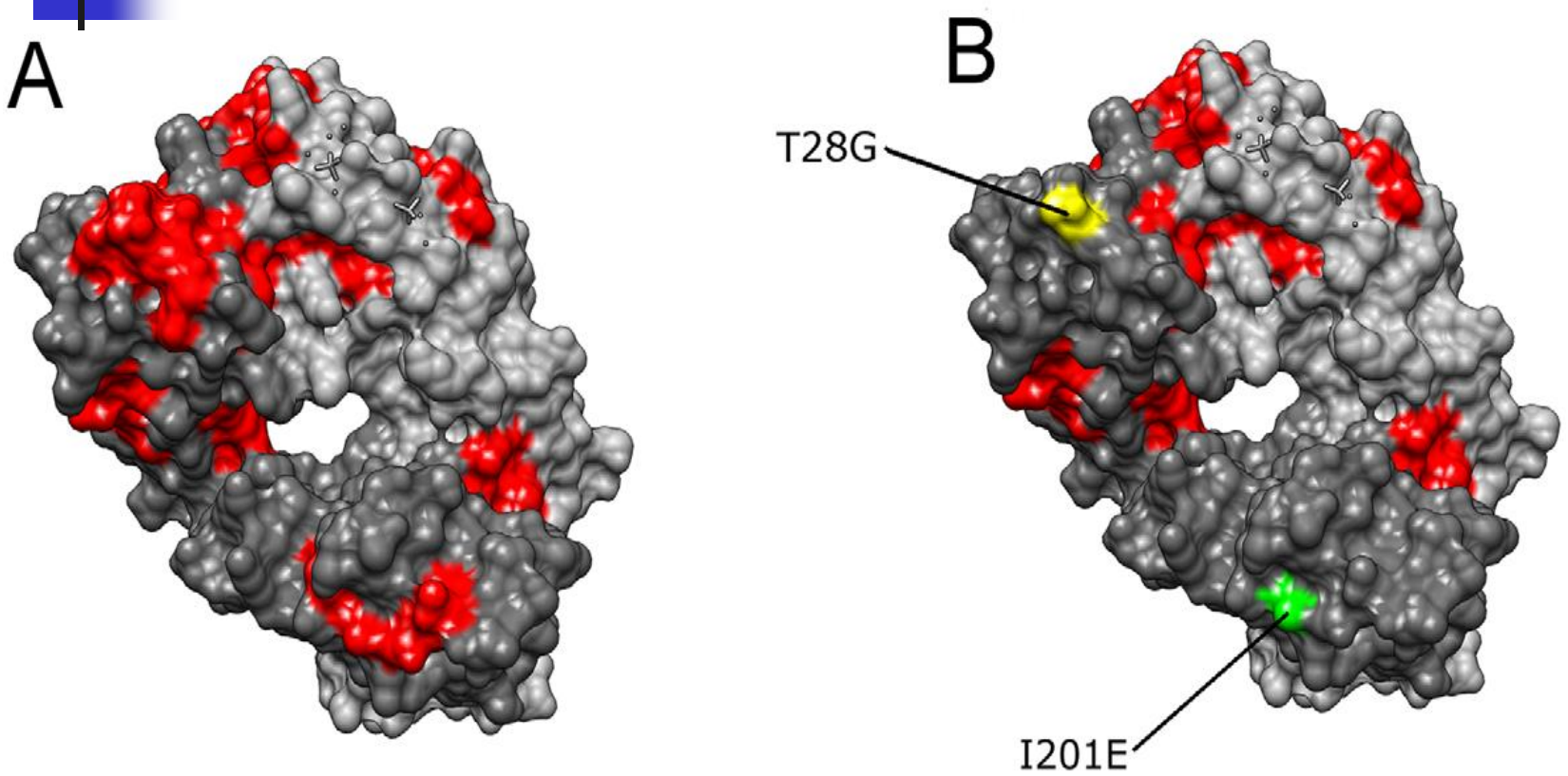
Evolutionary pressure against protein aggregation also results in the placement of amino acids that counteract aggregation at the flanks of protein sequences that are aggregation-prone. These so-called aggregation **gatekeepers** reduce aggregation by opposing nucleation of aggregates.

This disruption is achieved using the repulsive effect of charge (arginine [R], lysine [K], aspartate [D], glutamate [E]), the entropic penalty on aggregate formation (R and K) or incompatibility with β-structure backbone conformation (proline [P]).

The evolutionary enrichment of charged amino acids on the flanks of aggregating regions is coupled to chaperone specificity: studies have shown that chaperones recognize the pattern of charged residues followed by a hydrophobic region. As gatekeeper residues are enriched at the flanks of strongly aggregating hydrophobic sequences, chaperone binding occurs on average more tightly to strongly aggregating than to weakly aggregating sequences

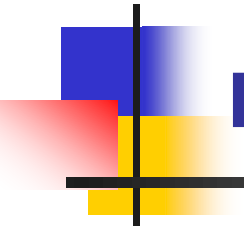
Joke Reumers et al., HUMAN MUTATION, Vol. 30, No. 3, 431–437, 2009

Protein aggregation: sequence and gatekeepers



The crystal structure (space-filling model) of the anti- ErbB2 Fab2C4 (PDB code: 1L7I) is shown. (A). This is a humanized monoclonal antibody fragment that binds to the extracellular domain of the human oncogene product ErbB2 (ErbB2 has been shown to play an important role in the pathogenesis of certain aggressive types of breast cancer). Computationally predicted 'aggregation-prone' regions by AMYLPRED2 are coloured red. Performing only two single amino acid substitutions (T28G and I201E), the AMYLPRED2 output suggests that the antibody has 'lost' two crucial 'aggregation-prone' regions and may, therefore, be more soluble, not forming aggregates (B).

Antonios C. et al, 2013, PLoS ONE 8(1): e54175. doi:10.1371/journal.pone.0054175

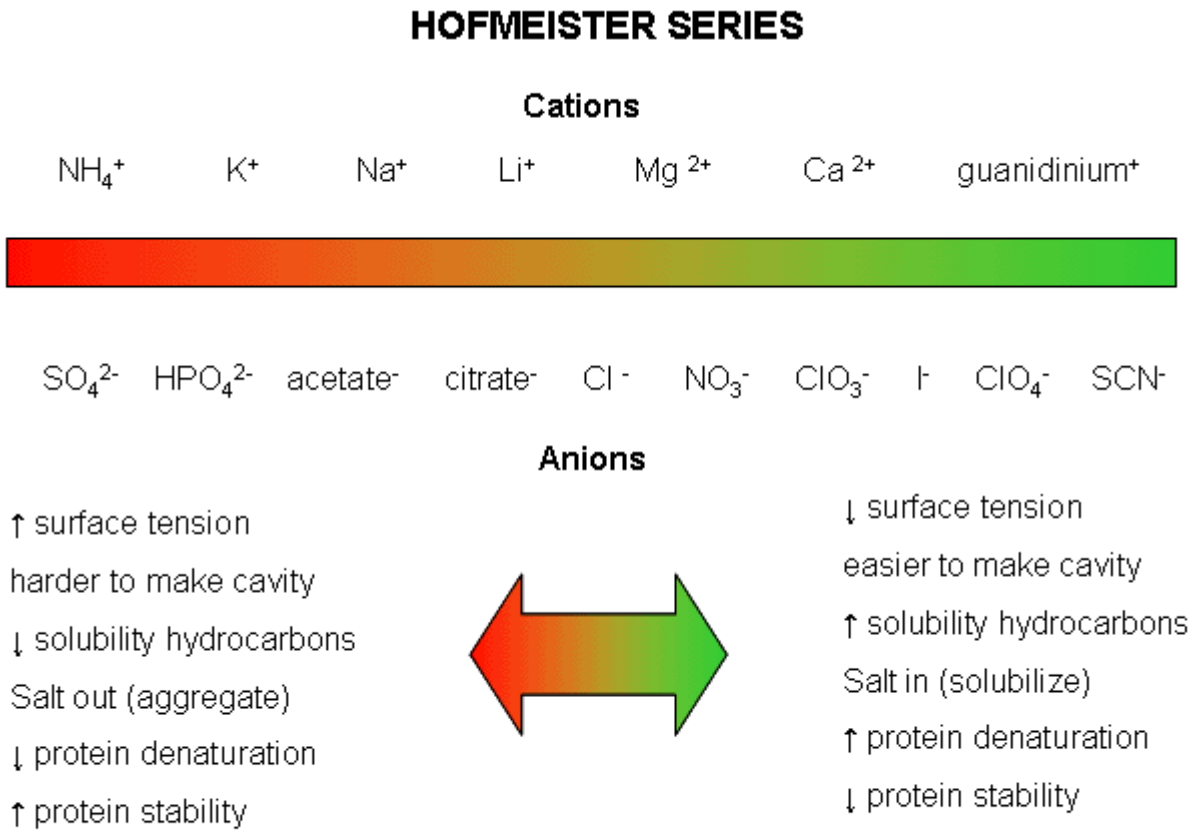


Proteins solubility

Proteins solubility

- Protein solubility is a thermodynamic parameter defined as the concentration of protein in a saturated solution that is in equilibrium with a solid phase, either crystalline or amorphous, under a given set of conditions
- Solubility can be influenced by a number of extrinsic and intrinsic factors.
  - Extrinsic factors that influence protein solubility include pH, ionic strength, temperature, and the presence of various solvent additives
  - The intrinsic factors that influence protein solubility are defined primarily by the amino acids on the protein surface

Solubility: ionic strength, Hofmeister series



Kosmotropes « water structure makers »

Chaotropes « water structure brakers »

Yanjie Zhang , Paul S Cremer, Current Opinion in Chemical Biology Volume 10, Issue 6 2006 658 - 663

Solubility: ionic strength

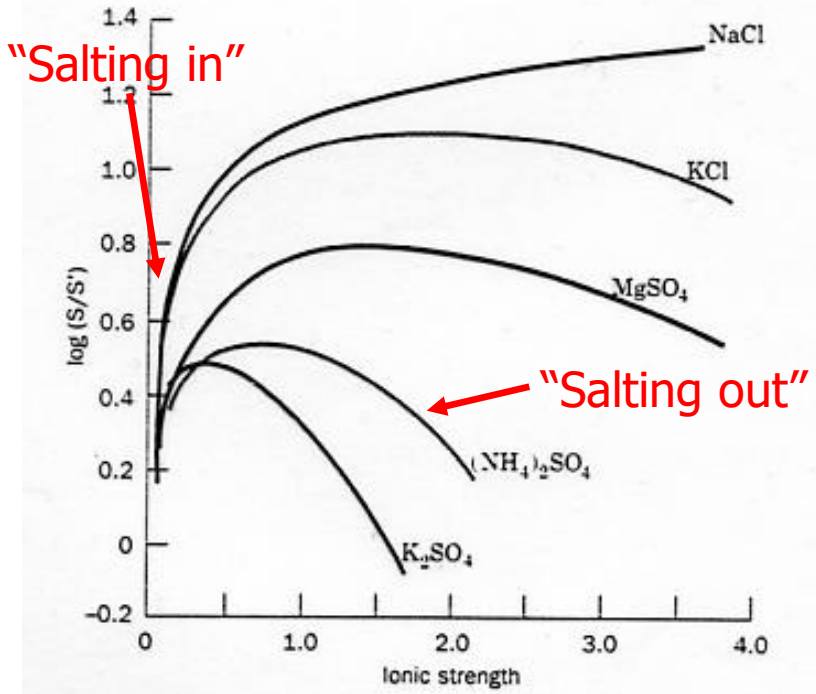


Figure 5-3
Solubility of carboxy-hemoglobin at its isoelectric point as a function of ionic strength and ion type. Here S and S' are, respectively, the solubilities of the protein in the salt solution and in pure water. The logarithm of their ratios is plotted so that the solubility curves can be placed on a common scale. [After Green, A. A., *J. Biol. Chem.* **95**, 47 (1932).]

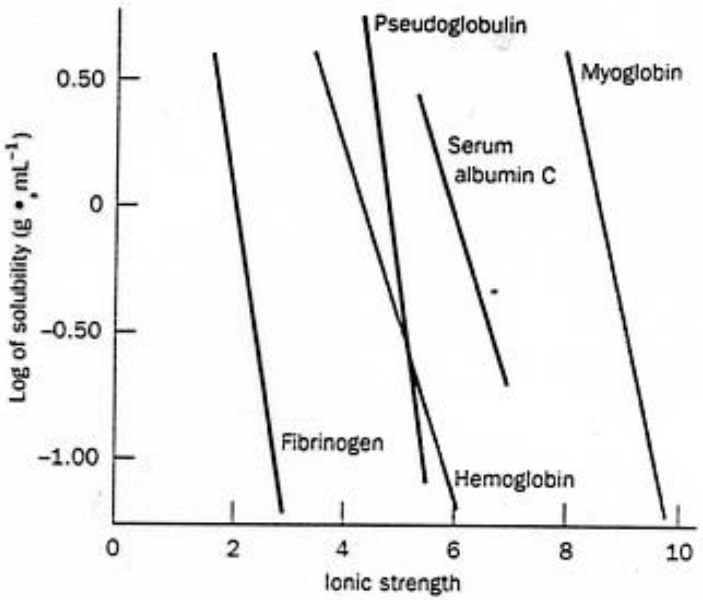
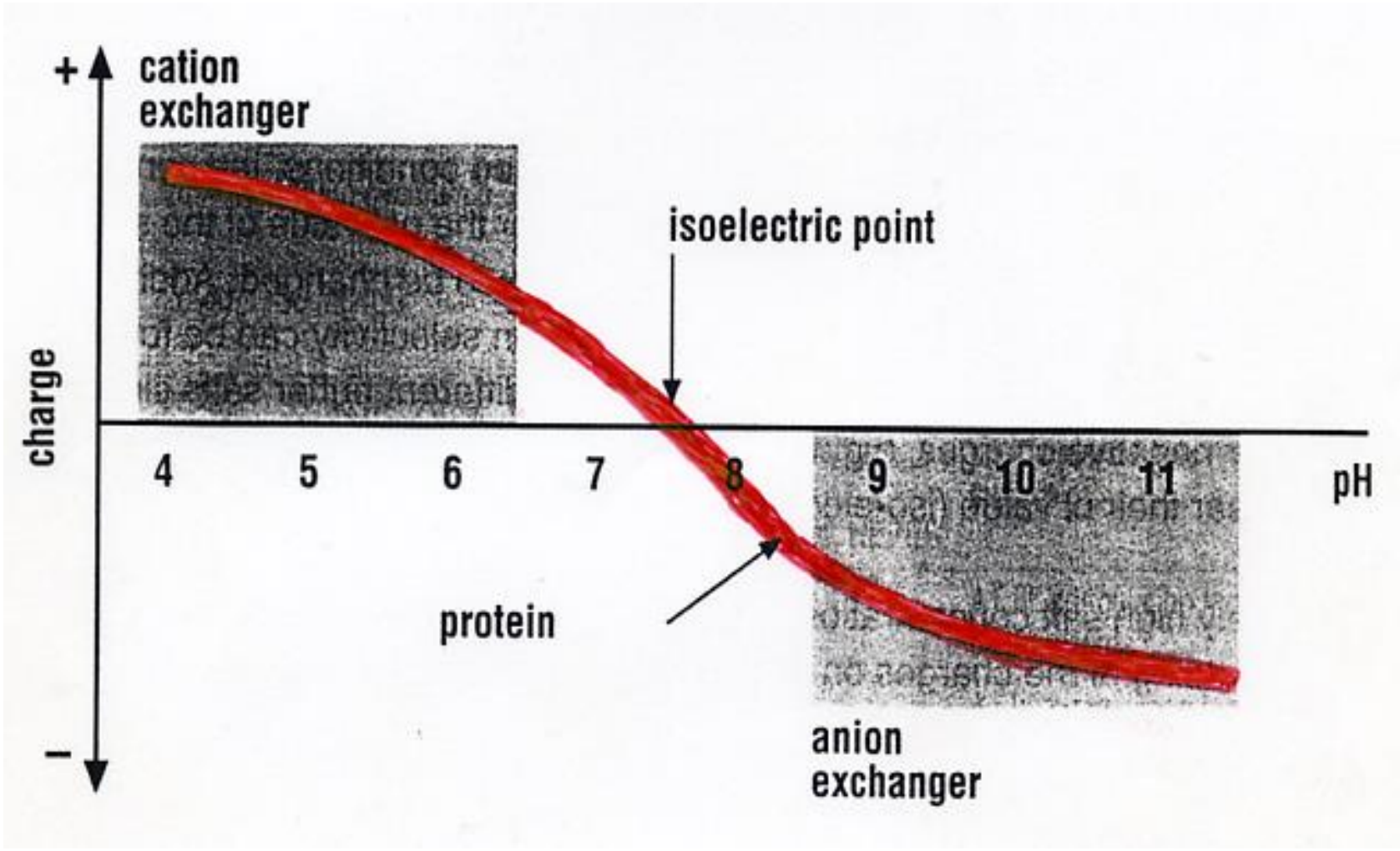


Figure 5-2
Solubilities of several proteins in ammonium sulfate solutions. [After Cohn, E. J. and Edsall, J. T., *Proteins, Amino Acids and Peptides, p.* 602, Academic Press (1943).]

Proteins solubility

Protein net charge in function of pH



Solubility: pH

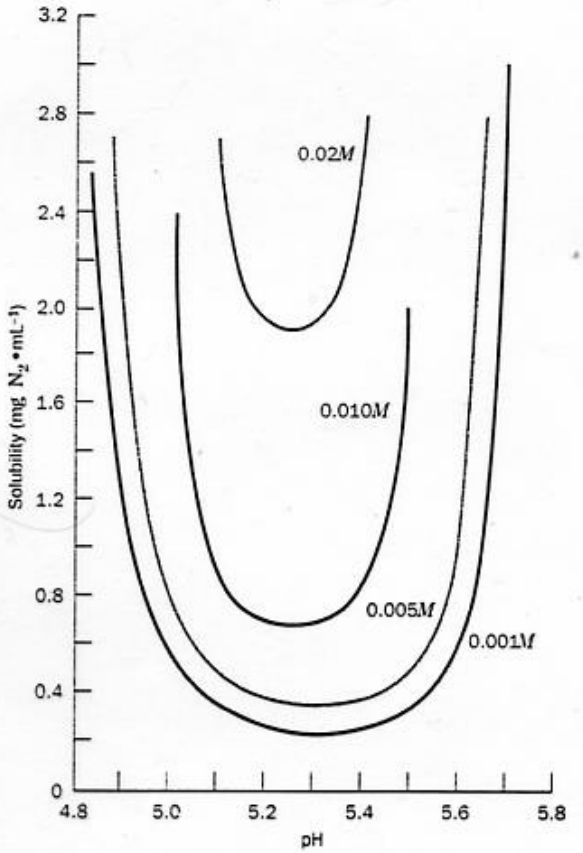


Figure 5-4
Solubility of β-lactoglobulin as a function of pH at several NaCl concentrations. [After Fox, S. and Foster, J. S., *Introduction to Protein Chemistry, p.* 242, Wiley (1957).]

