FROM UNSTABLE PROTEIN TO STABLE COMPLEXES
THE HIV-1 PRE-INTEGRATION COMPLEX
STRUCTURE, FUNCTION AND DYNAMICS

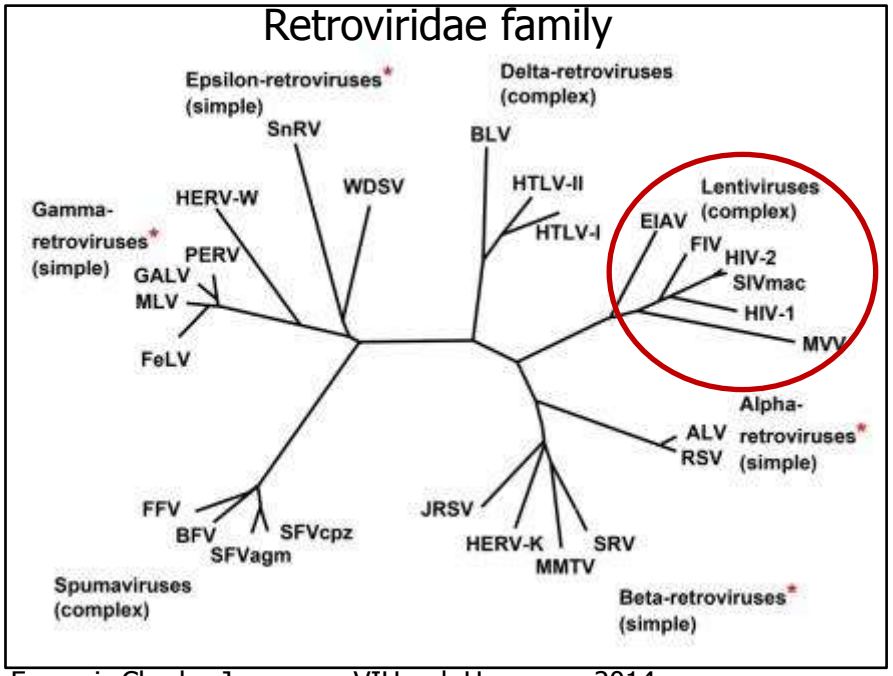
Marc Ruff

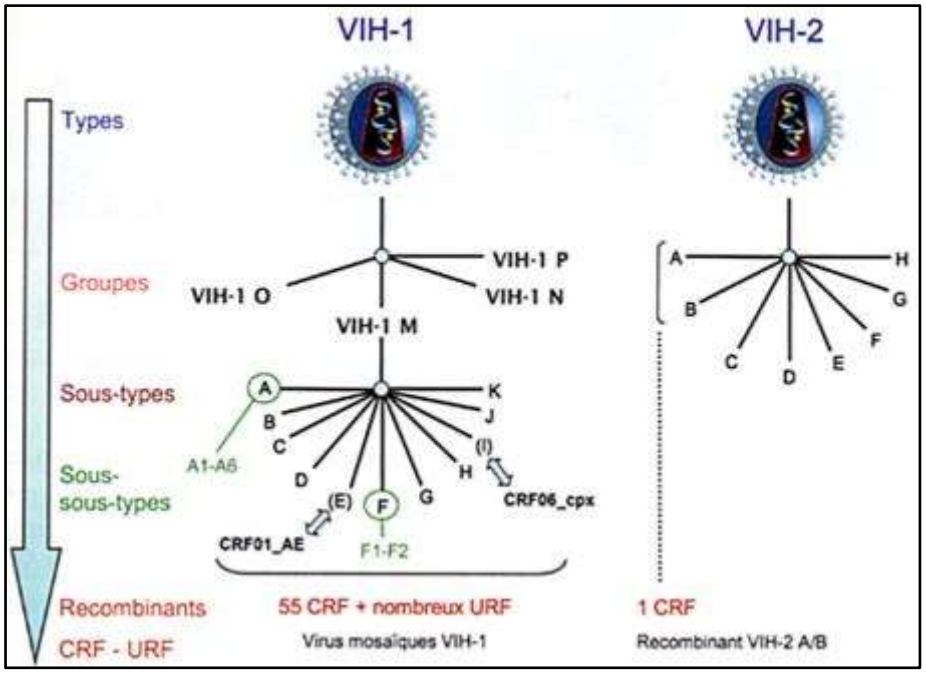
Integrated structural biology department
Chromatin stability and DNA mobility team

IGBMC, Illkirch, France

ruff@igbmc.fr

HIV phylogeny





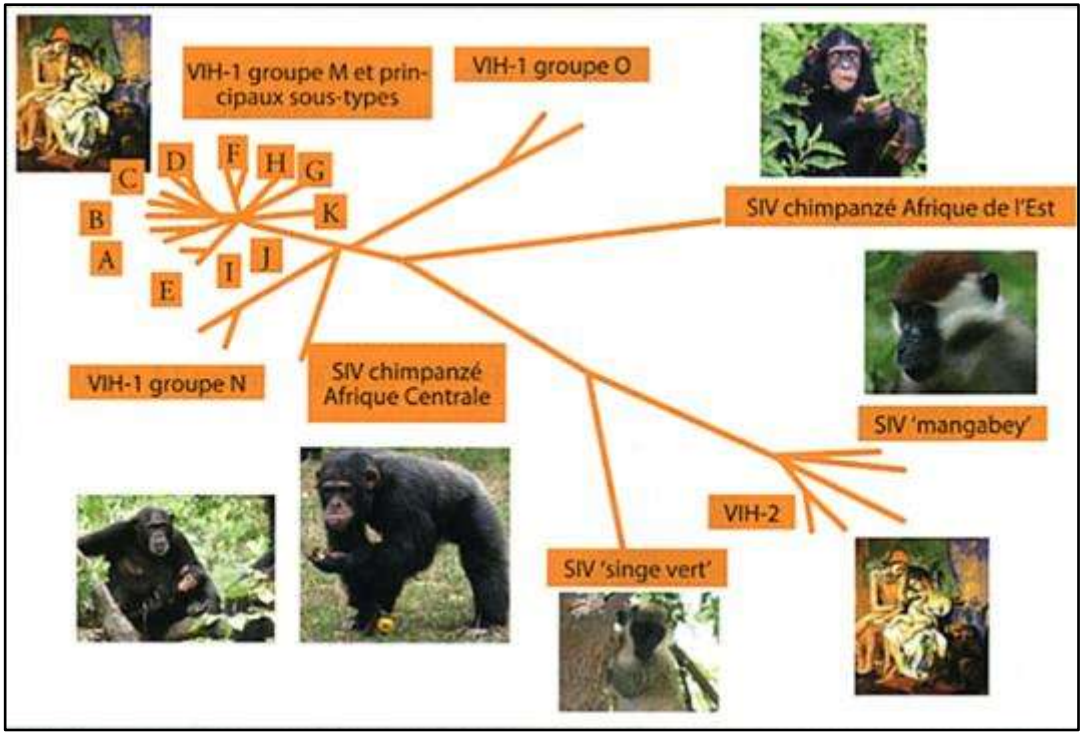
François Charles Javaugue, VIH, ed. Hermann, 2014

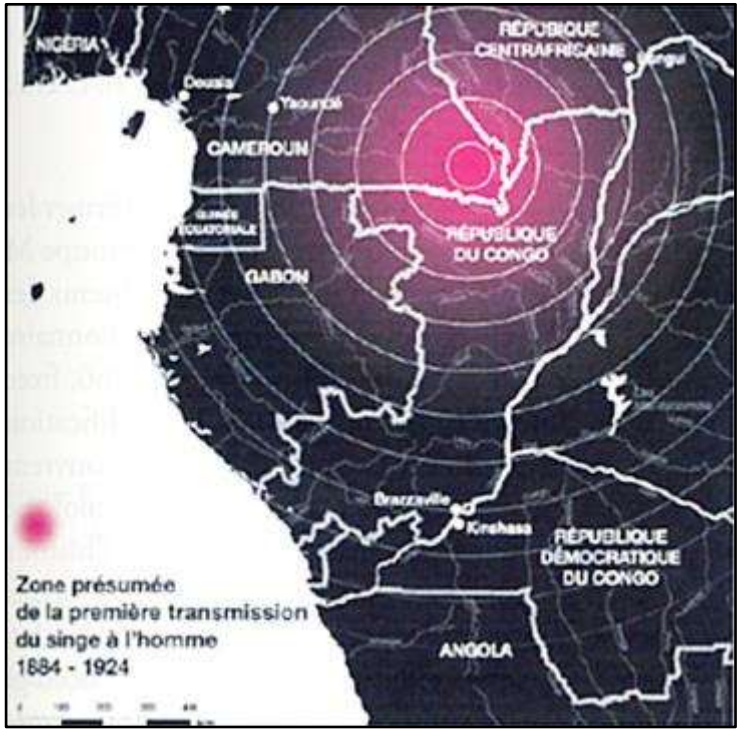
HIV-1 and 2 are lentiviruses, a genus of viruses of the Retroviridae family, characterized by a long incubation period.

There are many groups and subtypes of HIV-1 virus, the predominant form in Europe is the group M subtype B.

Enveloped virus (budding from the host cell enveloped by fragment of the cell membrane)

Origin of HIV pandemics



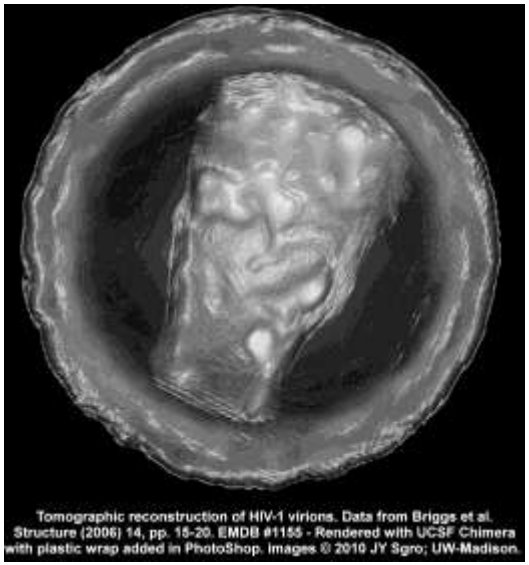


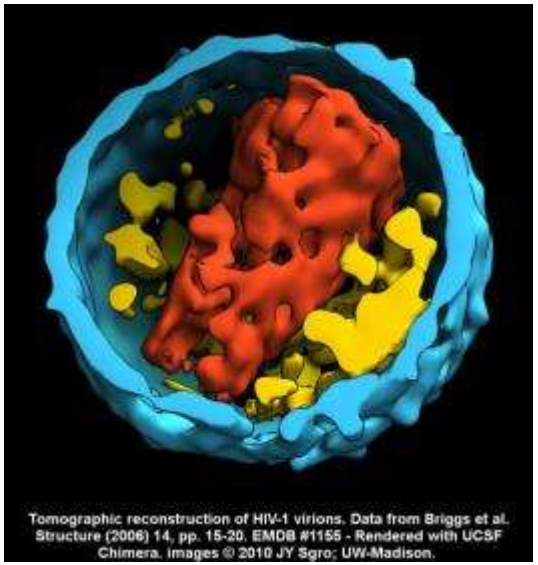
François Charles Javaugue, VIH, ed. Hermann, 2014

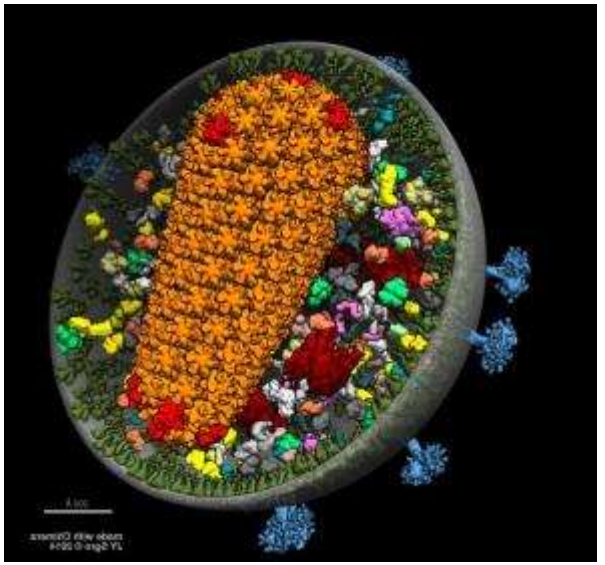
HIV emerged from SIV in late 19th / early 20th century through zoonosis between apes or monkeys and men:

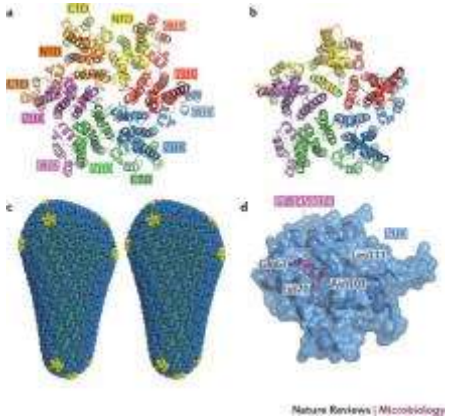
HIV-1 group M : chimpanzee (Central Africa : Cameroon, Equatorial Guinea, Gabon…)
HIV-1 group P : gorilla (Cameroon)
HIV-2 : sooty mangabey monkey (West Africa : Sierra Leone, Liberia, Ivory Coast…)

HIV-1 virion structure

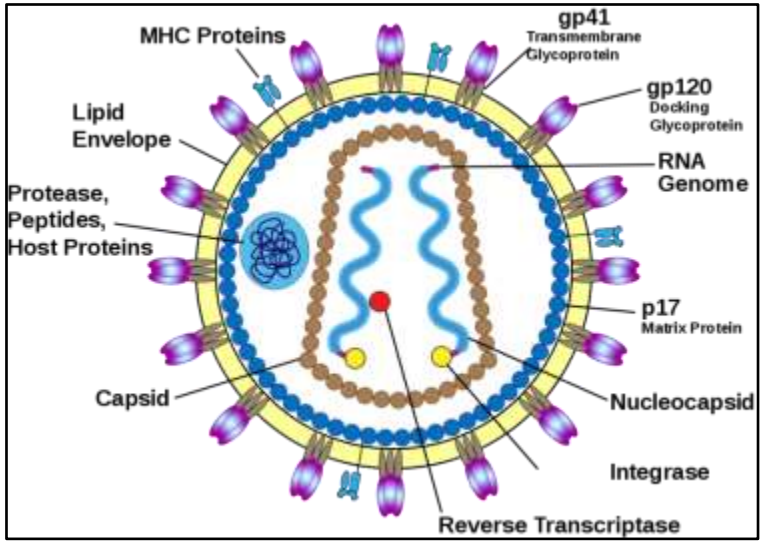


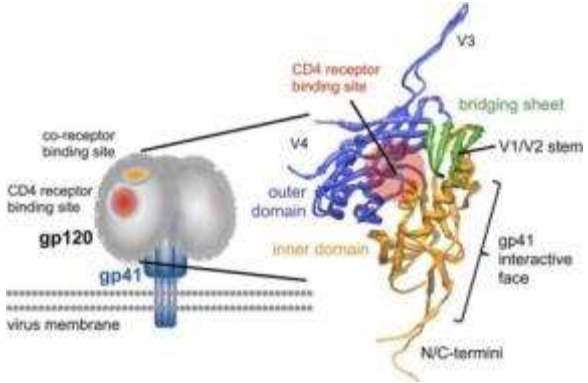






Alan Engelman and Peter Cherepanov, Nature Review in Microbiology, 2012, 10, 279-289





Miklos Guttman et al., J Virol. Aug 2012; 86(16): 8750–8764

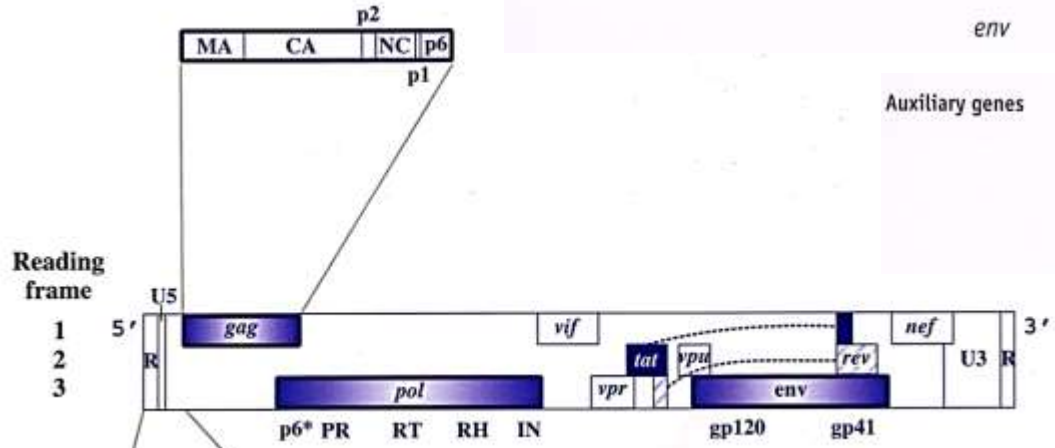
HIV-1 genome organization

Main genes

gag	*g*roup-specific *a*ntigen (encodes matrix, capsid, p2, nucleocapsid, p1 and p6)
pol	*pol*ymerase (encodes p6*, protease, reverse transcriptase, RNase H, integrase)
env	*env*elope

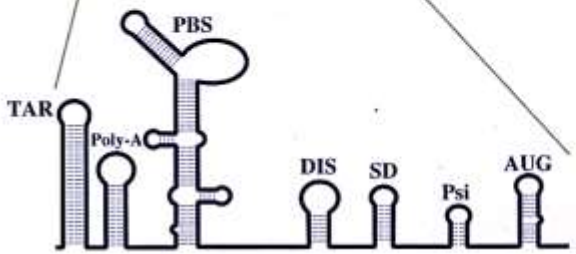
Auxiliary genes

nef	*n*egative regulatory *f*actor
rev	regulator of *e*xpression of *v*irion proteins
tat	*t*ransactivator of *t*ranscription
vif	*v*irion *i*nfectivity *f*actor
vpr	*v*iral protein *R*
vpu	*v*iral protein *U*



Non-coding sequences

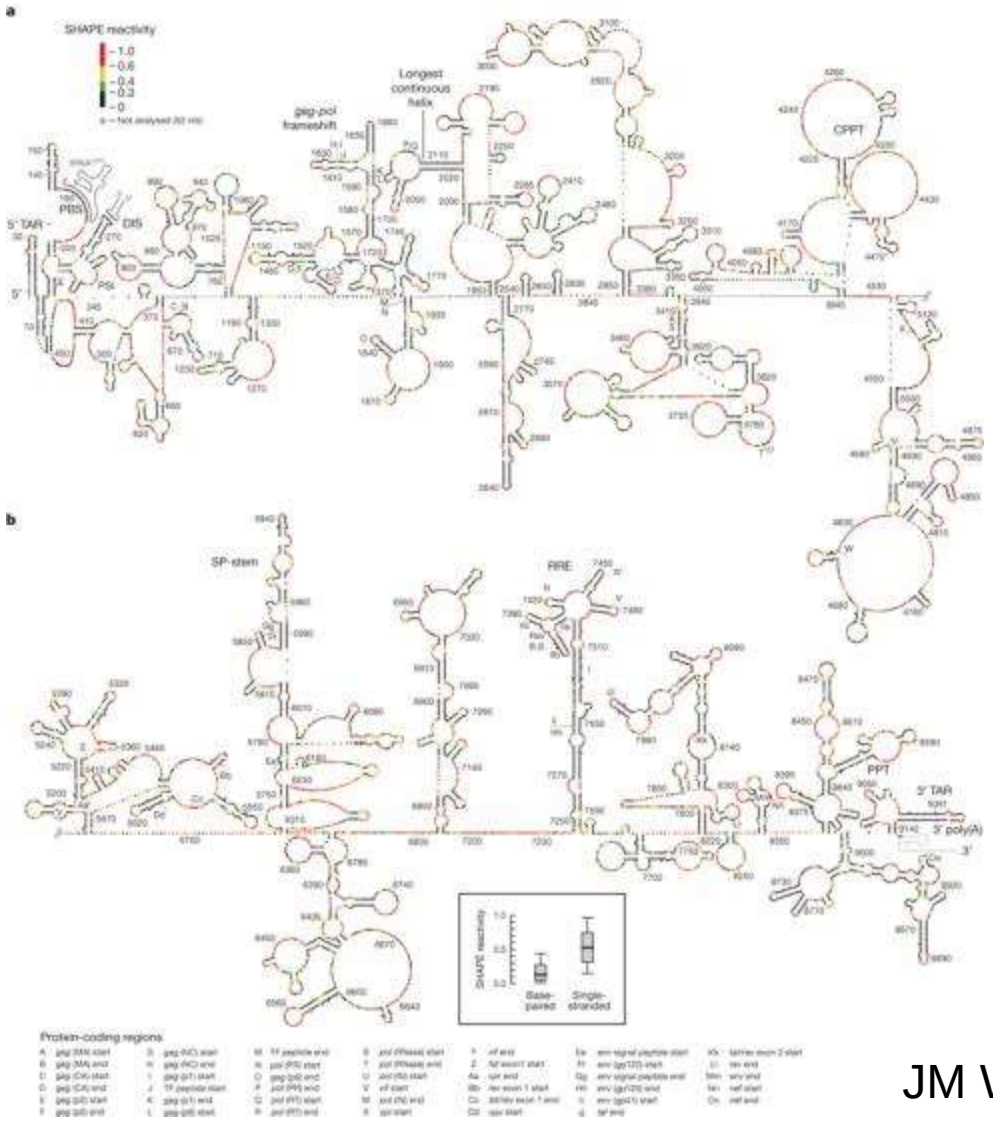
R	repeat sequence
U3	unique sequence at 3′ end of genome
U5	unique sequence at 5′ end of genome



Domains at the 5′ end of the genome

TAR	trans-acting response element
Poly-A	polyadenylation signal
PBS	primer-binding site
DIS	dimerization initiation site (involved in formation of kissing loop complex)
SD	splice donor site
Psi (ψ)	main part of the packaging signal
AUG	start codon of the *gag* gene

Structure of the HIV-1 NL4-3 genome

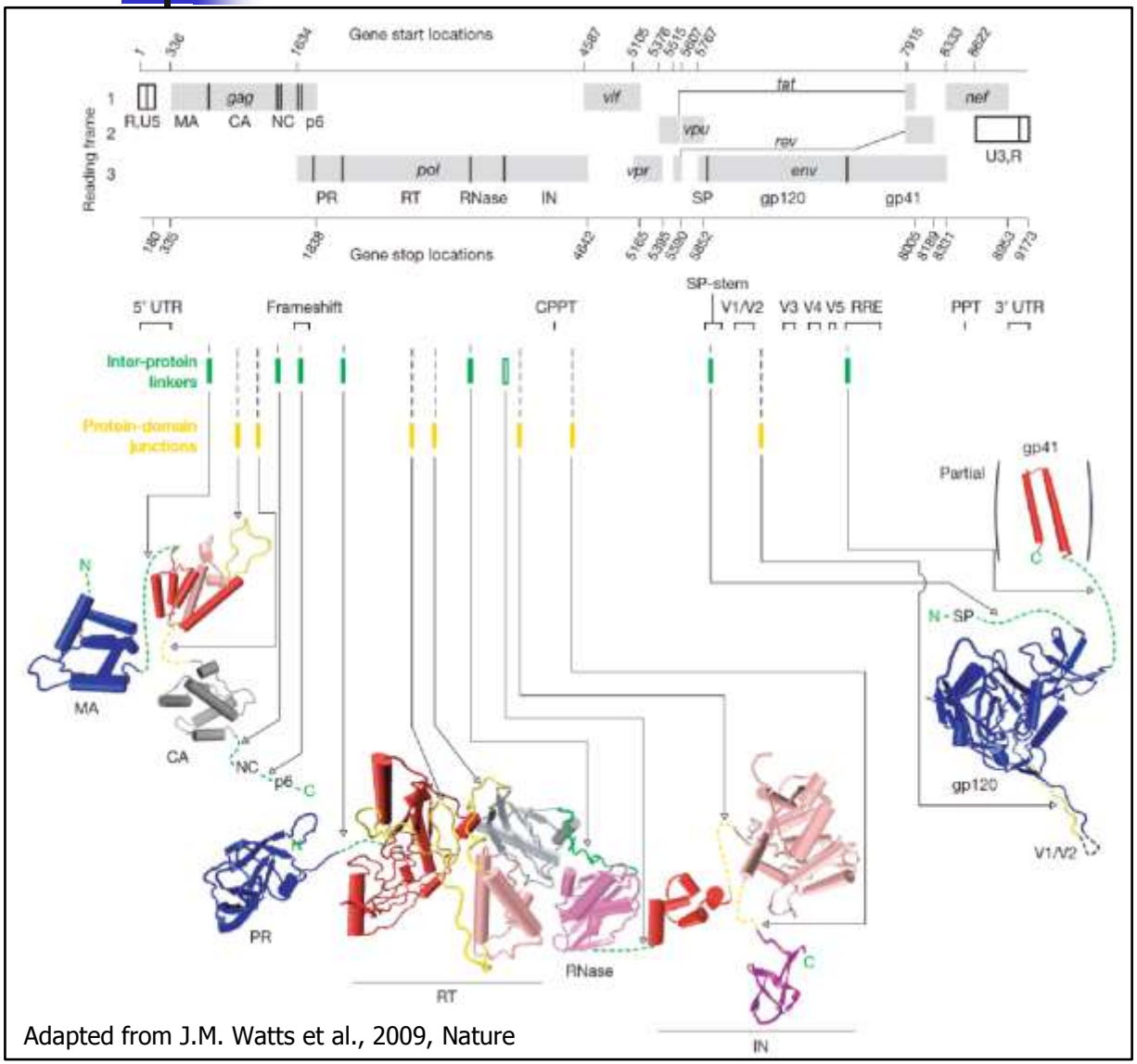


The 5' (**a**) and 3' (**b**) genome halves are shown. Nucleotides are coloured by their absolute SHAPE reactivities. Every nucleotide is shown explicitly as a sphere; base pairing is indicated by adjacent parallel orientation of the spheres.

Intermolecular base pairs involving the $tRNA^{Lys3}$ primer and the genomic dimer are shown in grey.

JM Watts *et al. Nature* **460**, 711-716 (2009)

HIV genome and viral protein structures



Adapted from J.M. Watts et al., 2009, Nature

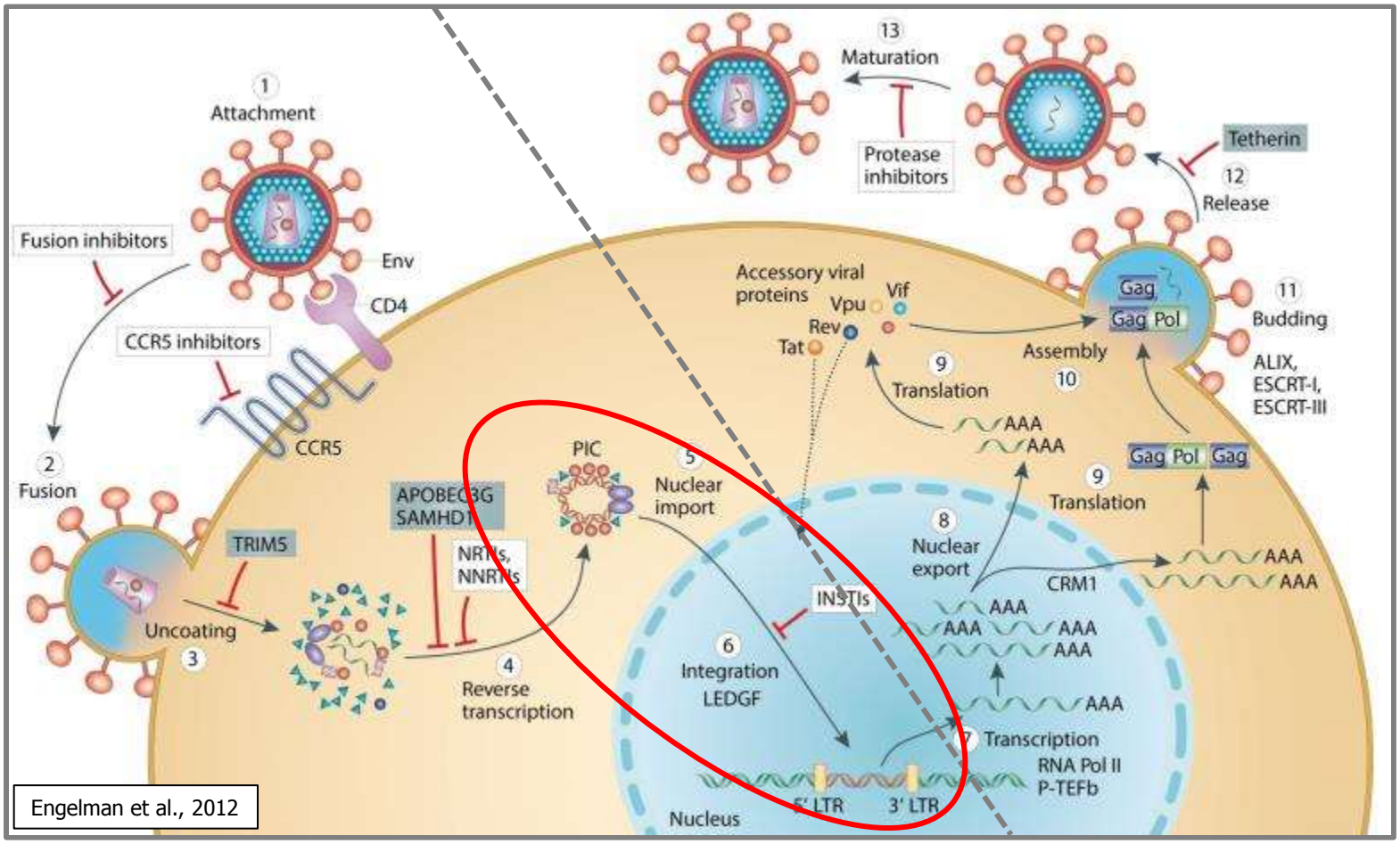
3 main genes coding for the viral polyproteins :
Gag, **Pol** and **Env**.
Gag => structural proteins
Pol => viral enzymes
Env => envelope proteins

2 regulatory genes:
Tat and **Rev** (activation of transcription and regulation of RNA splicing and export)

Accessory genes :
Vif, Vpr, Vpu, Nef
(regulation of synthesis and processing viral RNA and other functions)