The Isoelectric Points of Several Common Proteins

Protein	Isoelectric pH
Pepsin	<1.0
Ovalbumin (hen)	4.6
Serum albumin (human)	4.9
Tropomyosin	5.1
Insulin (bovine)	5.4
Fibrinogen (human)	5.8
γ-Globulin (human)	6.6
Collagen	6.6
Myoglobin (horse)	7.0
Hemoglobin (human)	7.1
Ribonuclease A (bovine)	7.8
Cytochrome *c* (horse)	10.6
Histone (bovine)	10.8
Lysozyme (hen)	11.0
Salmine (salmon)	12.1

Solubility: hydrophobicity

Capacity of non-polar areas of one or several molecules to come together to minimize the exposure of the hydrophobic area to solvent

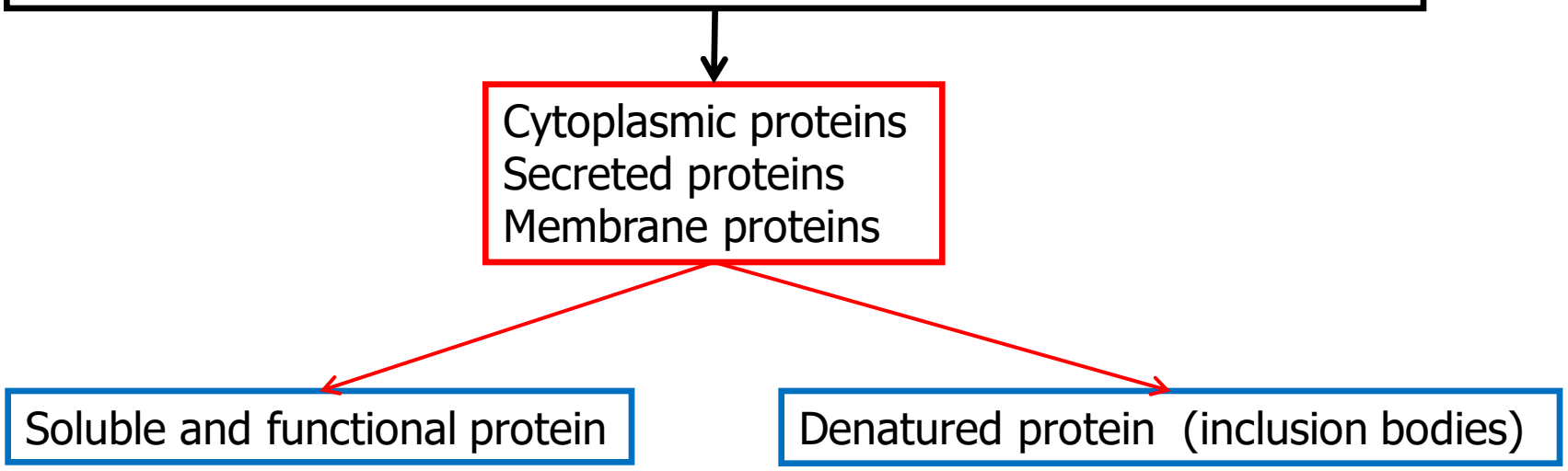
TABLEAU I — PROPRIÉTÉS DE CERTAINS DÉTERGENTS

Propriétés	Cholate de sodium	CHAPS	Octyl glucoside	Triton X100	Lubrol PX
Poids moléculaire des micelles	1 700	6 150	8 000	90 000	64 000
Concentration micellaire critique % (p/v)	0,36	0,49	0,73	0,02	0,006
"Dialysabilité"	oui	oui	oui	non	non
Présence de charges	oui	non	non	non	non

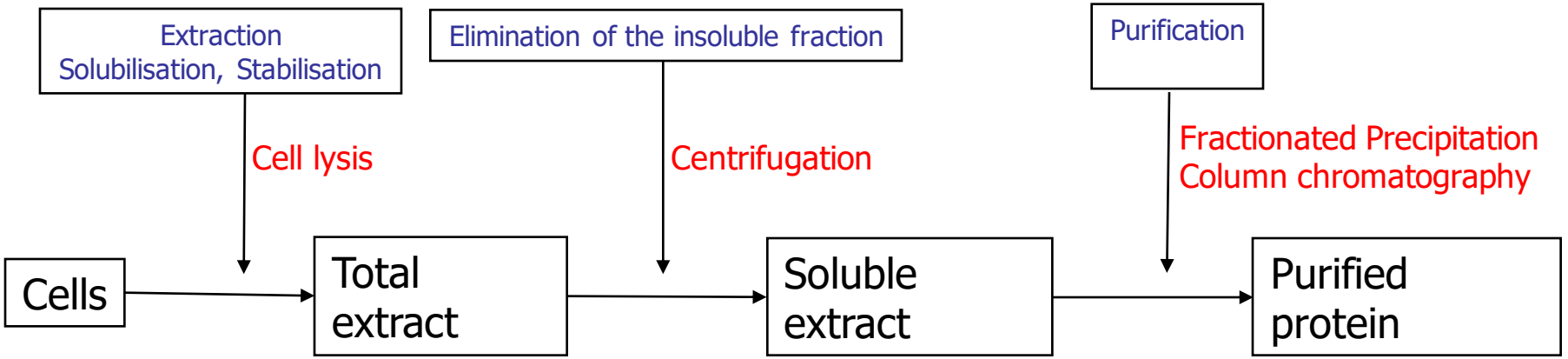
Protein purification: how to maintain a protein soluble and functional

Purification

Purification of a protein from a homologous or heterologous organism is to separate the protein of interest from a complex mixture consisting of all cellular components (fatty acids, lipids, carbohydrates, other proteins...)



Purification



Measuring the amount of:
Total protein
Specific protein

Purity analysis

Buffer exchange
Concentration and dilution

- Adapt the purification to the protein
- Adapt the protein to the purification: fusion proteins for affinity purification (HIS, GST, MBP, biotinilated peptide, STREP, FLAG, …)

Total extract → Affinity → Tag removal → Affinity → Gel filtration

Cloning and expression

- Protein production in prokaryotic cells
  - E. Coli cells
- Protein production in eukaryotic cells
  - Yeast cells
  - Insect cells
  - Mammalian cells
- Cell free systems

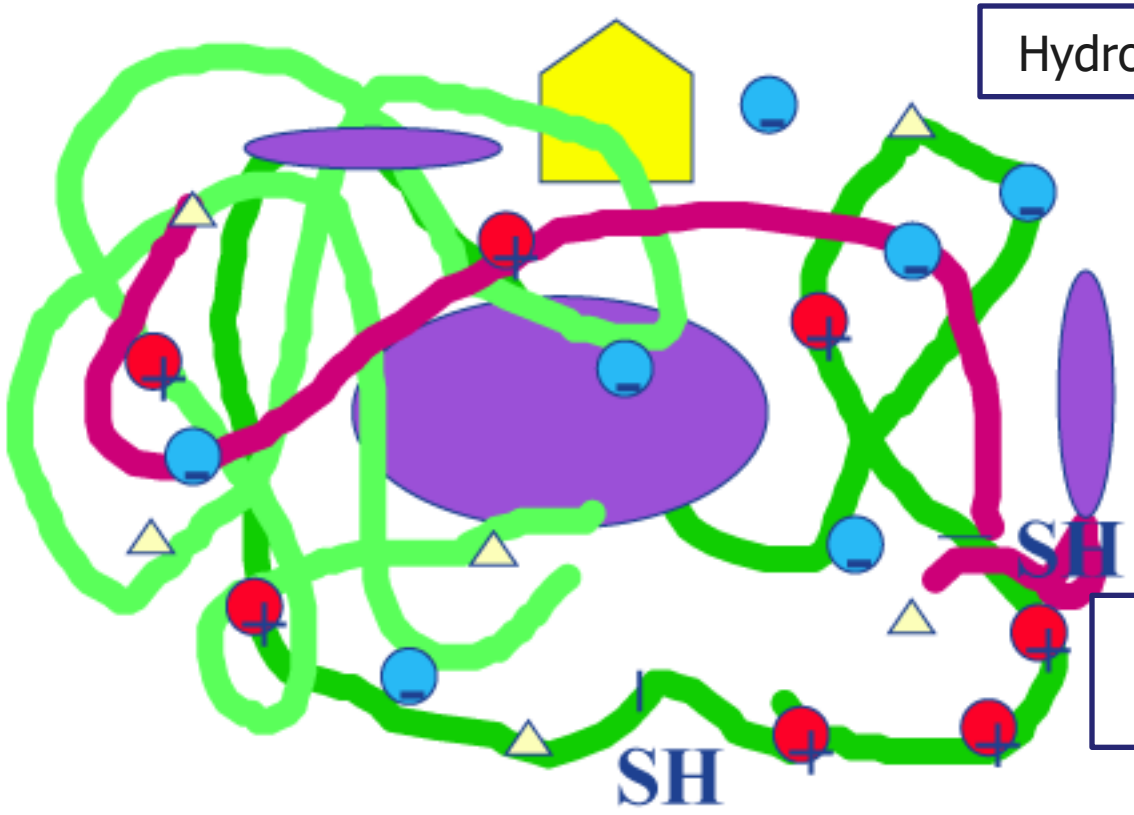
Purification

Size

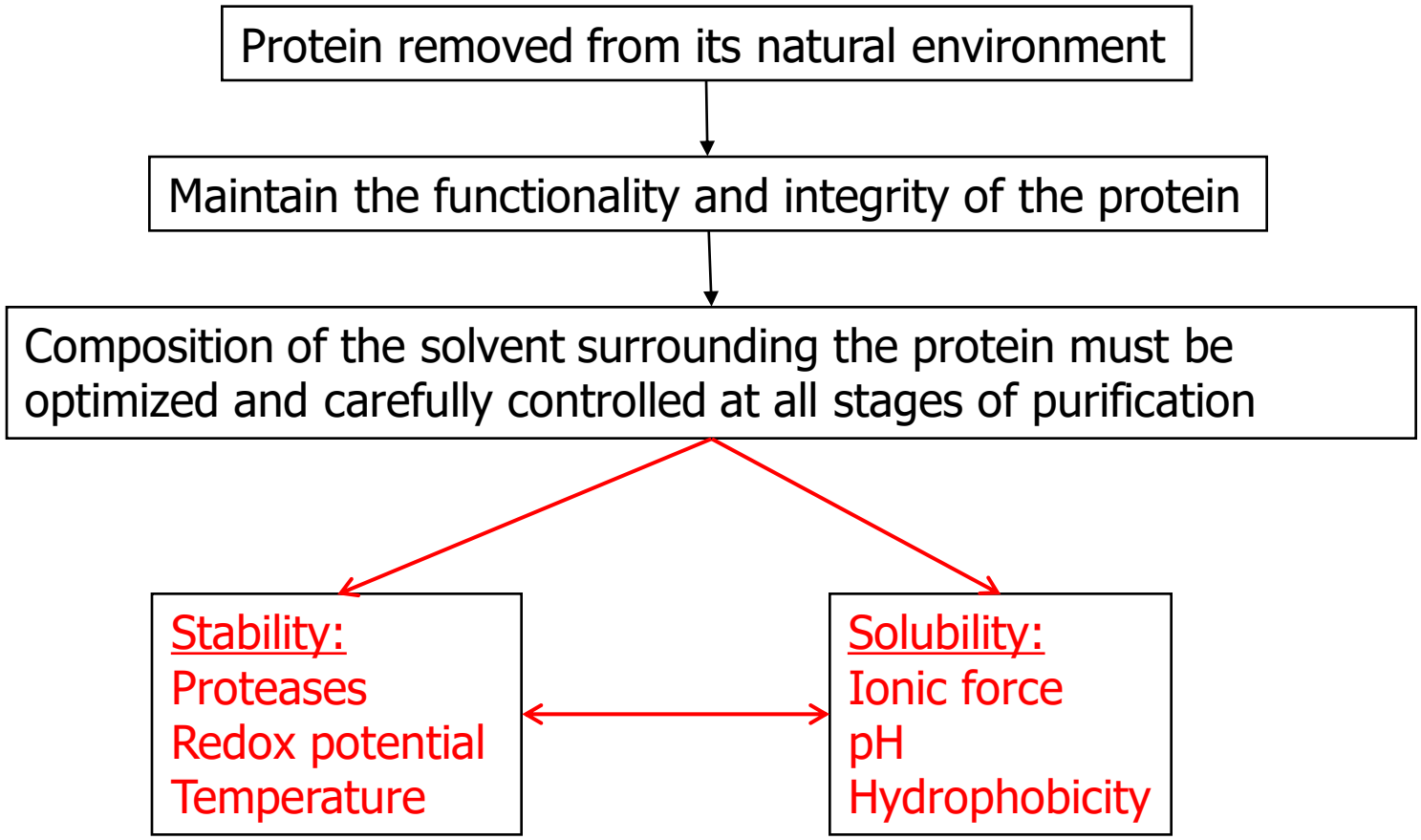
Charge

Hydrophobicity

Biological activity



Stabilisation, solubilisation



Stabilisation

- Find the right buffer composition to maintain the integrity of the purified protein
- The release of intra cellular components by grinding releases proteases
- Use of protease inhibitors
- Work at 4°C

Stabilization: REDOX potential

- Potential REDOX (use of reducing agent or oxydo/reducing mixtures)
  - Preventing the formation of inter molecules disulfides bridges
  - Preventing the formation of incorrect intra protein disulfide bridges
  - Maintaining correct disulfide bridge

Solubility

- Ionic force (solubility depending on the concentration and nature of salt)
- Detergents
- Additives (glycerol, fluoro-alcohol, Arg/Glu,…) for protein solubilization
- pH (solubility vs. pH, using buffers)

Solubility: ionic force

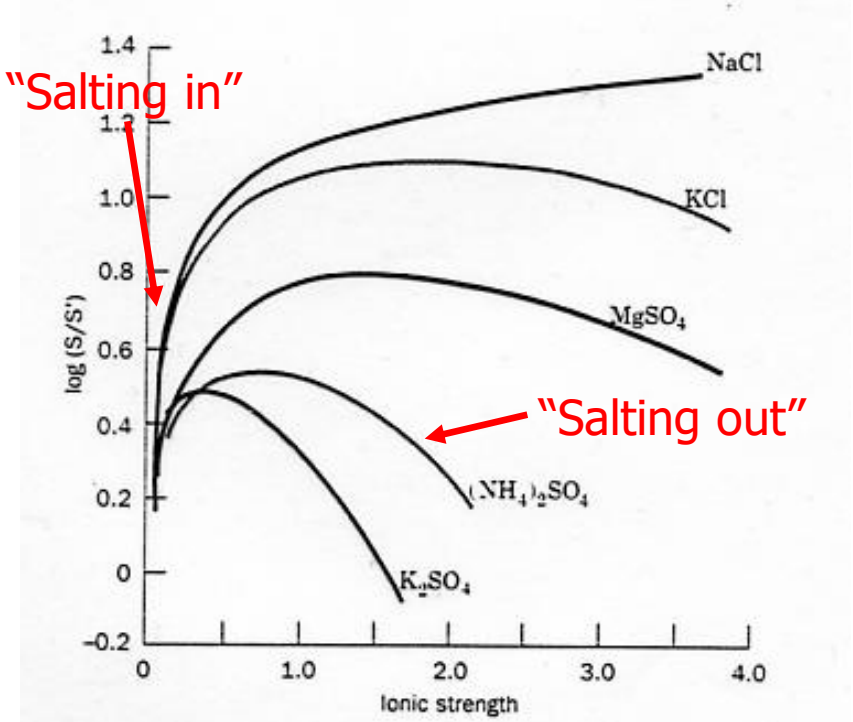


Figure 5-3
Solubility of carboxy-hemoglobin at its isoelectric point as a function of ionic strength and ion type. Here S and S' are, respectively, the solubilities of the protein in the salt solution and in pure water. The logarithm of their ratios is plotted so that the solubility curves can be placed on a common scale. [After Green, A. A., *J. Biol. Chem.* **95**, 47 (1932).]

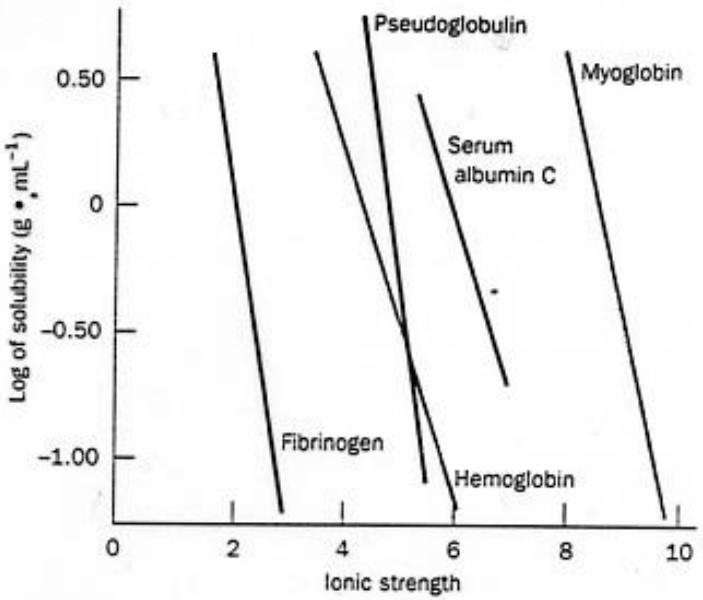


Figure 5-2
Solubilities of several proteins in ammonium sulfate solutions. [After Cohn, E. J. and Edsall, J. T., *Proteins, Amino Acids and Peptides, p.* 602, Academic Press (1943).]

Stabilisation, solubilisation

- Proteins denatured by contact with air - water interface (foam)
- If low amount, a significant fraction may be lost by adsorption to surfaces
- Minimize freezing
- Oxidation of cysteine: addition of reducers (βmercaptoethanol, DTT)
- pH stabilization by buffers
- Stabilization of ionic strength (NaCl, KCl)
- Addition of detergents in the case of aggregation due to hydrophobic interactions
- Proteins denature slowly by chemical or proteolytic degradation: addition of EDTA to remove heavy metals, inhibitors of proteases, purification at 4° C.
- Avoid bacterial contamination: NaN_3

Typical composition of a solution used for cells breaking:

NaCl 100-200 mM, EDTA 0.1 mM, TRIS/HCl pH=7.5, DTT 2 mM, PMSF 0.1 mM

IGBMC
Institut de génétique et de biologie moléculaire et cellulaire

PREPARATION AND CHARACTERIZATION OF BIOLOGICAL SAMPLES FOR STRUCTURAL STUDIES

FROM UNSTABLE PROTEIN TO STABLE COMPLEXES
THE HIV-1 PRE-INTEGRATION COMPLEX

Marc Ruff

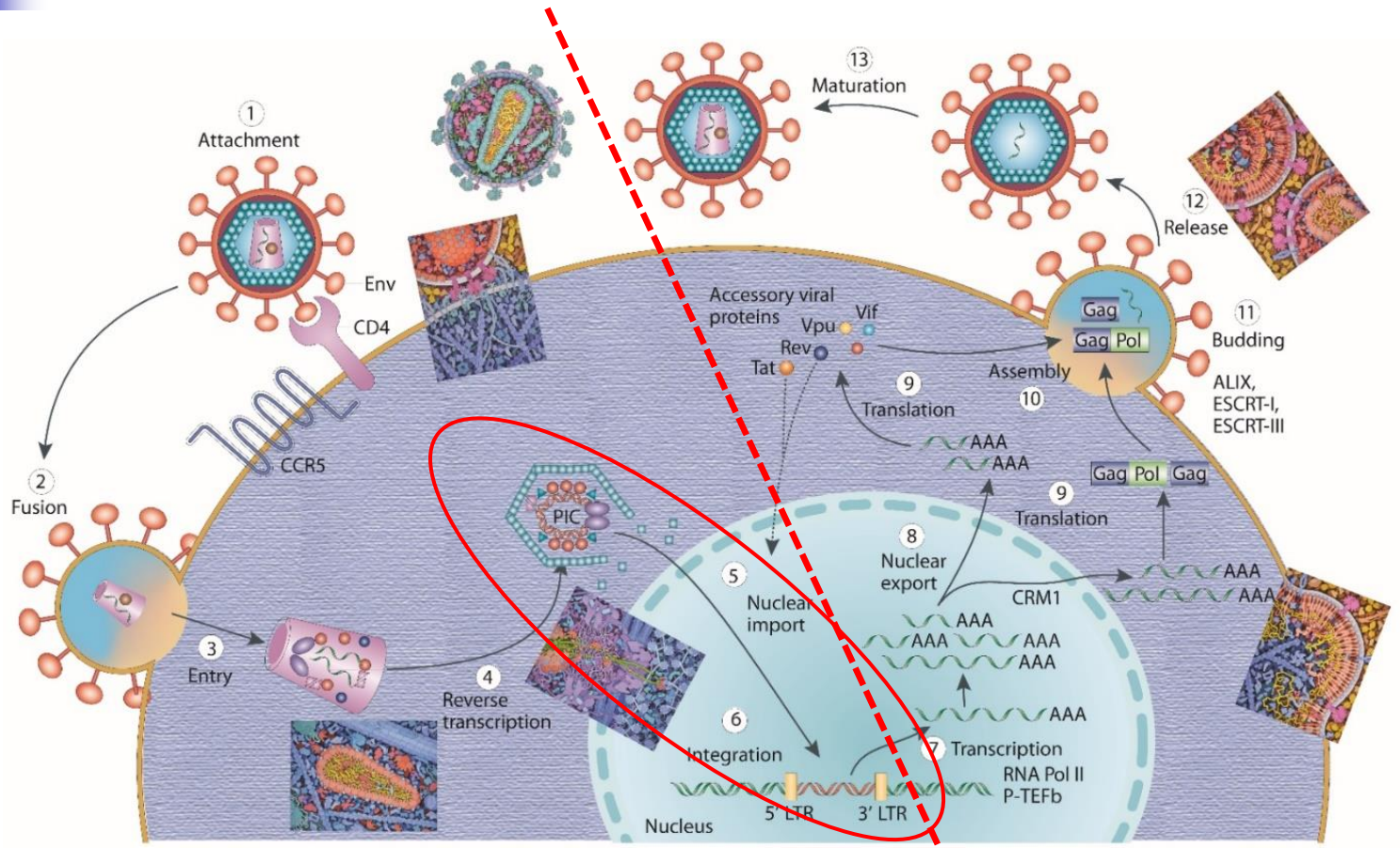
Integrated structural biology department
Chromatin stability and DNA mobility team

IGBMC, Illkirch, France

ruff@igbmc.fr

Schematic diagram of HIV replication cycle

Early phase



Late phase

Adapted from Engelman et al., 2012 and David S. Goodsell, http://hive.scripps.edu/resources.html

Constant need for new generations of inhibitors in AIDS treatment :
Need of precise knowledge of replication mechanisms

The HIV-1 viral DNA integration

David S. Goodsell, http://hive.scripps.edu/resources.html

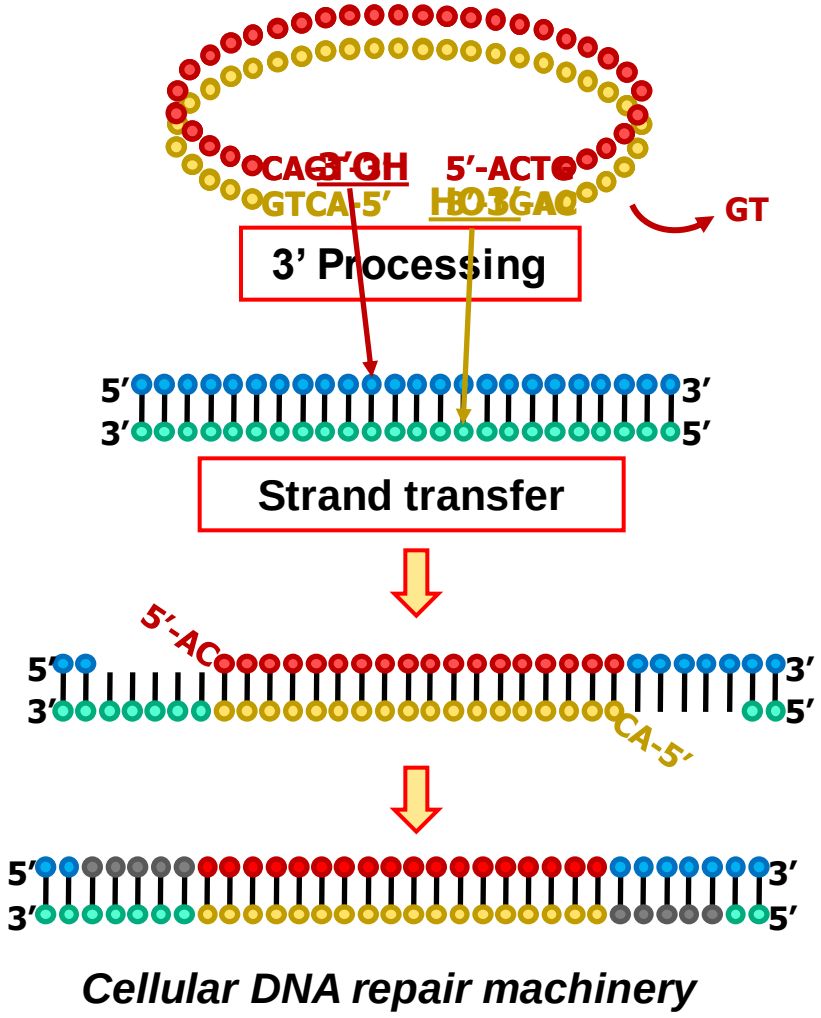
Decapsidation
Reverse transcription

Nuclear import

Continuous and dynamic process

Integration

Catalytic activities of HIV-1 integrase



Viral cDNA

Host cell target DNA

Integrated proviral DNA

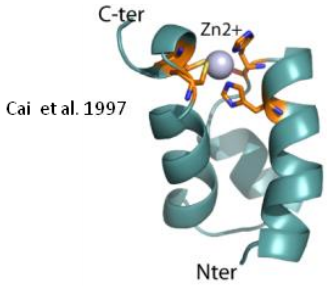
Structural domains of HIV-1 integrase

HIV integrases family

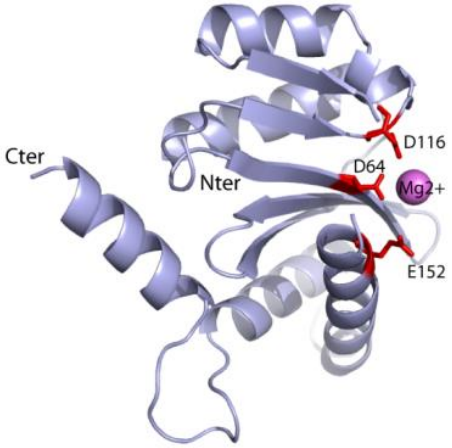
Ribonucleases H superfamily
Interaction with LEDGF

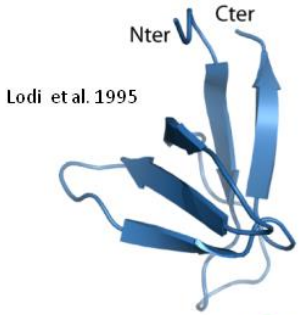
Retroviral integrases superfamily

N-terminal domain (NTD) | Catalytic Core Domain (CCD) | C-terminal domain (CTD)

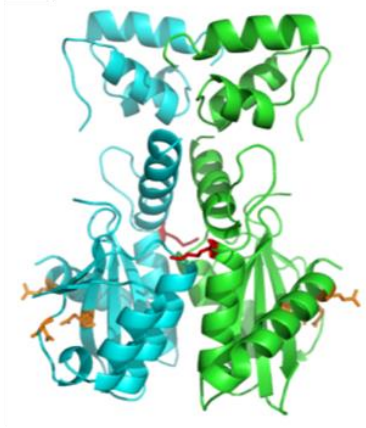


Cai et al. 1997

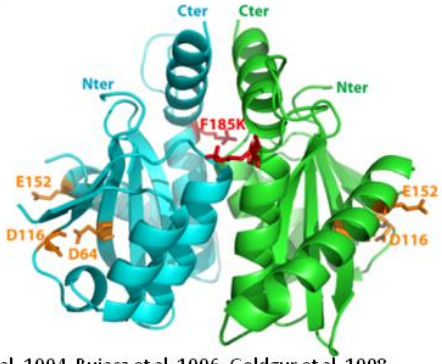




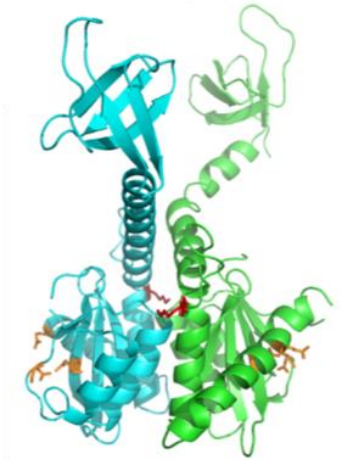
Lodi et al. 1995



Wang et al. 2001



Dyda et al. 1994, Bujacz et al. 1996, Goldgur et al. 1998
Maignan et al. 1998, Greenwald et al. 1999, Chen et al. 2000

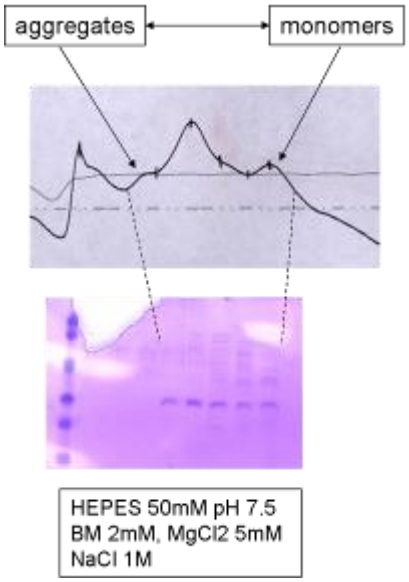


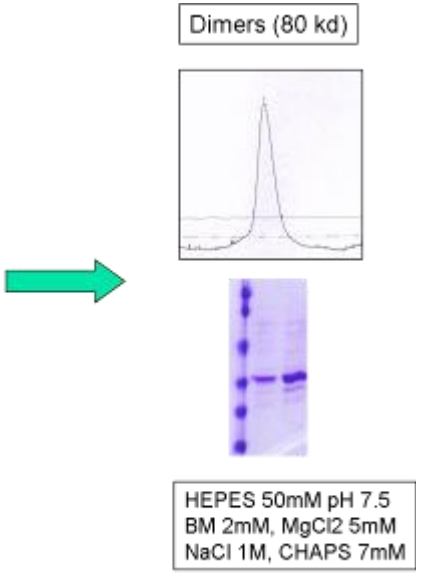
Chen et al. 2000

Full length Integrase wt : expression and solubilization

	37°		25°		18°	
	ET	EB	ET	EB	ET	EB
LB	+	-	+	+	+	+
LB/sucrose	+	-	+	+	++	++

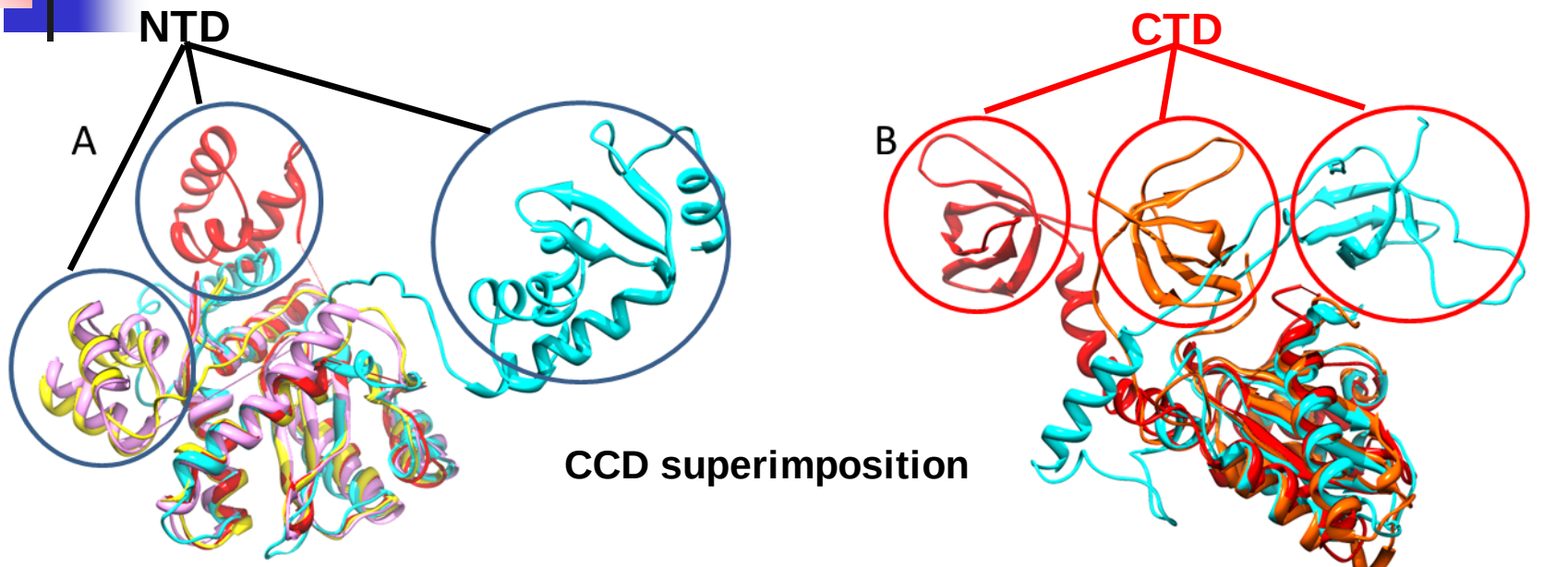
	-	Glycerol 10%	Chaps 10mM	Triton 0.1%	Sucrose 20%
50mM NaCl	+/-	+/-	++	+/-	+/-
1M NaCl	+	++	+++	+++	++





No crystals

HIV Integrase



A. Superimposition of NTD+CCD structures:
MVV (Hare et al., 2009a), PFV (Hare et al., 2010),
HIV-2 (Hare et al., 2009b) , HIV-1 (Wang et al., 2001)

B. Superimposition of CCD+CTD structures :
RSV (Yang et al., 2000), HIV-1 (Chen et al., 2000)
PVF (Hare et al., 2010).

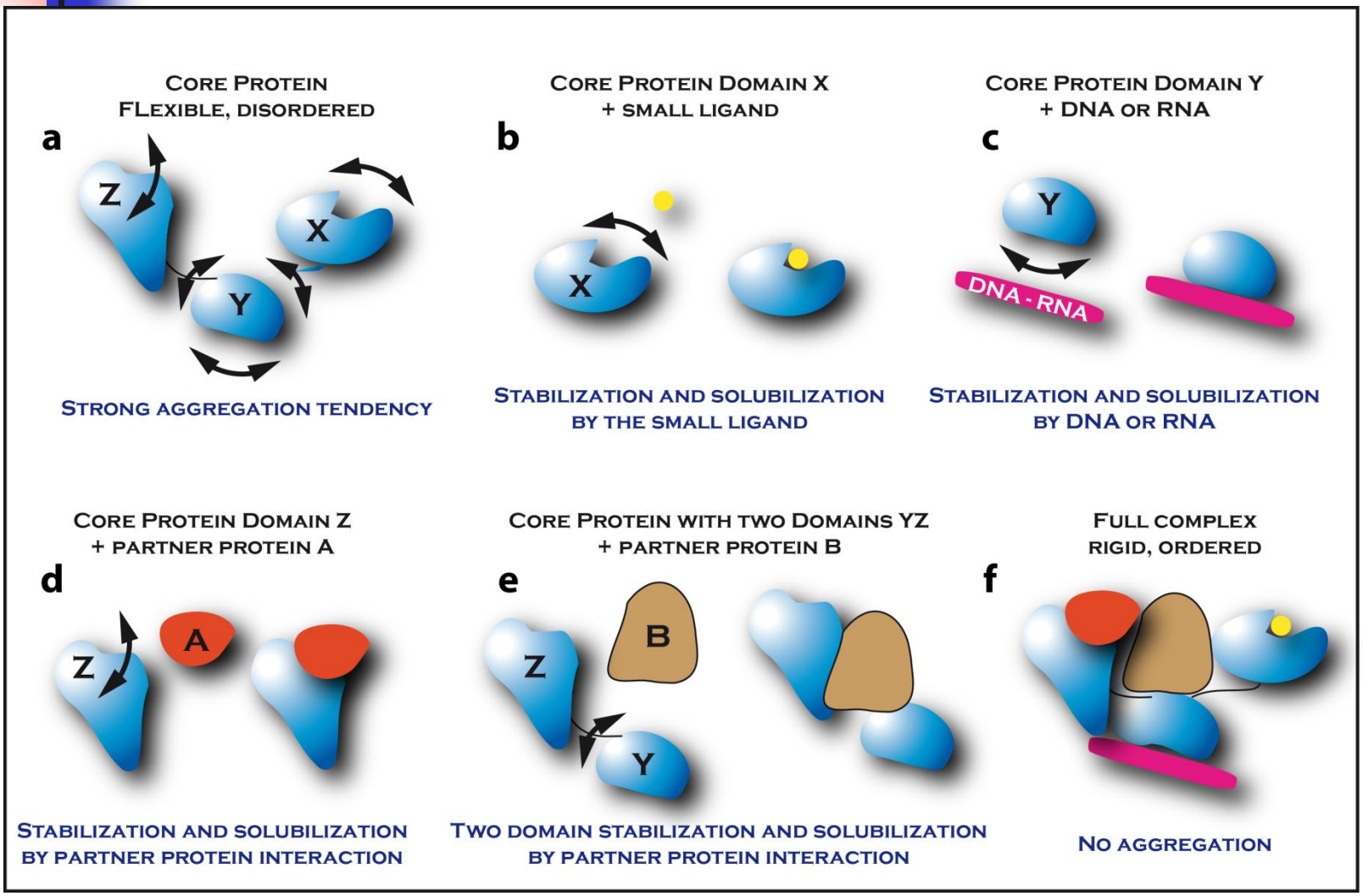
High flexibility allows to accommodate different partners and functions
No high resolution structure of full-length HIV integrase
=> Stabilization of integrase with partners/ligands for structural and functional studies

Structural and functional studies

Isolate stable IN complexes (VBP1, TRN-SR2, LEDGF, INI1….)

Isolate protein with PTMs:
Production in eukaryotic cells
Specific modifications

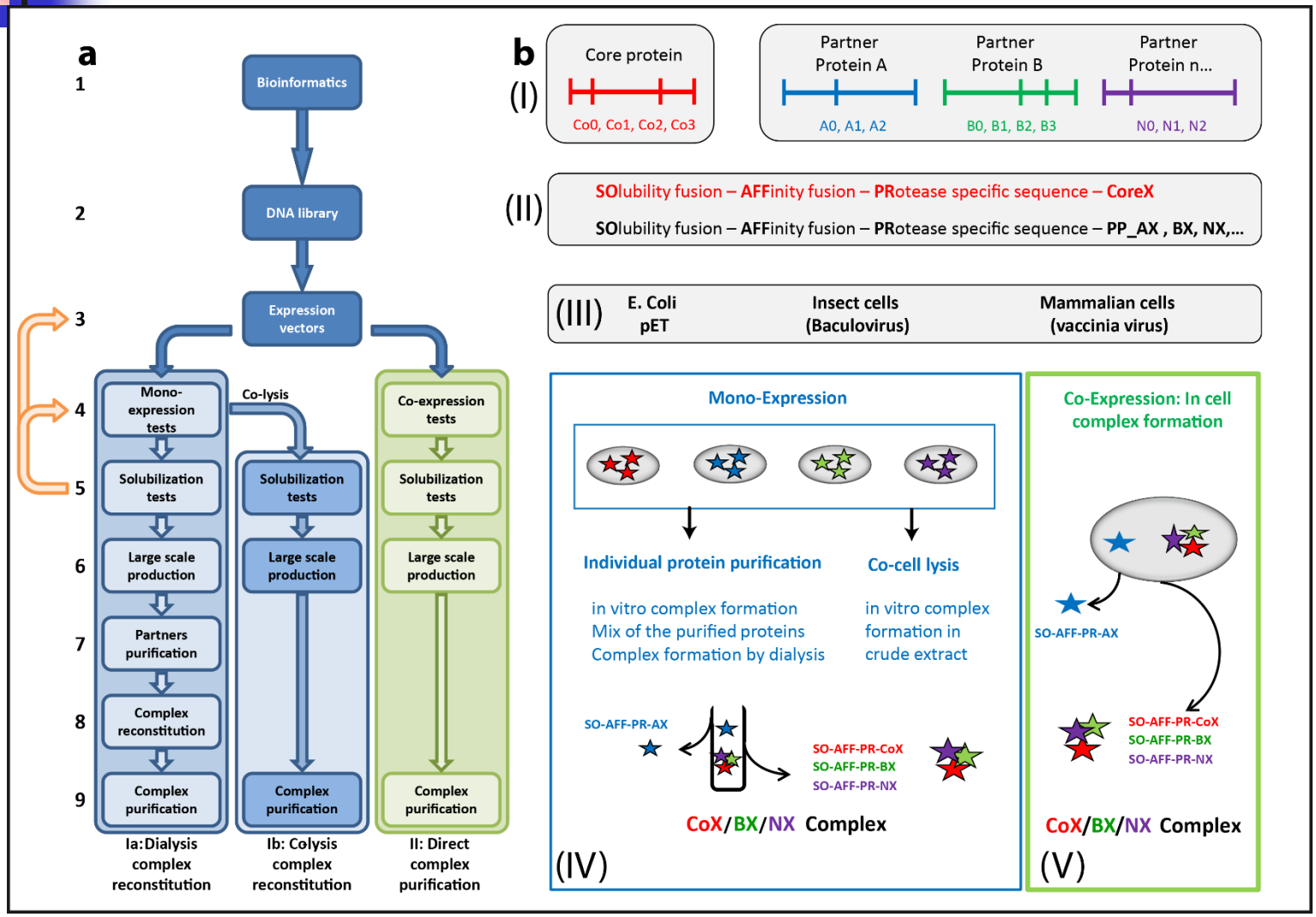
From unstable protein to stable complexes