Schematic diagram of HIV replication cycle

Early phase

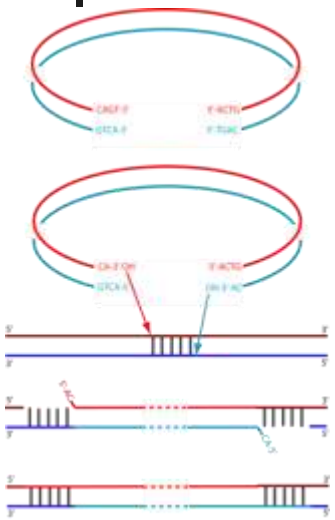


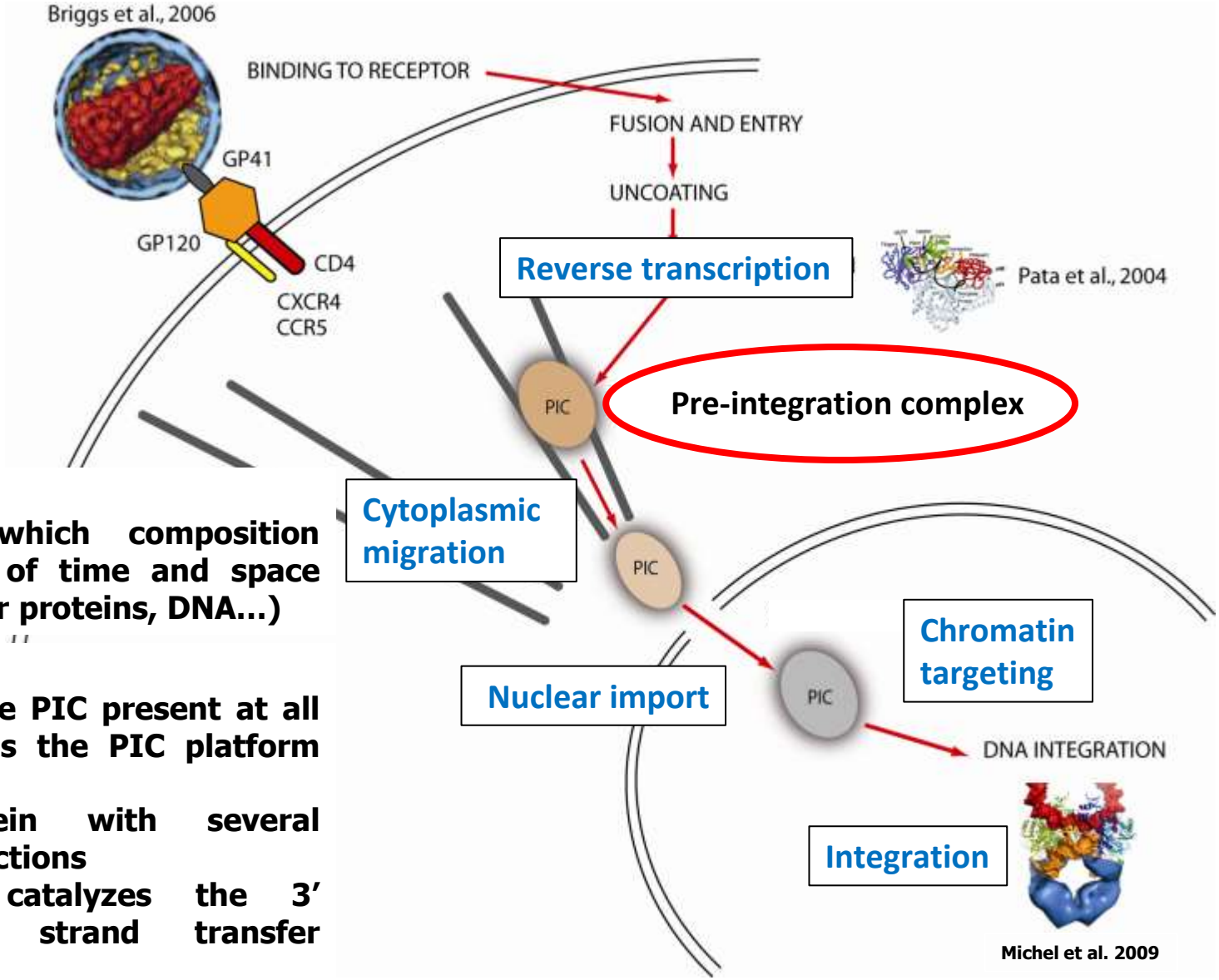
Engelman et al., 2012

Late phase

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

HIV-1 pre-integration complex





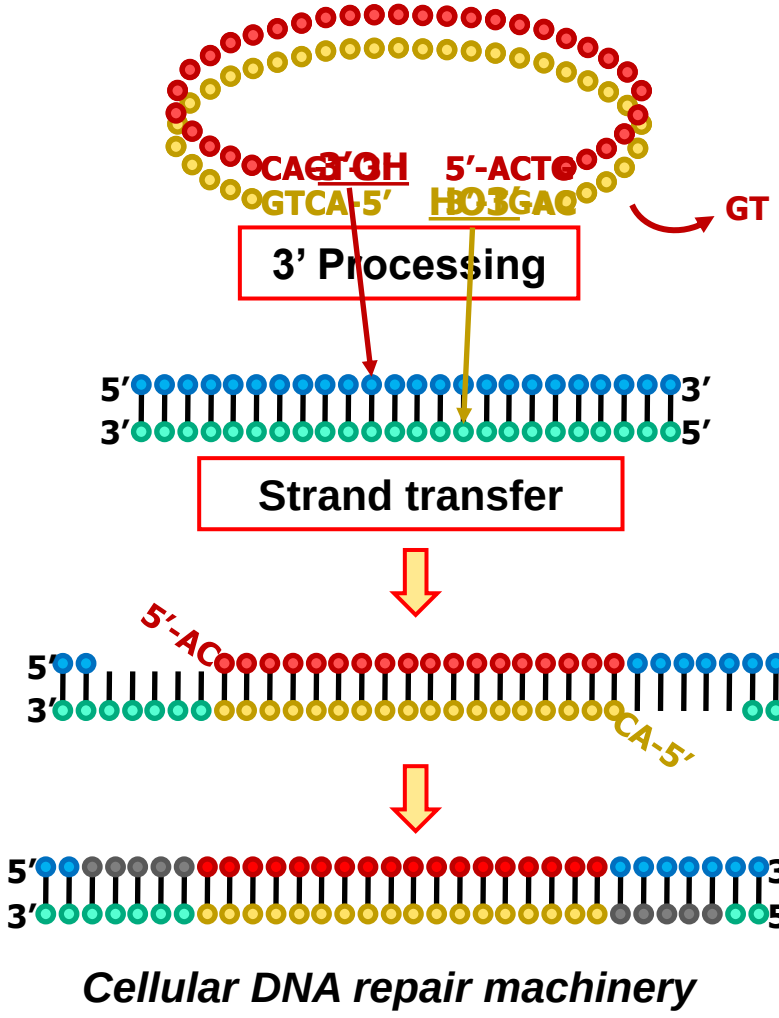
PIC:

Dynamic complex which composition changes in function of time and space (viral proteins, cellular proteins, DNA…)

Integrase:

- **Core protein of the PIC present at all steps: Integrase is the PIC platform protein**
- **Disordered protein with several structures and functions**
- **Enzyme which catalyzes the 3′ processing and strand transfer reaction**

Catalytic activities of HIV-1 integrase



Viral cDNA

Host cell target DNA

Integrated proviral DNA

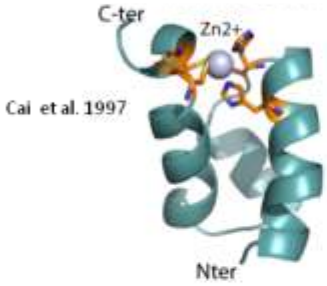
Structural domains of HIV-1 integrase

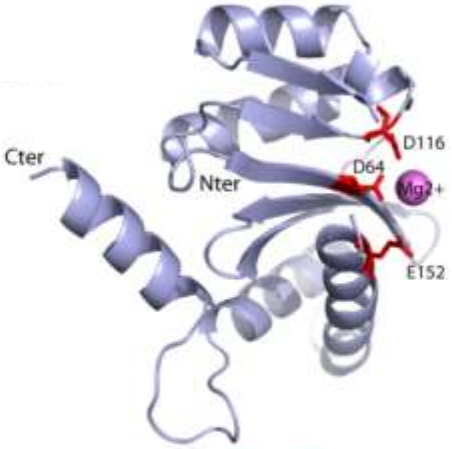
HIV integrases family

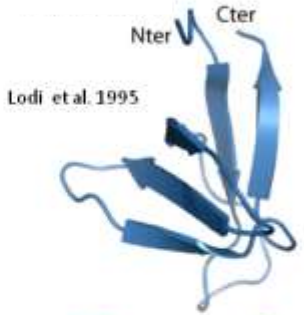
Ribonucleases H superfamily
Interaction with LEDGF

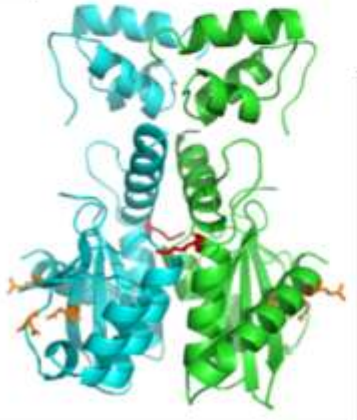
Retroviral integrases superfamily

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

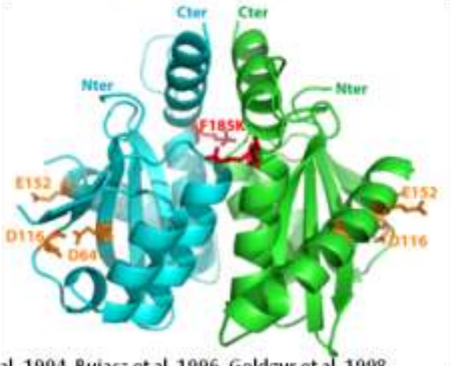




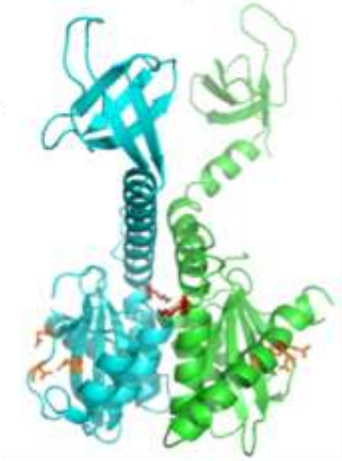




Wang et al. 2001

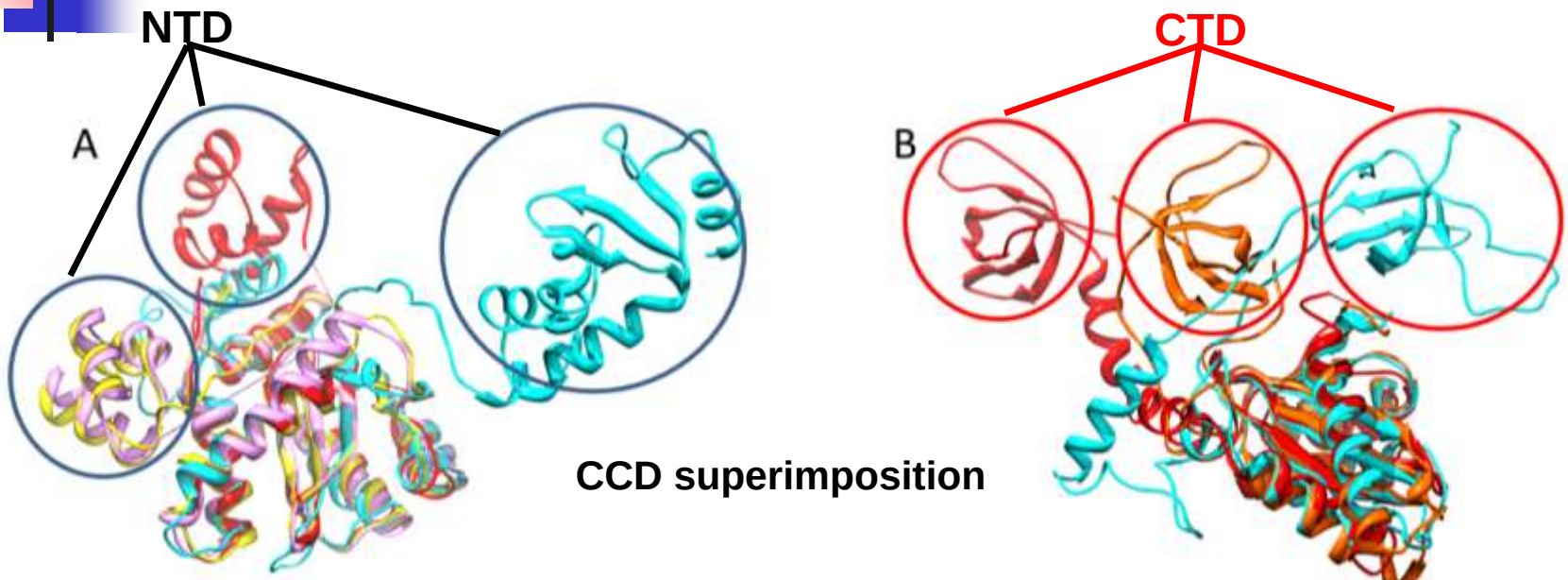


Dyda et al. 1994, Bujacz et al. 1996, Goldgur et al. 1998
Maignan et al. 1998, Greenwald et al. 1999, Chen et al. 2000



Chen et al. 2000

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)

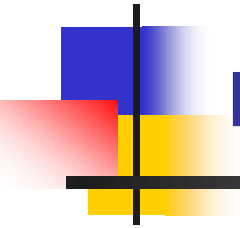
B. Superimposition of CCD+CTD structures :
RSV (Yang et al., 2000), HIV-1 (Chen et al., 2000)
PVF (Hare et al., 2010).

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

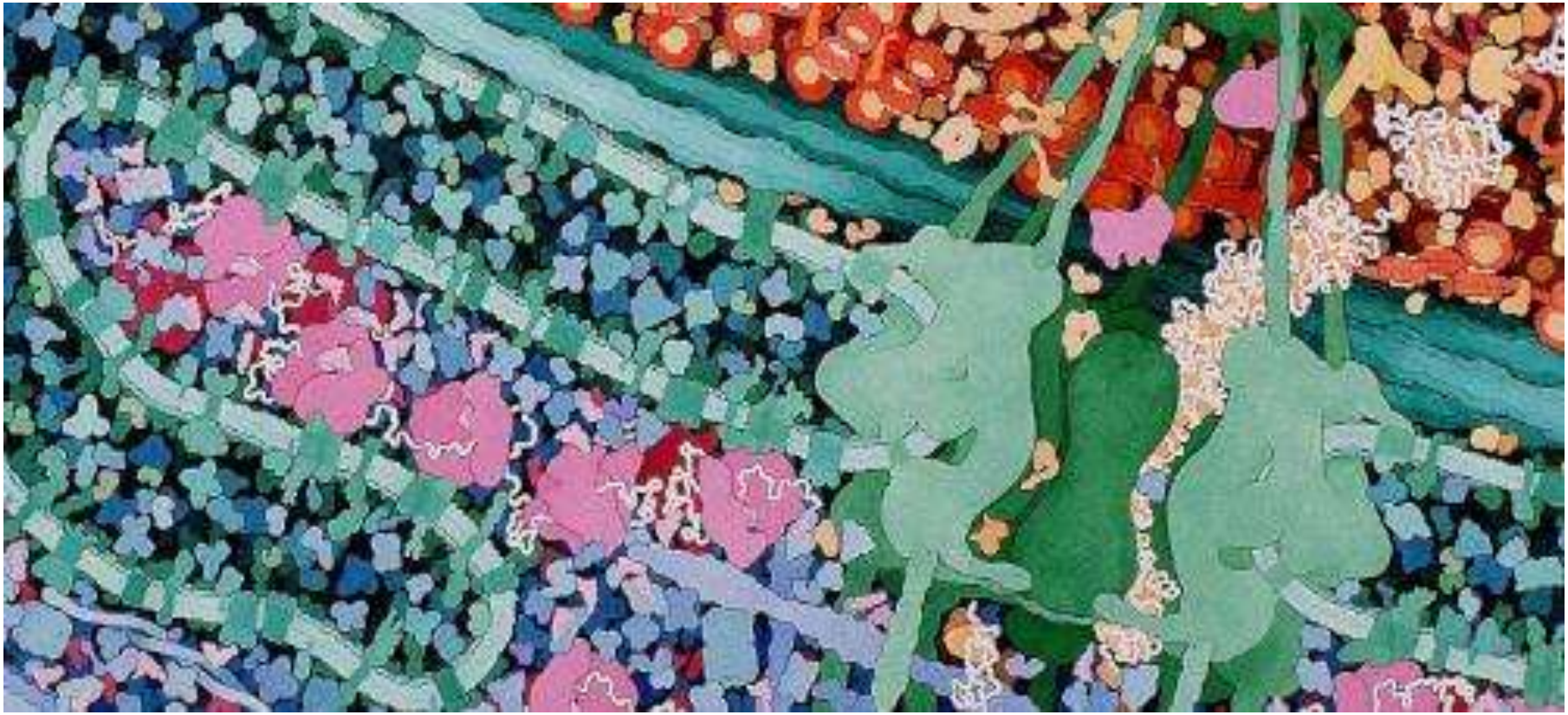
Isolate stable IN complexes (VBP1, TRN-SR2, LEDGF, INI1….)

Isolate protein with PTMs:
Production in eukaryotic cells
Specific modifications



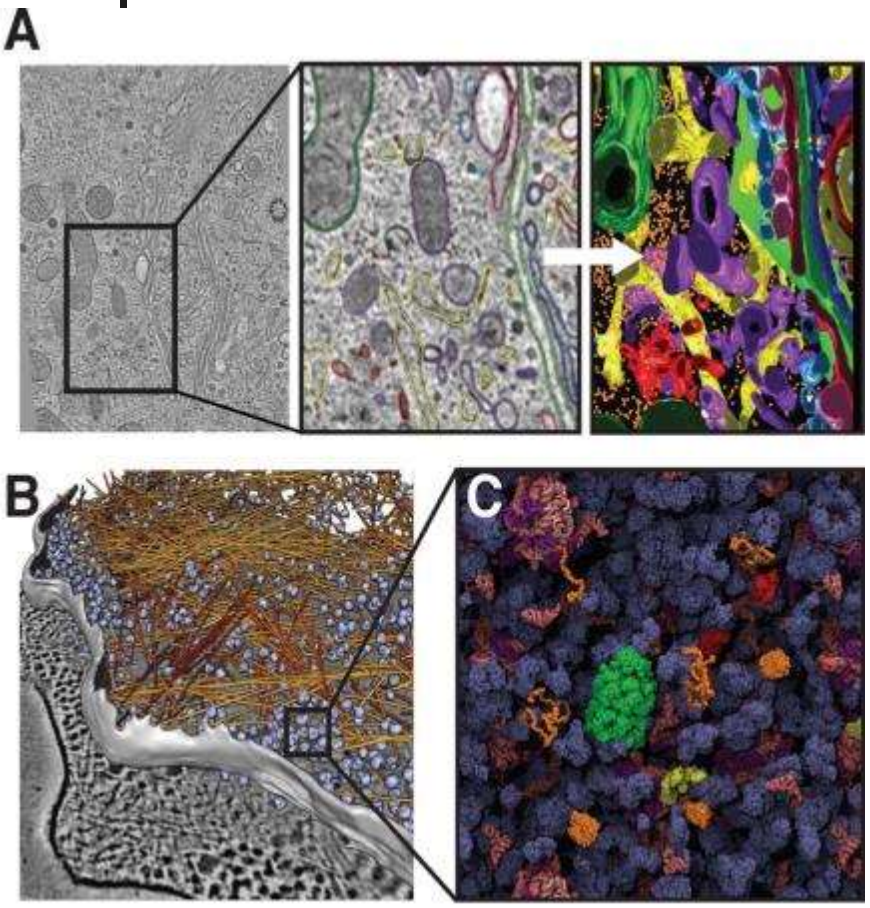
Protein and protein complexes purification

Physicochemistry of the intracellular medium



“Cellular crowding”

Intracellular complexity



(A) Left: Cryo-electron tomography slice of a mammalian cell. Middle: Close-up view of cellular structures colored according to their identities: Right: Three-dimensional surface representation of the same region. Yellow, endoplasmic reticulum; orange, free ribosomes; green, mitochondria; blue, dense core vesicles; red, clathrin-positive compartments and vesicles; purple, clathrin-negative compartments and vesicles.

(B) Tomography image of the interior of a *Dictyostelium* cell with actin filaments shown in orange and ribosomes in blue.

(C) Schematic representation of the *E. coli* cytosol. Ribosomes and tRNA are shown in pink, chaperones in green and red, disordered proteins in orange, and all other proteins in dark blue.

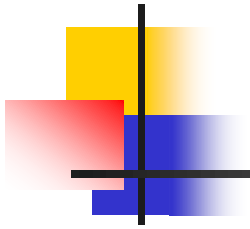
Francois-Xavier Theillet; Andres Binolfi; Tamara Frembgen-Kesner; Karan Hingorani; Mohona Sarkar; Ciara Kyne; Conggang Li; Peter B. Crowley; Lila Gierasch; Gary J. Pielak; Adrian H. Elcock; Anne Gershenson; Philipp Selenko; *Chem. Rev.* **2014,** 114, 6661-6714

Macromolecular crowding

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.

Obviously, 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, and the thermodynamic consequences of the unavailable volume are called **excluded volume effects**

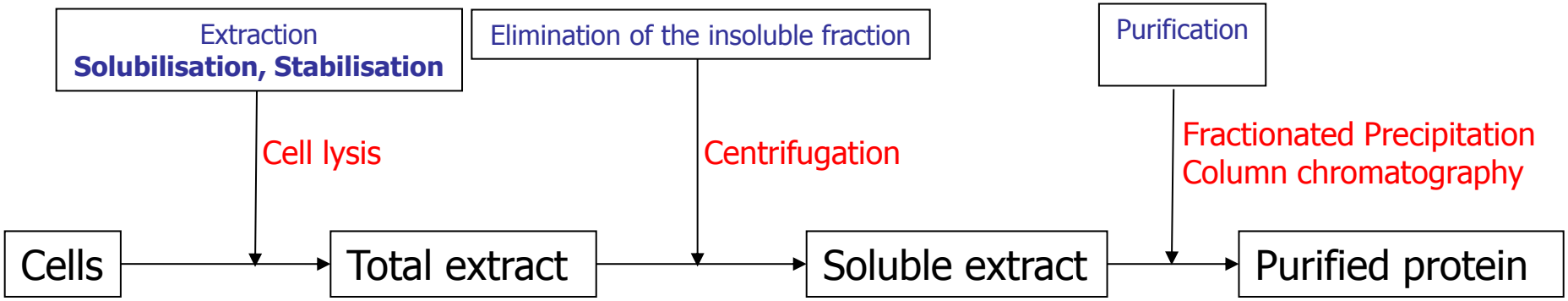
Irina M. Kuznetsova, Konstantin K. Turoverov and Vladimir N. Uversky, What Macromolecular Crowding Can Do to a Protein, Int. J. Mol. Sci. 2014, 15(12), 23090-23140



Solubility, Aggregation, Stability, Function

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

Purification



Measuring the amount of:
Total protein
Specific protein

Purity analysis

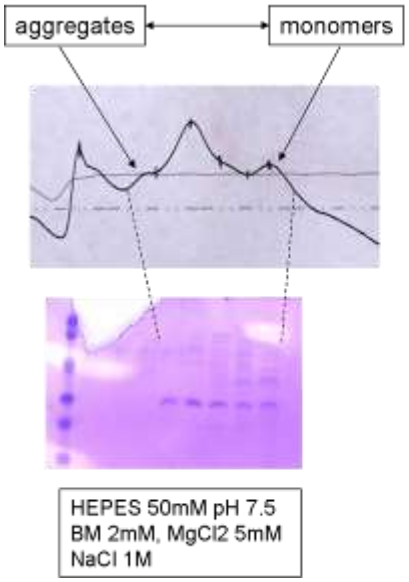
Buffer exchange
Concentration and dilution

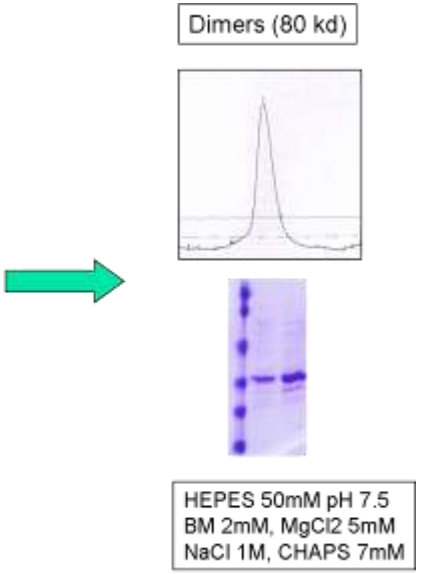
- Adapt the purification to the protein
- Adapt the protein to the purification: fusion proteins for affinity purification (HIS, GST, MBP, biotinilated peptide, STREP, FLAG, …)

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	+	++	+++	+++	++

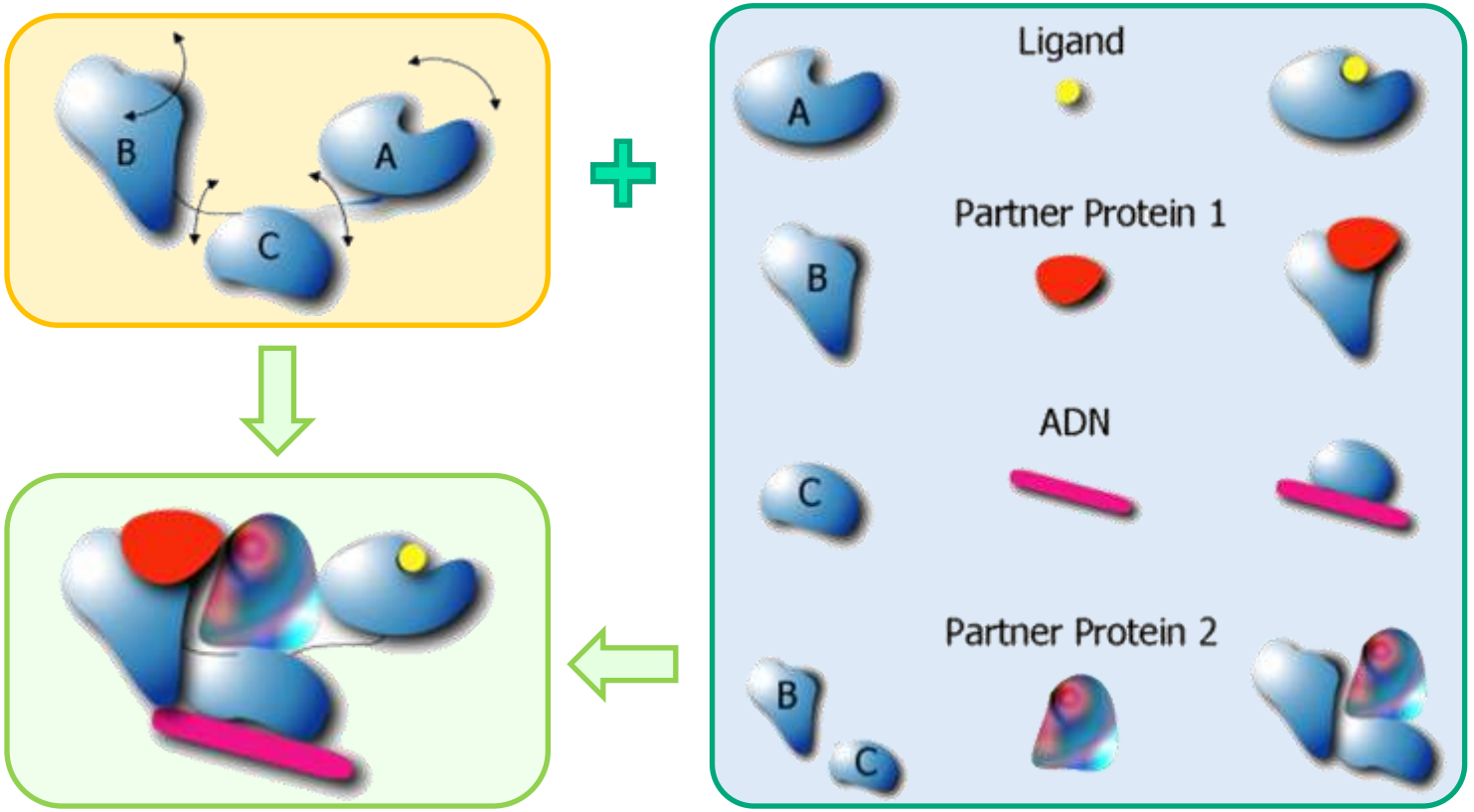




No crystals

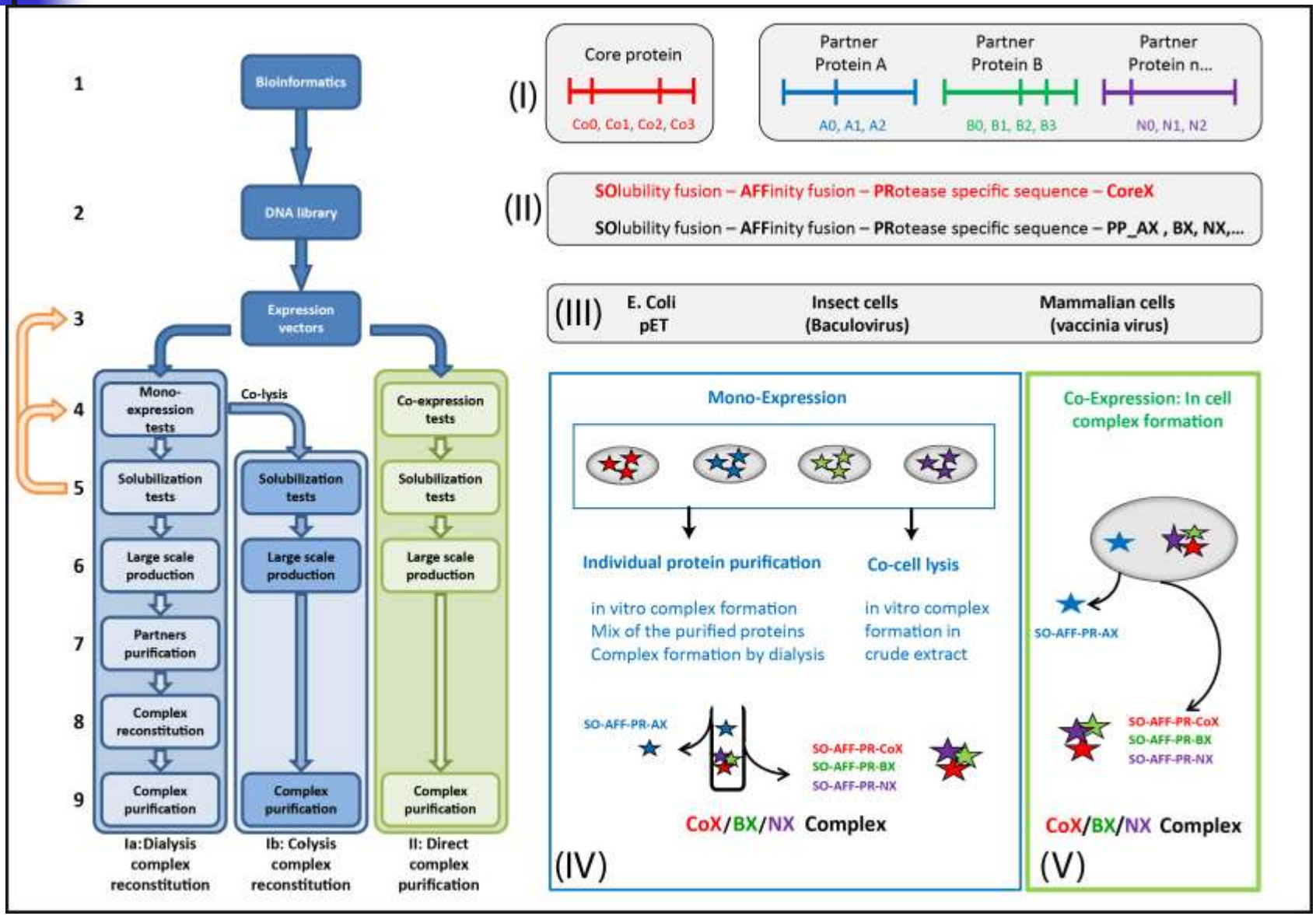
Strategy

- Partial non-structuration of the protein (disordered regions)
- High inter domain mobility



- Proteins domains, mutants, solubility fusions
- Stabilization by the interaction with partner proteins, ligands, DNA.