Levy N, Eiler S, Pradeau-Aubreton K, Maillot B, Stricher F, Ruff M (2016). Production of unstable proteins through the formation of stable core complexes. Nature Communications 7: 10932

Eiler S, Levy N, Maillot B, Batisse J, Pradeau Aubreton K, Oladosu O, Marc Ruff (2018). Unstable protein purification through the formation of stable complexes. Methods in Molecular Biology, 1764:315-328

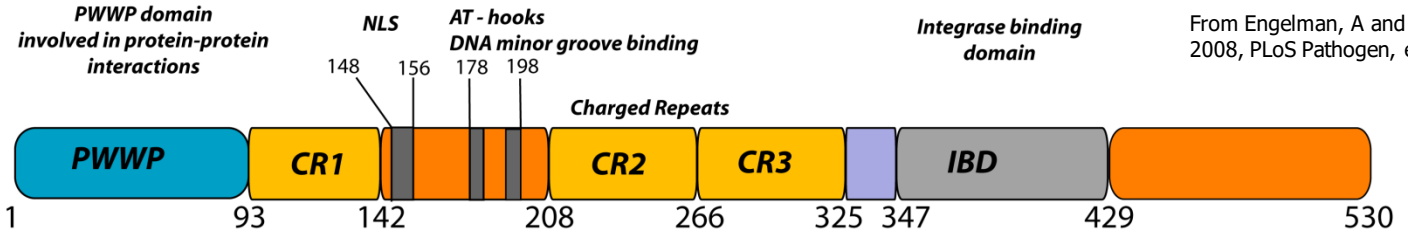
From unstable protein to stable complexes



Levy N, Eiler S, Pradeau-Aubreton K, Maillot B, Stricher F, Ruff M (2016). Production of unstable proteins through the formation of stable core complexes. Nature Communications 7: 10932

Eiler S, Levy N, Maillot B, Batisse J, Pradeau Aubreton K, Oladosu O, Marc Ruff (2018). Unstable protein purification through the formation of stable complexes. Methods in Molecular Biology, 1764:315-328

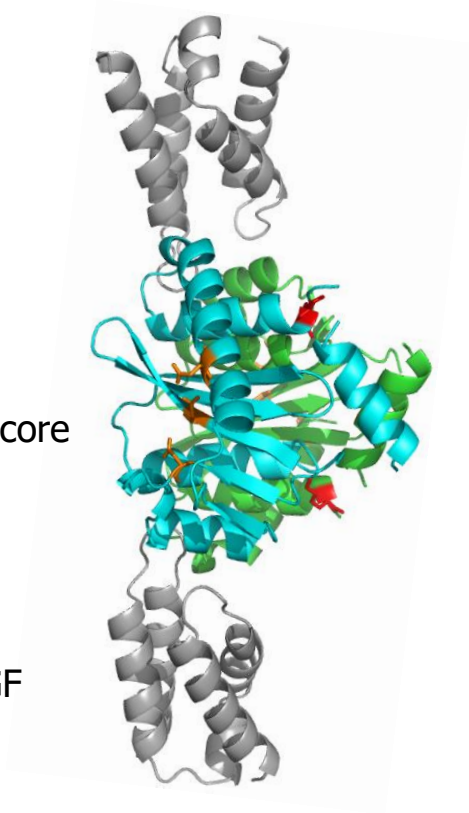
HIV-1 IN interact with LEDGF



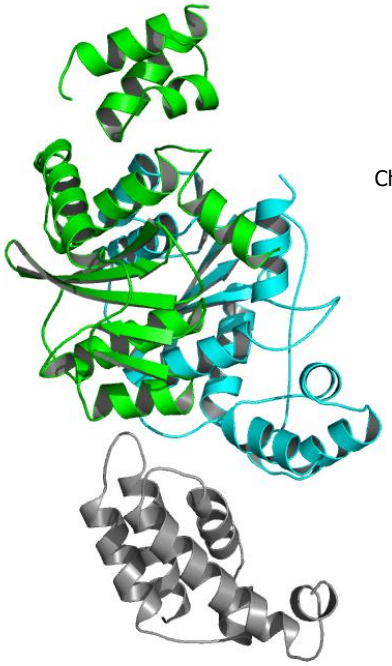
From Engelman, A and Cherepanov, P 2008, PLoS Pathogen, e1000046

IN catalytic core

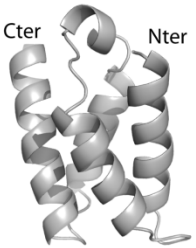
IBD LEDGF



Cherepanov et al. 2005



Hare et al. 2009

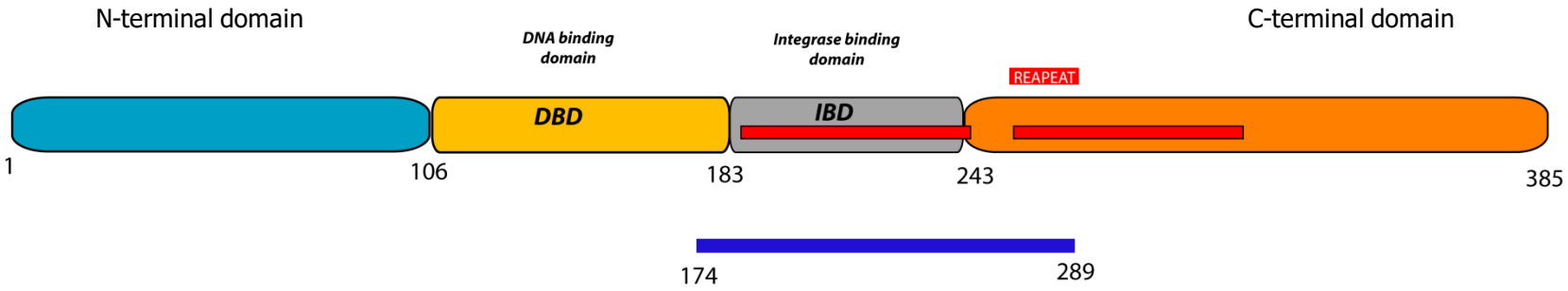


Cherepanov et al. 2005

LEDGF/P75

- PC4 and SFRS1 interacting protein (Lens epithelium-derived growth factor) (Transcriptional coactivator p75/p52) (Dense fine speckles 70 kDa protein) (DFS 70) (CLL-associated antigen KW-7).
- Transcriptional coactivator involved in neuroepithelial stem cell differentiation and neurogenesis. Involved in particular in lens epithelial cell gene regulation and stress responses. May play an important role in lens epithelial to fiber cell terminal differentiation. May play a protective role during stress-induced apoptosis.
- Length: 530 aa, molecular weight: 60103 Da, pI = 9.85

IN interact with INI1/SNF5



SNF5/Ini1, a subunit of the SWI/SNF chromatin remodeling complex, is the first cofactor identified to interact with IN.

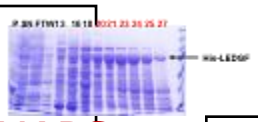
SNF5/Ini1 is one of the core subunits of the ATP-dependent chromatin remodeling complex SWI/SNF that regulates expression of numerous eukaryotic genes by altering DNA/histone interactions

It has been postulated that SNF5/Ini1 could target PICs to regions of the genome that are enriched for the SWI/SNF complex

IN/LEDGF : complex formation and purification: E. Coli

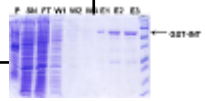
Purification GST-P3C-INT and HIS-Thr-LEDGF

Purification His-Thr-LEDGF



1M NaCl, 7mM CHAPS

Purification GST-P3C-INT



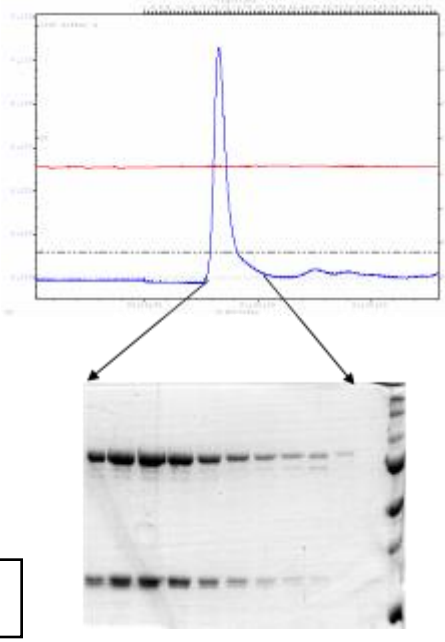
Complex formation by dialysis

GST Affinity

P3C cut

GST Affinity

Gel filtration G200



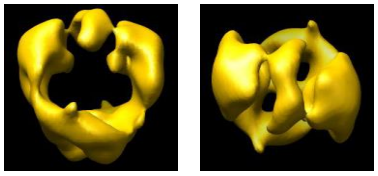
Functional and biochemical characterization

HTS Screening of inhibitors

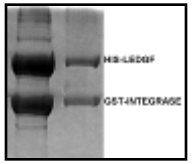
Crystallization Assays

SAXS, SANS studies

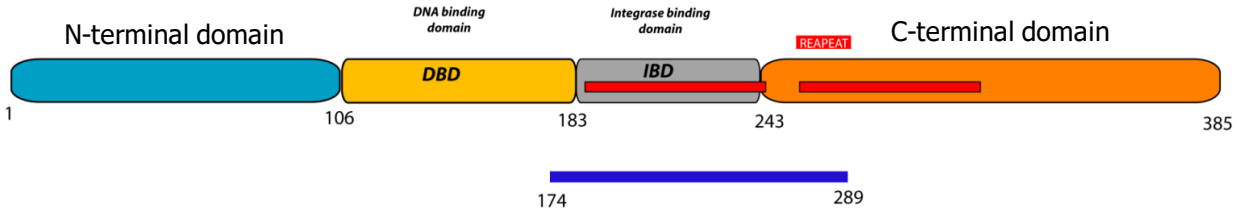
EM studies



Yield: 5 mg of complex (INT 5L, LEDGF 6L)



IN/LEDGF/INI1 : complex formation and purification



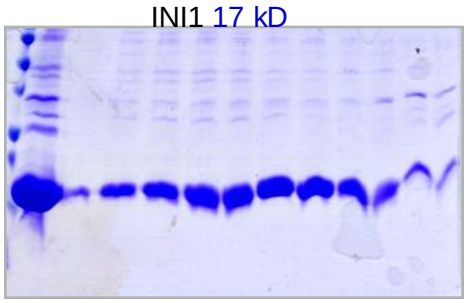
HIS-INI1(174-289) → His Affinity → Gel filtration G200

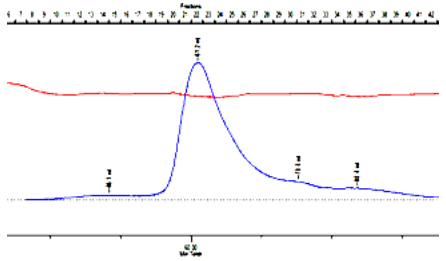
+ IN/LEDGF

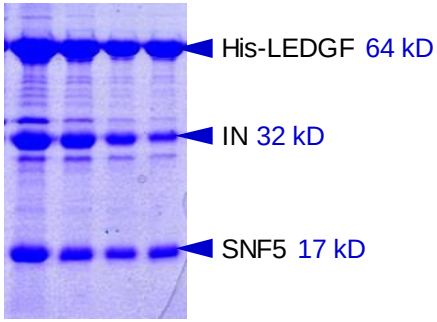
High salt detergent

Gel filtration G200

- Functional and biochemical characterization
- HTS Screening of inhibitors
- Crystallization Assays
- SAXS studies
- EM studies



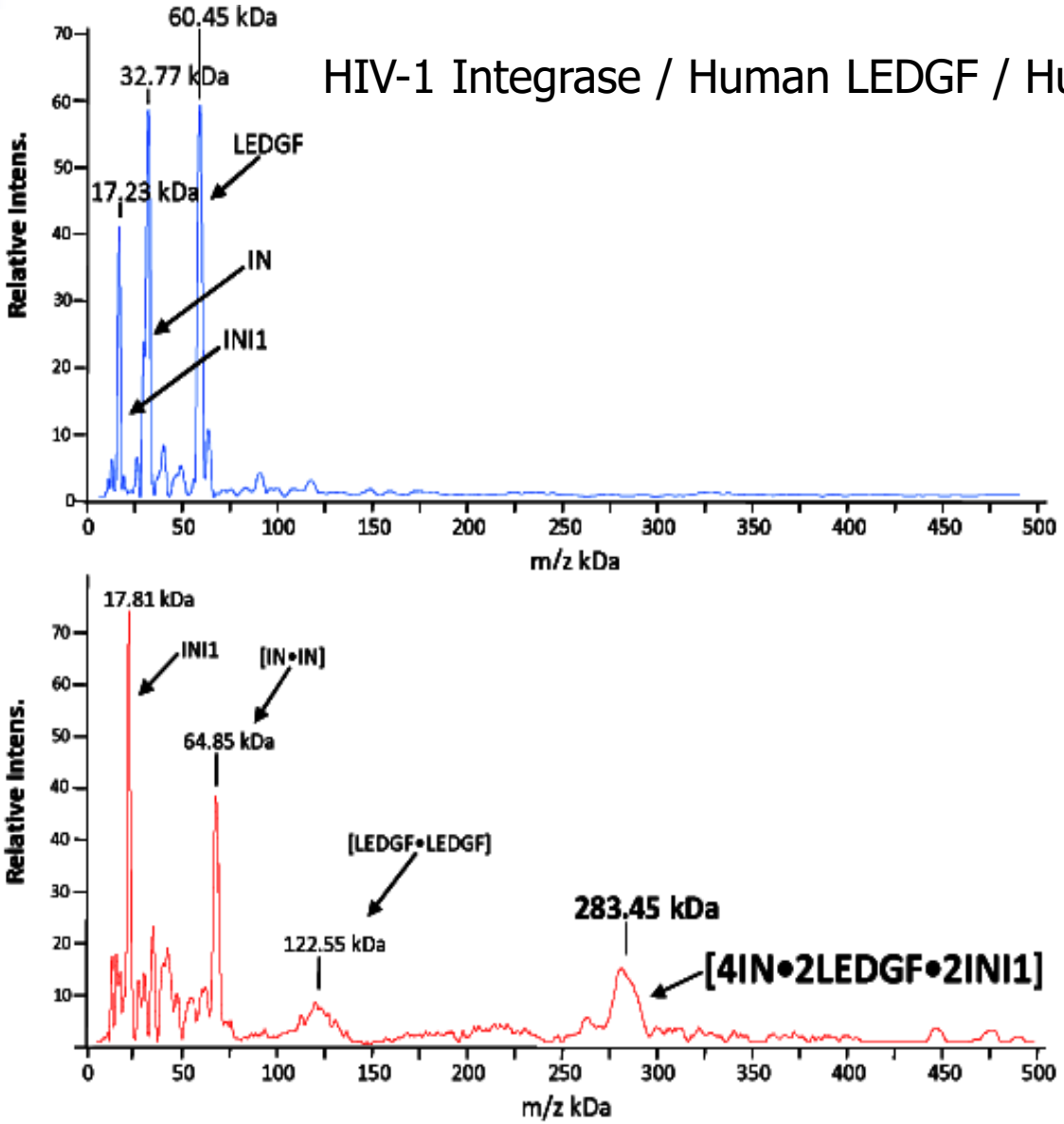




Yield: 2.5 mg of complex (INT 0.5L, LEDGF 0.5L, SNF5 1.0L)

Protein complexes analysis : High Mass MALDI-ToF

HIV-1 Integrase / Human LEDGF / Human INI1 (173-290)



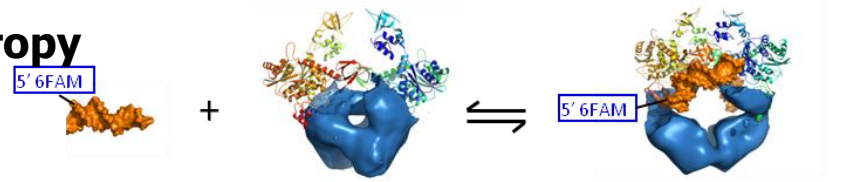
IN/LEDGF and IN/LEDGF/INI1: Functional characterization

Viral DNA Binding assay by fluorescence anisotropy

DNA: 40-mer duplex that mimic the HIV-1 U5 viral DNA end

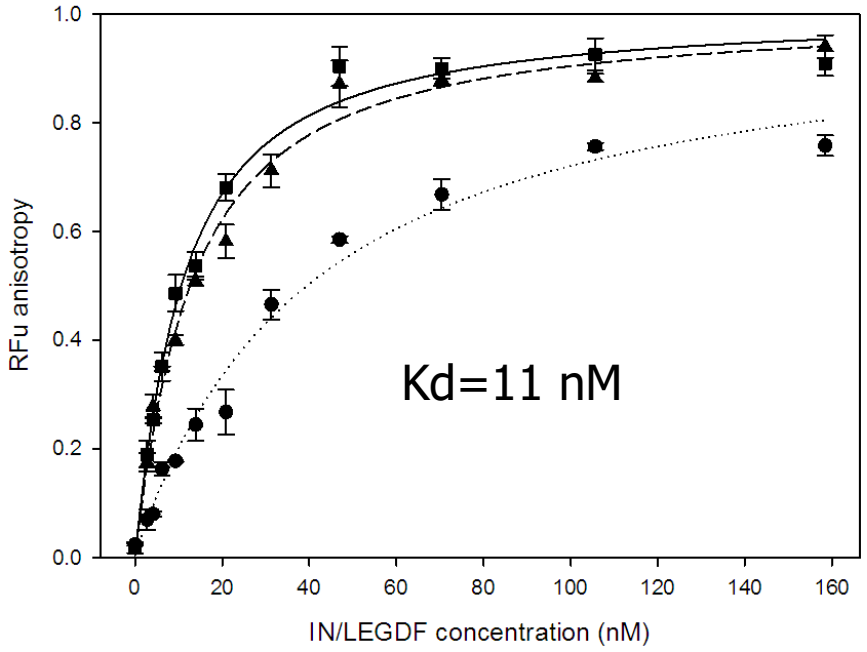
5′-GACTACGGTTCAAGTCAGCGTGTGGAAAATCTCTAGCAGT-3′

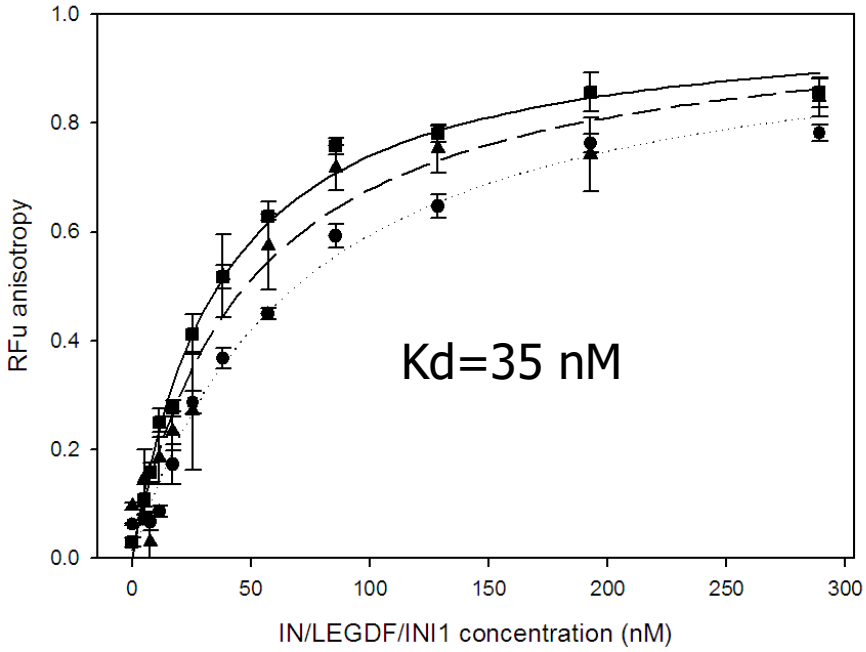
3′-CTGATGCCAAGTTCAGTCGCACACCTTTTAGAGATCGTCA-5′



Fast rotation: low anisotropy

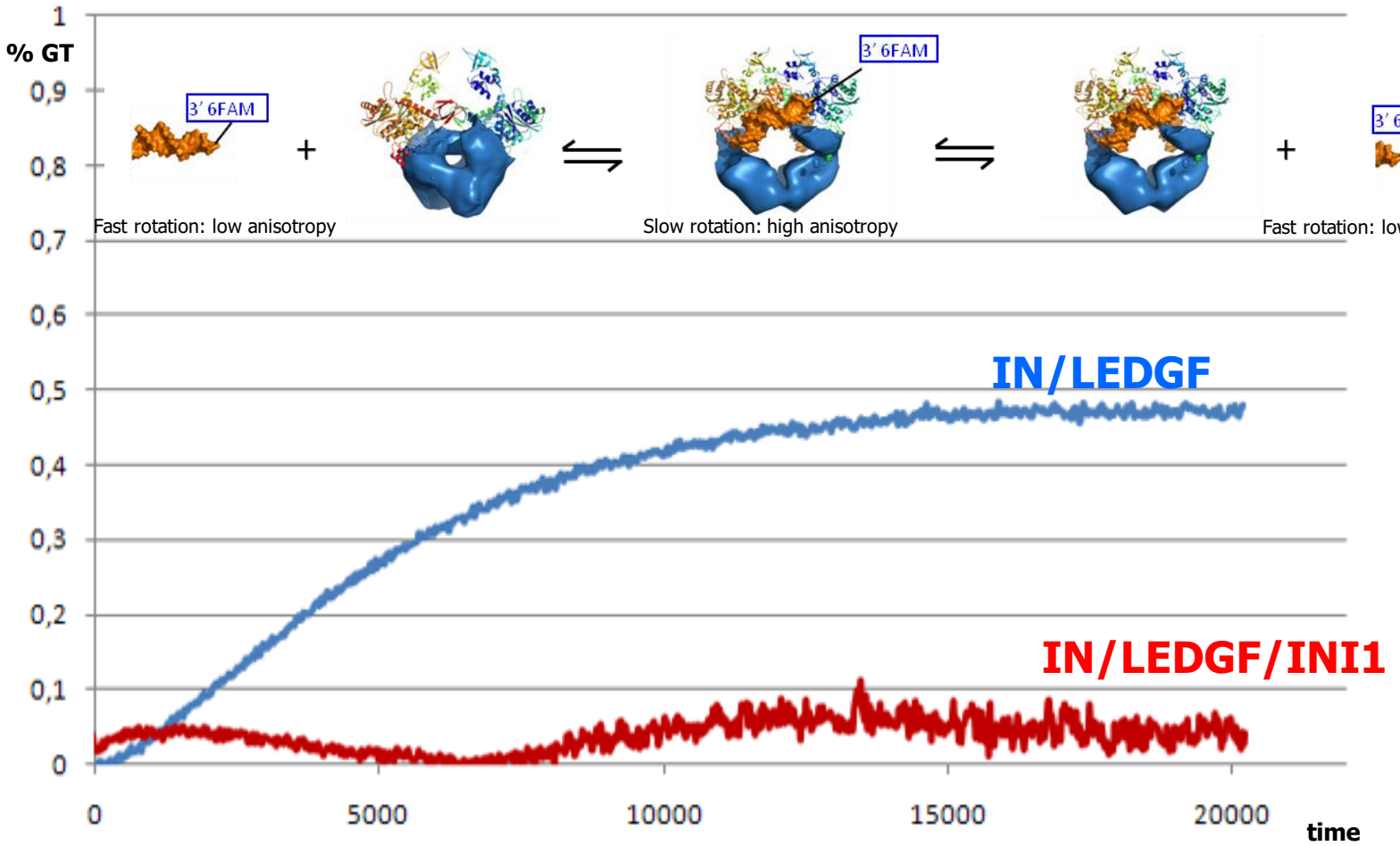
Slow rotation: high anisotropy





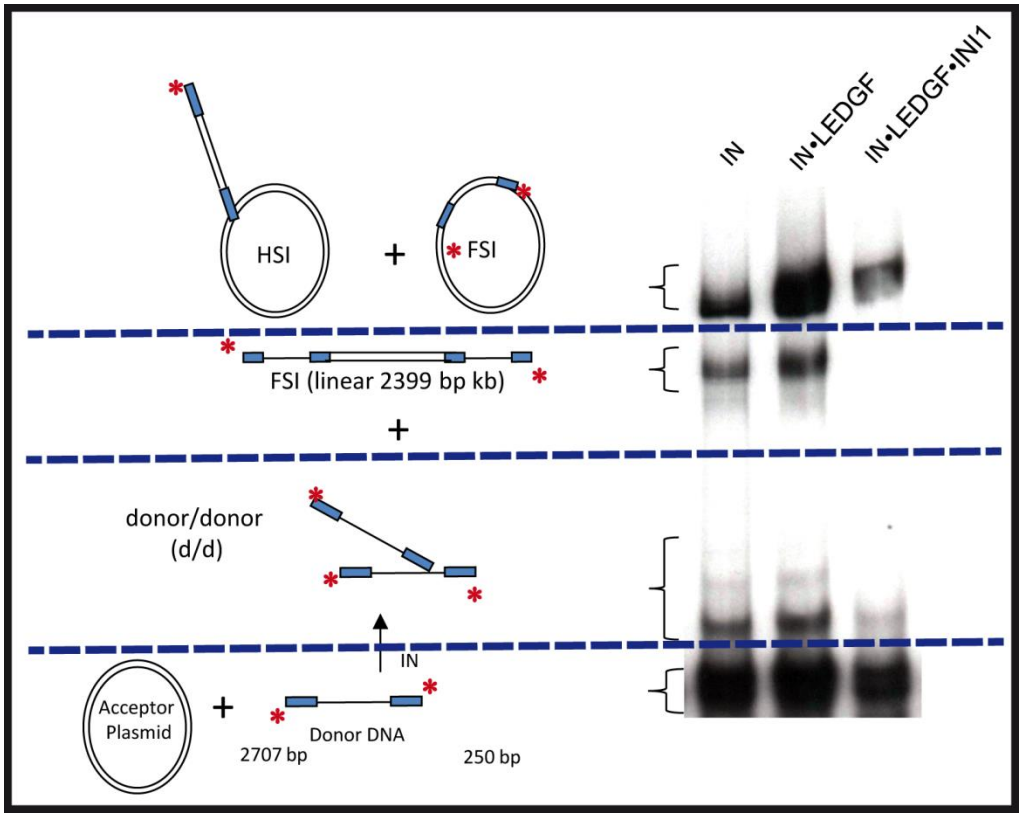
IN/LEDGF and IN/LEDGF/INI1: Functional characterization

The 3′ Processing assay by fluorescence anisotropy



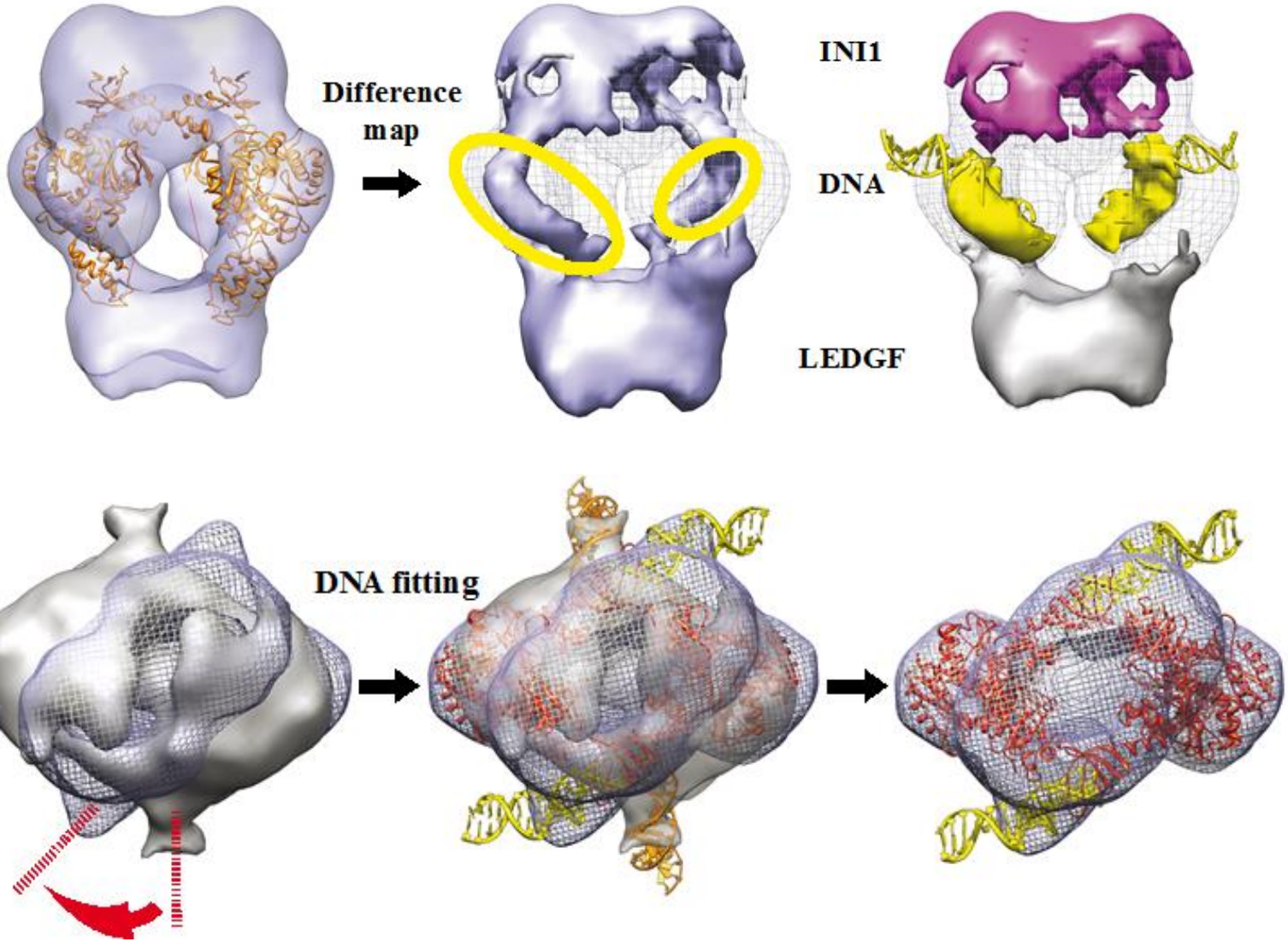
INI1 inhibits the *3′ processing* activity of IN/LEDGF

IN/LEDGF and IN/LEDGF/INI1: Functional characterization

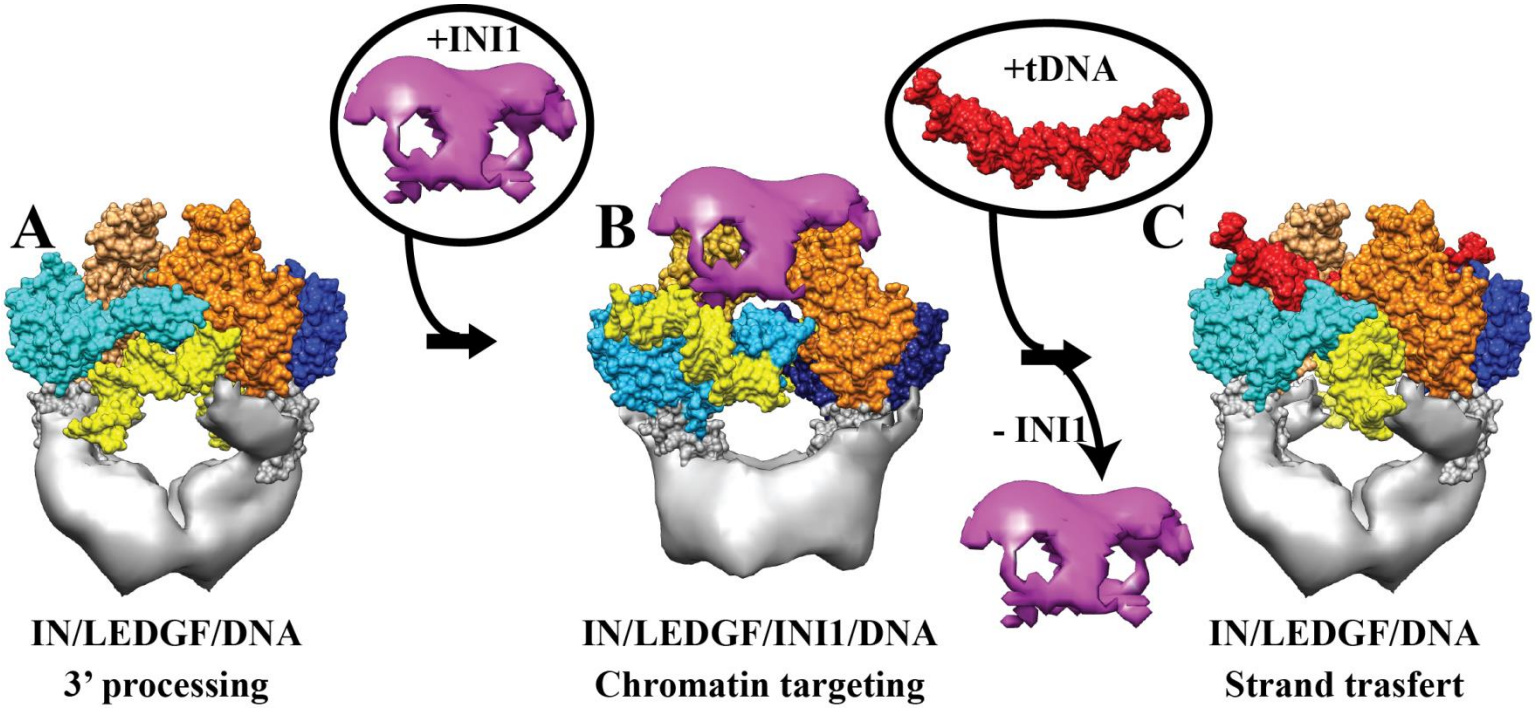


In the presence of INI1, integration occurs with reduced kinetics compared to IN alone or to the IN/LEDGF complex with strongly reduced by-products formation

IN/LEDGF/INI1/DNA: Cryo-EM structure



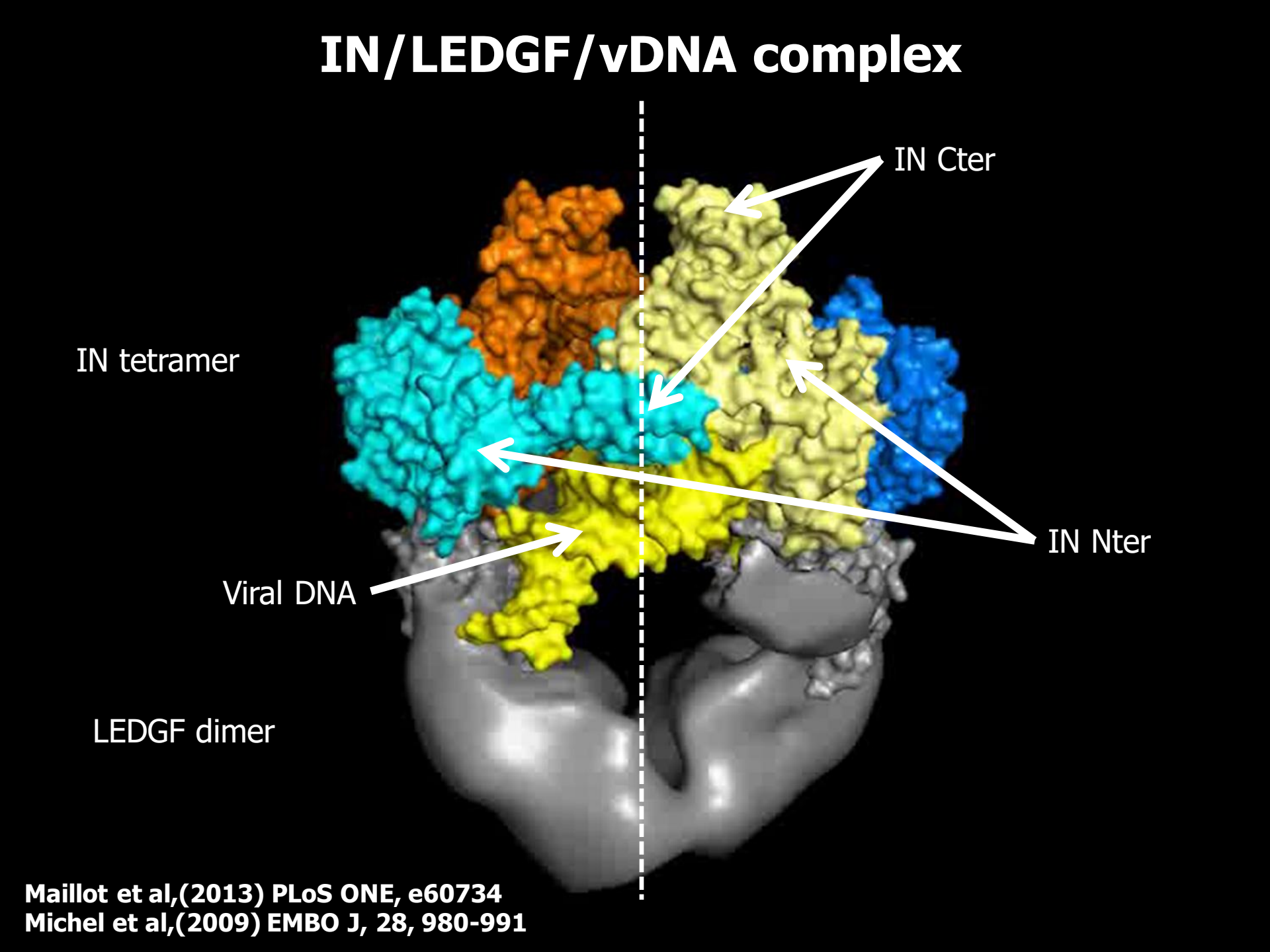
Function of INI1 in HIV-1 infection



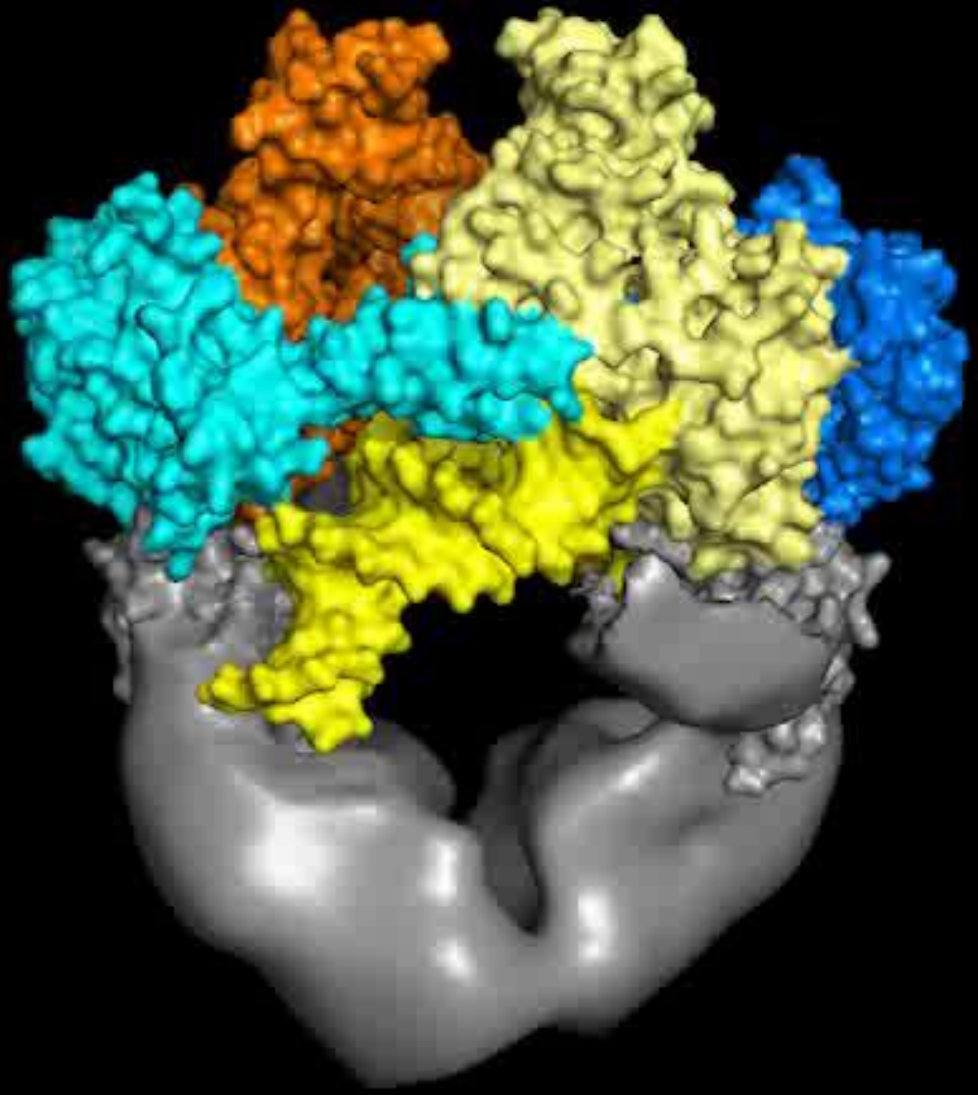
Benoit Maillot, Nicolas Lévy, Sylvia Eiler, Corinne Crucifix, et al., (2013), Structural and functional role of INI1 and LEDGF in the HIV-1 preintegration complex, PlosOne, In Press