From unstable protein to stable complexes: Pipeline procedure for stable complex characterization and production



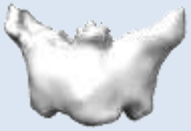
HIV Integrase structural and functional studies

DNAs

Viral cofactors

vDNA, RT, CA, NC, Ma, Vpr, Rev …

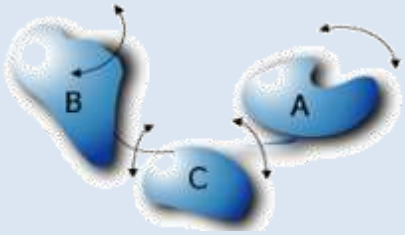
Cellular cofactors



LEDGF INI1

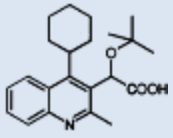
VBP1, TRN-SR, GCN5, P300…

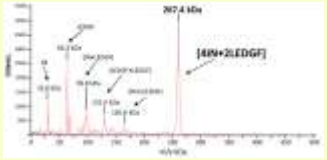
Post-translational modifications



HIV integrase

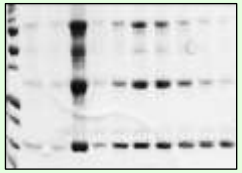
Small molecules



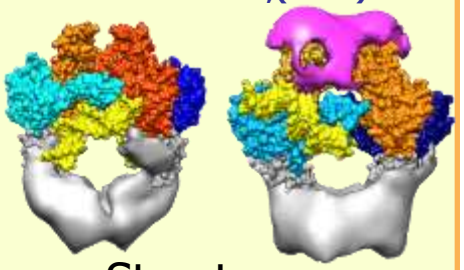


Mass spectrometry

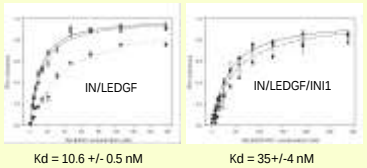
Stable complexes



Maillot B. et al,(2013)
Michel F. et al,(2009)



Structures (Xray, CryoEM)



Functional validation *in vitro*

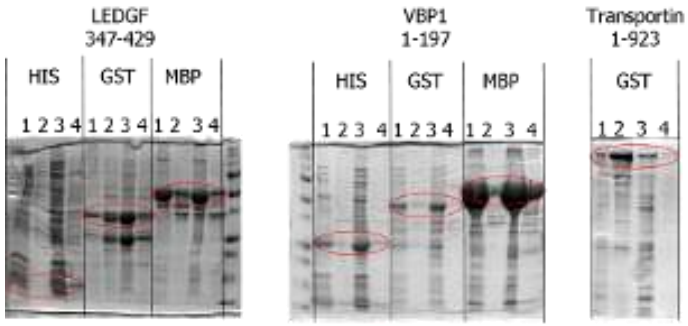
Functional validation *in vivo*
Complexes in the cellular context

Cloning, expression and solubility tests (E. Coli)

Protein	Fragment	Tag N_{term}	Protease site	Antibiotic	Expression	S_HS	S_LS	Notes
LEDGF	1-530	No	No	Chlo	++	++	Nd	
		No	Thrombin	Amp	++	+++		
		6xHis		Amp	+++	+++		
		6xHis		Zeocin	+++	+++		
		GST		Amp	+	+++		
		MBP		Amp	+	+++		
		6xHis-Nus		Amp	+	+++		
		Flag		Amp	Nd			
		No	Tev	Amp	Nd			
		6xHis		Amp	+++	++	++	
		GST		Amp	+	++	++	
		No	Protease 3C	Amp	Nd			
		6xHis		Amp	+++	++	++	
		GST		Amp	+	++	++	
		6xHis		Spec	+++	++	++	Bicistronic (IN)
	347-429	No	No	Chlo				
		6xHis	Protease 3C	Amp	+++	++	++	
		GST		Amp	+++	+++	+++	
		MBP		Amp	+++	+++	++	
		6xHis		Spec				Bicistronic (IN)
	347-442	No	No	Chlo	Nd			
		6xHis	Tev	Amp	Nd			
		GST		Amp	Nd			
		6xHis-MBP		Amp	Nd			
		6xHis	Protease 3C	Amp	Nd			
		GST		Amp	Nd			
		6xHis-MBP		Amp	Nd			
	347-471	No	No	Chlo	Nd			
		6xHis	Tev	Amp	Nd			
		GST		Amp	Nd			
		6xHis-MBP		Amp	Nd			
		6xHis	Protease 3C	Amp	Nd			
		GST		Amp	Nd			
		6xHis-MBP		Amp	Nd			

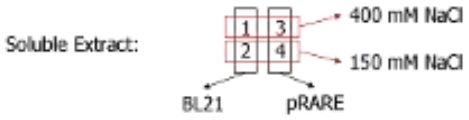
Protein	Fragment	Tag N_{term}	Protease site	Antibiotic	Expression	Solubility	Notes
Integrase	1-288	No	No	Chlo	+	+	Only in co-expression
		No	Thrombin	Amp	+	+	
		6xHis		Amp	++++	+++	pET15b
		6xHis		Amp	+	++	
		6xHis		Zeocin	++	++	
		6xHis		Kana	++	++	
		GST		Amp	++	+++	
		MBP		Amp	+++	+++	Affinity Pb
		6xHis-Nus		Amp	++	+++	
		Flag		Amp	Nd		
		No	Tev	Amp	Nd		
		6xHis		Amp	Nd		
		GST		Amp	-		Only GST exp
		No	Protease 3C	Amp	Nd		
		6xHis		Amp	++	+++	
		GST		Amp	+++	+++	
		6xHis		Chlo	+	+	Bicistronic (LEDGF)

Protein	Fragment	Tag N_{term}	Protease site	Antibiotic	Expression	Solubility HS	Solubility LS	Notes
VBP1	1-197	6xHis	P3C	Amp	++	+++	Nd	
		GST			+++	+++		
		MBP			+++	+++		
		6xHis-MBP			+++	+++		
		6xHis	Tev		+++	+++		
		GST			+++	+++		
		6xHis-MBP			+++	+++		
		6xHis	P3C	spec	Nd	Nd		Bi-cistronic
		0		Chlo	Nd	Nd		
SNF5	1-385	6xHis	Tev	Amp	-	-		
		GST			+++	+++		
		6xHis-MBP			+++	+++		
		6xHis	P3C		-	-		
		GST			+++	+++		
		6xHis-MBP			+++	+++		
		0		Chlo	Nd	Nd		
Transportin	1-923	GST	Thrombin	Amp	+++	++	++	pGEX



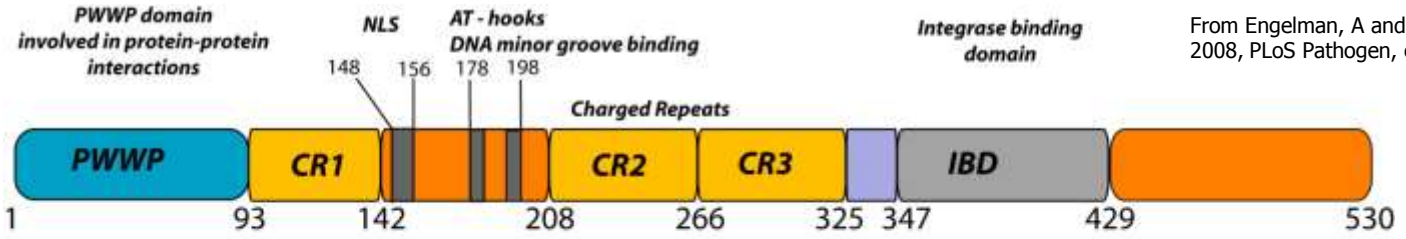
	37°		25°		18°	
	ET	EB	ET	EB	ET	EB
LB	+	-	+	+	+	+
LB/sucrose	+	-	+	+	++	++

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



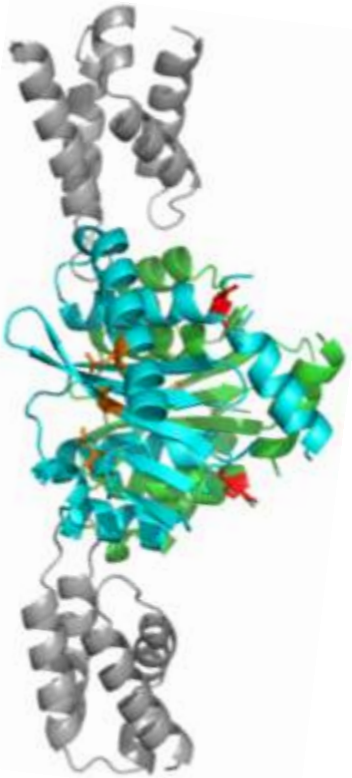
HIV-1 IN / LEDGF complex

HIV-1 IN interact with LEDGF



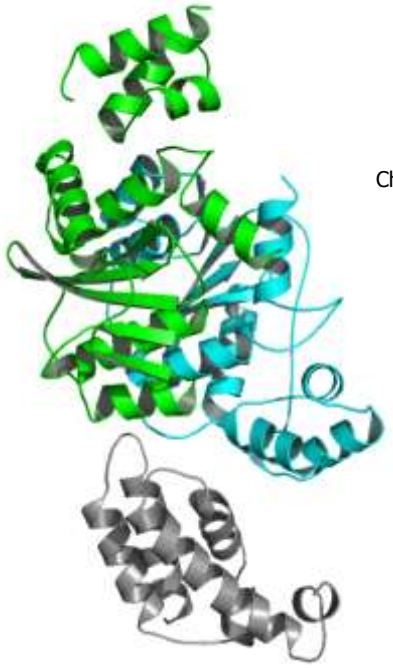
From Engelman, A and Cherepanov, P 2008, PLoS Pathogen, e1000046

IN catalytic core

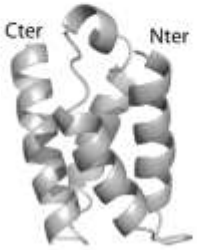


Cherepanov et al. 2005

IBD LEDGF



Hare et al. 2009



Cherepanov et al. 2005

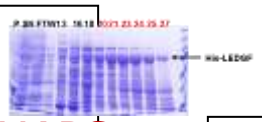
LEDGF/P75

- 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/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

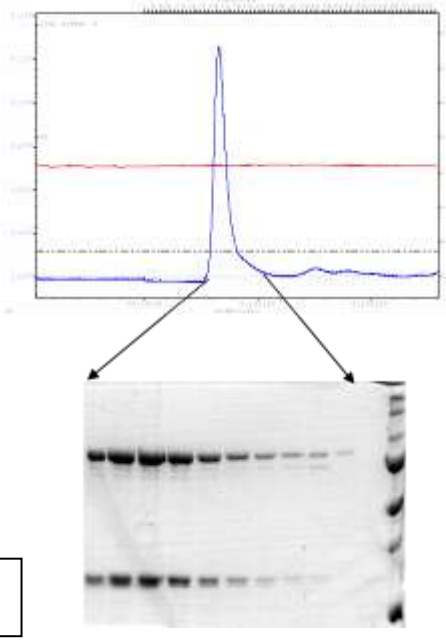
Complex formation by dialysis

GST Affinity

P3C cut

GST Affinity

Gel filtration G200



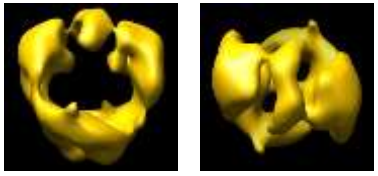
Functional and biochemical characterization

HTS Screening of inhibitors

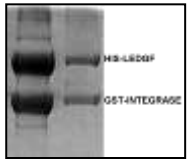
Crystallization Assays

SAXS, SANS studies

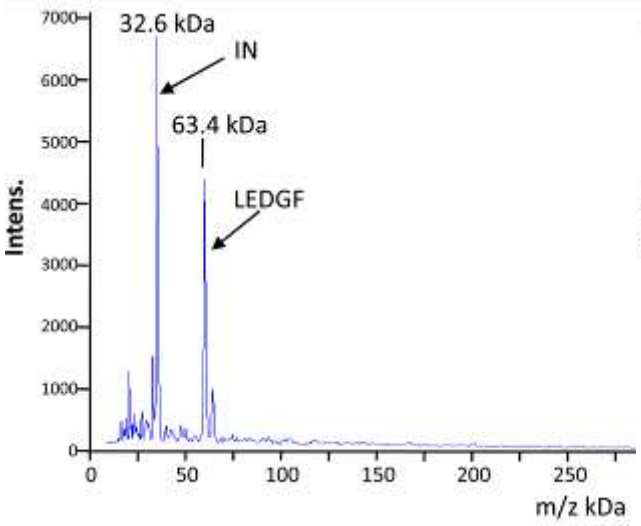
EM studies



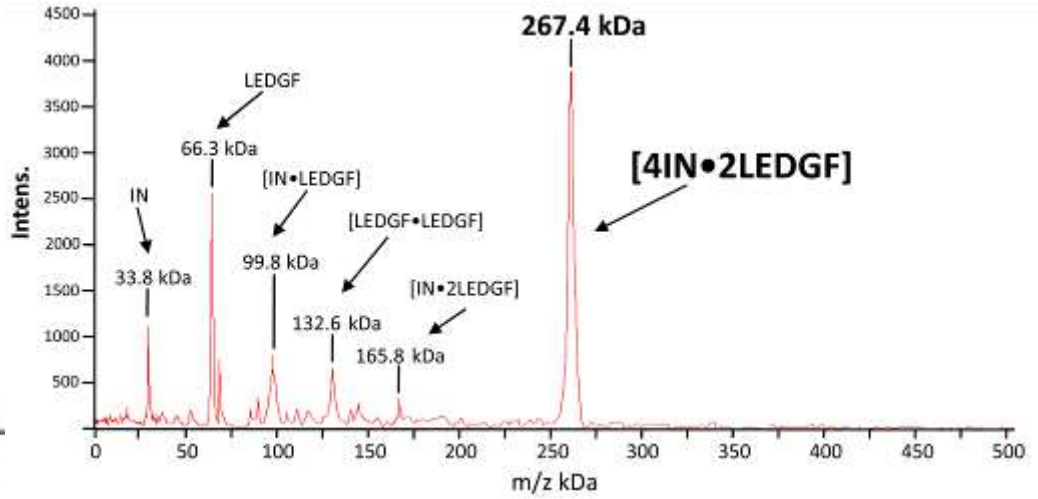
Yield: 5 mg of complex (INT 5L, LEDGF 6L)



IN/LEDGF : characterization by High Mass MALDI-TOFF



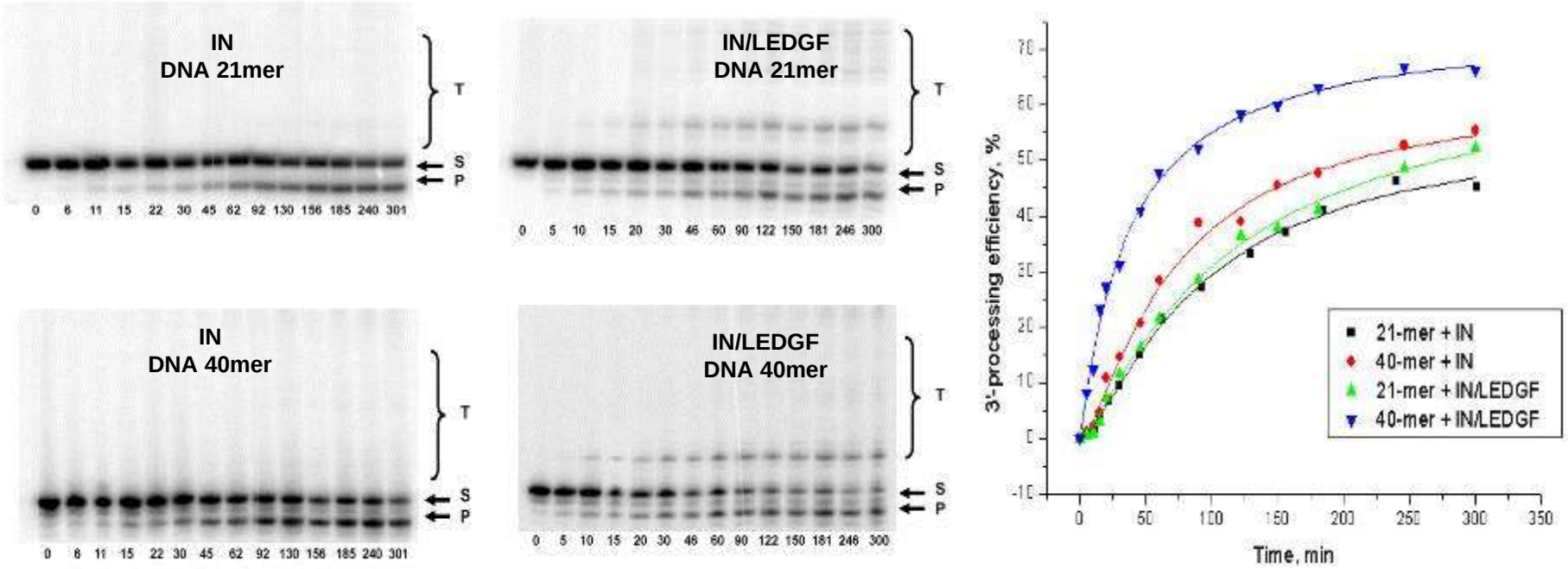
Absence of cross linking agents



Presence of cross linking agents

The cross-linking reactions were realised using a solution containing different cross-linkers specific for amino and sulfhydryl groups. The cross linking reactions were performed using a reagent composed of iodoacetec acid N-hydoxysuccinimide ester, Octaneodic acid di-N-hydroxysuccinimide ester and ethylene glycol bis-succinimidylsuccinate. (K200 MALDI MS analysis Kit, CovalX AG, Zürich, Switzerland).

IN/LEDGF : Functional characterization

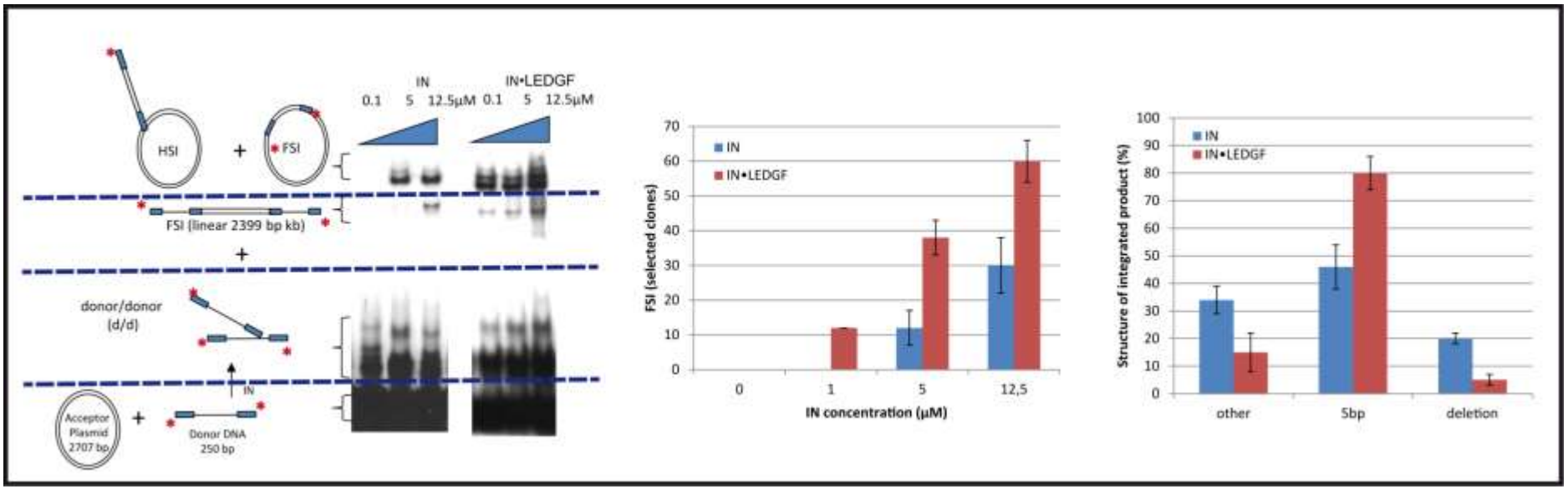


DNA: 21 or 40-mer duplex that mimic the HIV-1 U5 viral DNA end

5′-**GACTACGGTTCAAGTCAGC**GTGTGGAAAATCTCTAGCAGT-3′
3′-**CTGATGCCAAGTTCAGTCG**CACACCTTTTAGAGATCGTCA-5′

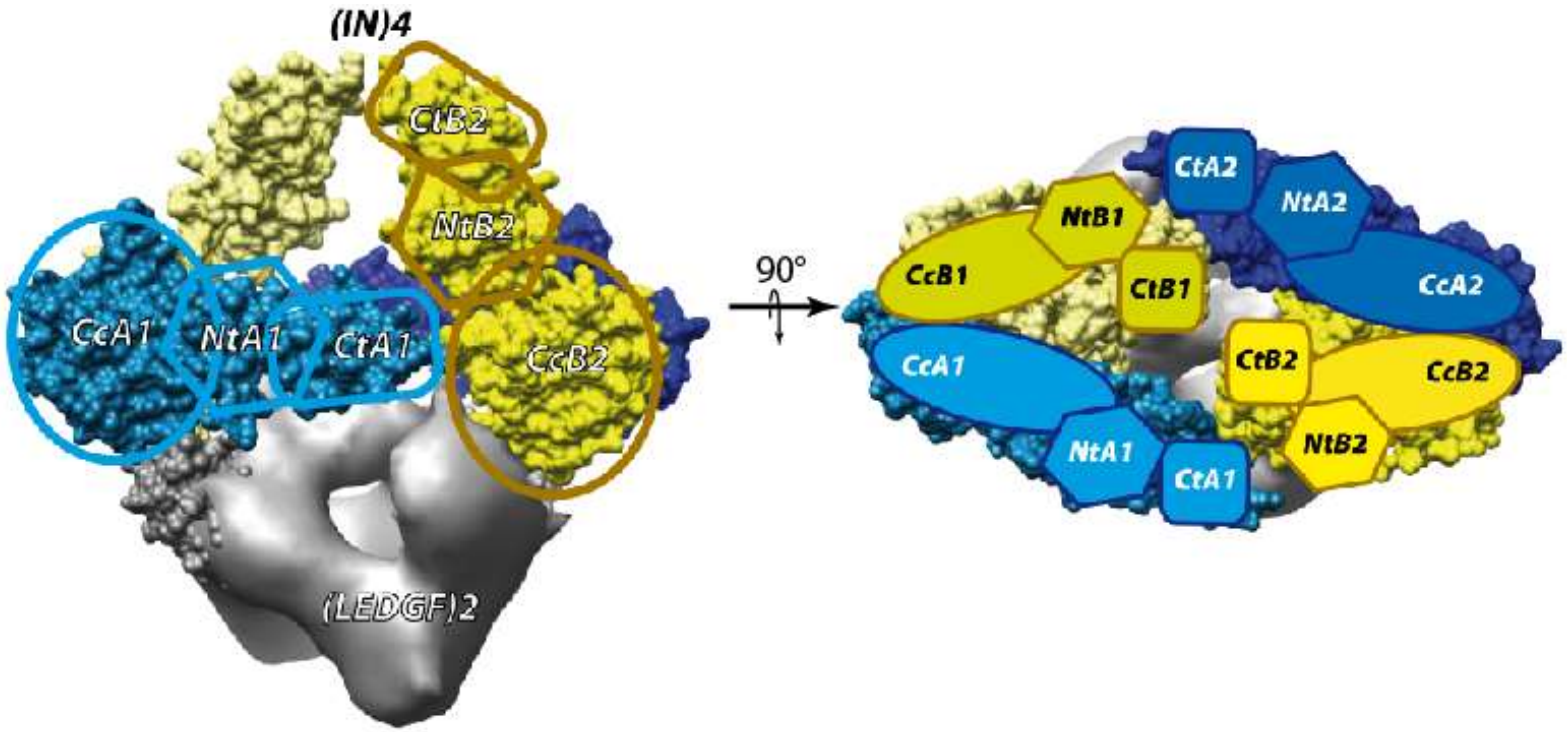
LEGDF stimulates the 3′ processing activity of IN with a 40-mer DNA.
In the presence of LEGDF, the strand transfer efficiency is strongly enhanced for both the 21- and the 40-mer DNA.

IN/LEDGF : Functional characterization



The global integration efficiency is higher for the IN/LEDGF complex than for isolated IN molecules. Specific cloning and quantification of the circular FSI products attested that the IN/LEDGF complex catalyzes more concerted integration events than isolated IN molecules. The integration reaction catalyzed by the IN/LEDGF complex is closer to the expected physiological reaction than IN alone (5bp staggered cuts of the target DNA).

IN/LEDGF EM Structure: Domain organization

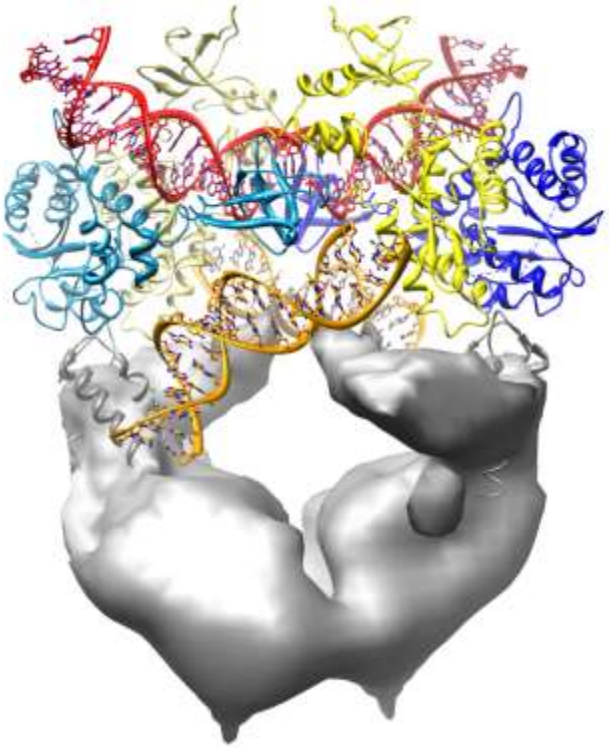


IN-LEDGF complex contains 4 IN molecules (A1, A2, B1 and B2) organized in two IN dimers

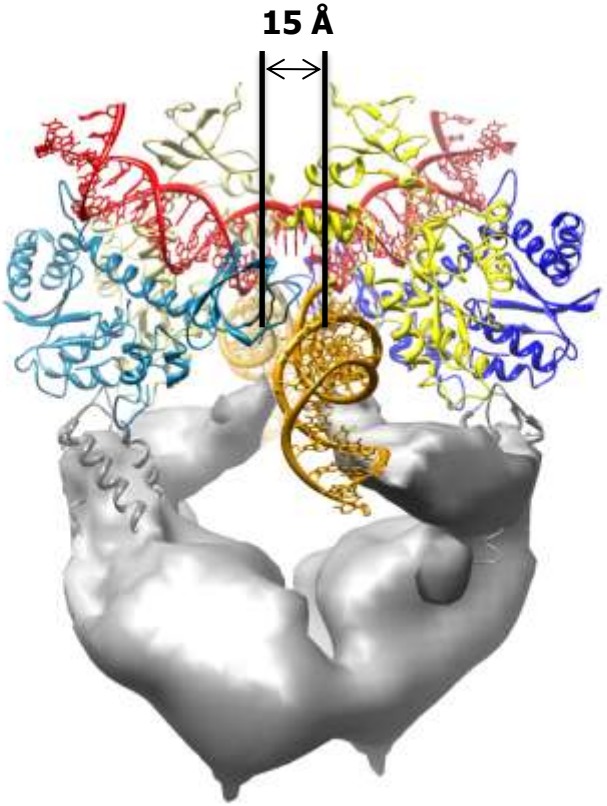
Each IN monomer within the IN dimer has a different conformation → the IN dimer is asymmetric

Each IN molecule has a distinct function within the dimer.

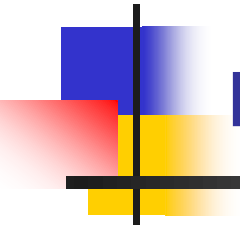
IN/LEDGF/DNA EM structure



3′ Processing

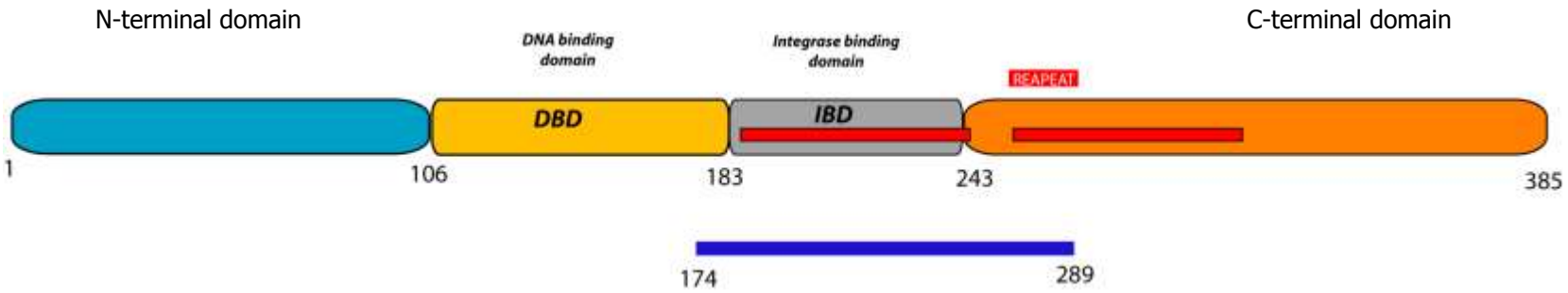


Integration



HIV-1 IN / LEDGF / INI1 complex

IN interact with INI1/SNF5

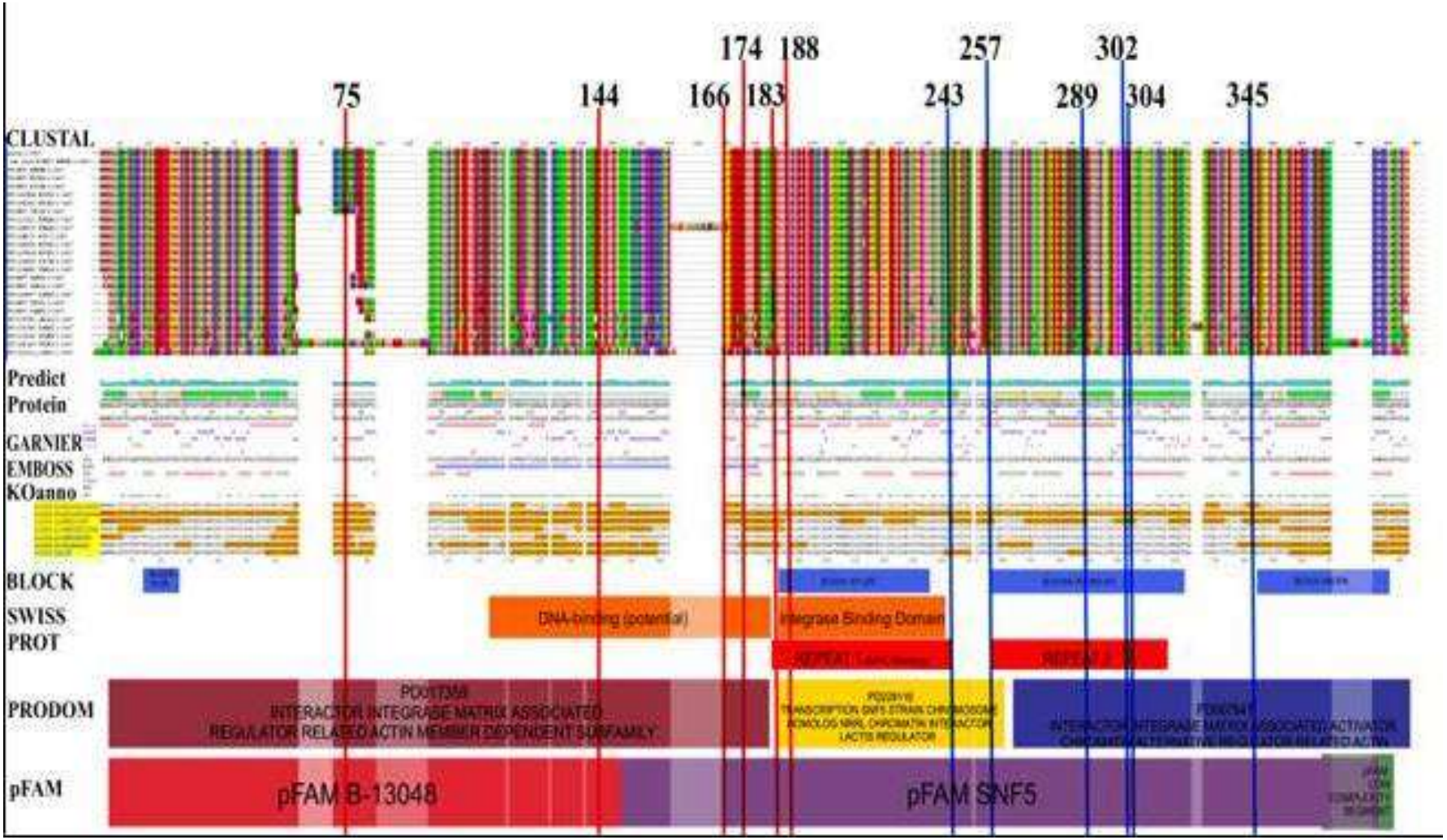


SNF5/Ini1, a subunit of the SWI/SNF chromatin remodeling complex, is the first cofactor identified to interact with IN.