Michel, F., Crucifix, C., Granger, F., et al., (2009). Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. EMBO J., 28, 980-991

IN/LEDGF/vDNA complex

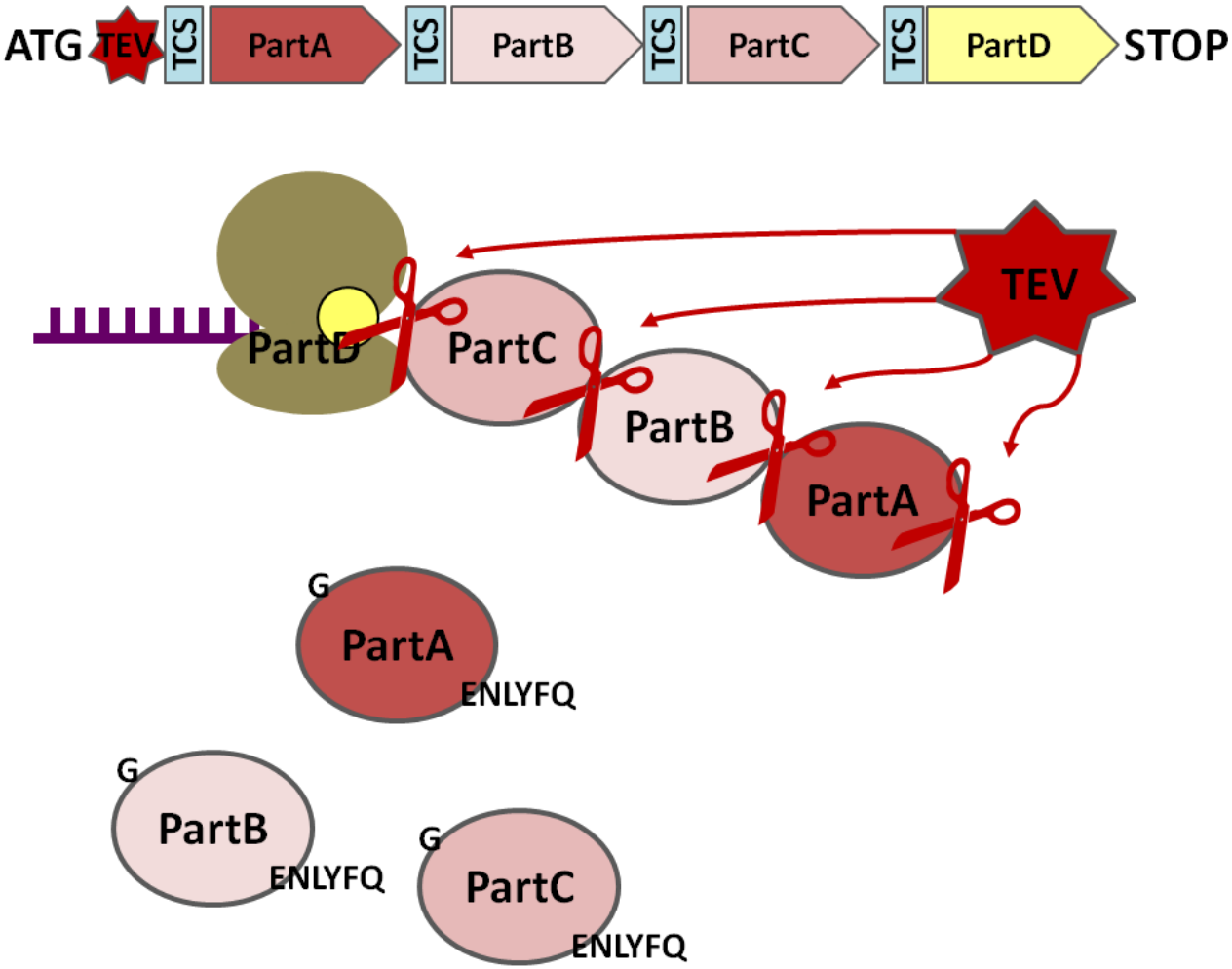


Maillot et al,(2013) PLoS ONE, e60734
Michel et al,(2009) EMBO J, 28, 980-991



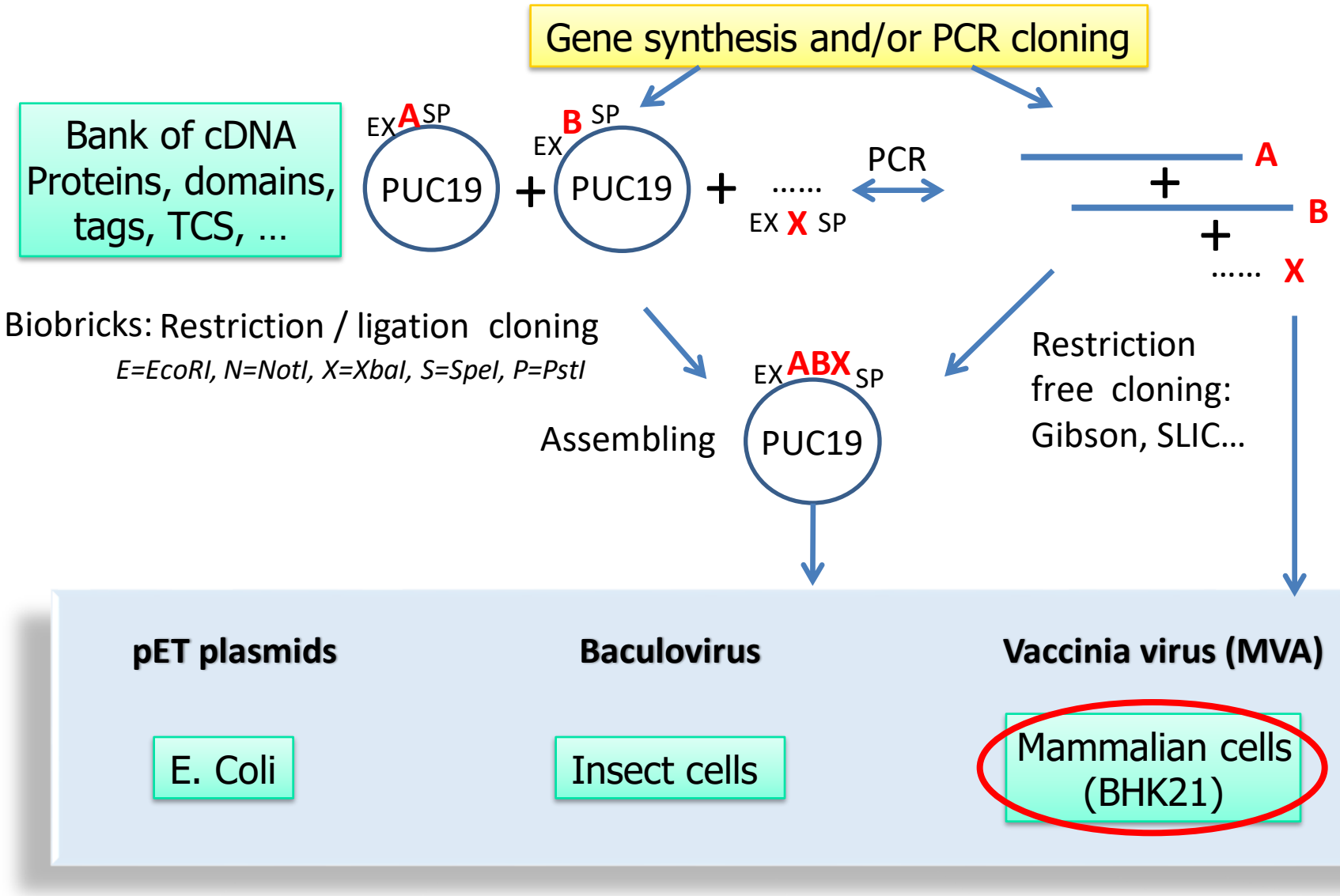
New setup for protein complexes production and purification

Co-expression of the different partners together in the same organism



Bieniossek et al., Trends in Biochemical Sciences 2012;37(2):49-57 (insect cell expression)

Protein complexes expression in E. Coli, Insect and mammalian cells



cDNA collection for protein expression in E. Coli, Insect and mammalian cells

Purification tags	H6P	6xHis - P3C cleavage site
	H10P	10xHis - P3C cleavage site
	H10FP	10xHis - Flag- P3C cleavage site
	FH10P	Flag - 10xHis - P3C cleavage site
	RP	Strep - P3C cleavage site
	RRP	Strep - Strep - P3C cleavage site
	RRFP	Strep - Strep - Flag - P3C cleavage site
	FRRP	Flag - Strep - Strep - P3C cleavage site
TEV cleavage sites	TT1	Twin TEV cleavage site (TCS) 1 : TCS A – TCS B
	TT2	Twin TEV cleavage site (TCS) 2 : TCS C – TCS D
	TT3	Twin TEV cleavage site (TCS) 3 : TCS E – TCS F
	TT4	Twin TEV cleavage site (TCS) 4 : TCS G – TCS H
Linker STOP	STOP	STOP codon
Proteins	TEV	Tobacco Etch Virus protease
	LEDGF	Human Lens epithelial derived growth factor
	IN	HIV-1 Integrase
	IN*	Degenerated HIV-1 Integrase
	TRNSR2	Human Transportin
	TT8-eYFP	Twin TEV cleavage site 8 – enhanced Yellow Fluorescent Protein
	……	

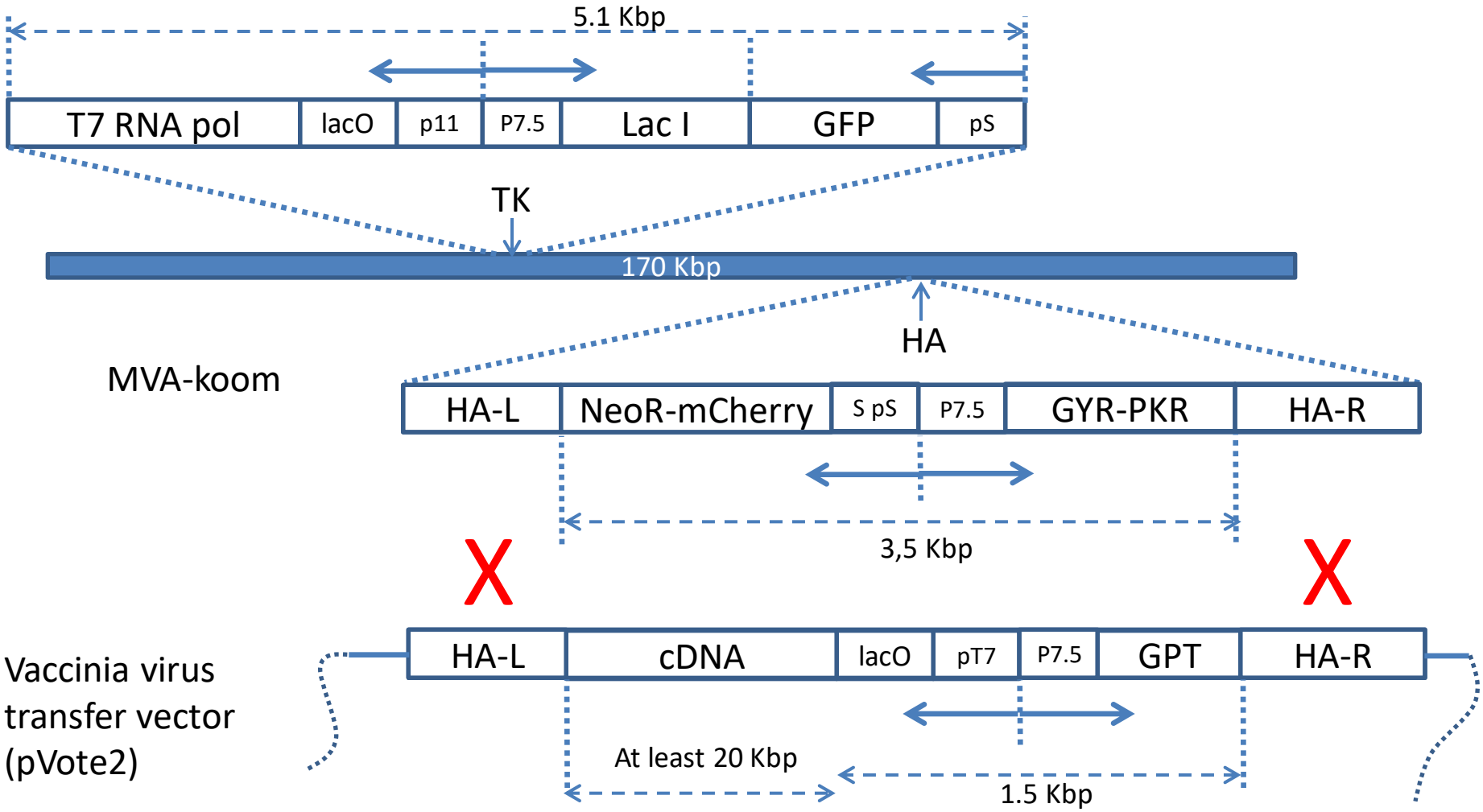
Vaccinia virus as an expression vector for mammalian cells

- **Vaccinia virus**: poxvirus family
  - dsDNA virus (≈ 200 kb) encoding its own transcription and replication machinery
  - viral multiplication in the cell cytoplasm: no RNA splicing
  - viral infection diverts the cellular machinery in its favour
  - at least 25 kb of foreign DNA

- **MVA**: Modified Vaccinia virus Ankara
  - non replicative in human cells
  - safe for people with immunodeficiency or skin disorders
  - manipulation is authorized under BSL1 containment

- **Mammalian cells**: BHK21 (baby hamster kidney cells)
  - authentic post-translational modifications
  - proper folding
  - protein function and structure analysis

- **Encode T7 RNA polymerase, IPTG inducible**

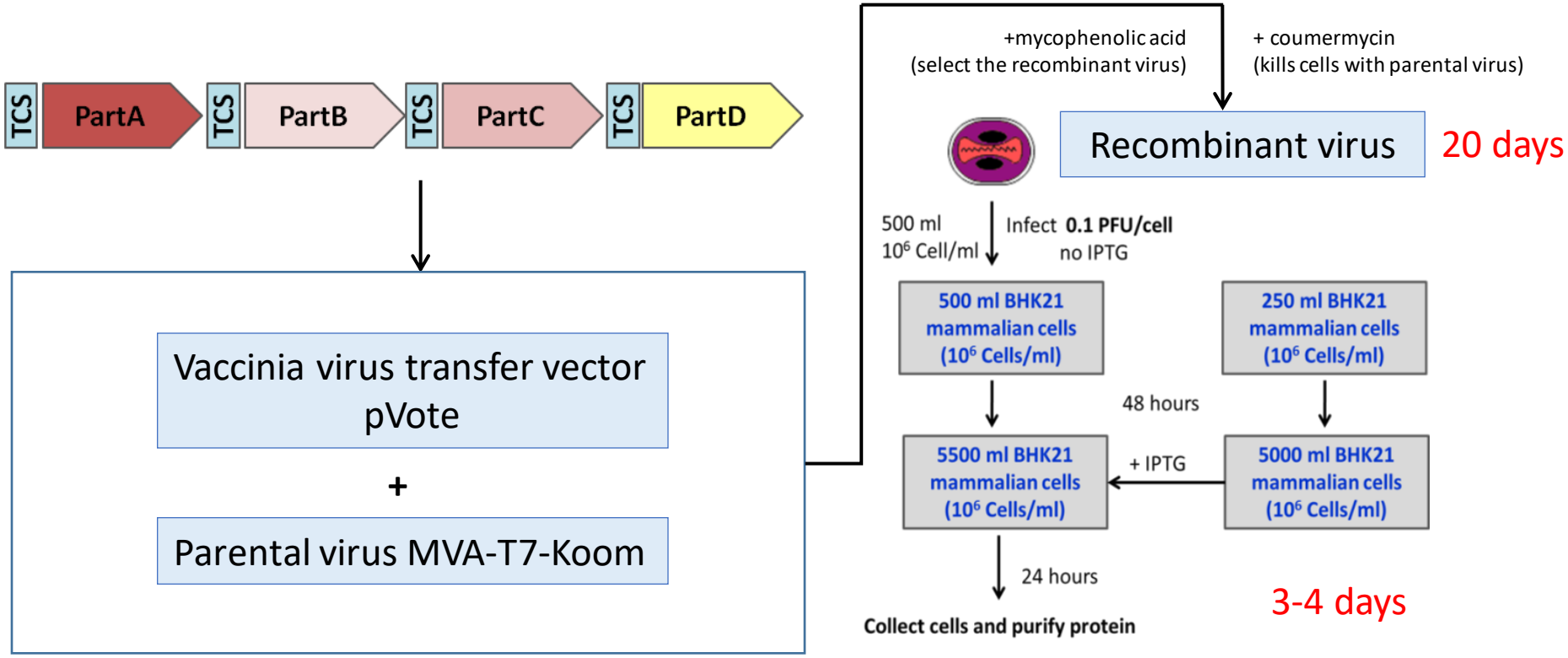


Before recombination: GFP(+), mCherry(+), NeoR(+), GYR-PKR(+) (Coumermycin Sensible), GPT(-) (Mycophenolic acid sensible)

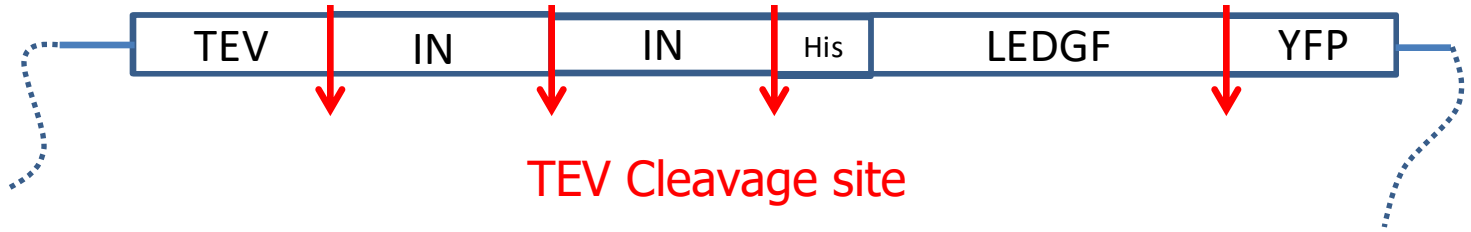
After recombination: GFP(+), mCherry(-), NeoR(-), GYR-PKR(-) (Coumermycin Resistant), GPT(+) (Mycophenolic acid resistant)