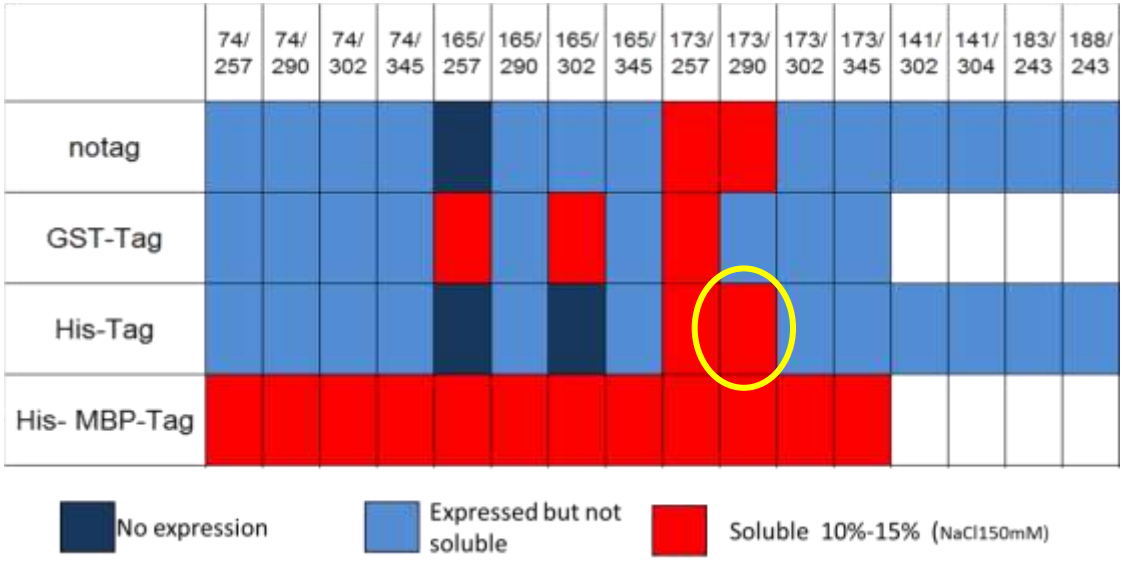
SNF5/Ini1 is one of the core subunits of the ATP-dependent chromatin remodeling complex SWI/SNF that regulates expression of numerous eukaryotic genes by altering DNA/histone interactions

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

INI1/SNF5: bioinformatic analysis

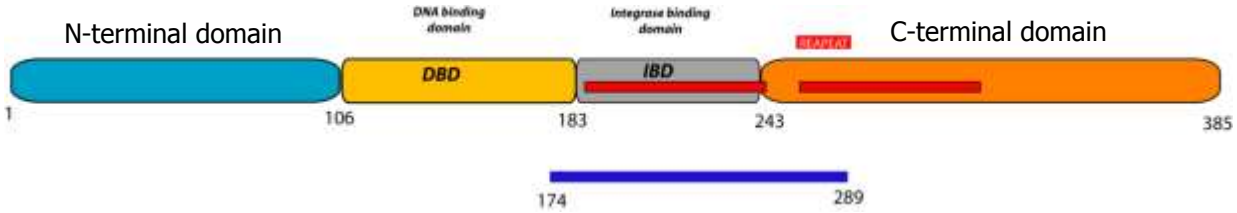


INI1: Cloning, expression and solubility tests (E. Coli)



	NaCl				
pH7,5	150mM	500mM	1M	2M	2,5M
0	-	-	-	-	-
CHAPS 7mM	-	-	-	-	-
CHAPS 10mM	-	-	-	-	+
CHAPS 20mM		-	-	+	++
Z[3-12] 4mM		+++			

IN/LEDGF/INI1 : complex formation and purification



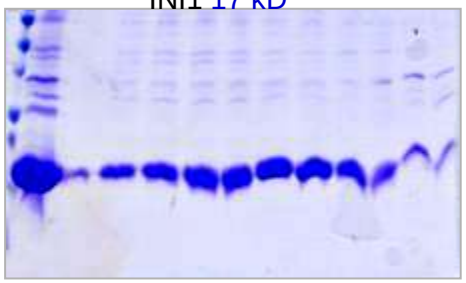
HIS-INI1(174-289) → His Affinity → Gel filtration G200

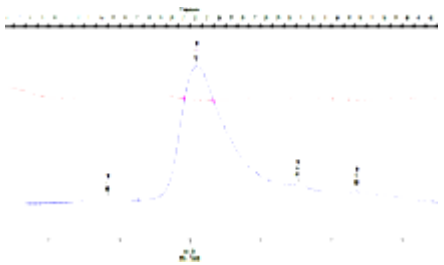
+ IN/LEDGF

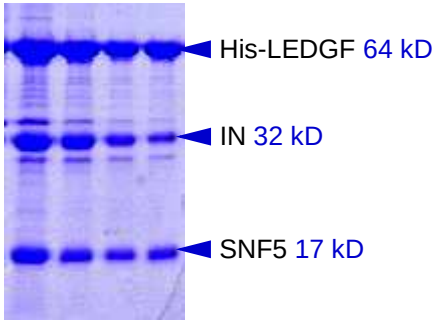
High salt detergent

Gel filtration G200

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



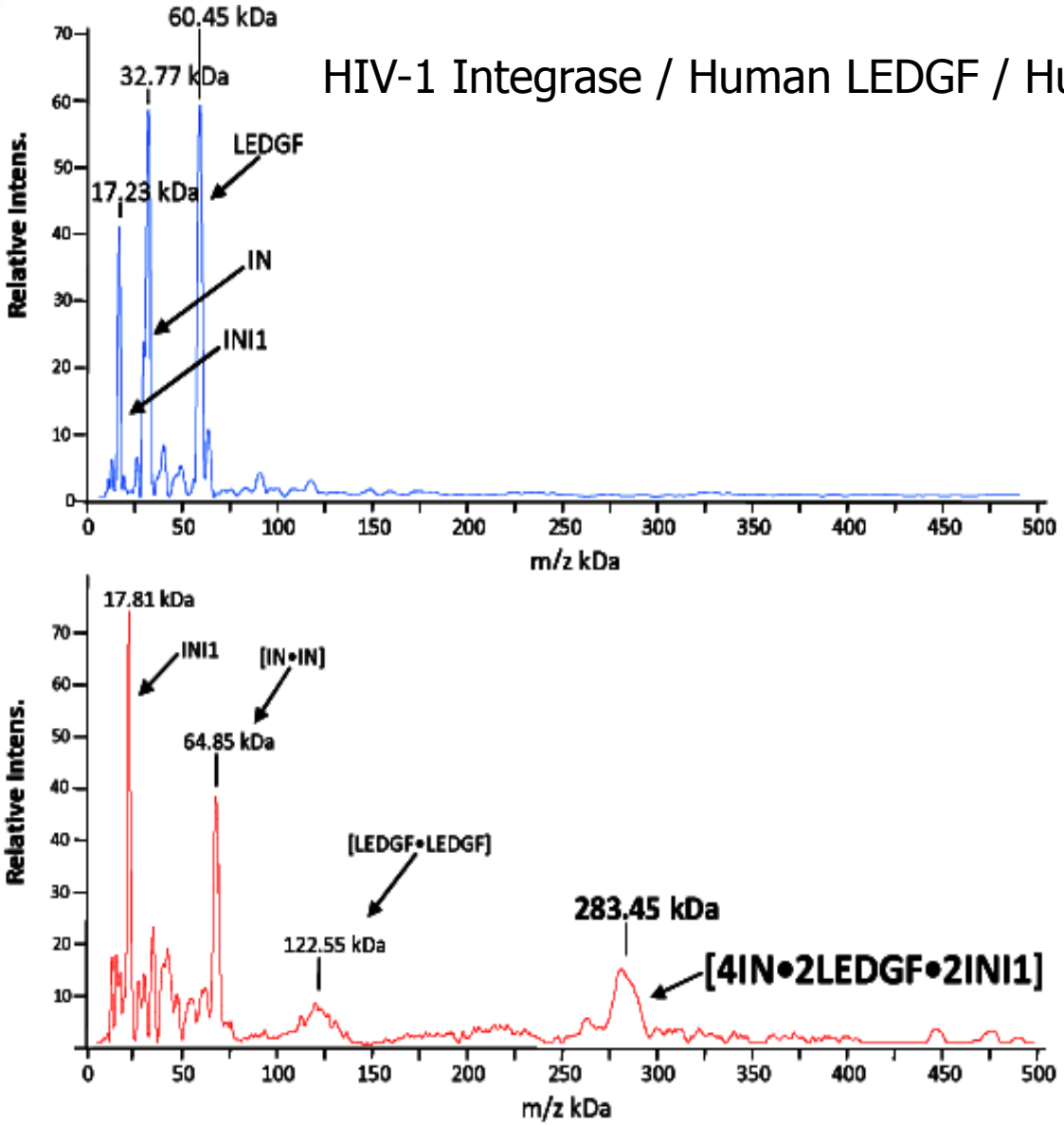




Yield: 2.5 mg of complex (INT 0.5L, LEDGF 0.5L, SNF5 1.0L)

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

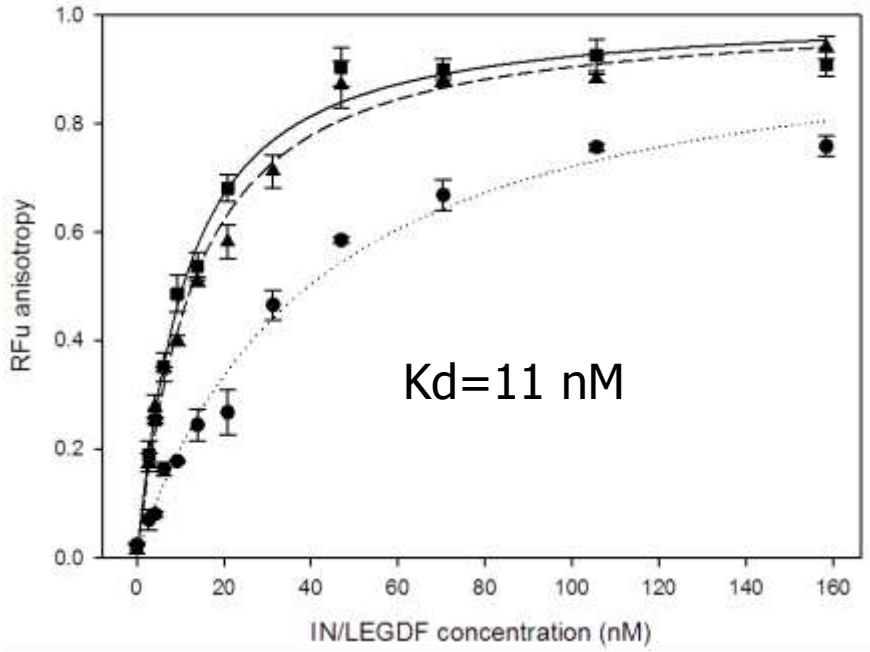
DNA: 40-mer duplex that mimic the HIV-1 U5 viral DNA end

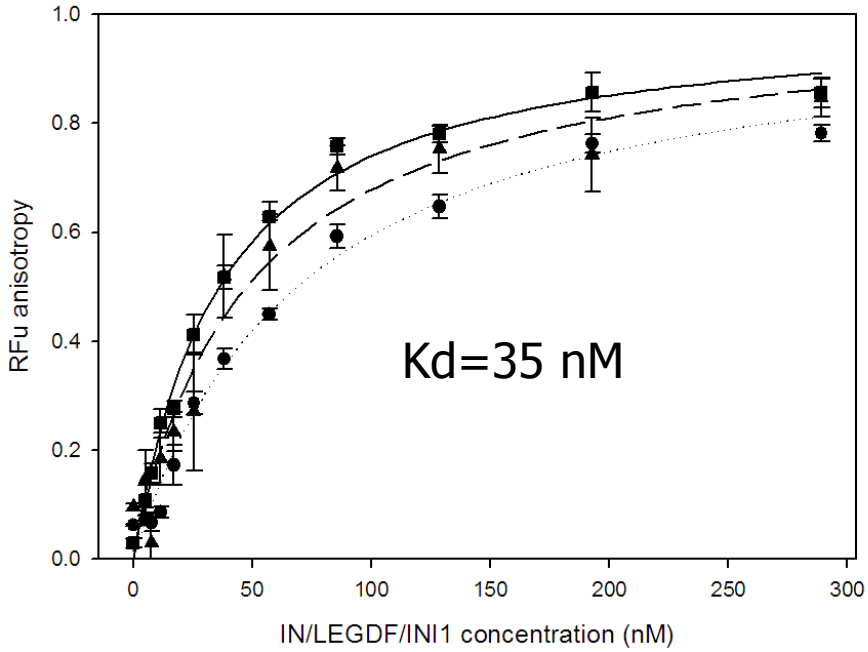
5′-GACTACGGTTCAAGTCAGCGTGTGGAAAATCTCTAGCAGT-3′

3′-CTGATGCCAAGTTCAGTCGCACACCTTTTAGAGATCGTCA-5′

Fast rotation: low anisotropy

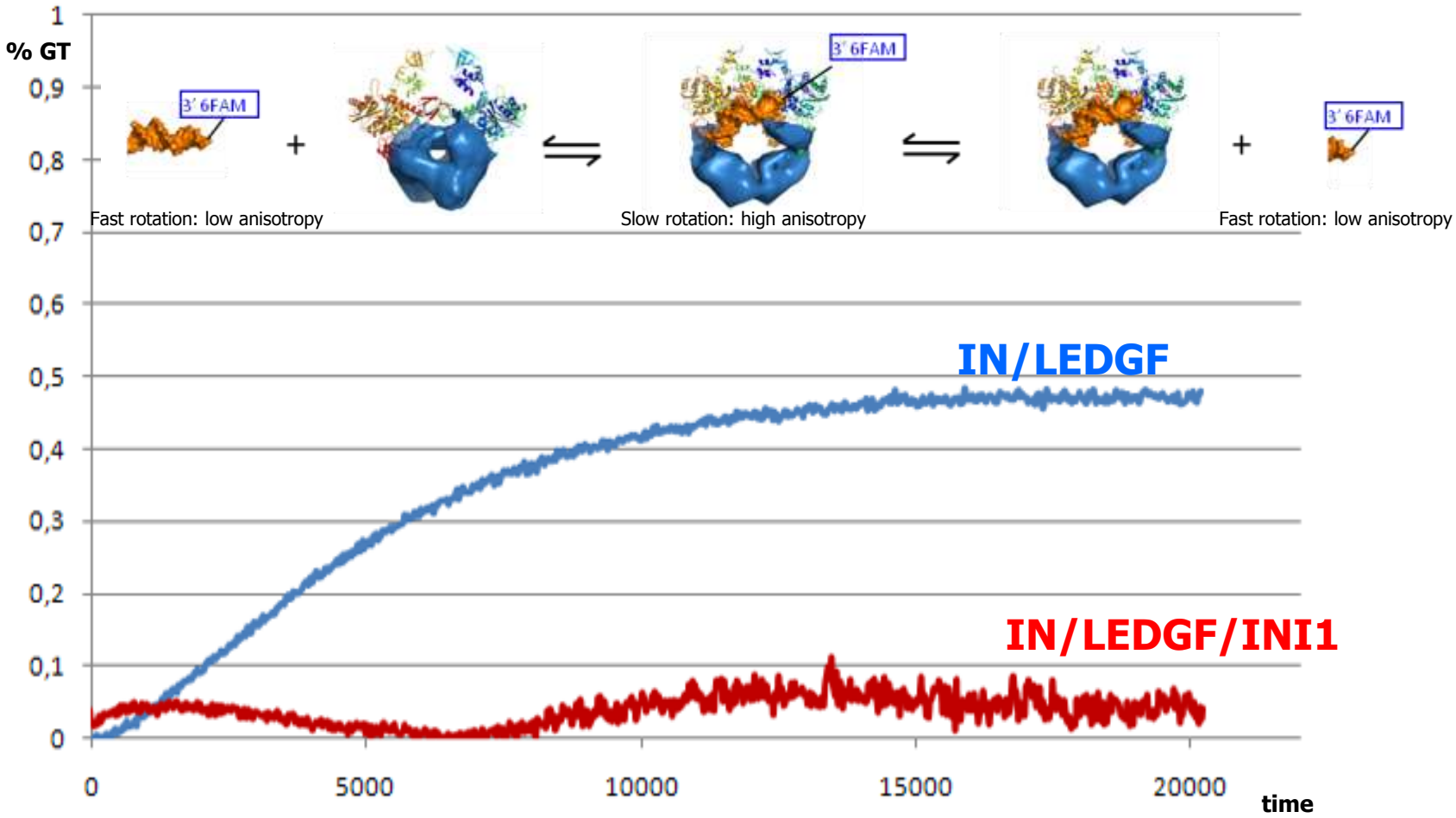
Slow rotation: high anisotropy





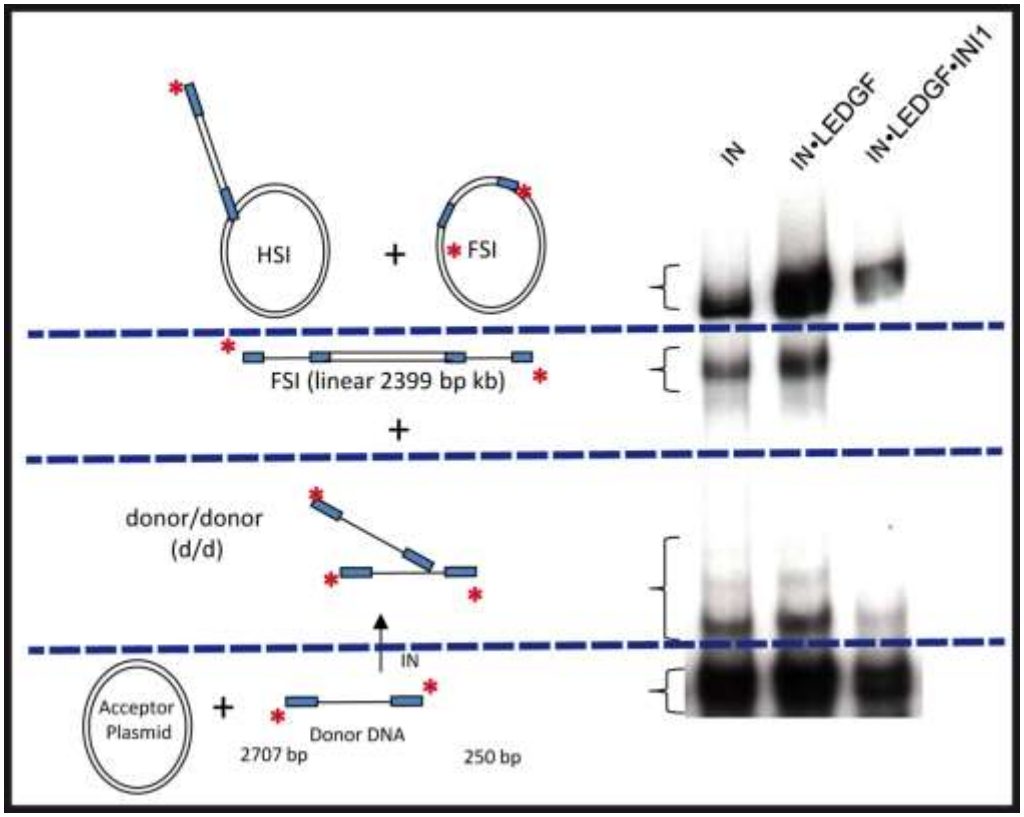
IN/LEDGF and IN/LEDGF/INI1: Functional characterization

The 3′ Processing assay by fluorescence anisotropy



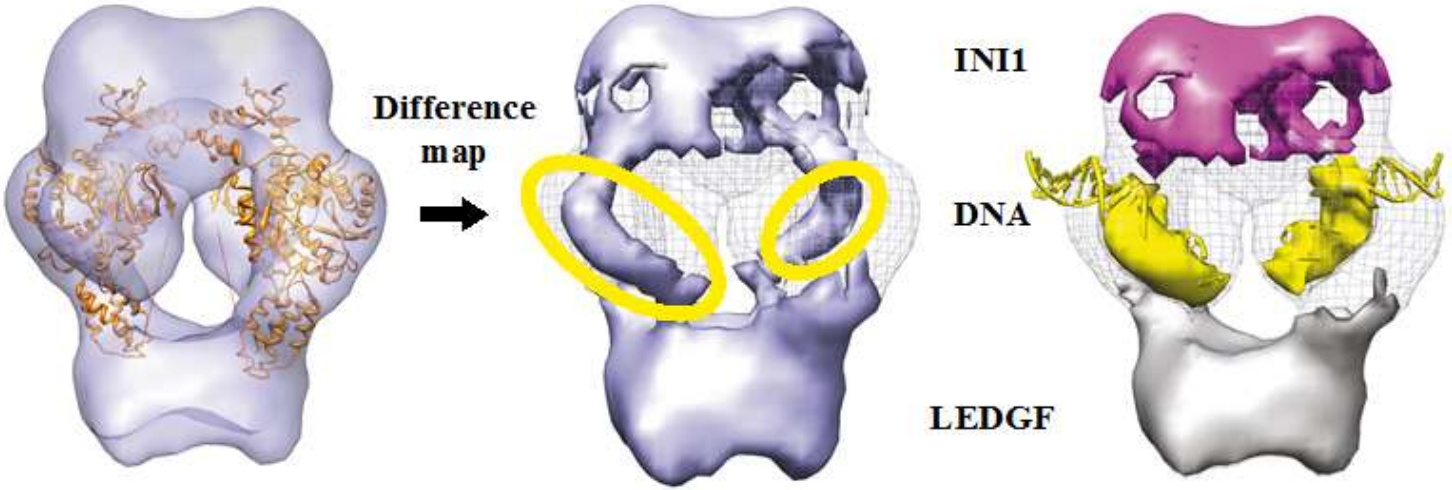
INI1 inhibits the *3′ processing* activity of IN/LEDGF

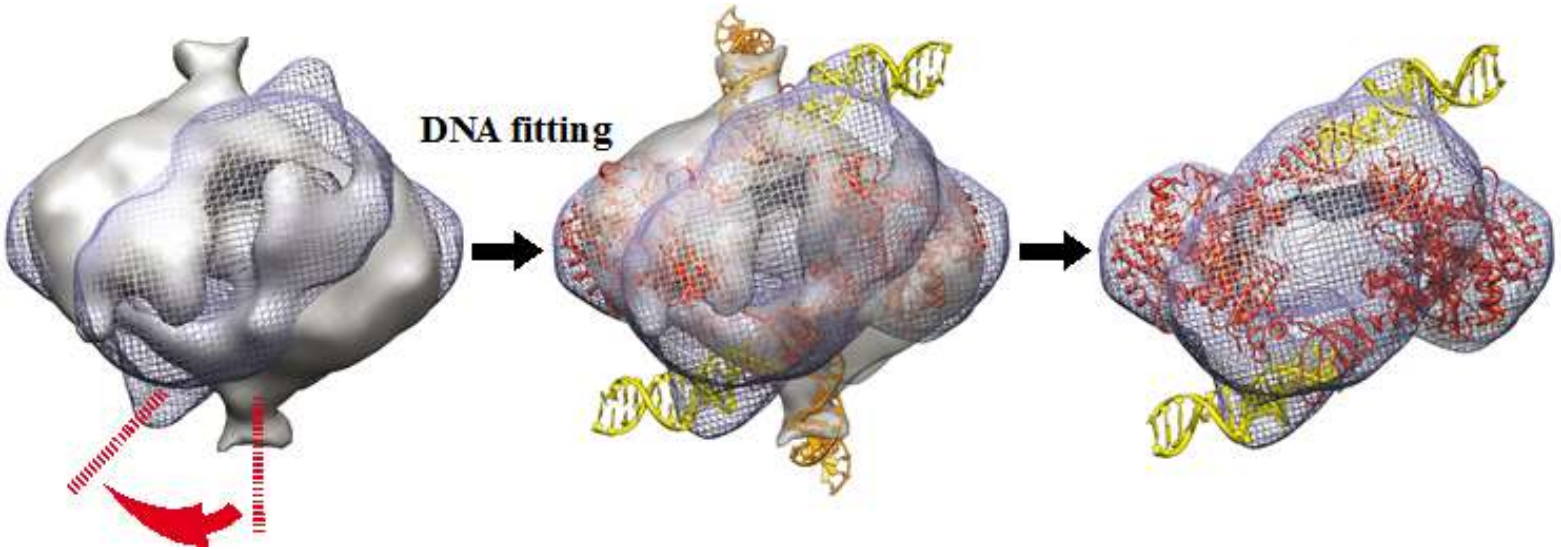
IN/LEDGF and IN/LEDGF/INI1: Functional characterization



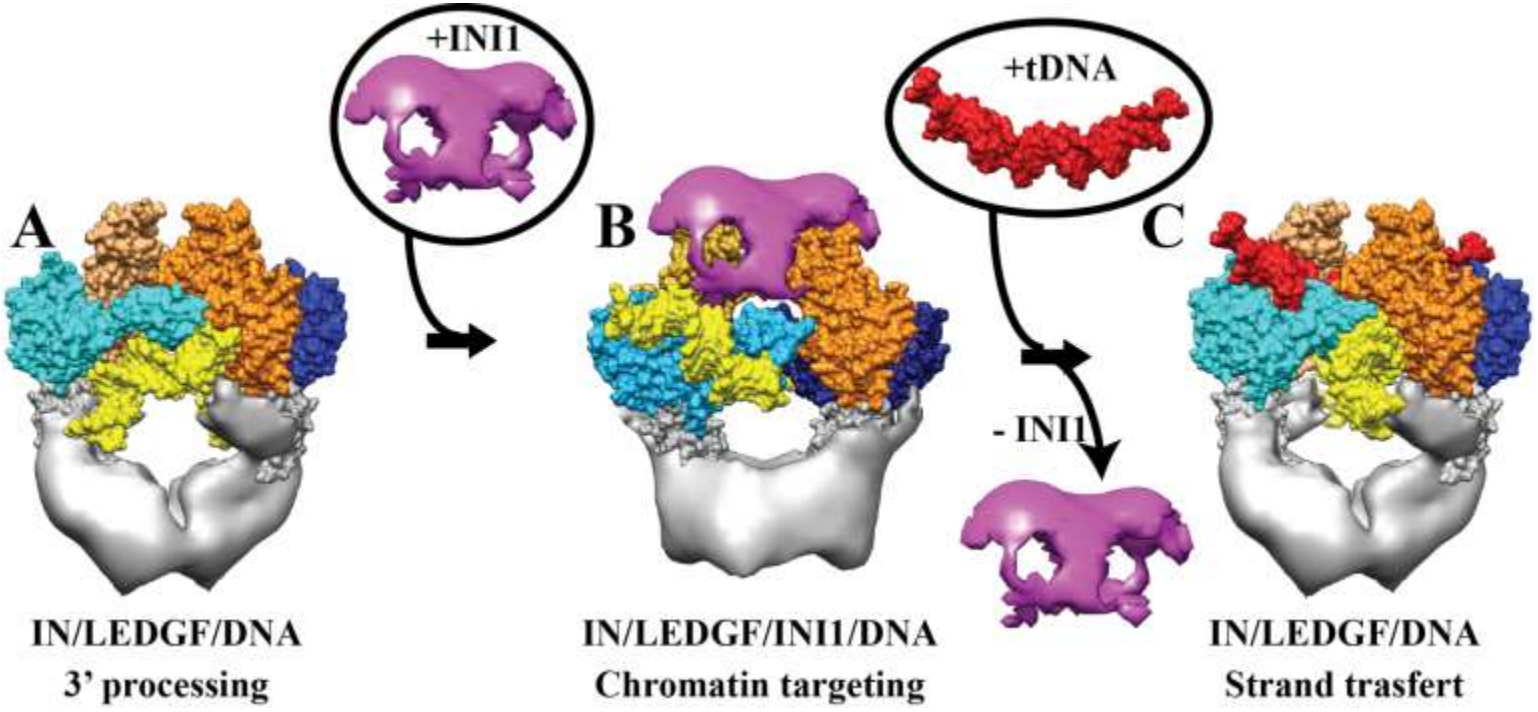
In the presence of INI1, integration occurs with reduced kinetics compared to IN alone or to the IN/LEDGF complex with strongly reduced by-products formation

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





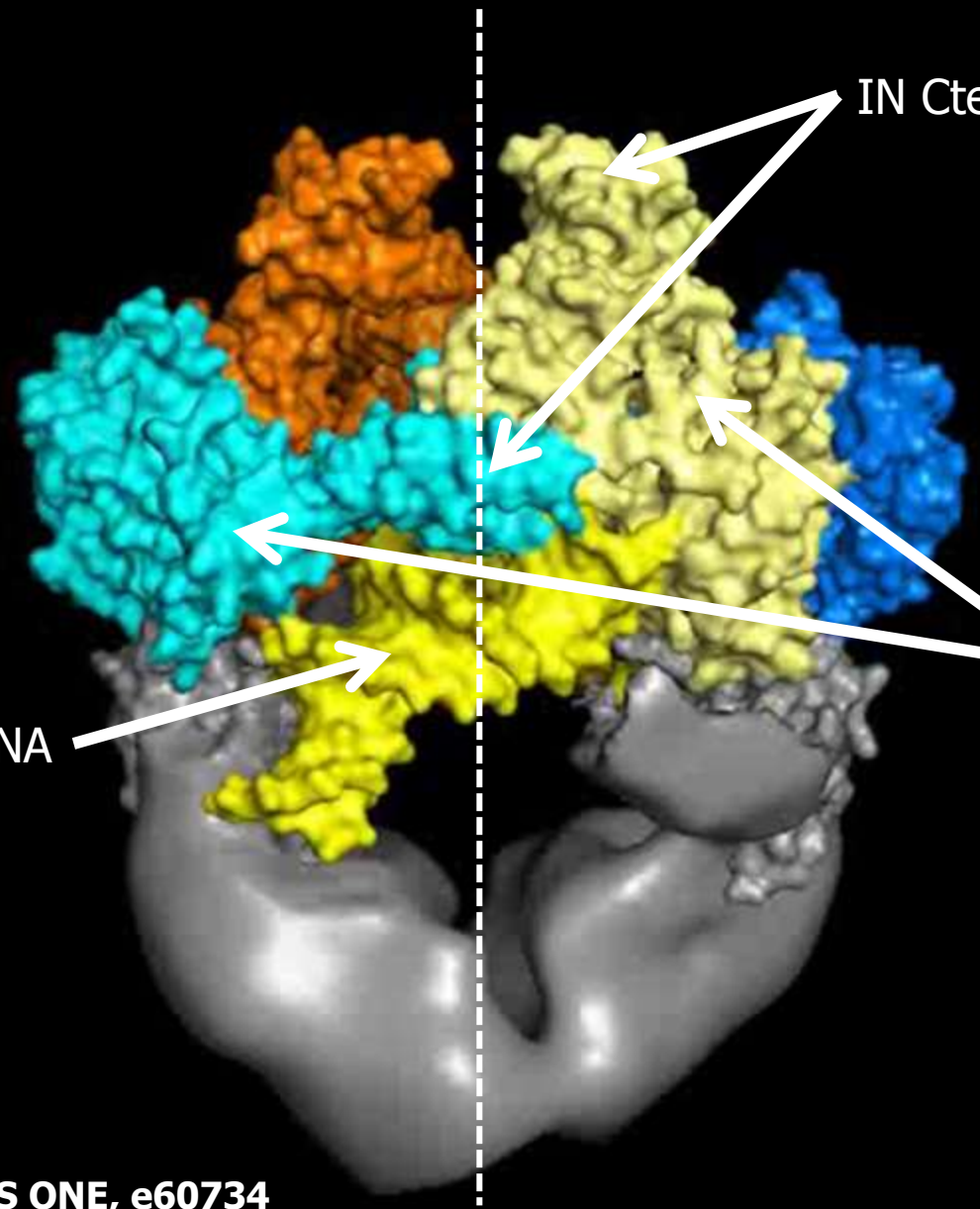
Function of INI1 in HIV-1 infection



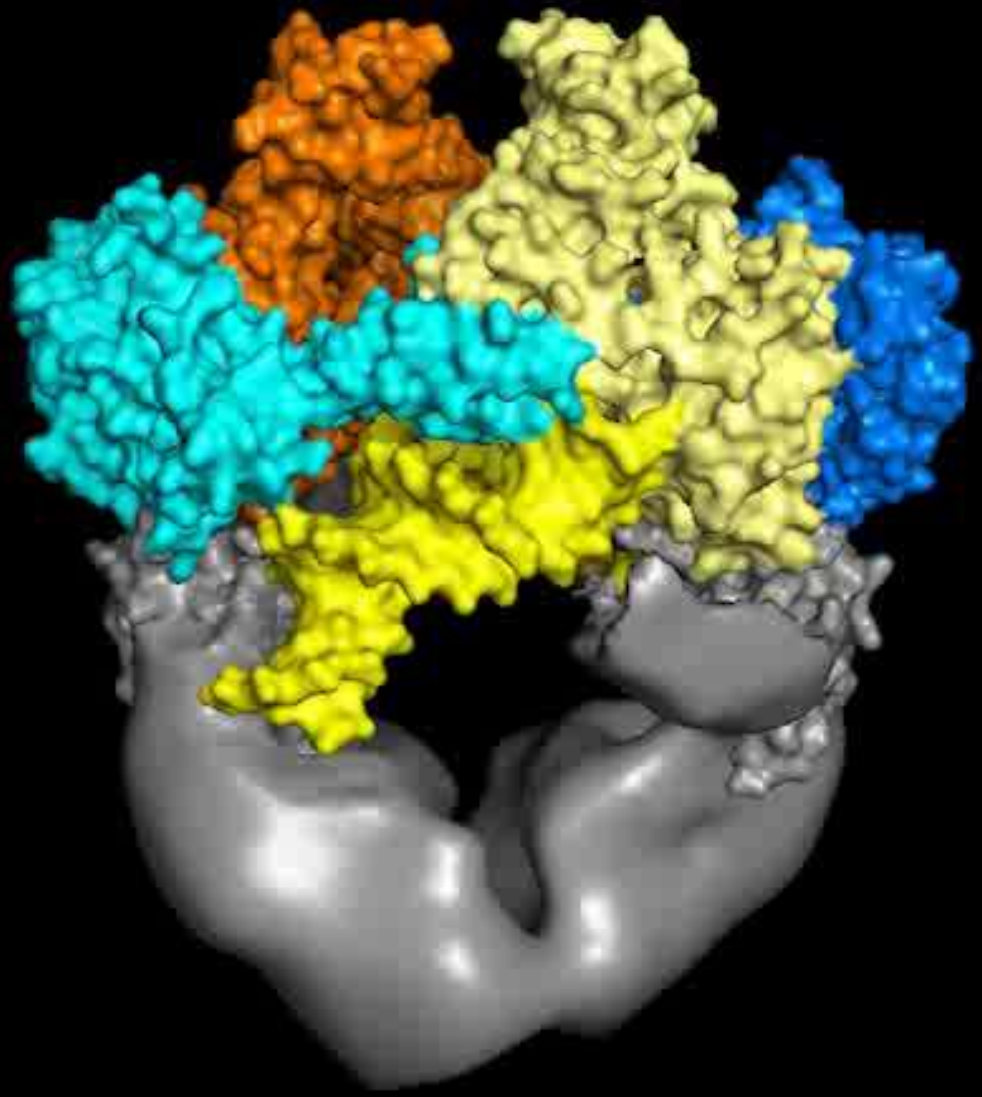
Benoit Maillot, Nicolas Lévy, Sylvia Eiler, Corinne Crucifix, et al., (2013), Structural and functional role of INI1 and LEDGF in the HIV-1 preintegration complex, PlosOne, In Press

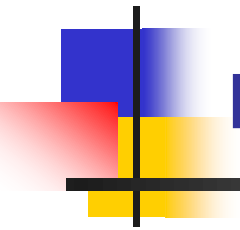
Michel, F., Crucifix, C., Granger, F., et al., (2009). Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. EMBO J., 28, 980-991

IN/LEDGF/vDNA complex



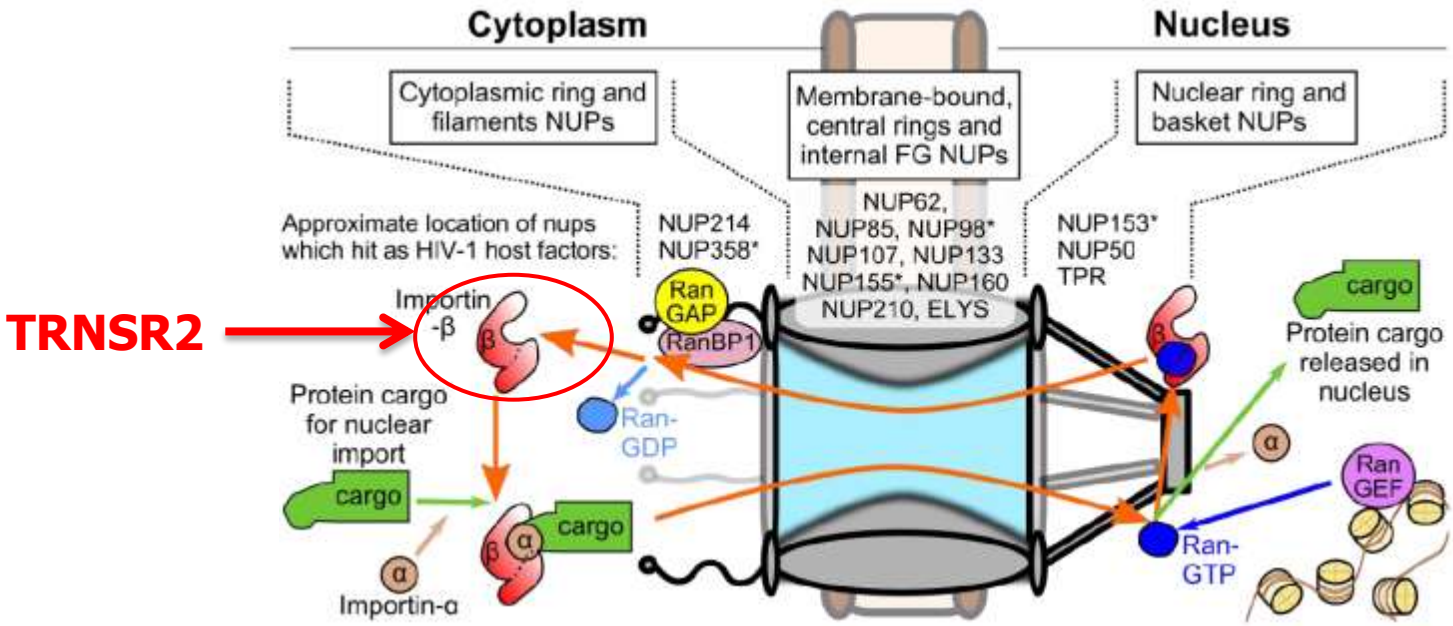
Maillot et al,(2013) PLoS ONE, e60734
Michel et al,(2009) EMBO J, 28, 980-991





HIV-1 IN / TRN-SR2 / VBP1 complex

PIC Nuclear import



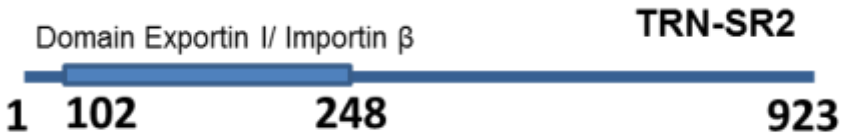
From K.A Matreyek and A. Engelman, 2013, Viruses, 5, 2483-2511

Capsid dependent nuclear transport factors: NUP153, NUP358, CPSF6, CypA

We have characterized two stables complexes: IN/TRNSR2/VBP1 and IN/TRNSR2/LEDGF

IN interact with TRNSR2 and VBP1

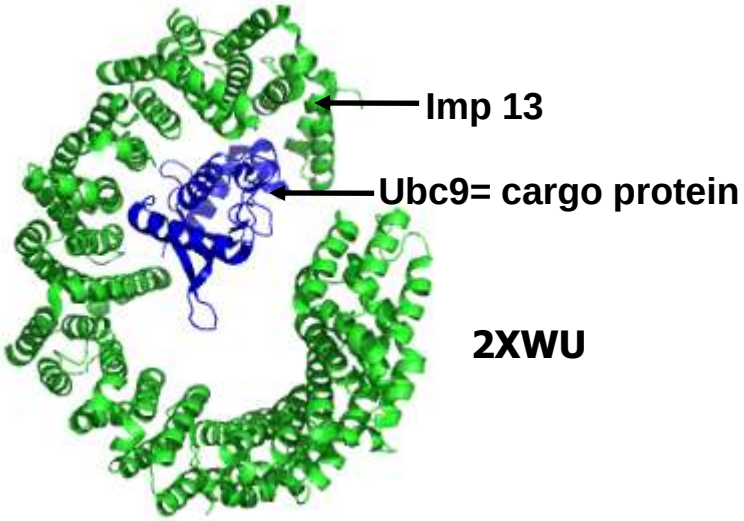
TRN-SR2



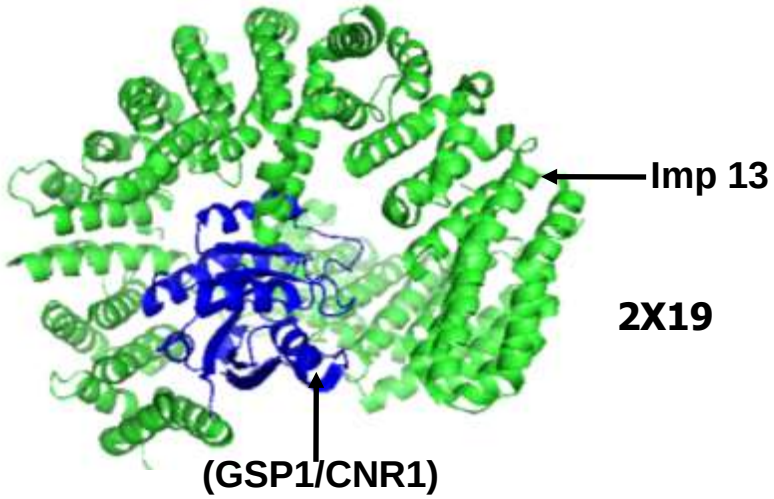
Importin β family (Lai et al, 2000), SR proteins (involved in RNA splicing) nuclear import receptor (Lai et al, 2001)

TRN-SR2 siRNA studies showed inhibition of HIV-1 replication after reverse transcription and before integration (Brass et al, 2008), PIC nuclear maturation (Zhou L, 2011), PIC nuclear import with CA (De Iaco A, 2011, Yamashita M,2007) and IN (Christ F, 2008)

22% identity with Importin 13



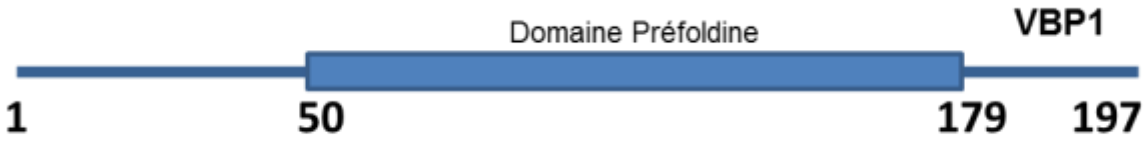
Gruenwald et al, 2011



Bono et al, 2010

IN interact with TRNSR2 and VBP1

VBP1: Von Hippel Lindau Binding Protein1



Interaction with pVHL, E3 ubiquitin ligase tumor suppressor protein in complex with Elongin B, ElonginC, Cullin2 et Rbx-1 (Tsuchiva et al, 1996)

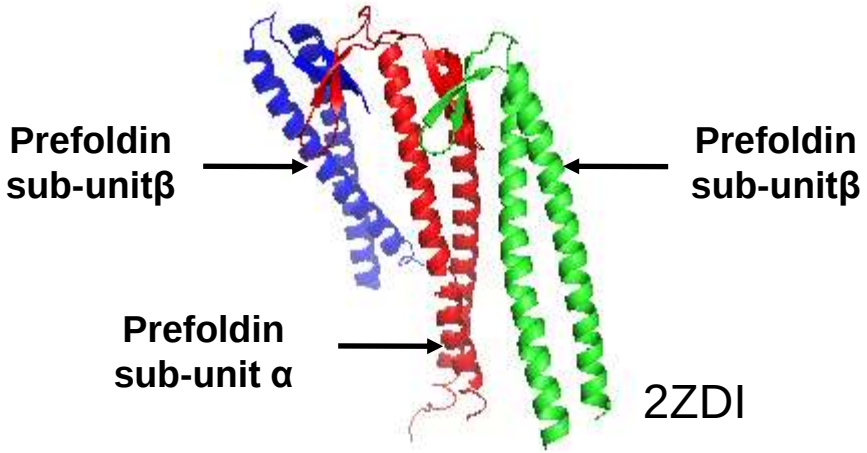
Subunit 3 of the Prefoldin heterohexameric complex involved in the actin, tubulin α, tubulin β and other cellular proteins through the chaperonin TRiC (Martin-Benito et al, 2002)

Integrase degradation through the ubiquitin/proteasome pathway via VHL E3 ubiquitine ligase complex (Mousnier et al., 2007)

VBP-1 implication in the integration and transcription transition in the viral replication cycle

Role in the pre-integration complex stabilization?

29% identity with the α subunit of Pyrococcus horikoshii prefoldin



Ohtaki et al, 2008

IN/TRN-SR2/VBP1 : complex formation and purification: E. Coli

High salt, detergent

Purification His-Thr-IN

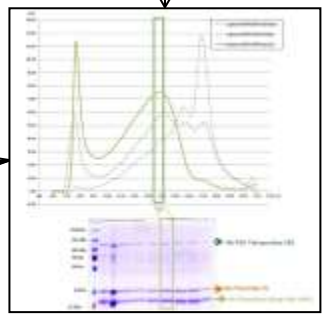
Purification HIS-Strep-VBP1

Purification HIS-P3C-TRN-SR2

Complex formation by dialysis

Strep Affinity

Gel filtration G200



Functional and biochemical characterization

HTS Screening of inhibitors

Crystallization Assays

SAXS, SANS studies

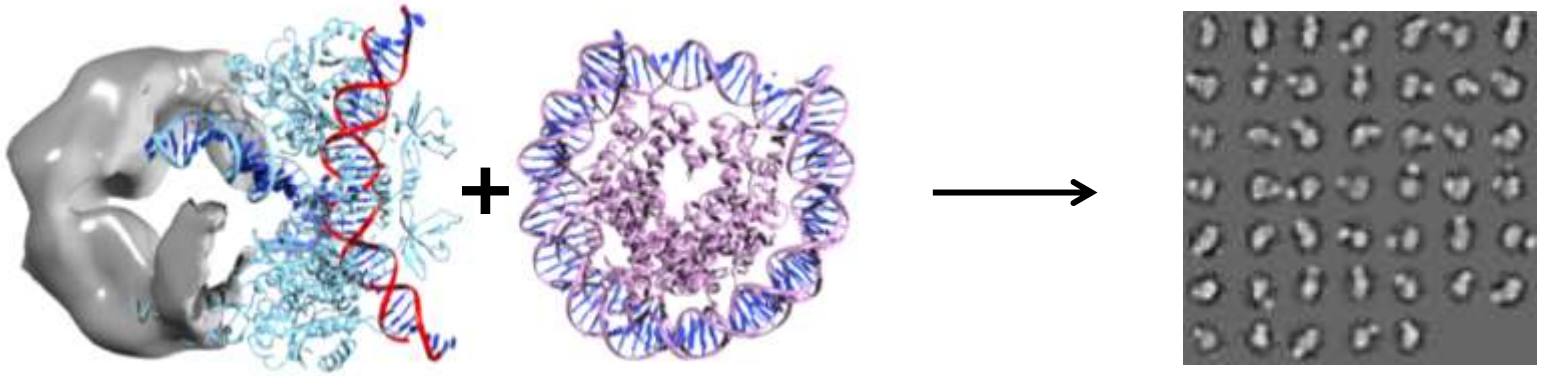
EM studies

Cryo-EM analysis

Work in progress

Other PIC complexes….. In progress

IN interacts with the nucleosome



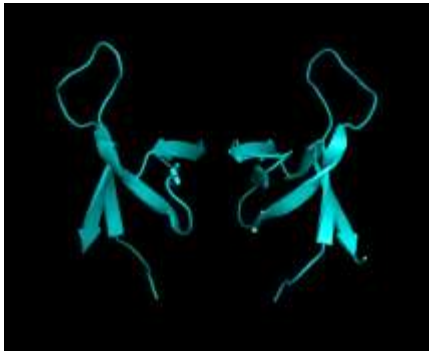
IN/LEDGF/DNA complex

Nucleosome

Preliminary Negative stained EM

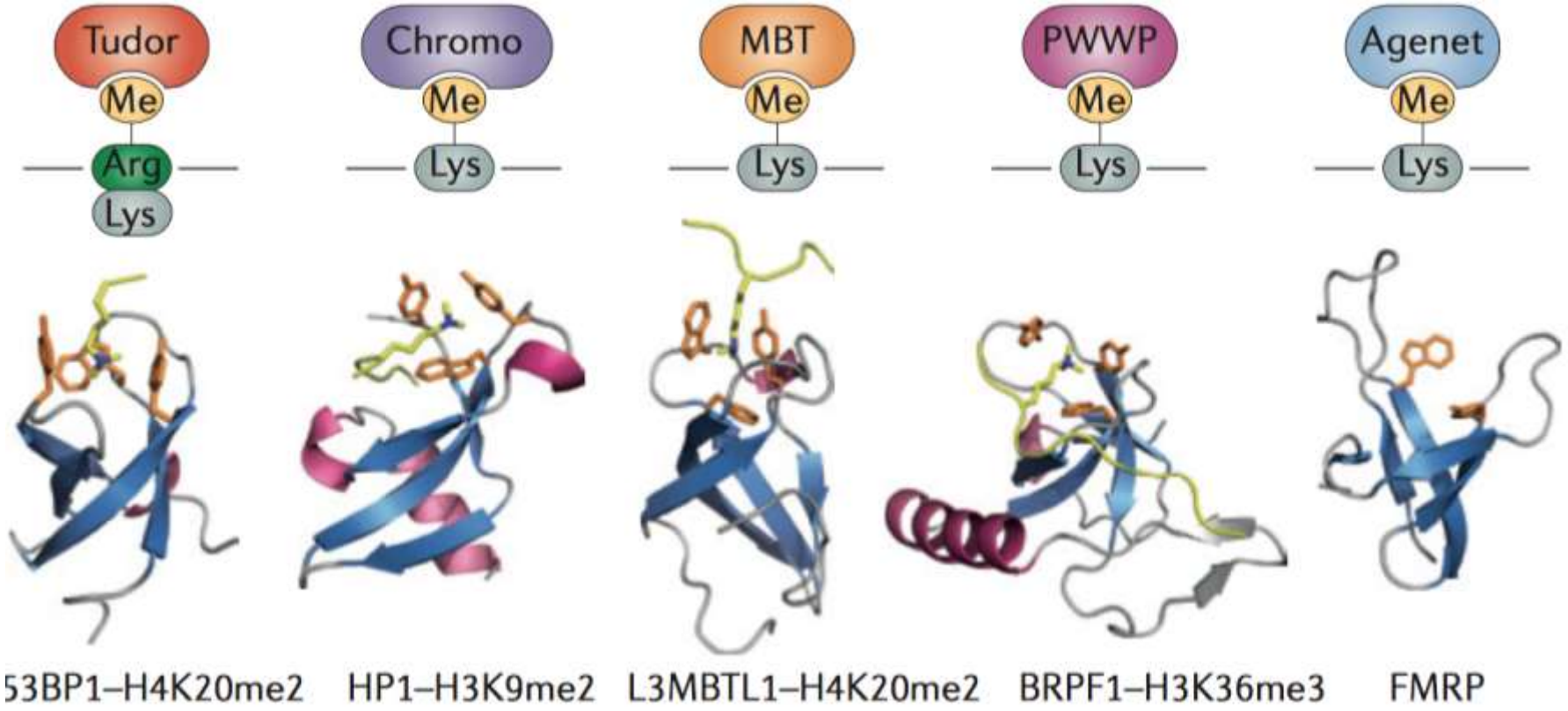
Structure and function of HIV Integrase domains

Histone H4K20Me1 interacts with HIV-1 IN C-terminal domain



1QMC, NMR structure

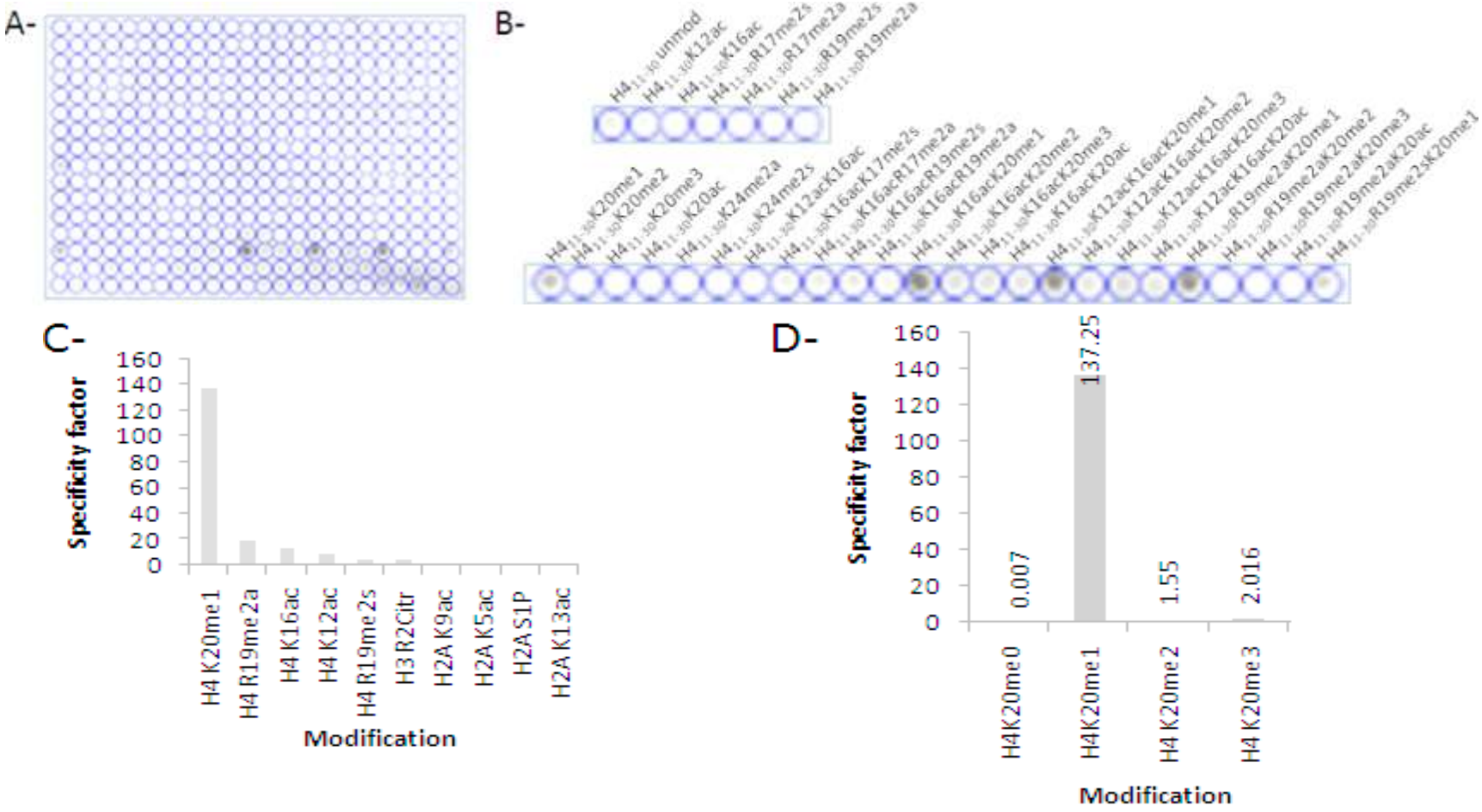
Royal Domain Superfamily



Chen et al, Nature Mol Cell Bio, 2011

The Royal Domain Superfamily is known to recognize post-translationally modified histone tails. The C-terminal Domain of HIV-1 Integrase has a SH3 fold consistent with the Royal Domain Superfamily, especially the Tudor domains. However, the CTD lacks the aromatic cage consistent with members of this Superfamily.

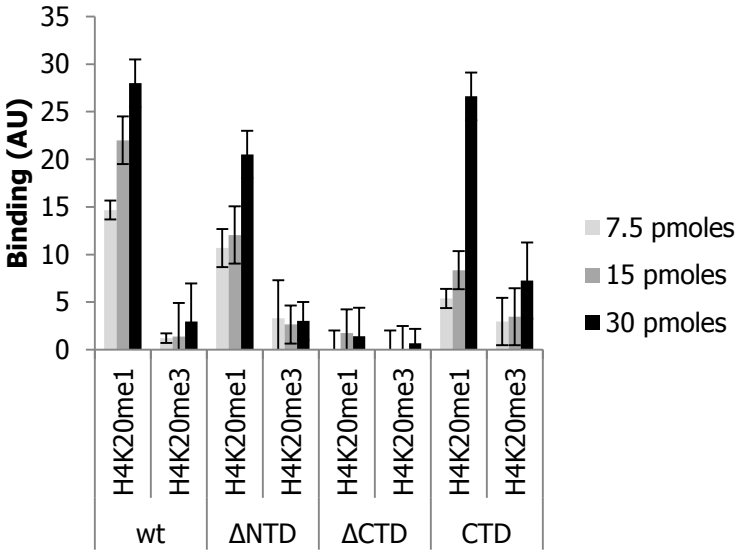
Histone peptide array: H4K20Me1 interacts with HIV-1 IN



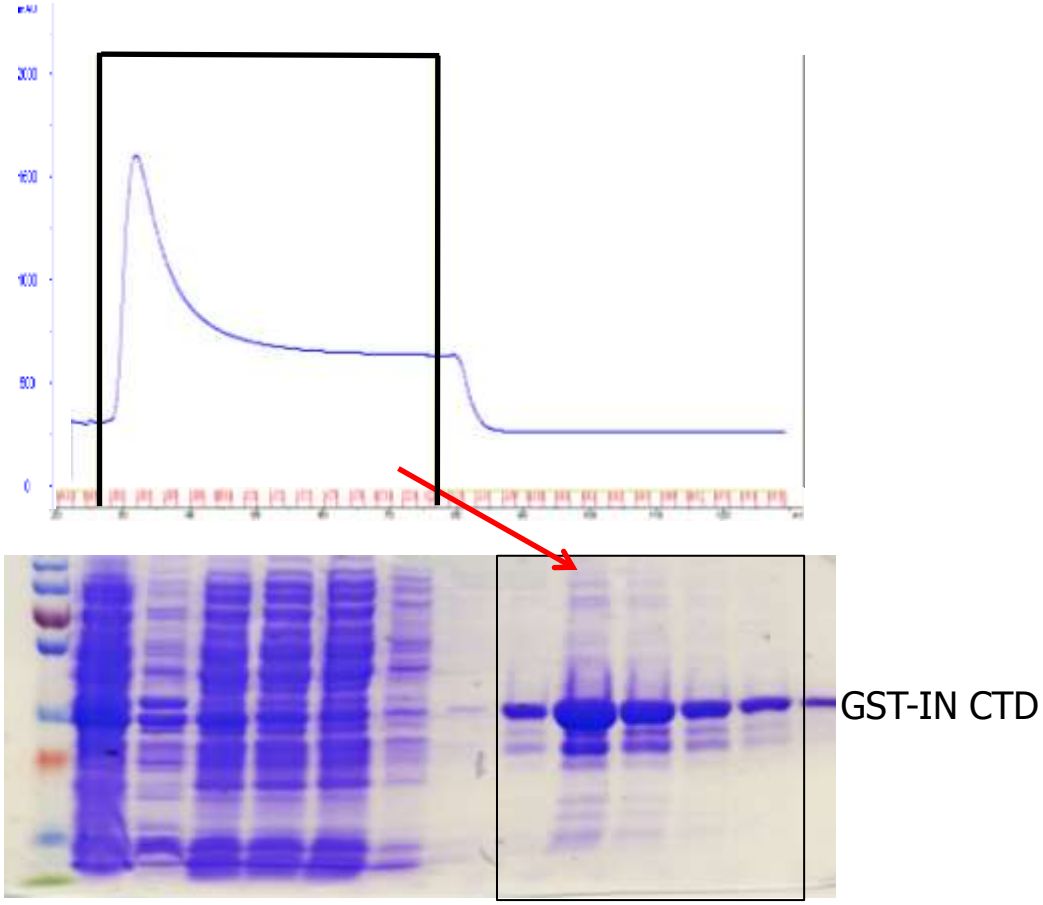
Matysiak et al., submitted

The amino terminal tail of H4K20Me1 interacts specifically with full length HIV-1 IN.
The specificity factor is higher in H4K20Me1 than in other peptides/other methylation states.

GST-CTD HIV-1 Integrase purification

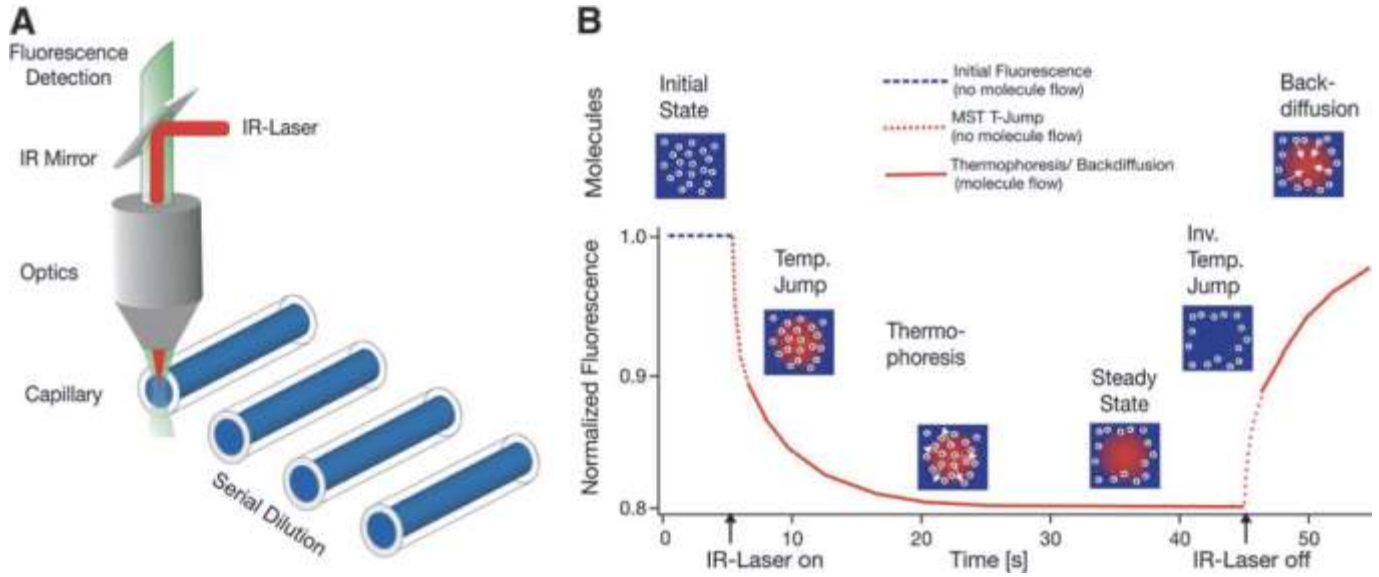


Far dot blot assay



Protein purified using GSTrap column on the AKTA system.

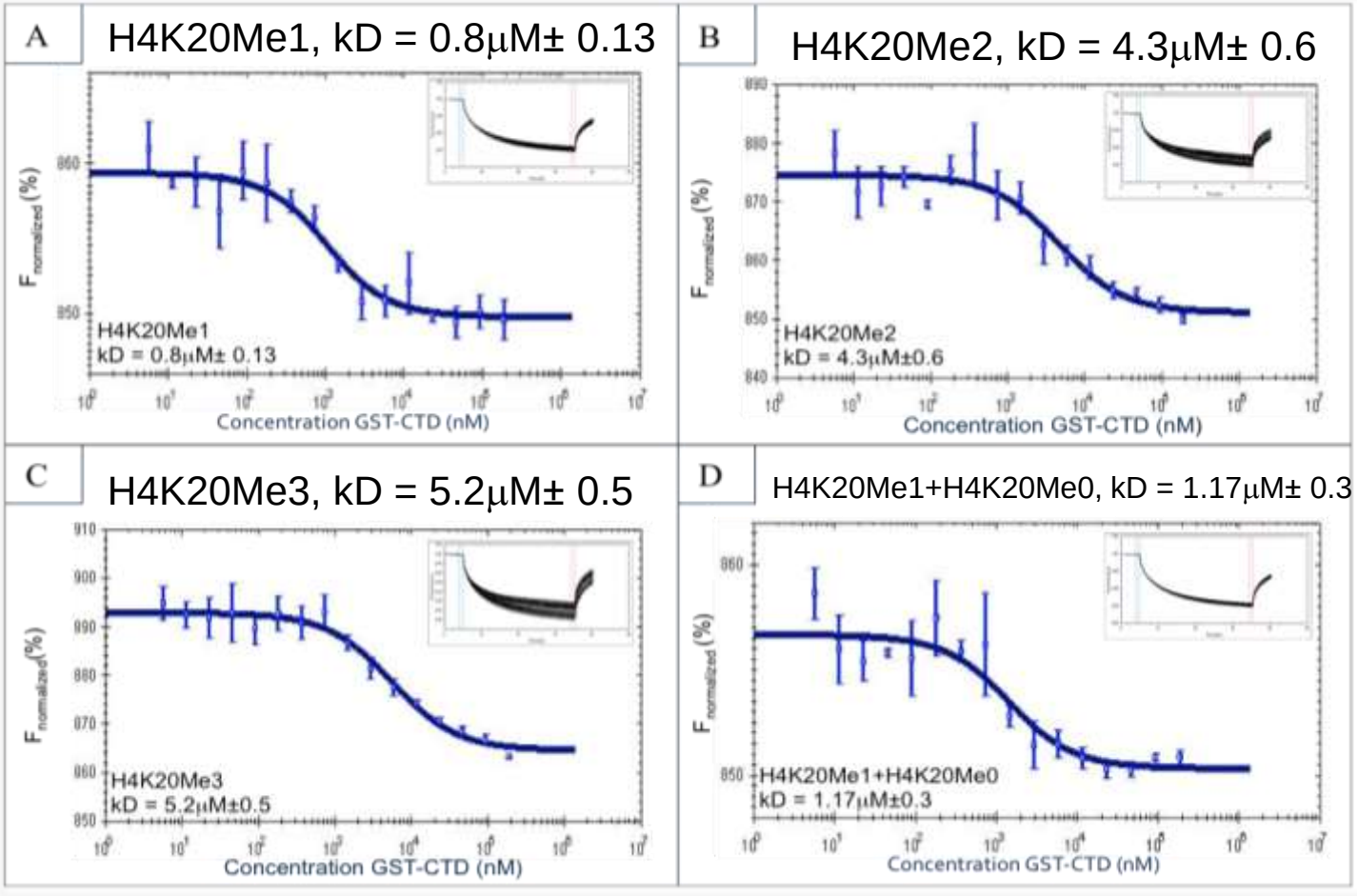
Microscale thermophoresis for Kd measurement



(A) MST is measured in a capillary with a total volume of 4 mL. The fluorescence within the capillary is excited and measured through the same optical element. An IR-Laser is used to locally heat the sample volume that is observed by fluorescence. T-Jump and thermophoresis are directly observed as a change in fluorescence at different time scales.