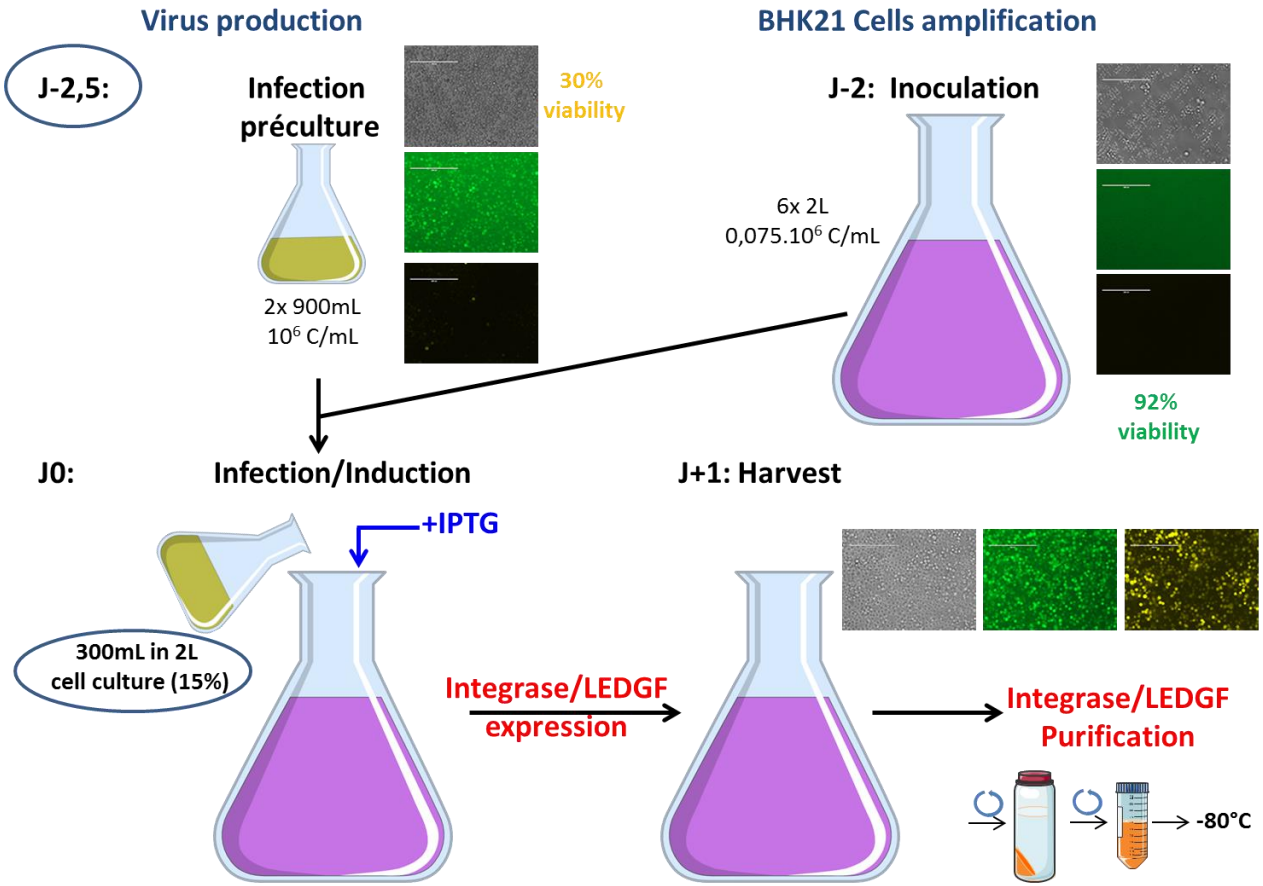
Mammalian cell protein complexes production

Optimization of the Modified Ankara Vaccinia virus as an expression vector for protein production in BHK21 mammalian cells



Large scale production of Integrase/LEDGF complex in mammalian cells

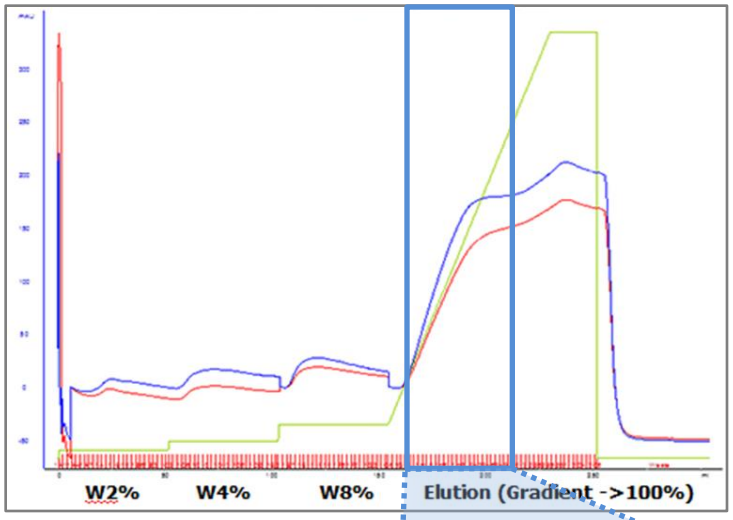




Large scale production of Integrase/LEDGF complex in mammalian cells

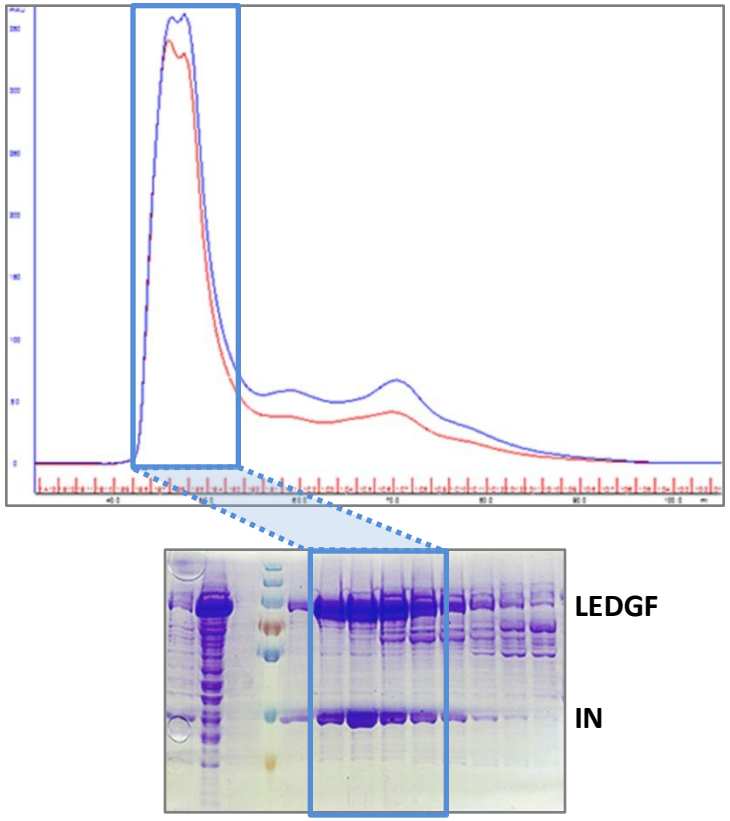
12L of BHK21 cell culture

Affinity chromatography (HisTrap Excel 5ml)



Concentration on Amicon 100KDa

Size exclusion chromatography (Highload 16/60 Superdex 200)



10 mg of IN/LEDGF complex

Towards high resolution structures

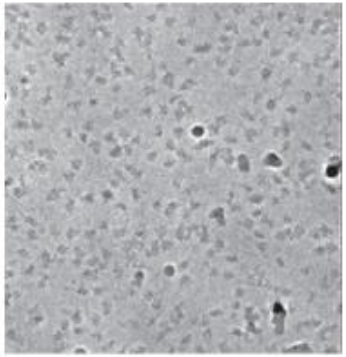
IN/LEDGF/DNA: High resolution CryoEM data collection in progress

Microscope: FEI Titan Krios

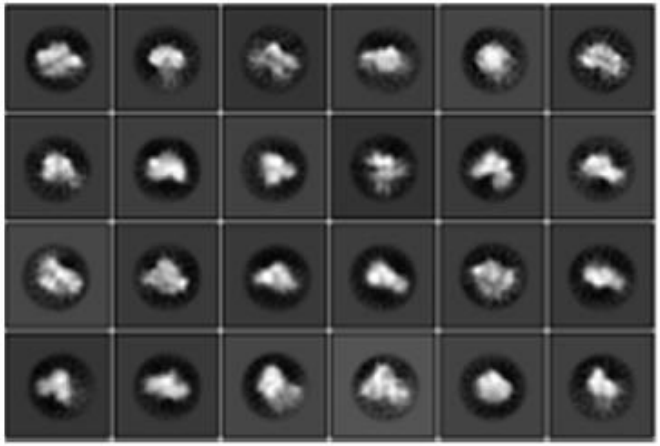
Accelerating Voltage: 300 kV
Spherical Abberation: 0.01 mm
Amplitude Contrast: 10 %
Pixel Size: 1.09 A
Phase Plate shift: from 10 to 175 A
Defocus: from 0.3 to 1.2 µm
Defocus step: 5 A
Camera: Gatan CMOS K2 summit (4Kx4K)

Sample : IN/LEDGF/DNA
Cross-linked with Glutaraldehyde 0,1%
Purified on GF, concentrated 3x
Load on quantifoil EM grids R1,2/1,3
Glow discharged (ELMO)
Vitrified on Vitrobot (FEI)

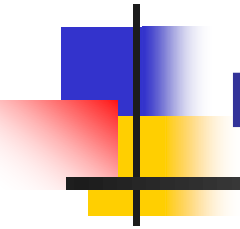
2614 micrographs
Motioncor & GCTF treated



2288 micrographs selected (CTF validated)
420057 auto picked particles
1st run 2D classification: 205227 selected particles
2nd run 2D classification: 138195 selected particles
3rd run 2D classification: 131758 selected particles



Eman2 (automatic picking), Relion2 (2D Classification), CryoSPARC (3D reconstruction)

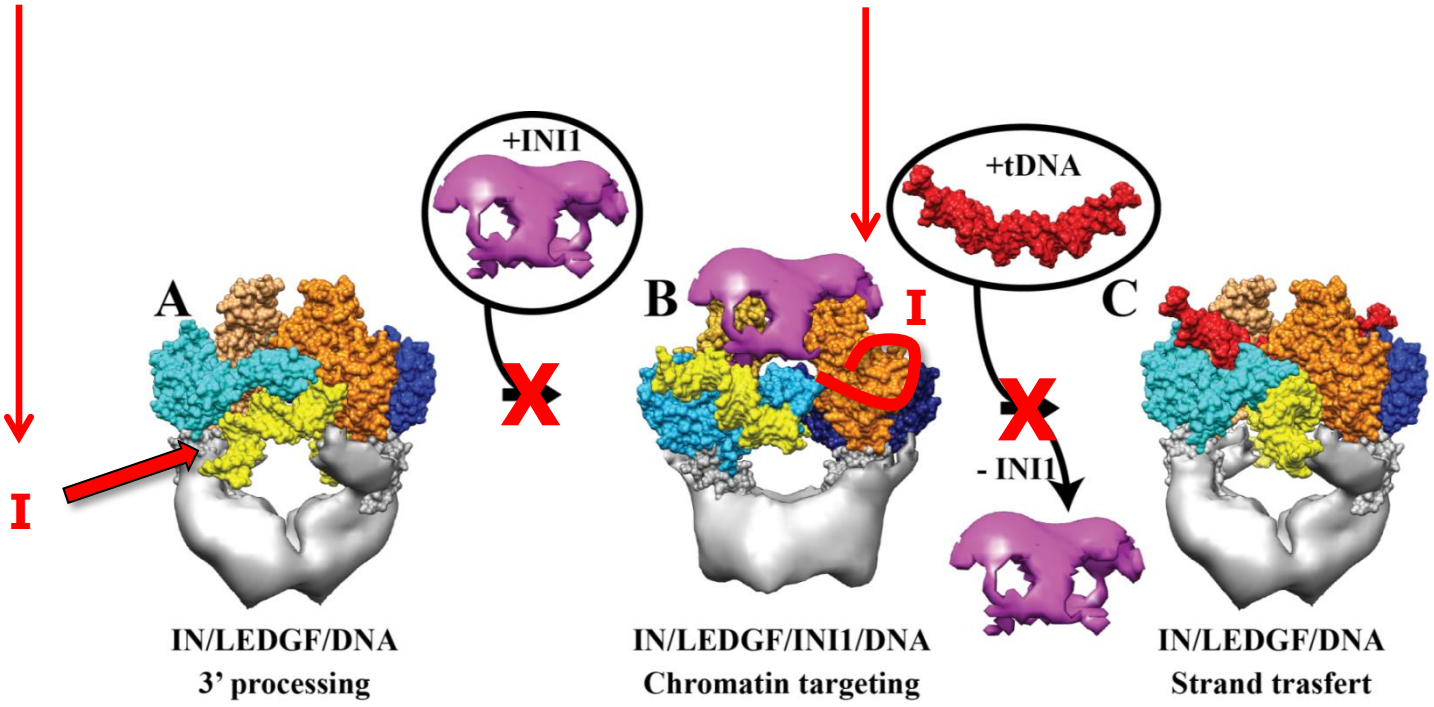


DRUG-DESIGN

Protein – Protein interaction and allosteric inhibitors

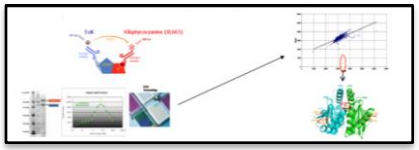
IN/LEDGF interaction inhibitors
Mutabilis

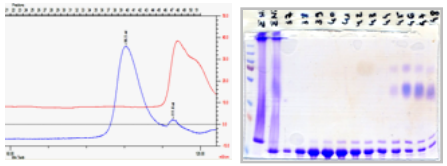
IN/LEDGF allosteric inhibitors
Shionogi



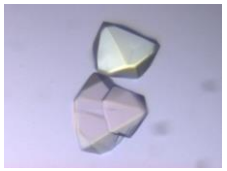
Protein – Protein interaction and allosteric inhibitors

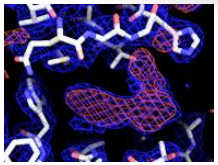
HTP IN – LEDGF interaction inhibitors screening (HTRF)





IN CCD Production and crystallization

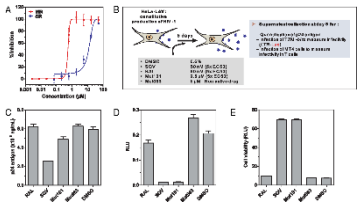




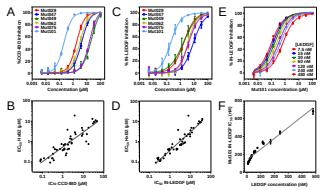
HTP inhibitors crystal soaking, data collection, structure

~100 structures solved

Structure based drug design

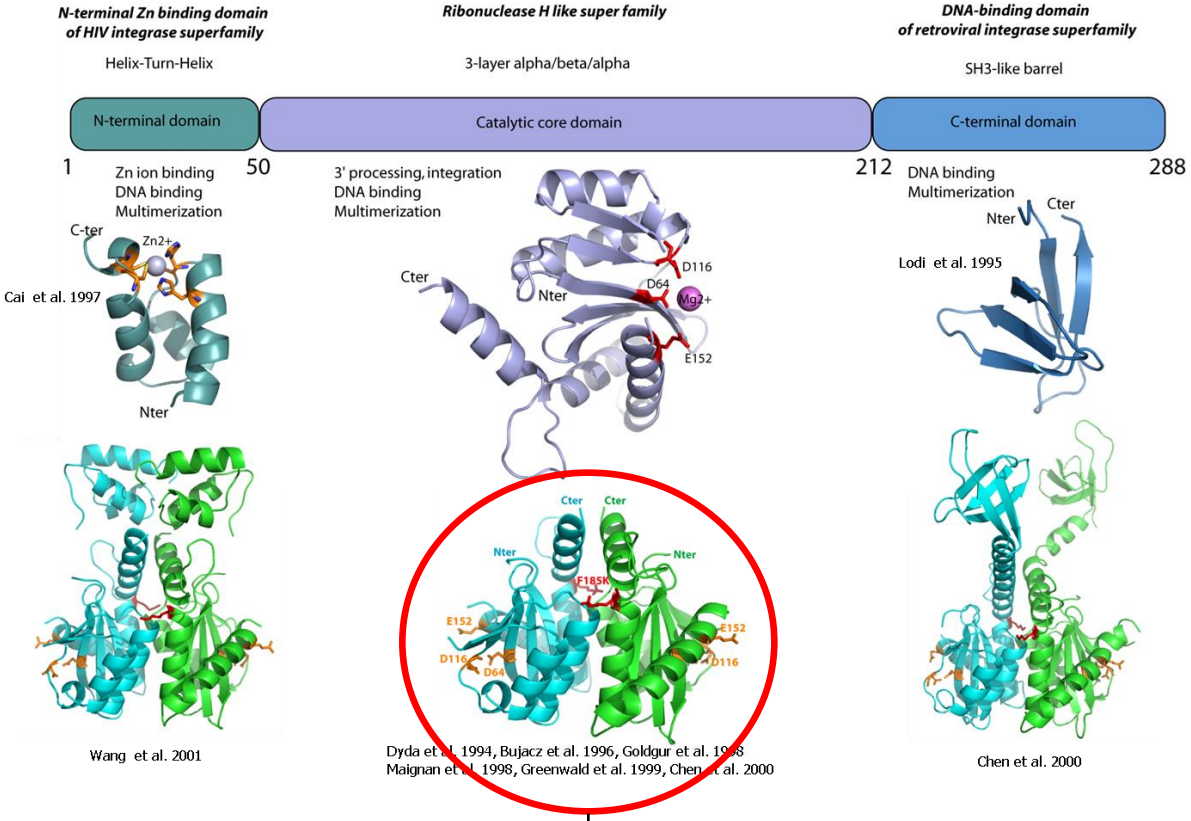


in-vitro functional assay



In cellulo functional assay

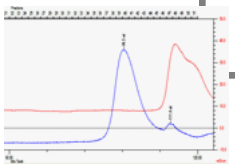
HIV Integrase



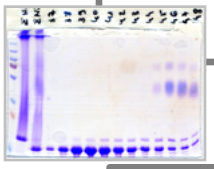
CONSTRUCT	MUTANT	SEQ.	VECTOR	MW	MW fusion	MW protein	PI	PI fusion
His-tb-INTEGRASE	F185K	50-212	pEt15b	19,9	1,9	18,0	8,05	10,54

Crystallization of IN CCD for Drug Design

Production of IN CCD [50-212]
E.Coli – 4h at 37°C



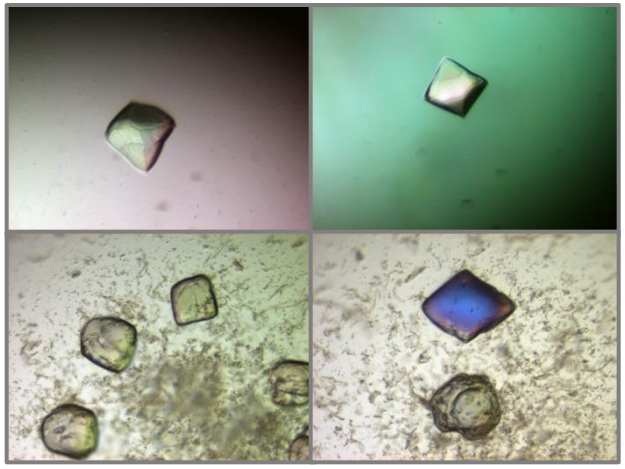
Affinity purification
Nickel column



Size exclusion chromatography
Superdex 200

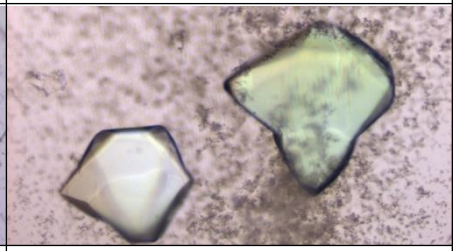
Crystallization
1.26 M AmSO4
100 mM NaCacodylate pH 6.5
Hanging drops
@ 22-23 °C

22-23°C
Diffraction
~2Å

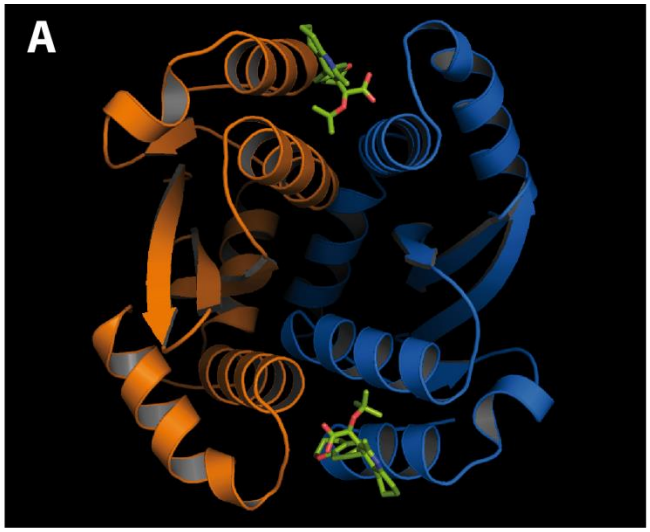


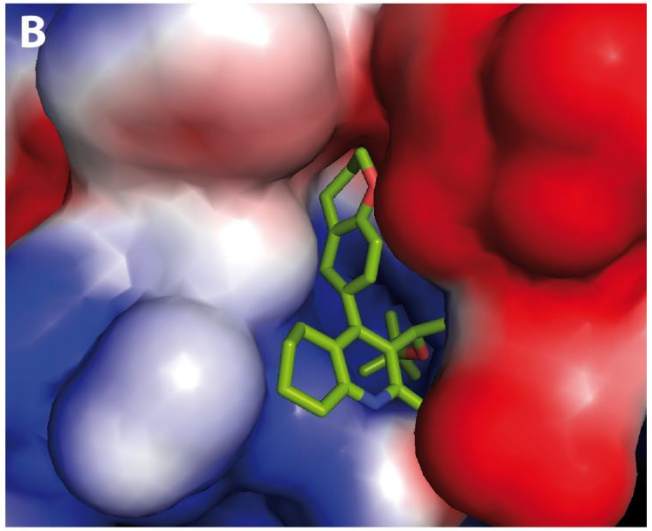
24-25°C
No diffraction !