(B) A typical MST signal for a given capillary. Initially, the molecules are homogeneously distributed and a constant "initial fluorescence" is measured. As soon as the IR-Laser is turned on, a fast T-Jump is observed, followed by thermophoretic molecule motion. The fluorescence decrease is measured for about 30 s. When the IR-Laser is turned off, an inverse T-Jump is observed, followed by the "backdiffusion" of molecules, which is purely driven by mass diffusion, allowing to deduce information on the molecule size.

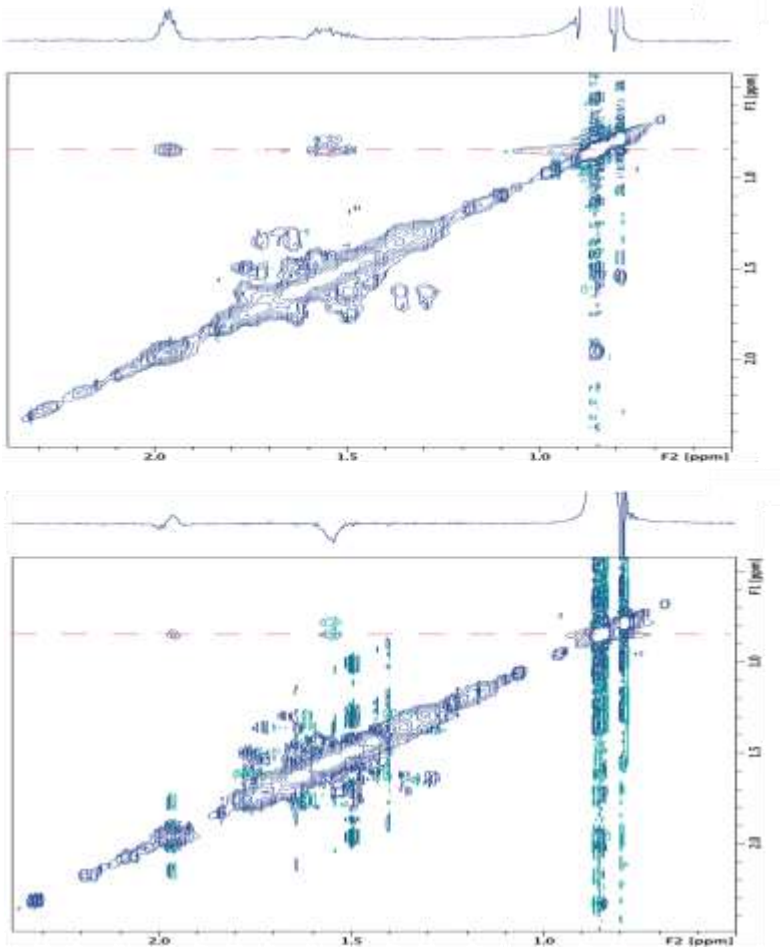
Microscale Thermophoresis confirm a specific interaction between H4K20Me1 and GST-CTD HIV-1 Integrase



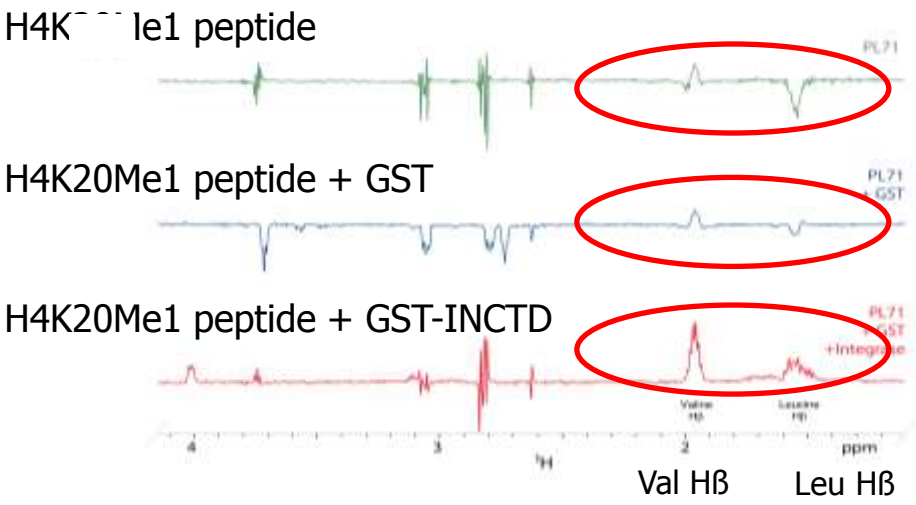
Matysiak et al., submitted

GST-CTD HIV-1 Integrase preferentially binds to the monomethylated H4K20 with a kD of 0.8μM, when compared to the di- and tri- methylated H4K20 (4.3μM and 5.2μM respectively). In the presence of a competing peptide (H4K20Me0), the monmethylated peptide binds with a similar kD of 1.17μM±0.3.

NOESY-NMR confirms the interaction between GST-CTD HIV-1 Integrase and H4K20Me1



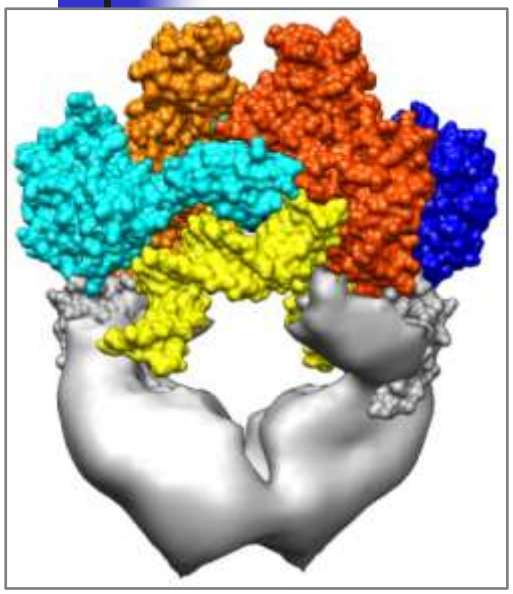
A change in NOE signal indicates an interaction between GST-CTD HIV-1 Integrase and H4K20Me1



Interaction is specific to CTD as there is no change in signal in the sample with GST only

NMR and crystal structure in progress

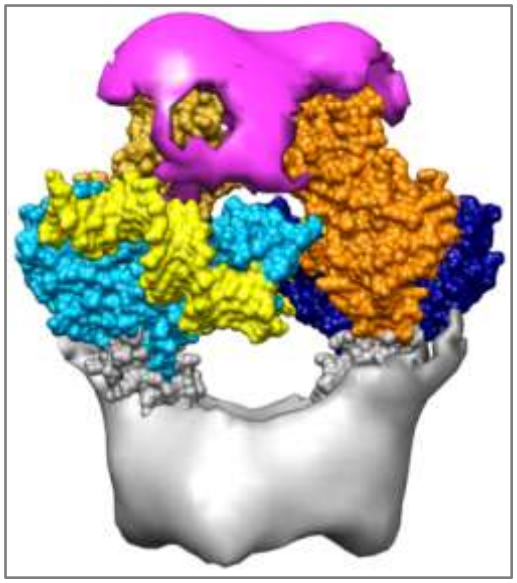
Towards high resolution structure



Michel et al. EMBO J. 2009

IN/LEDGF/DNA

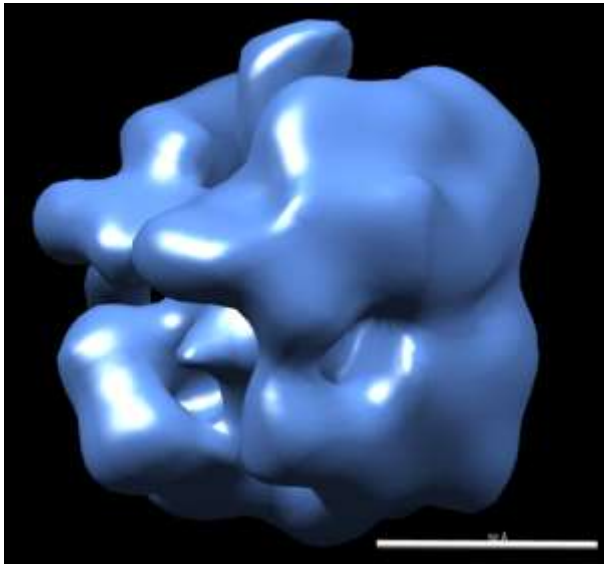
Organization and stabilization of an active IN tetramer



Maillot et al., PLOSONE, 2013

IN/LEDGF/INI1/DNA

Prevention of non-specific integration



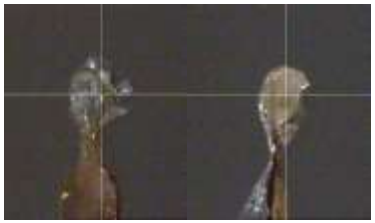
IN/VBP1/TRN-SR2

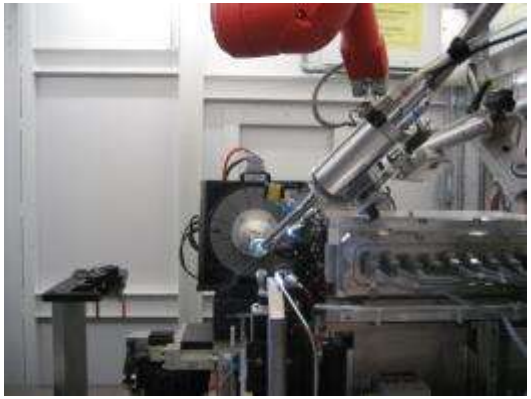
In progress

Complexes suitable for low resolution EM studies but do not crystallize : too many flexible / unstructured parts

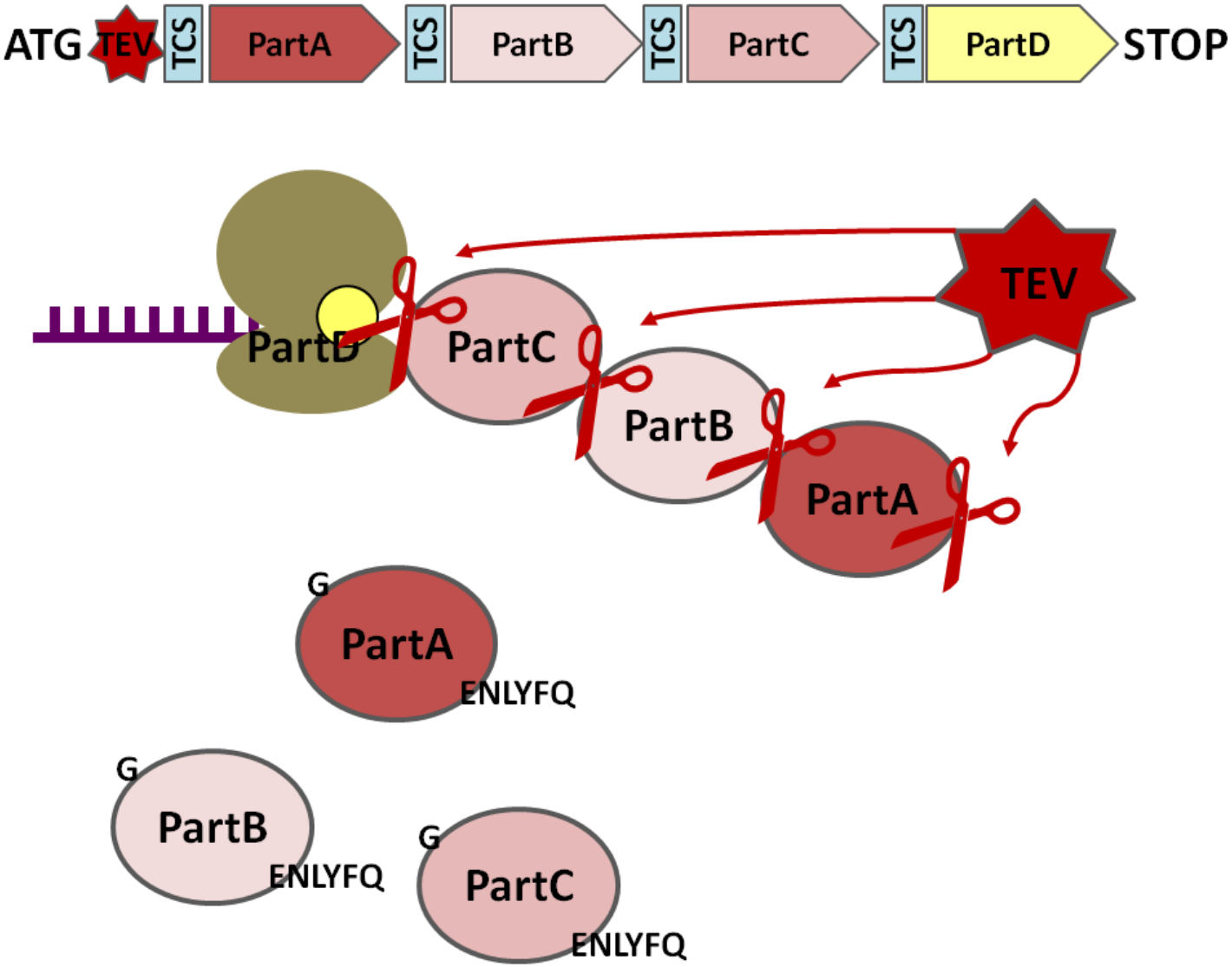
Towards high resolution structure

- **Increase solubility, stability and homogeneity**
  - Produce stable complexes in cell (Production without solubilizing agents (small detergents, high salt))
  - Effect of PTMs on stability (Setup efficient production in mammalian cells)
  - Others partners (full length, domains, ligands, …)
  - Complex reconstitution procedures
- **Improve crystallization setups**
  - Setup analysis procedures to decipher the best physico-chemical conditions (DNA stoechiometry, pH, ionic force…) for the stabilization of non-covalent complexes in homogeneous conformation (thermofluor, DLS, ITC)
  - Fast screening of crystals (in crystallization drop screening)





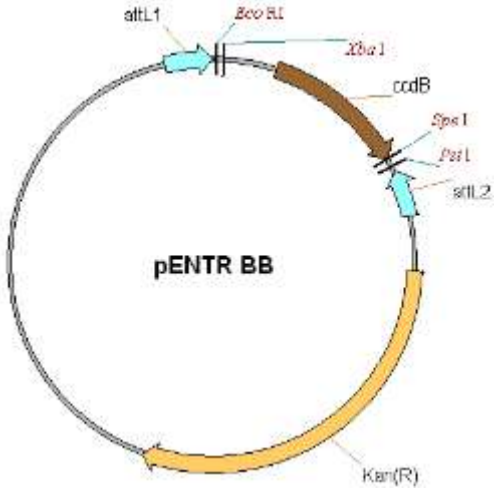
Protein complexes expression in E. Coli, Insect and mammalian cells

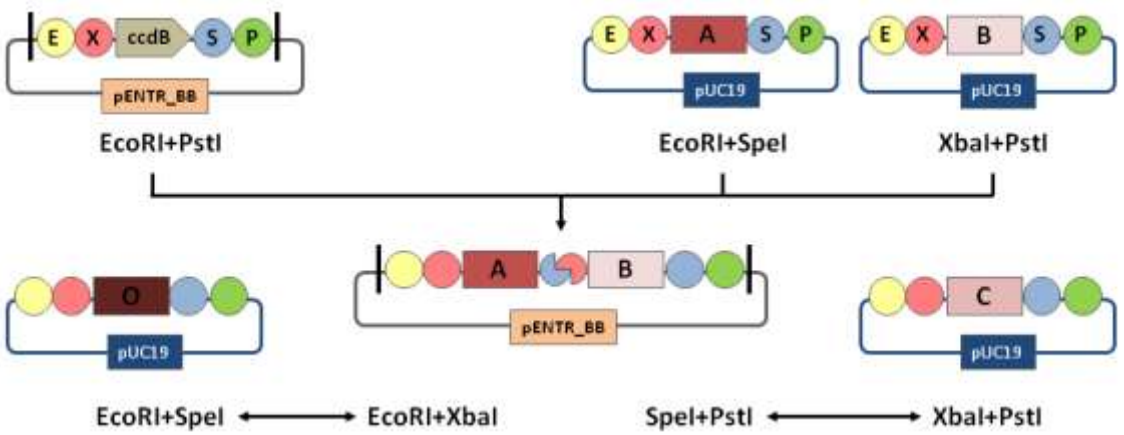


Bieniossek et al., Trends in Biochemical Sciences 2012;37(2):49-57

Synthetic biology: biobricks for multifactorial gene assembly in expression vectors

Using synthetic biology to built complex, multi-factor expression plasmids for the optimization of protein and protein complexes expression in heterologous organisms by the combinatorial use of different DNA parts.





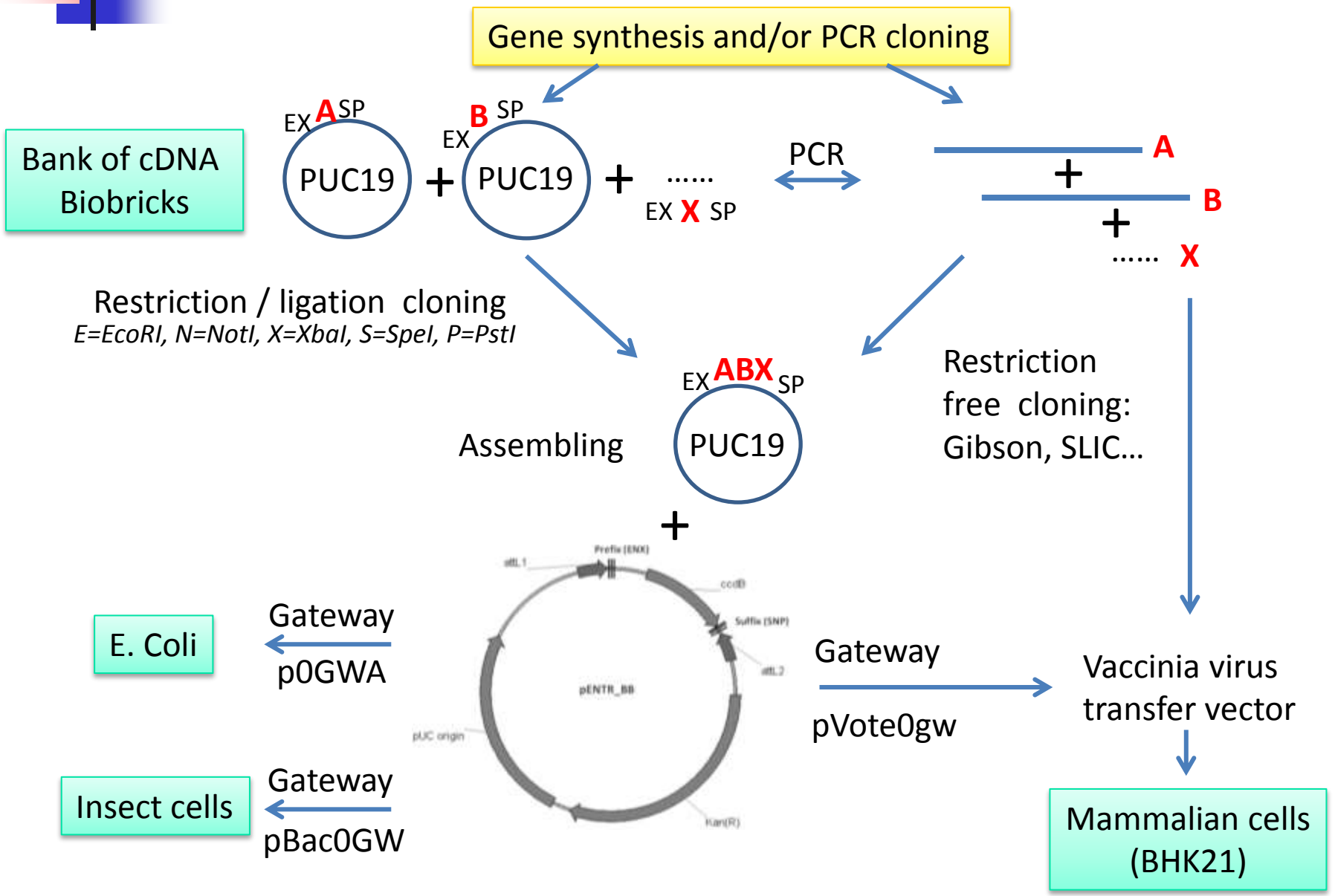
Assembly vector: New pENTR

E=EcoRI, X=XbaI, S=SpeI, P=PstI, N=Not1

Ho-Shing et al., 2012, Methods in Molecular Biology, 852, 61-76

http://partsregistry.org
http://biobricks.org/

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



Biobrick 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
	……	

IN/LEDGF in mammalian cells (vaccinia virus)

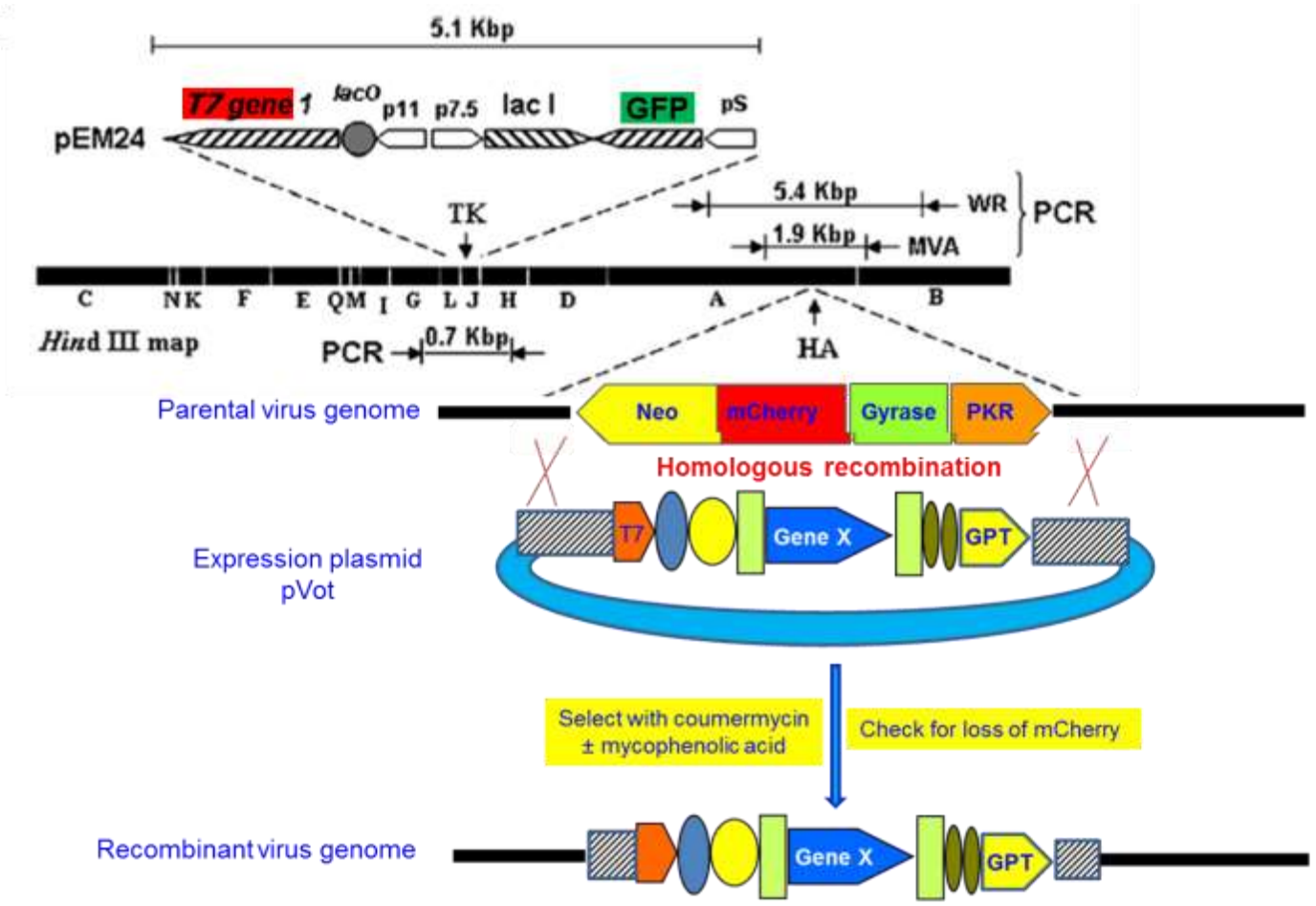
- **Vaccinia virus**: poxvirus family
  - dsDNA virus (≈ 200 kb) encoding its own transcription and replication machinery
  - viral multiplication in the cell cytoplasm: no RNA splicing
  - viral infection diverts the cellular machinery in its favour
  - at least 25 kb of foreign DNA

- **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

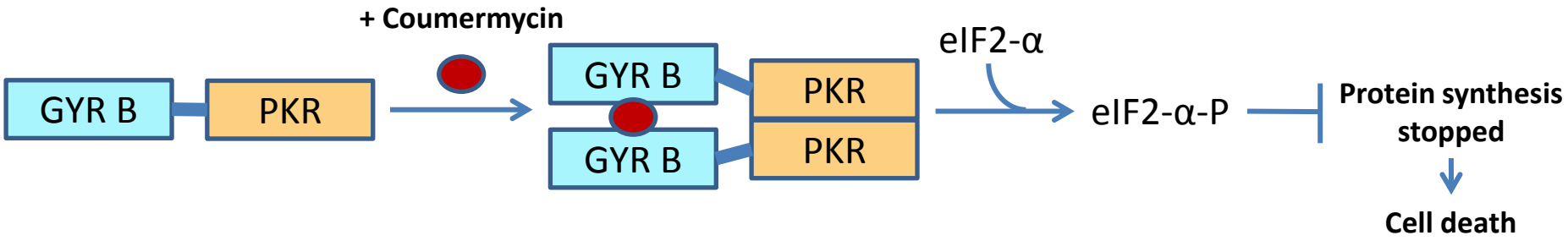
- **Encode T7 RNA polymerase, IPTG inducible**

Modified Vaccinia virus Ankara (MVA) genome



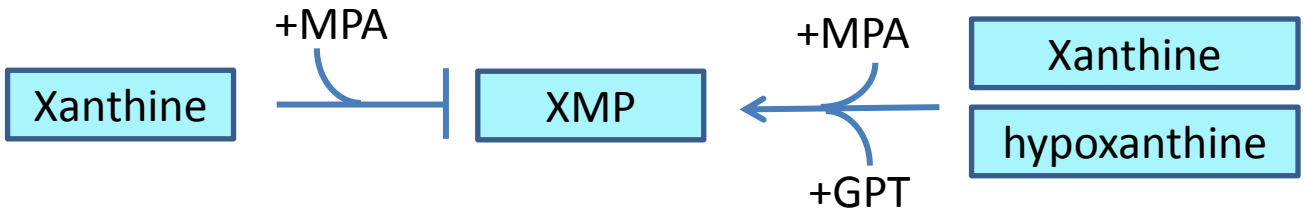
IPTG Induction => No more *LacI* inhibition
=> T7 gene expression => Target gene expression

Coumermycin mode of action



In presence of coumermycin, the GyrB domain of the bacterial gyrase dimerizes. This domain is fused to the PKR (double stranded RNA dependent protein kinase) which is activated upon dimerization. This leads to the phosphorylation of eIF2 and the arrest of protein synthesis and death of infected cells before virus production if the recombination fails.

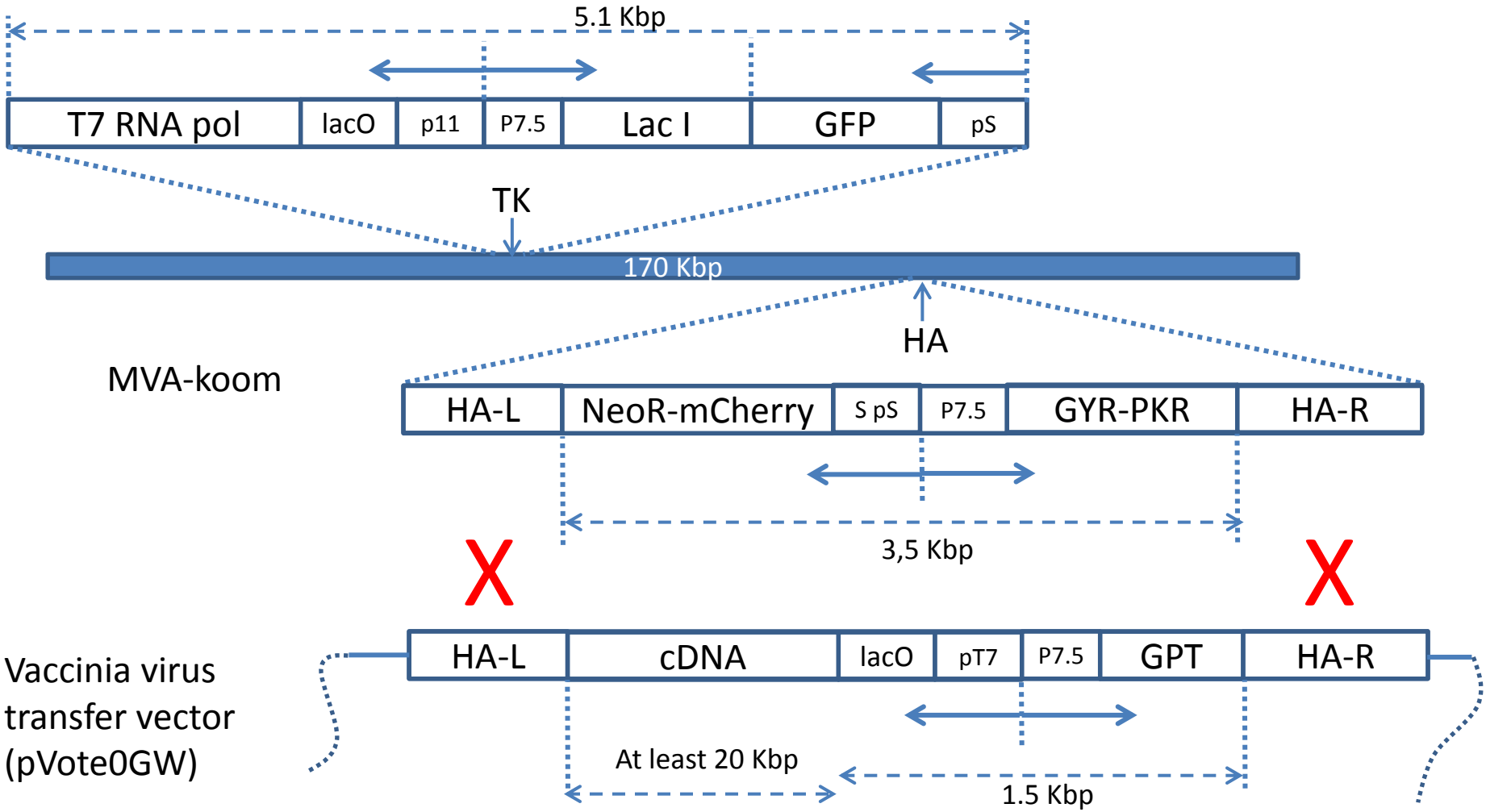
Mycophenolic acid mode of action



Cell death → viral death

Cell growth → viral replication

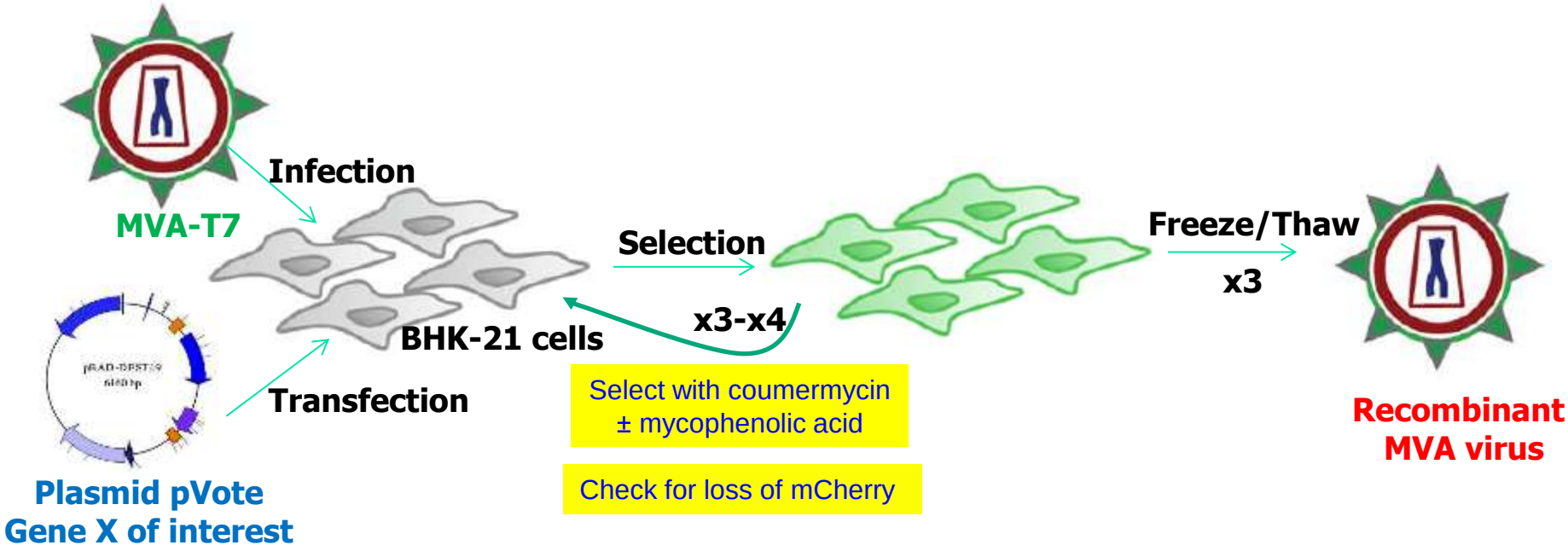
The mycotoxin MPA inhibits the enzyme inosine monophosphate dehydrogenase and thereby prevents the formation of xanthine monophosphate (XMP). This results in the intracellular depletion of purine nucleotides and in an inhibition of cell growth. The presence of xanthine-guanine phosphoribosyltransferase from *E. coli* (GPT) enable the synthesis of XMP in a selective medium containing mycophenolic acid, xanthine, and hypoxanthine and allows cell division if recombination occurs.



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)

Recombinant MVA selection



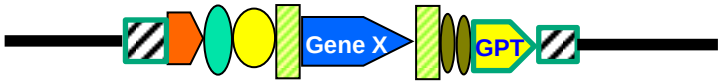
Parental virus genome

Negative selection: (monitored by mCherry loss) **+ coumermycin**

=> Gyrase dimerisation => active PKR => eif2α phosphorylation

=> translation inhibition => cell death

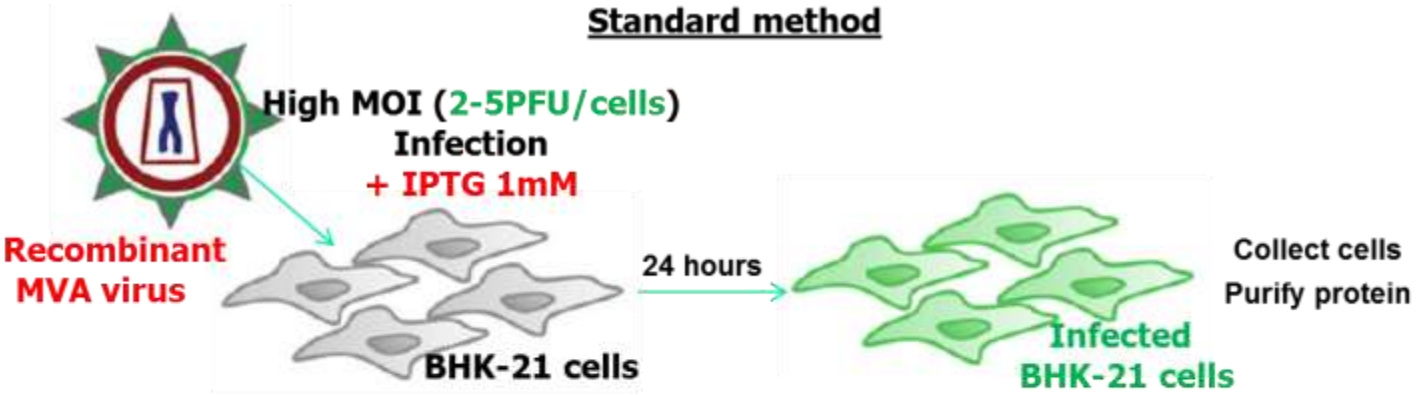
Recombinant virus genome

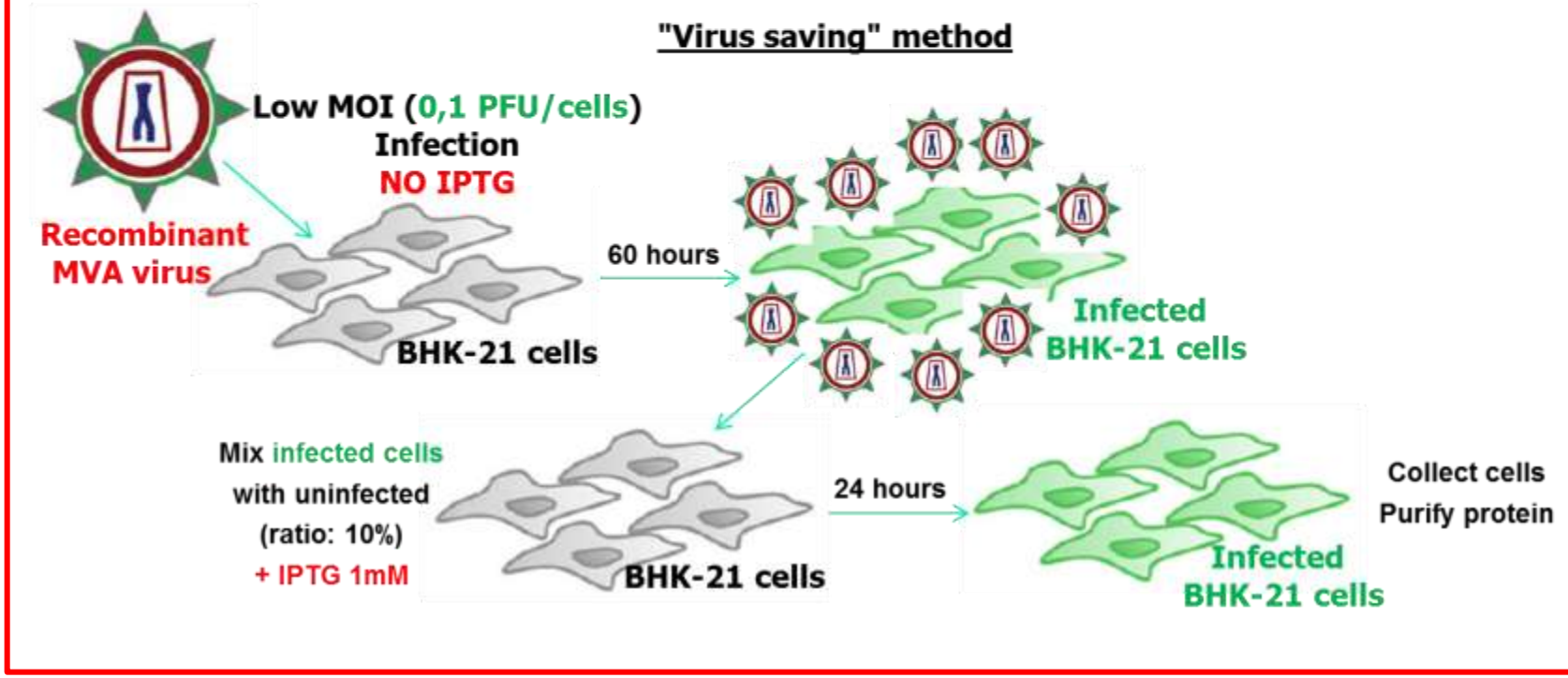


Positive selection: **+ mycophenolic acid**

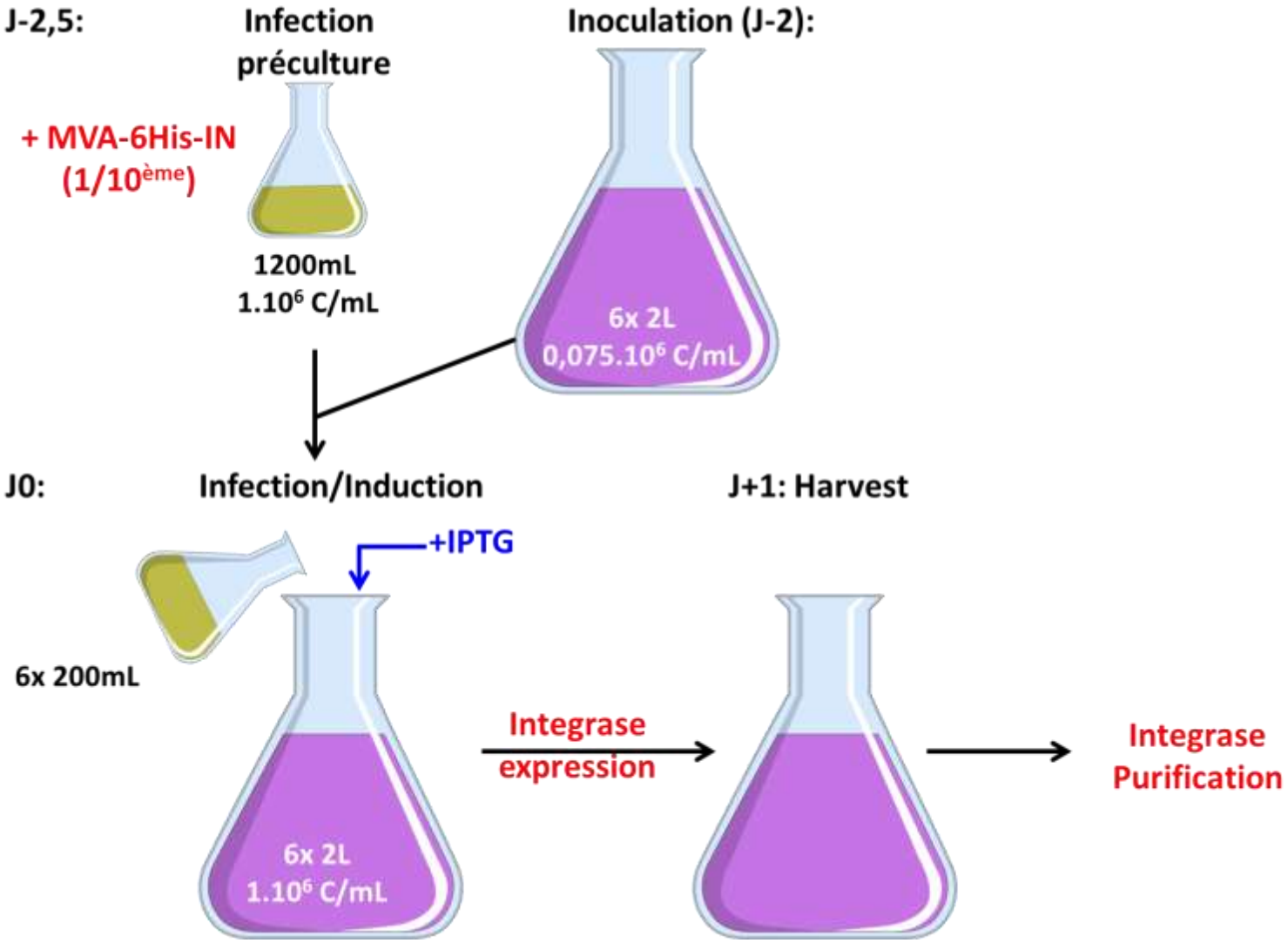
=> GPT integration selection

MVA Infection methods



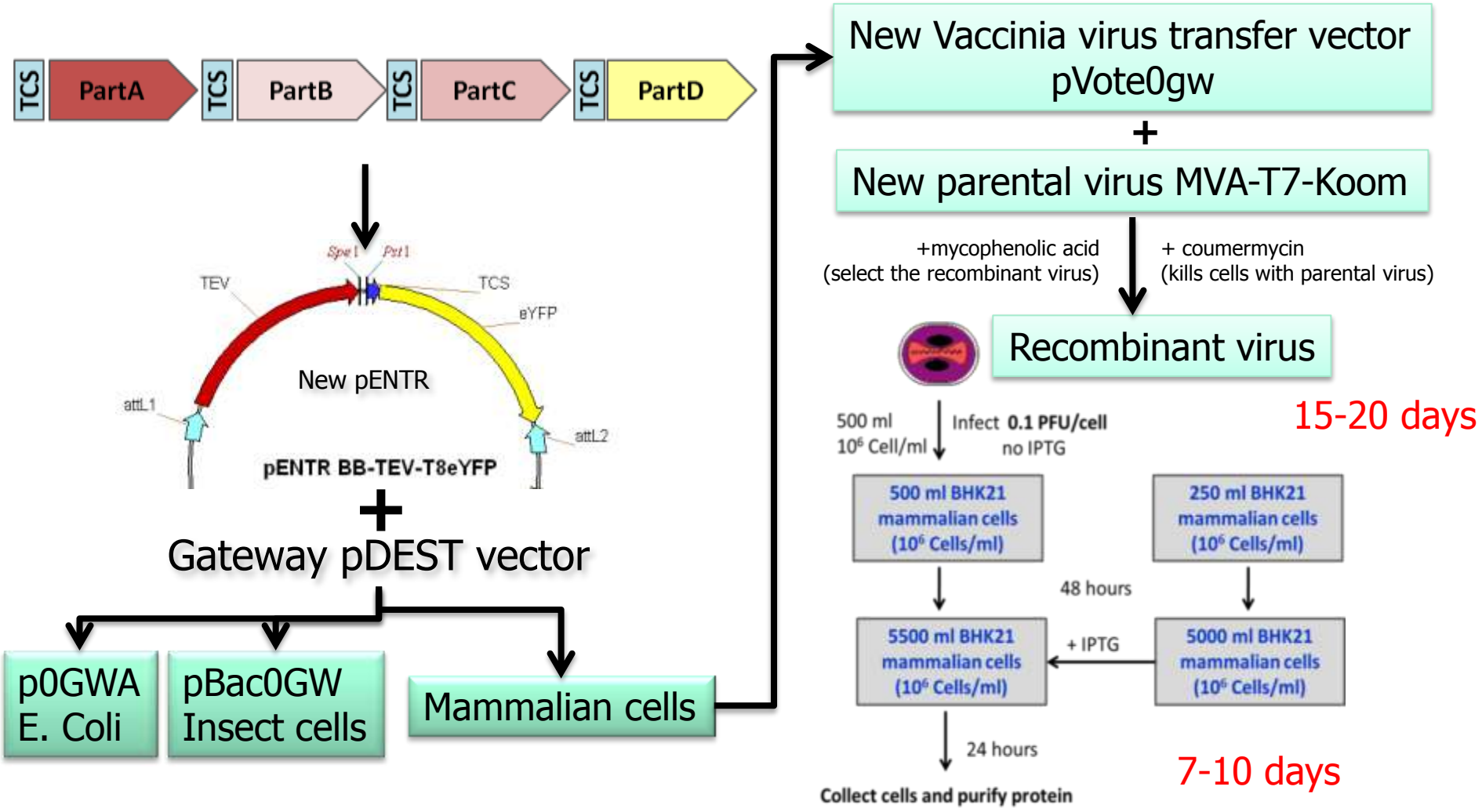


Large scale production of Integrase



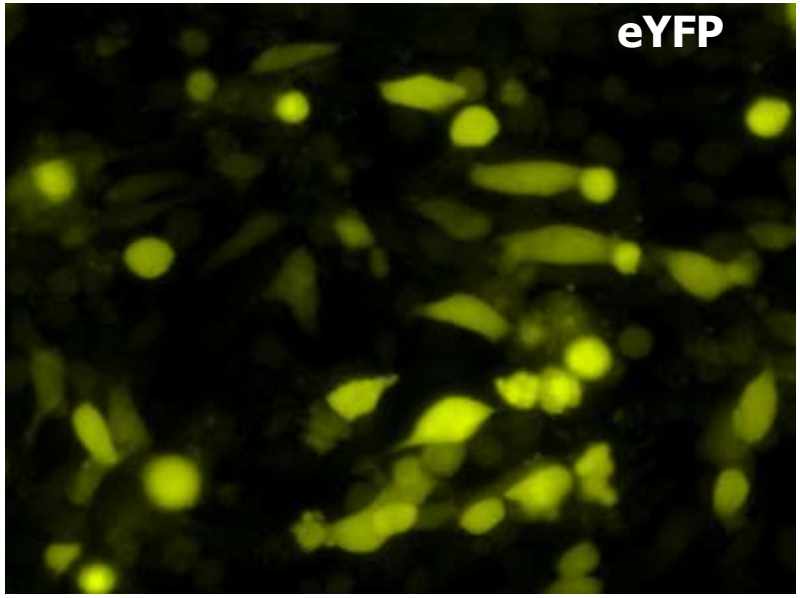
Mammalian cell protein complexes production

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

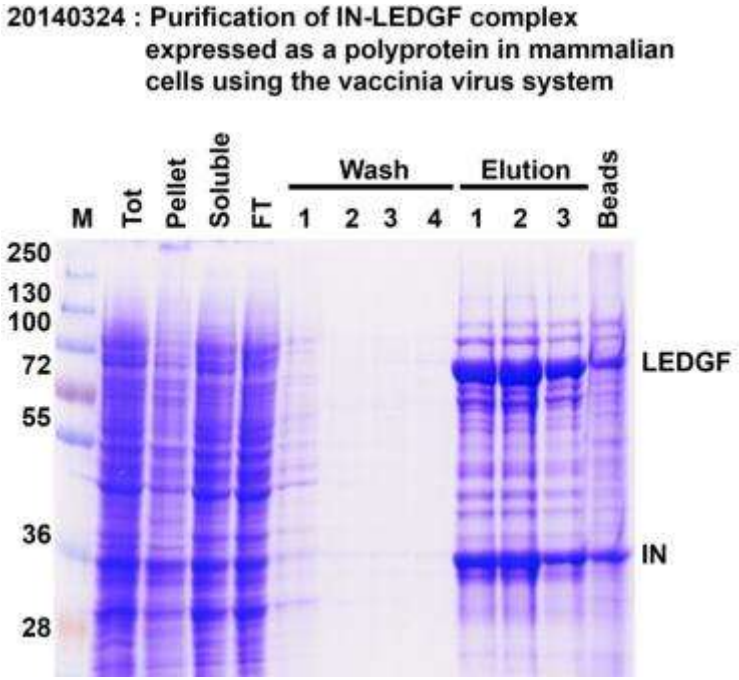


IN/LEDGF complex production in mammalian cells: polycistronic vector

Poly-protein expression validation
eYFP expression



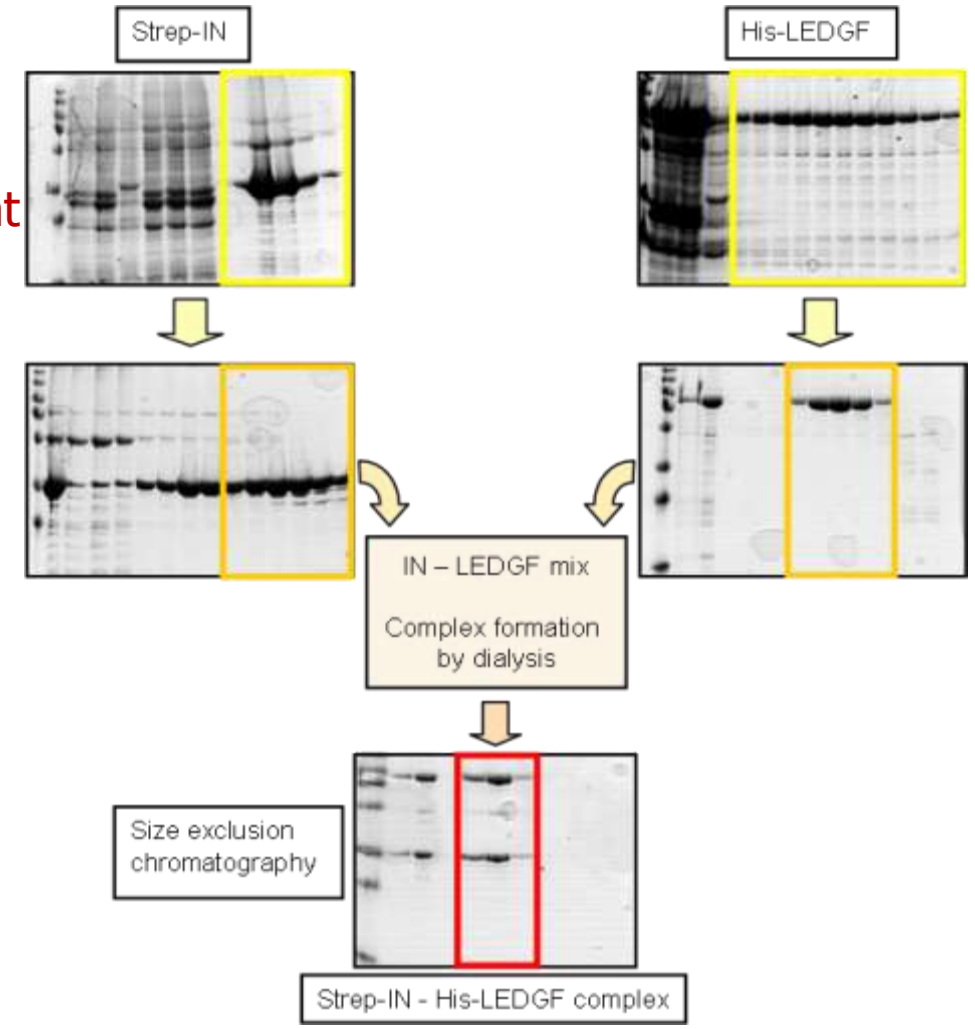
IN-LEDGF complex purification



Comparison of HIV-1 IN produced in E. Coli, Insect and mammalian cells

IN/LEDGF : complex formation and purification in Insect cells

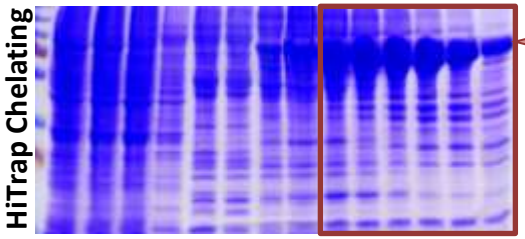
Expression vector: Baculovirus



IN/LEDGF complex production in mammalian cells: mono expression

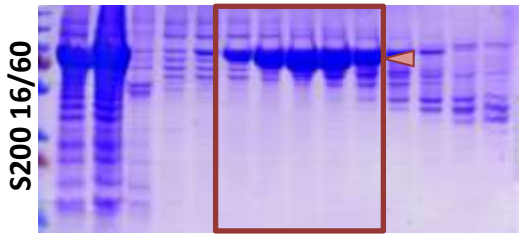
His-LEDGF
Mammalian suspension cells
Lysis, sonication, centrifugation
→ Soluble extract

Affinity chromatography



5-10mg / 10^9 cells (1L)

Gel filtration



LEDGF

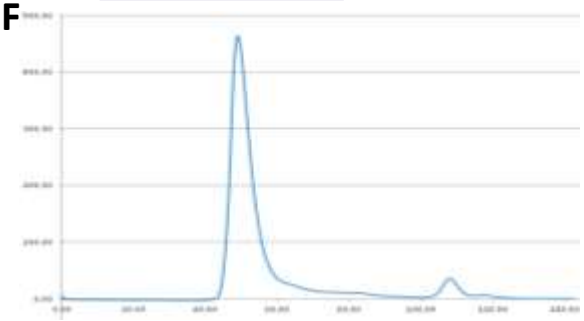
Mix
LEDGF + IN

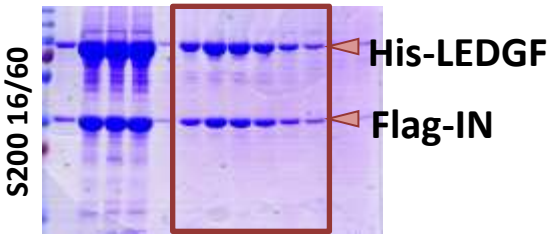
7mM CHAPS → 0mM

1M NaCl → 0,25M

Complex formation by dialysis

Gel filtration



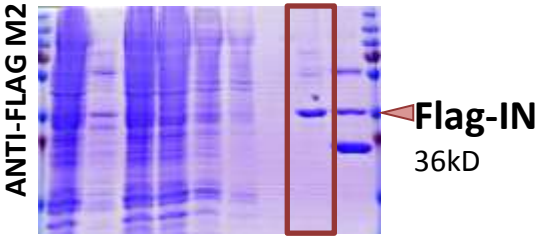


3mg/L of mammalian cells

IN-LEDGF complex

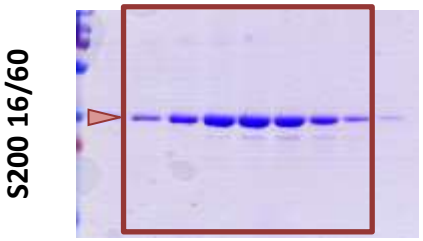
FLAG-IN
Mammalian suspension cells
Lysis, sonication, centrifugation
→ Soluble extract

Affinity chromatography



8mg / 10^9 cells (1L)

Gel filtration



IN

Comparison of HIV-1 IN produced in E. coli, insect and mammalian cells

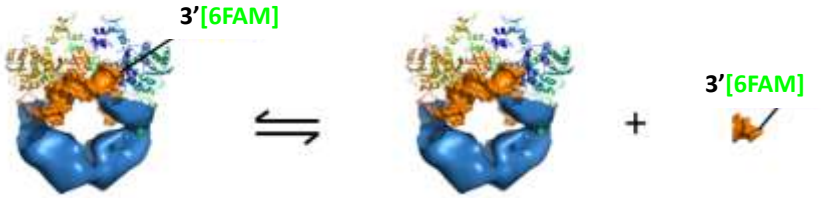
Solubility analysis

INTEGRASE	1M NaCl 7mM CHAPS	1M NaCl Ø CHAPS	0.5M NaCl Ø CHAPS
Ecoli	+	-	-
Insect cells	+	-	-
Mammalian cells	+	+	+

Solubility increase for IN produced in mammalian cells

Comparison of HIV-1 IN produced in E. coli, insect and mammalian cells

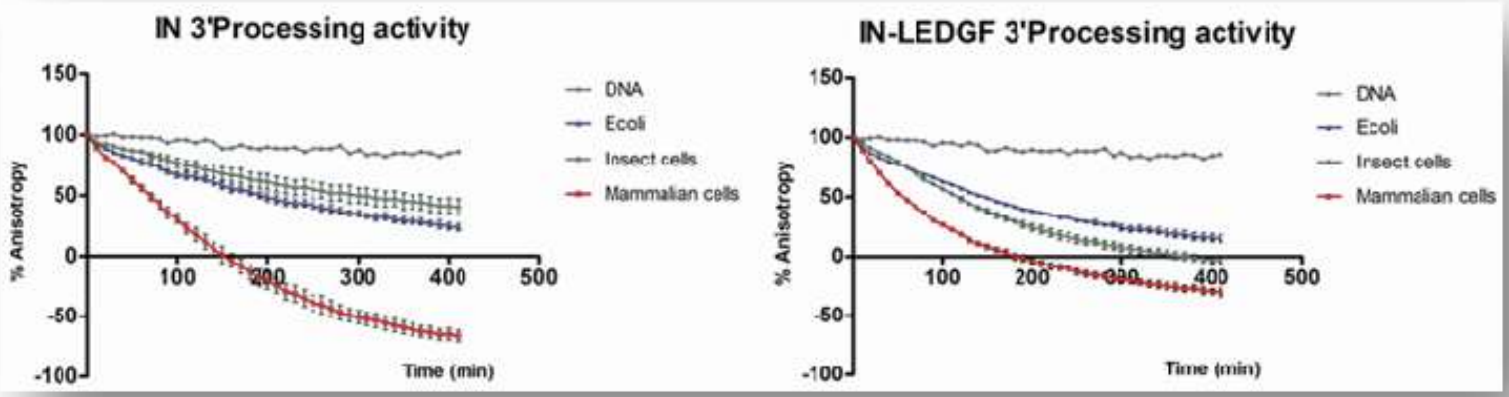
3′ processing IN activity by fluorescence anisotropy measurements



Slow rotation
High anisotropy

Fast rotation
Low anisotropy

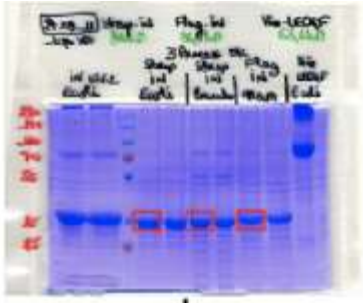
5′-GACTACGGTTCAAGTCAGCGTGTGGAAAATCTCTAGCAGT-3′ [6FAM]
3′-CTGATGCCAAGTTCAGTCGCACACCTTTTAGAGATCCTCA-5′



Increase of the 3′ processing activity for IN produced in mammalian cells

Comparison of HIV-1 IN produced in E. coli, insect and mammalian cells

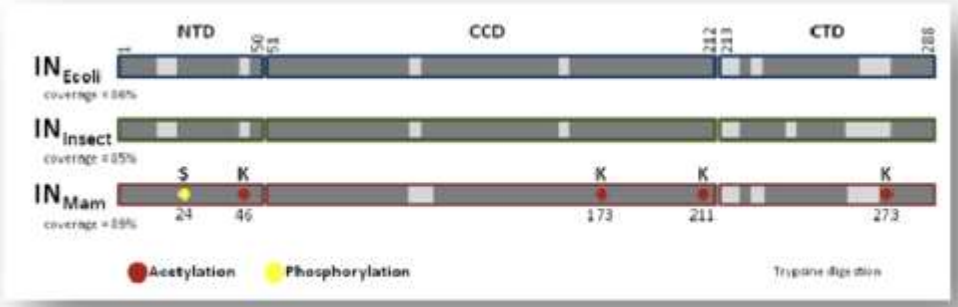
Mass spectrometry and sequence analysis

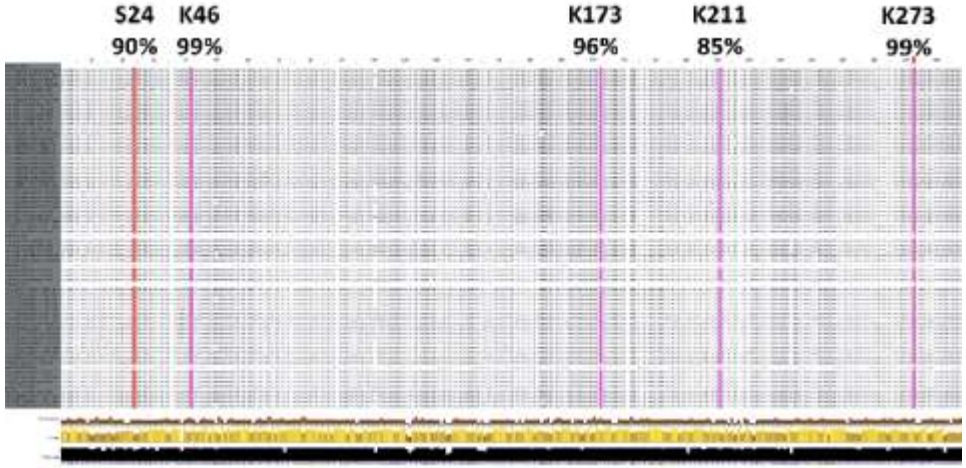


Band Excision

Reduction (DTT)
Alkylation (iodoacetamide)
Digestion (Trypsin, AspN)