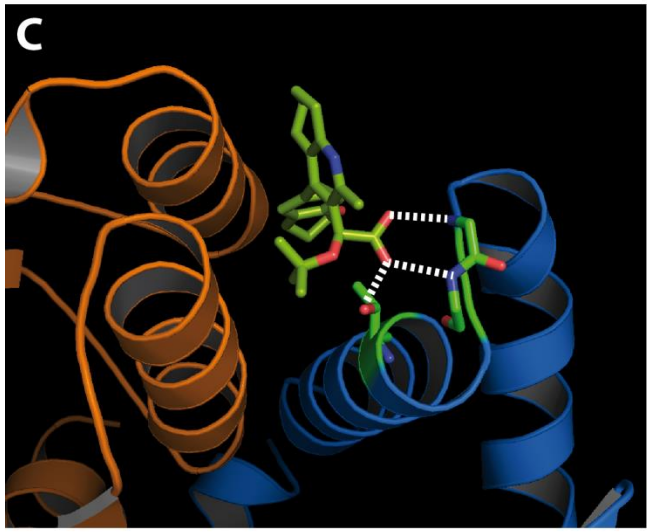
HIV Integrase 50 - 212 (F185K), Crystallization

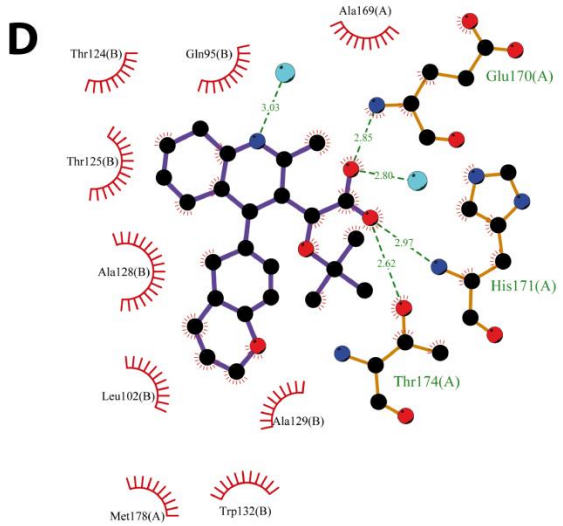
Crystal name	P1_B2_L11	P1_D4_L12	P1_C6_L13
Crystal source	Plate1_B2	Plate1_D4	Plate1_C6
Crystal image			
Protein concentration	3.1 mg/ml	3.1 mg/ml	3.1 mg/ml
Reservoir composition	1.26 M AS	1.26 M AS	1.50 M AS
Initial drop composition	2µl prot + 2µl res	2µl prot + 2µl res	2µl prot + 2µl res
Ligand soaking	LIG11	LIG12	LIG13

HIV Integrase CCD – ligand structure

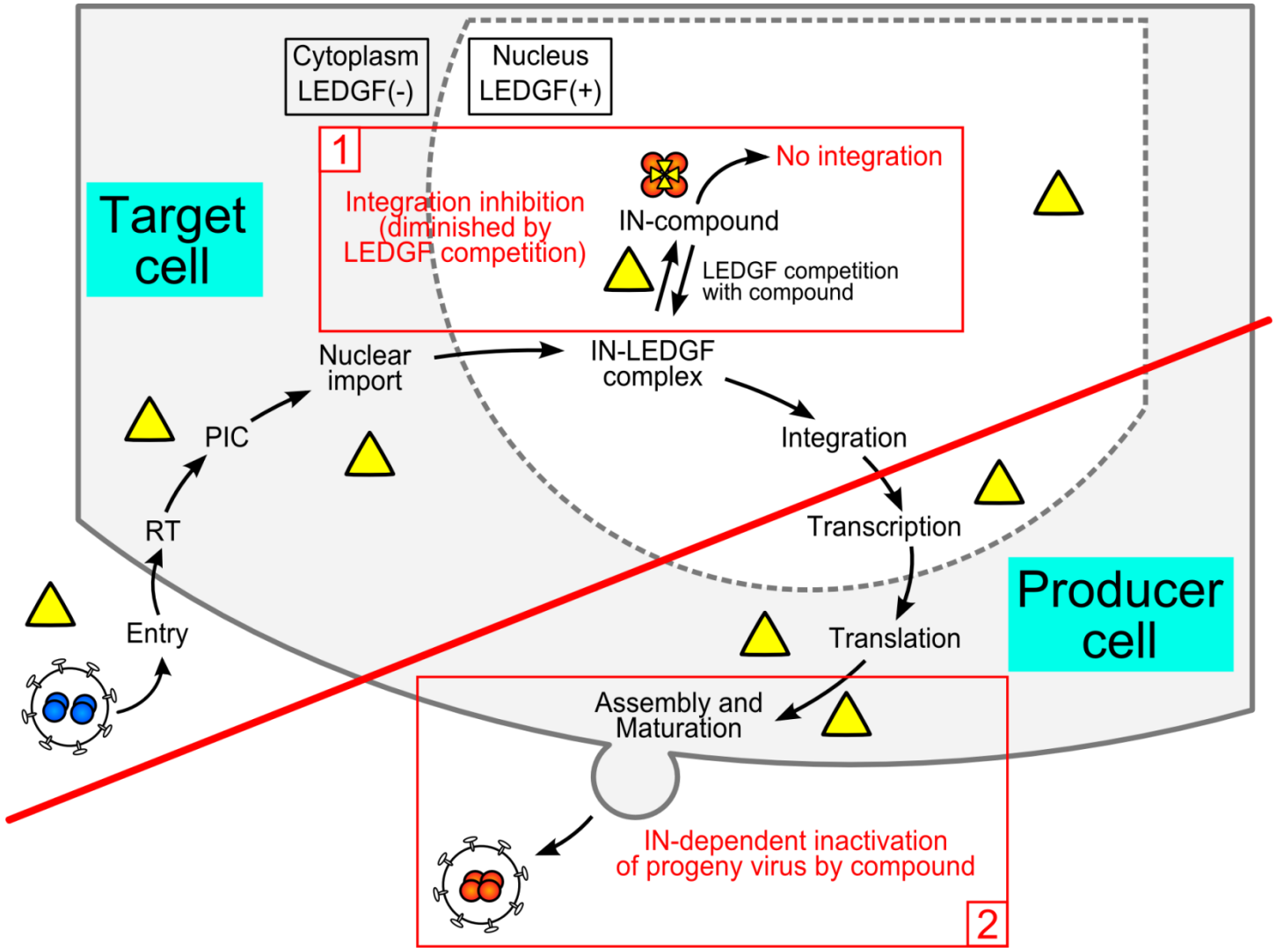






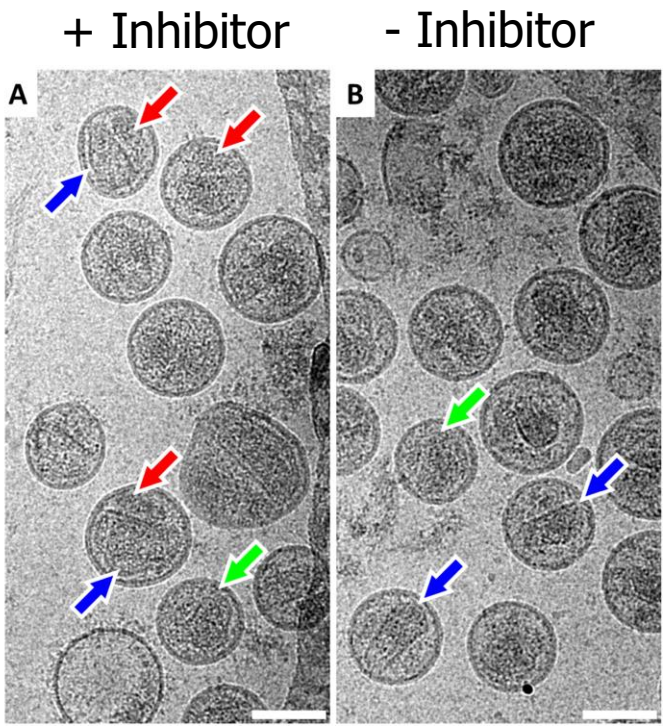


IN – LEDGF interaction and IN allosteric inhibitors



Le Rouzic et al., (2013), Retrovirology, 10, 144

IN allosteric inhibitors



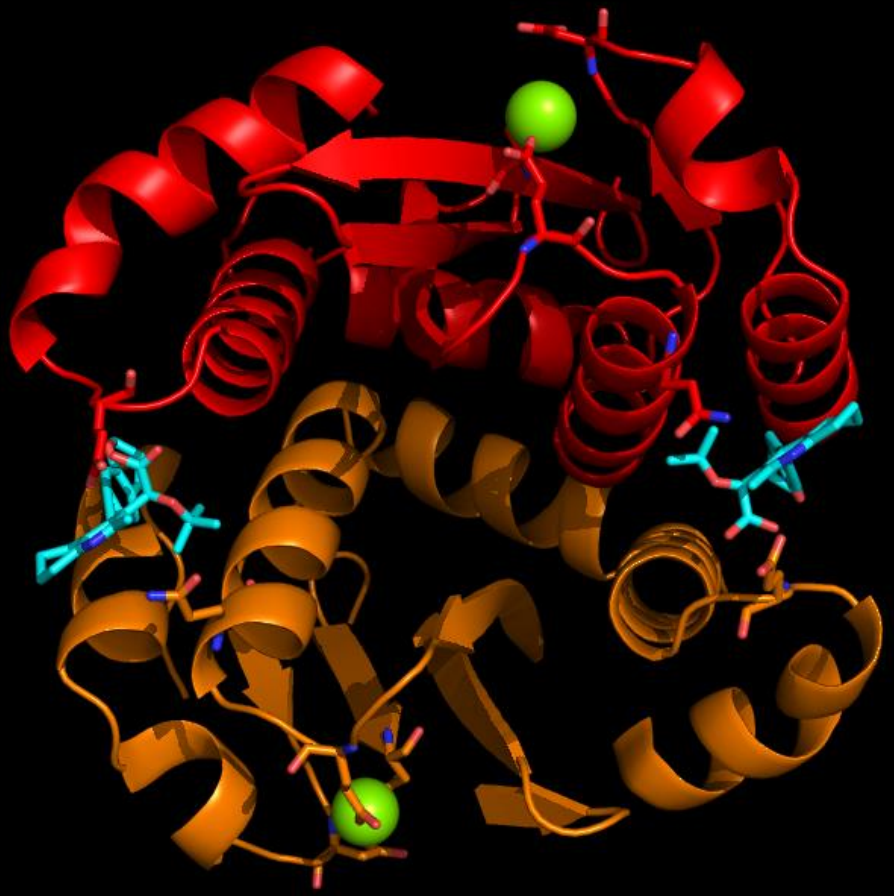
A. Viruses produced from 293T cells transfected with pNL4-3 in the presence of MUT-A.

B. NL4-3 viruses produced in the absence of MUT-A.

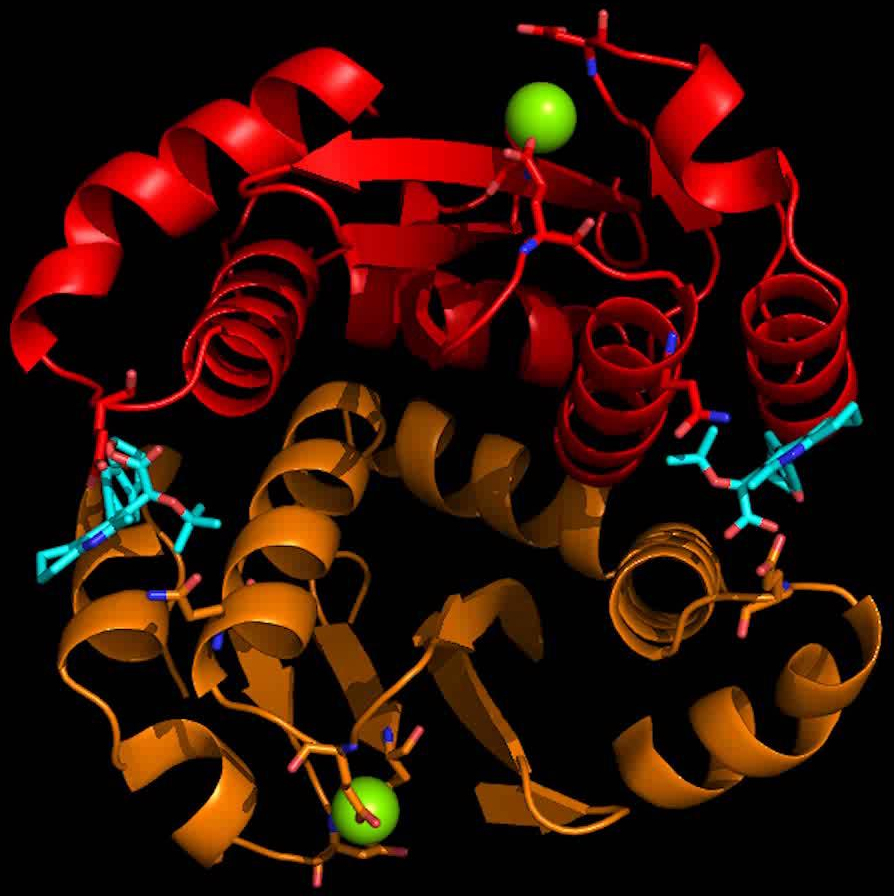
Red arrows: formation of eccentric condensates
Blue arrows: normal conical cores
Green arrows: non-conical cores.

Scale bars = 100 nm.

Bonnard, D. et al., J Biol Chem, 2018, 293(16):6172-6186
Amadori, C. et al., Retrovirology, 2017, 14:50.
Le Rouzic E, et al. Retrovirology. 2013 Nov 21;10(1):144.



Le Rouzic et al., (2013), Retrovirology, 10, 144



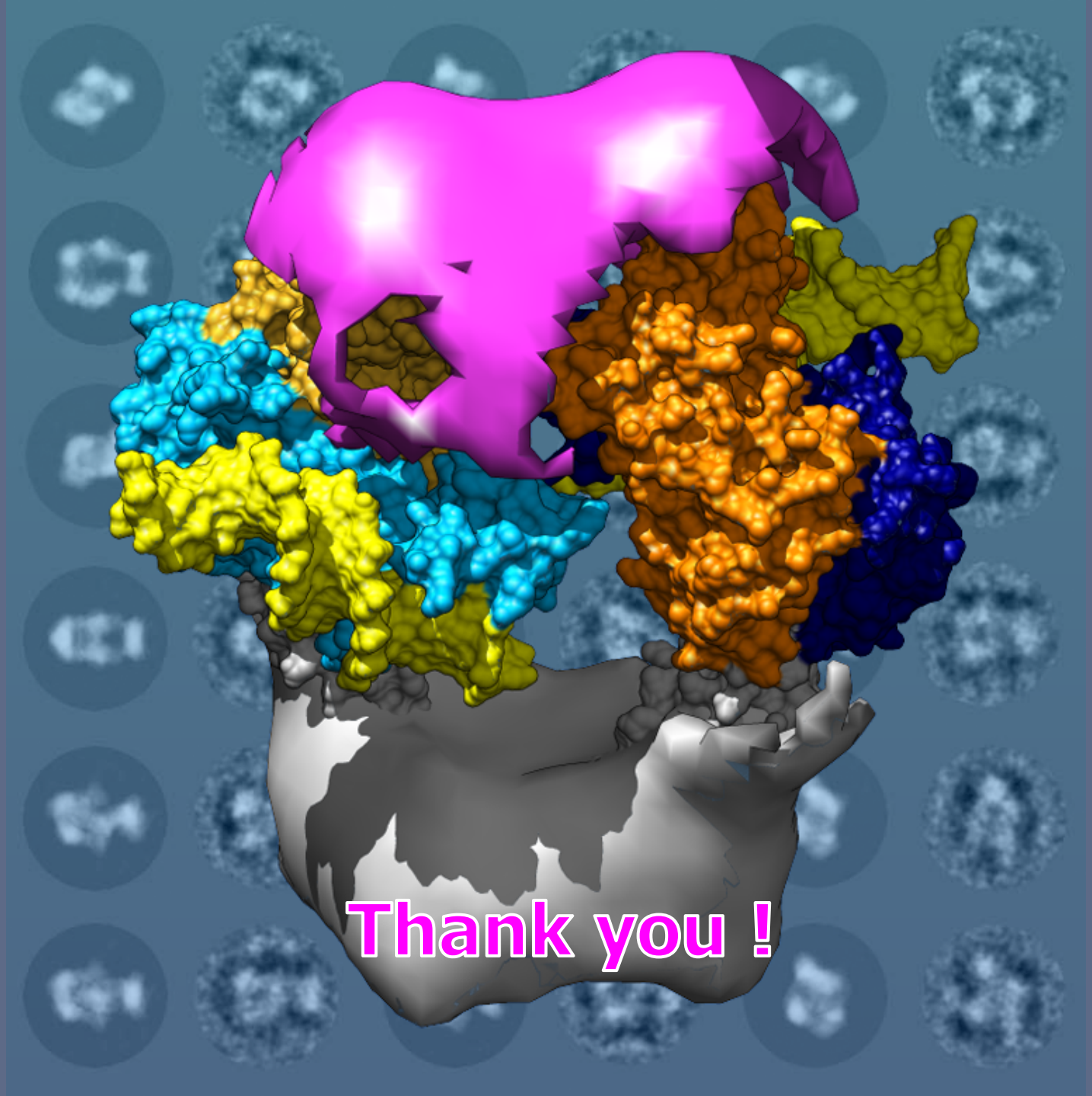
Acknowledgments

HIV-1 IN Drug design, conformational inhibitors
Tomokazu Yoshinaga, Shionogi, Osaka Japan
Richard benarous, Mutabilis, Paris
Yves Mely, University of Strasbourg
Molecular mechanisms of Viral DNA integration
Stéphane Emiliani, Cochin institute, Paris
Marc Lavigne, Pasteur Institute, Paris
Olivier Delelis, ENS, Cachan
Vincent Parissi, University of Bordeaux 2
Patrice Gouet, IBCP, Lyon
Ganjam V. Kalpana, Albert Einstein, NY, USA
IN Co-evolution
Matteo Negroni, IBMC, Strasbourg
IN Restriction factors (APOBEC 3G)
Eric Ennifar, IBMC, Strasbourg
Viral assembly and budding
Jean Christophe Paillart, IBMC, Strasbourg
Roland Marquet, IBMC, Strasbourg
Methods development
Yves Mely, University of Strasbourg
Olivier Haeberlé, MIPS, University of Haute-Alsace
Mass spectrometry
Alexis Nazabal, CovalX, Zurich
Sarah Cianferani, University of Strasbourg
CryoEM
Patrick Schultz, IGBMC
Julio Ortiz Espinoza, IGBMC
NMR
Bruno Kieffer, IGBMC
Serge Bouaziz, Paris Descartes
Synchrotron
Vincent Olieric, SLS, Villigen

Robert Drillien
Sylvia Eiler
Nicolas Levy
Julien Batisse
Eduardo Bruch

Valérie Lamour
Arnaud Vanden Broeck
Christophe Lotz

Oyindamola Oladosu
Karine Pradeau
Benoit Maillot


Thank you !