LTQ XL ETD
(Thermo Fisher Scientific)
nanoLC-nanoESI-CID/MS-MS
nanoLC-nanoESI-ETD/MS-MS

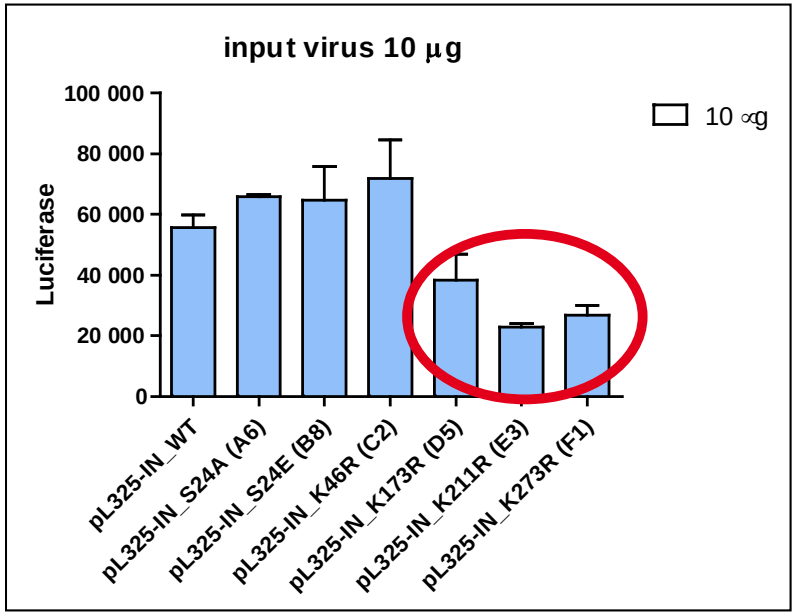


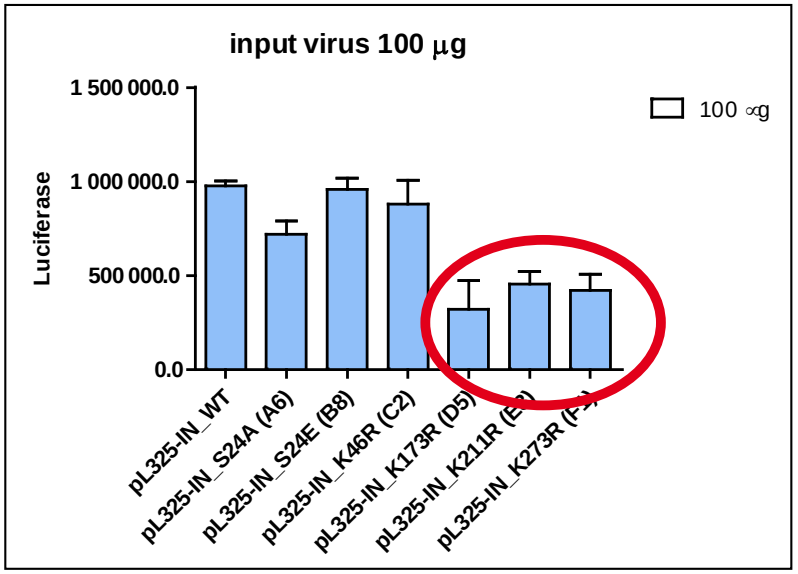


Phosphorylation S24, Acetylation K46, K173, K211, K273

Comparison of HIV-1 IN produced in E. coli, insect and mammalian cells

Effect on viral replication (THP1 cells)

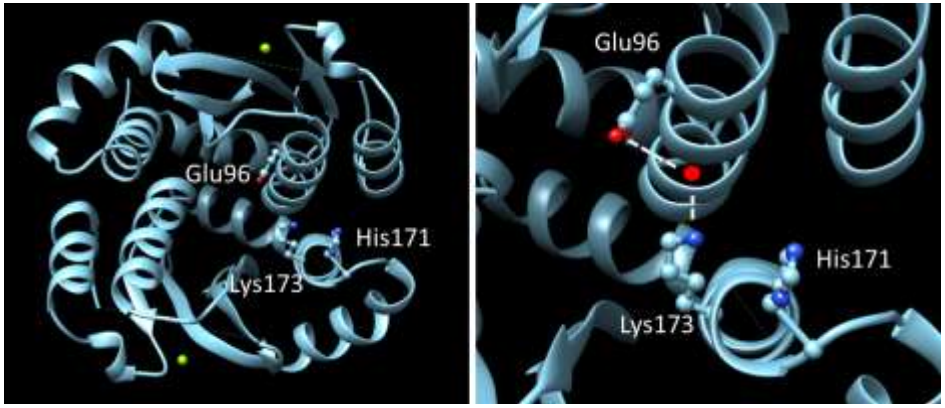




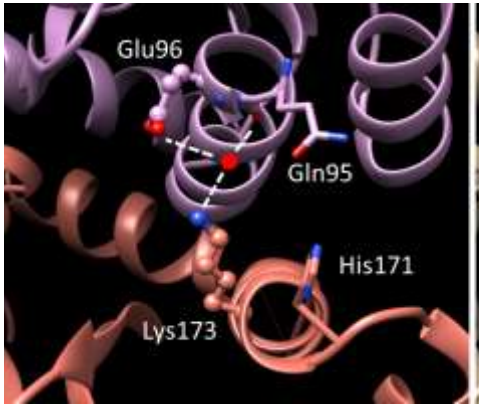
Lys --> Arg mutations: effect on viral replication for K173, K211 and K273

Structural analysis of HIV-1 IN K173 acetylation

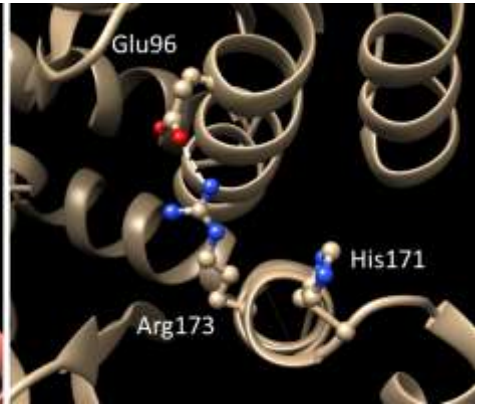
HIV-1 IN CCD X-ray structure:
K173 involved in the dimer stabilization



HIV-1 IN CCD + IN/LEDGF
interaction inhibitor X-ray Structure

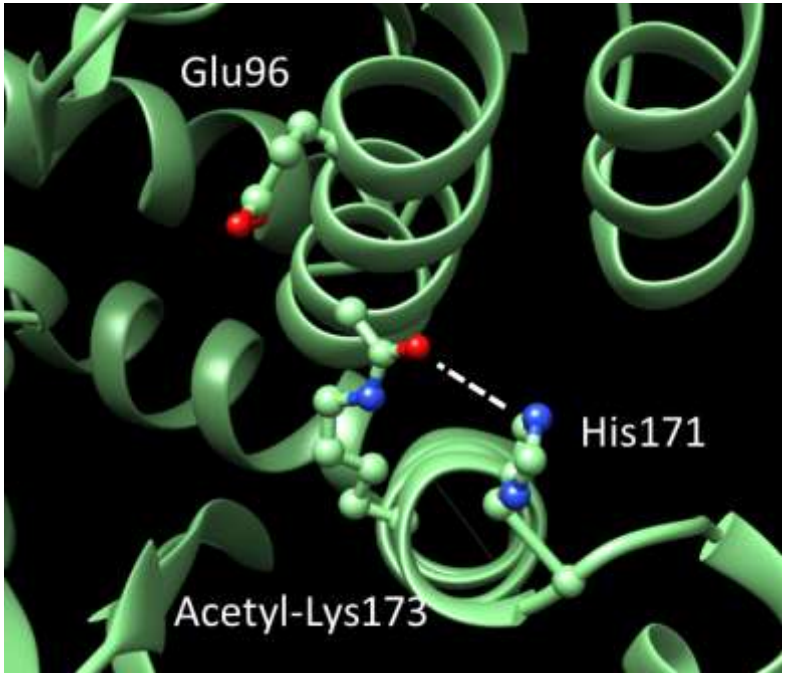


HIV-1 IN CCD K173R
X-ray Structure



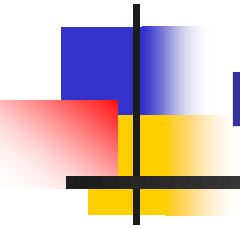
Stabilization of the HIV-1 IN dimer

HIV-1 IN CCD K173-NAcetyl
Model



Destabilization of the HIV-1 IN dimer

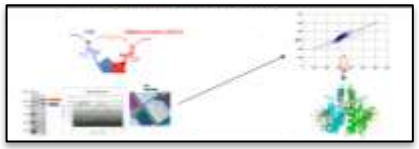
K173 acetylation weaken the IN - IN interaction in the dimeric interface resulting in increased flexibility and structural adaptability

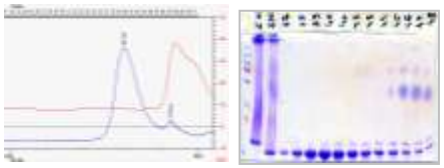


DRUG-DESIGN

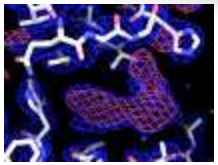
Protein – Protein interaction and allosteric inhibitors

HTP IN – LEDGF interaction inhibitors screening (HTRF)





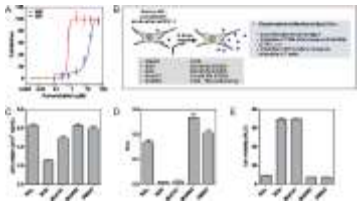
IN CCD Production and crystallization



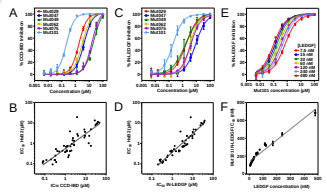
HTP inhibitors crystal soaking, data collection, structure

~100 structures solved

Structure based drug design

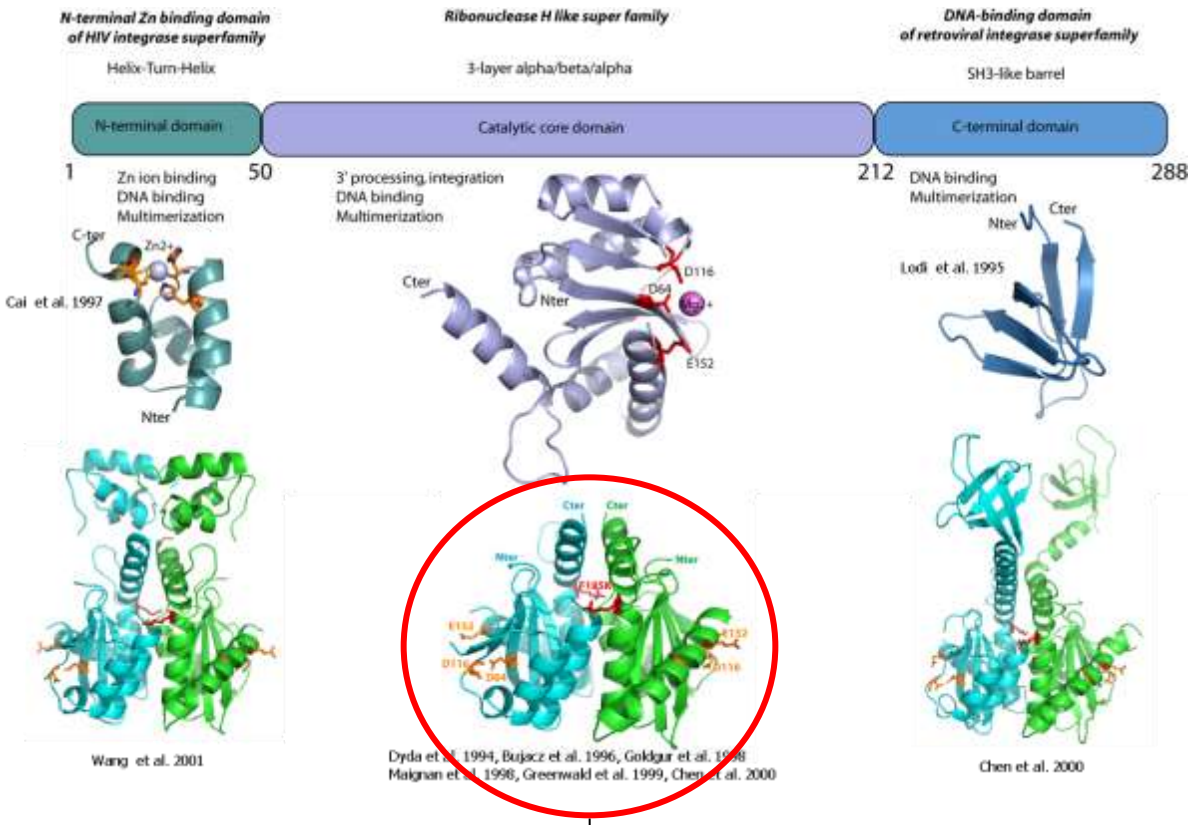


in-vitro functional assay



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

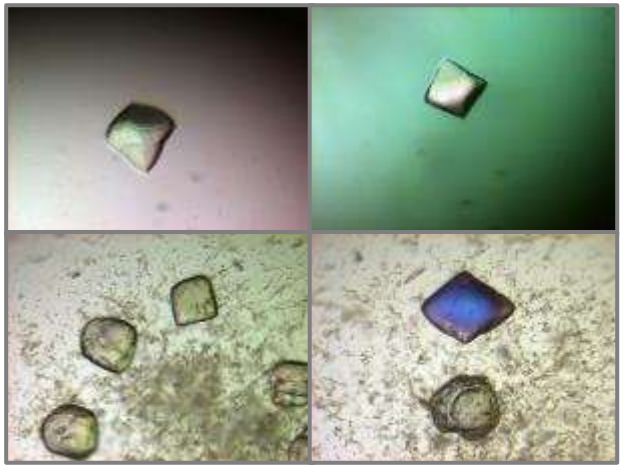
↓

Size exclusion chromatography
Superdex 200

↓

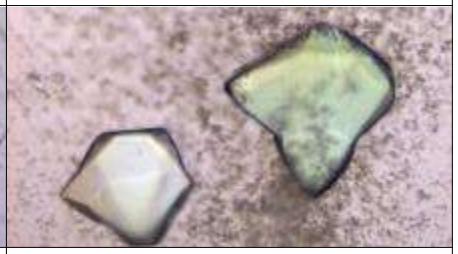
Crystallization
1.26 M AmSO4
100 mM NaCacodylate pH 6.5
Hanging drops
@ 22-23 °C

22-23°C
Diffraction
~2Å

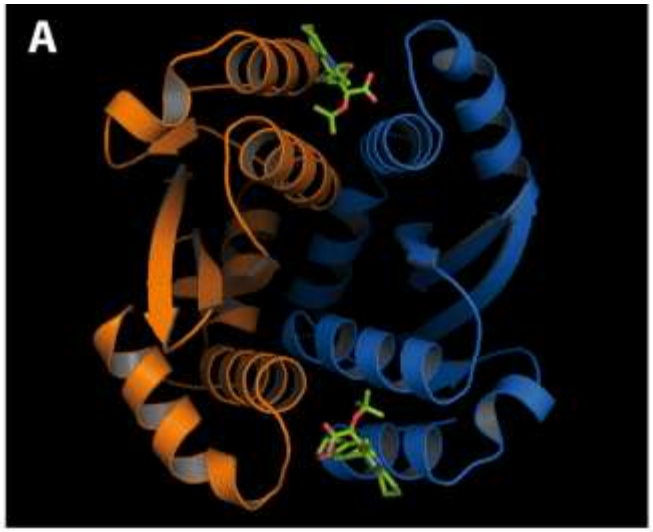


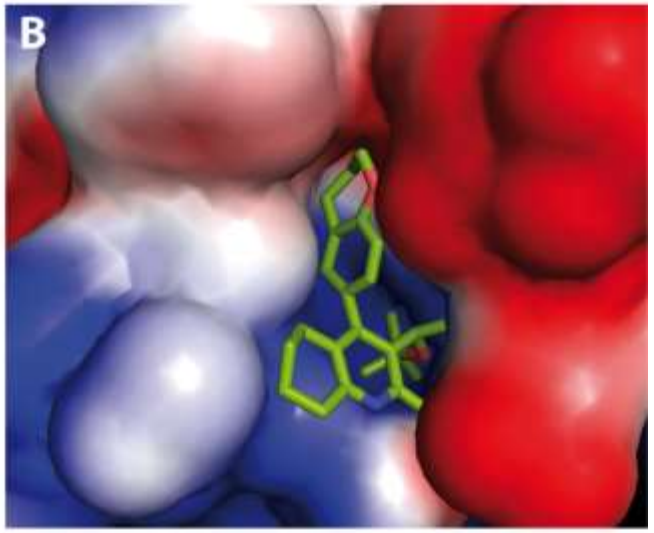
24-25°C
No diffraction !

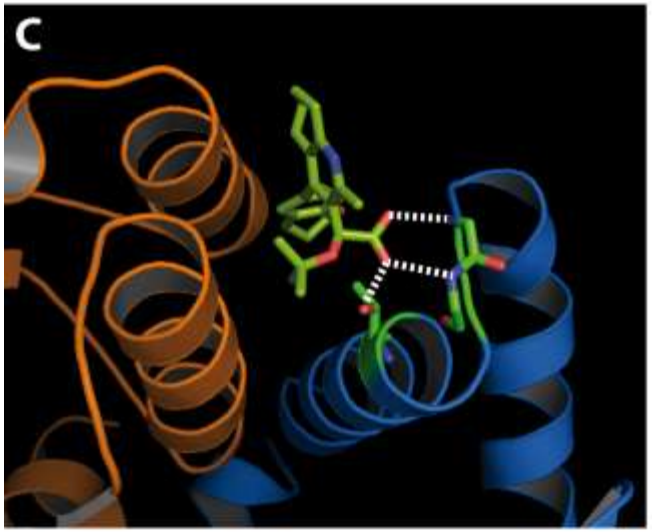
HIV Integrase 50 - 212 (F185K), Crystallization

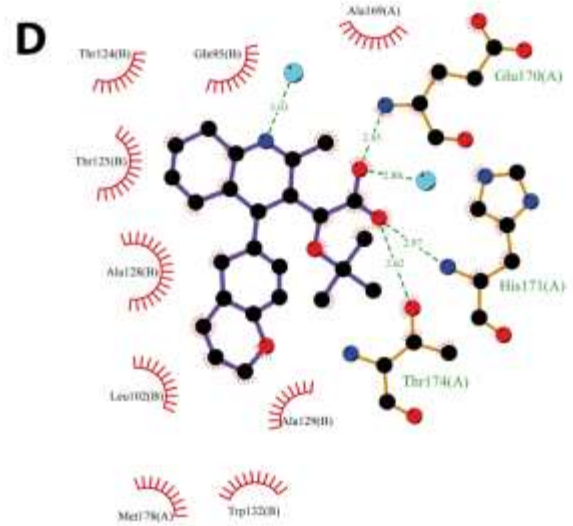
Crystal name	P1_B2_L11	P1_D4_L12	P1_C6_L13
Crystal source	Plate1_B2	Plate1_D4	Plate1_C6
Crystal image			
Protein concentration	3.1 mg/ml	3.1 mg/ml	3.1 mg/ml
Reservoir composition	1.26 M AS	1.26 M AS	1.50 M AS
Initial drop composition	2µl prot + 2µl res	2µl prot + 2µl res	2µl prot + 2µl res
Ligand soaking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

HIV Integrase CCD – ligand structure

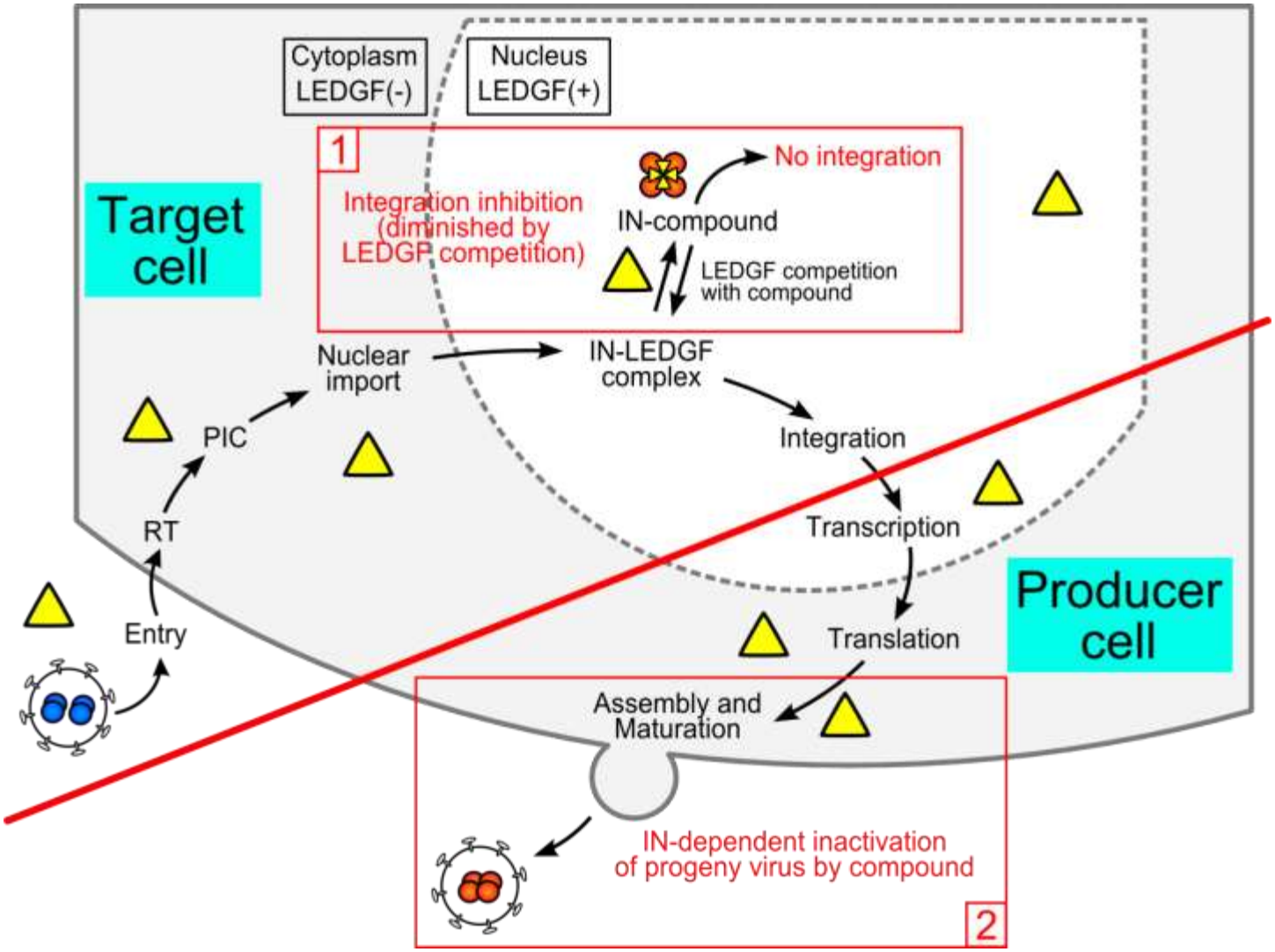




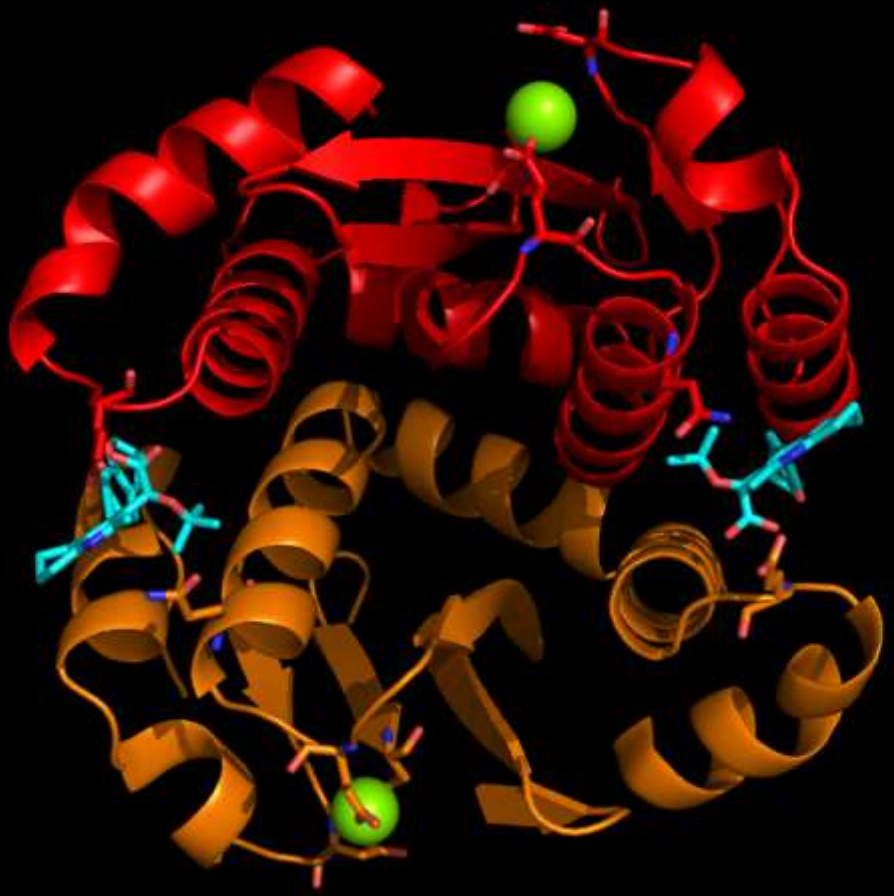




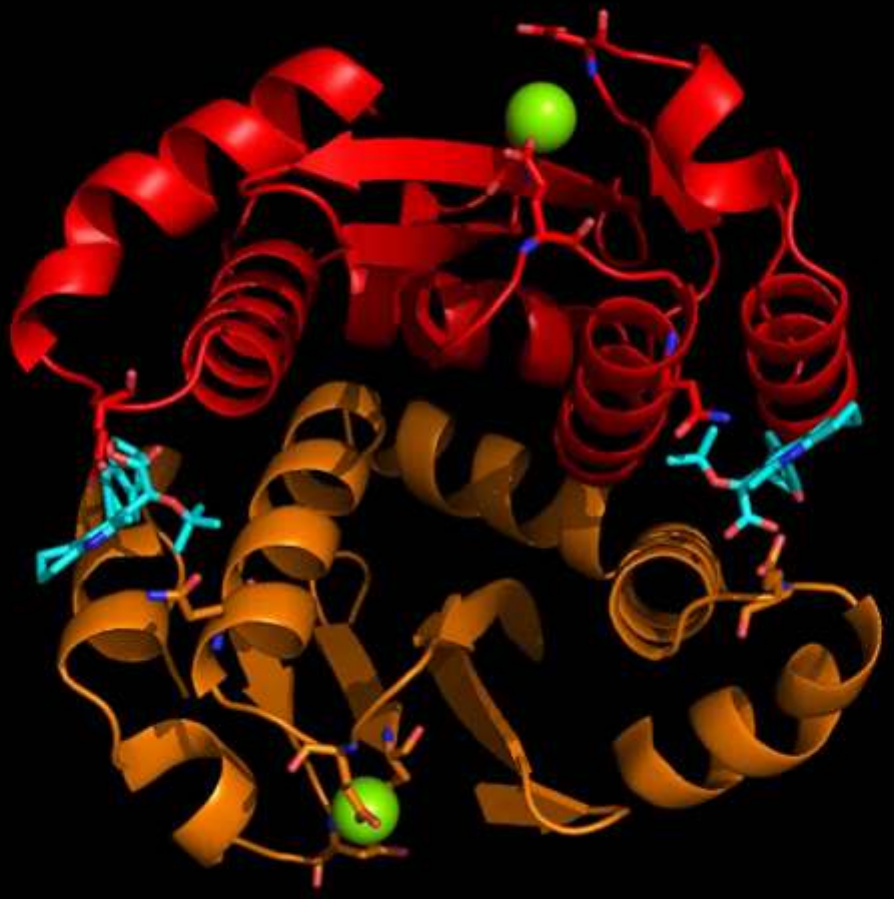
IN – LEDGF interaction and IN allosteric inhibitors



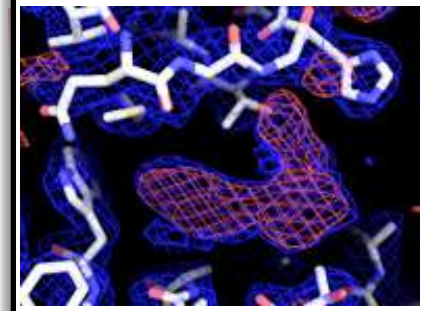
Le Rouzic et al., (2013), Retrovirology, 10, 144



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



INLAIs inhibitors



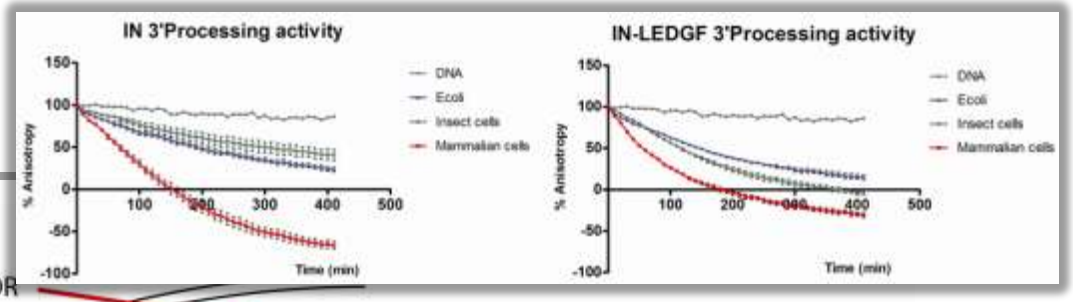
~100 structures solved

Allosteric inhibitors. The allosteric effects are mediated by structural changes in the active site, dimeric interface, accessible surface.

Mammalian cells production: increase integrase stability, solubility and 3'processing activity. Presence of acetylation and phosphorylation

LEDGF stabilizes an IN tetramer and increases its integration and 3'processing activity

INI1 prevents non specific aggregation and auto integration on the way to nucleosomes in the nucleus



PIC Pre-Integration Complex

PIC

Integrase VBP1 TRN-SR2

Integrase LEDGF TRN-SR2

PIC

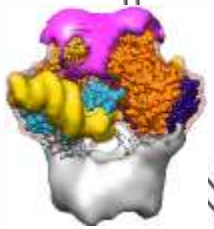
NUCLEAR IMPORT

Integrase LEDGF

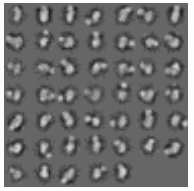
3' PROCESSING

Integrase LEDGF DNA

INTEGRATION



Integrase LEDGF INI1



Integrase LEDGF Nucleosome

Acknowledgments

Sylvia Eiler
Nicolas Levy
Julien Batisse
Oyindamola Oladosu
Karine Pradeau
Robert Drillien
Valérie Lamour
Claire Bedez
Arnaud Vanden Broeck

Fabrice Michel
Benoit Maillot
Aurélie schaetzel

BTS & IUT trainee
-Betty ORY
-Loïc Duffet
-Kevin Letscher

Patrick Schultz, IGBMC
Bruno Kieffer, IGBMC
Stéphane Emiliani, Cochin institute, Paris
Olivier Delelis, ENS, Cachan
Yves Mely, University of Strasbourg
Richard benarous, Mutabilis, Paris
Sarah Sanglier, University of Strasbourg
Alexis Nazabal, CovalX, Zurich
Vincent Parissi, University of Bordeaux 2
Patrice Gouet, IBCP, Lyon
Marc Lavigne, Pasteur Institute, Paris
Matteo Negroni, IBMC, Strasbourg
Eric Ennifar, IBMC, Strasbourg
Jean Christophe Paillart, IBMC, Strasbourg
Roland Marquet, IBMC, Strasbourg

Vincent Olieric, SLS, Villigen

Members of Structural Biology and Genomics platform IGBMC, Illkirch

Members of the IGBMC Mass Spectrometry Facility

Members of the imaging platform

Members of the IGBMC's common services Cell, baculovirus and cloning facilities

Thank you !