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, STEROID NUCLEAR RECEPTORS
STRUCTURE, FUNCTION AND DYNAMICS

Marc Ruff

Integrated structural biology department
Chromatin stability and DNA mobility team

IGBMC, Illkirch, France

ruff@igbmc.fr

INTRINSICALLY DISORDERED PROTEINS

Intrinsically Disordered Proteins (IDP)

Definitions: (Keith Dunker - Indiana University)

Many proteins contain regions that lack specific 3-D structure; indeed some proteins lack specific 3-D structure in their entireties under physiological conditions and yet carry out biological function. Such proteins and regions have been called natively unfolded, intrinsically unstructured, naturally disordered, …

Disorder increases in example proteomes in the order of multicellular eukaryotes > single cellular eukaryotes > archaea > prokaryotes, with significant associations of disorder with signaling, regulation, and posttranslational modification.

Bacteria: 4.5%, Archea: 6.6%, Eukaryote: 25%

Intrinsically disordered proteins

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 occupy a continuum of "conformational space", ranging from highly unstructured, through molten globule, to local disorder within an otherwise folded domain.

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

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)**. Because IDRs generally lack bulky hydrophobic amino acids, they are unable to form the well-organized hydrophobic core that makes up a structured domain and hence their functionality arises in a different manner as compared to the classical structure–function view of globular, structured proteins.

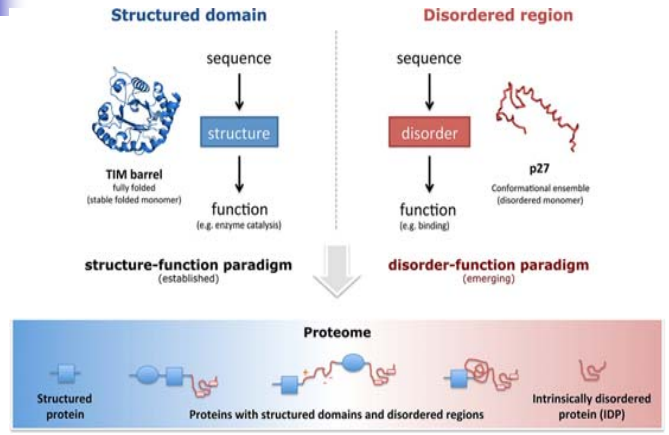
In this framework, **protein sequences in a genome can be viewed as modular because they are made up of combinations of structured and disordered regions**.

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.

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 regions



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.

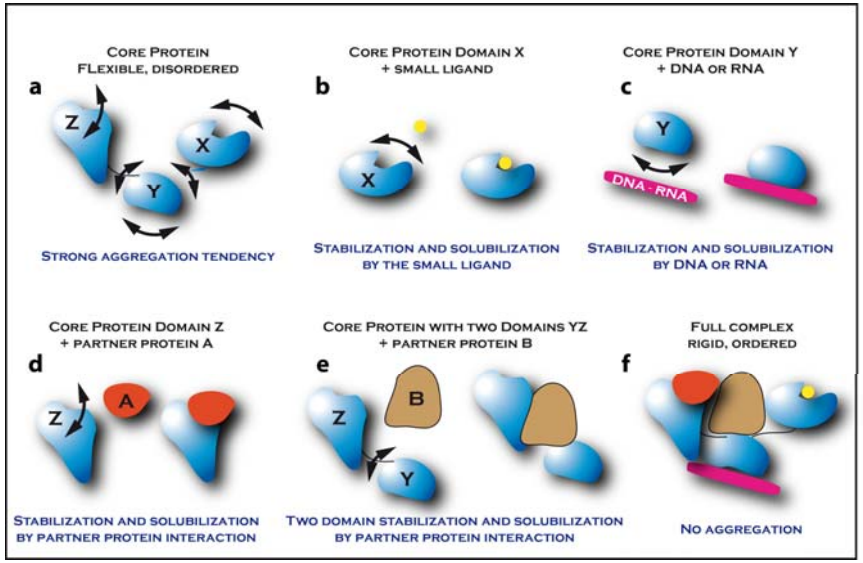
Robin van der Lee; Marija Buljan; Benjamin Lang; Robert J. Weatheritt; Gary W. Daughdrill; A. Keith Dunker; Monika Fuxreiter; Julian Gough; Joerg Gsponer; David T. Jones; Philip M. Kim; Richard W. Kriwacki; Christopher J. Oldfield; Rohit V. Pappu; Peter Tompa; Vladimir N. Uversky; Peter E. Wright; M.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

FROM UNSTABLE PROTEIN TO STABLE COMPLEXES

THE HIV-1 PRE-INTEGRATION COMPLEX, STEROID NUCLEAR RECEPTORS

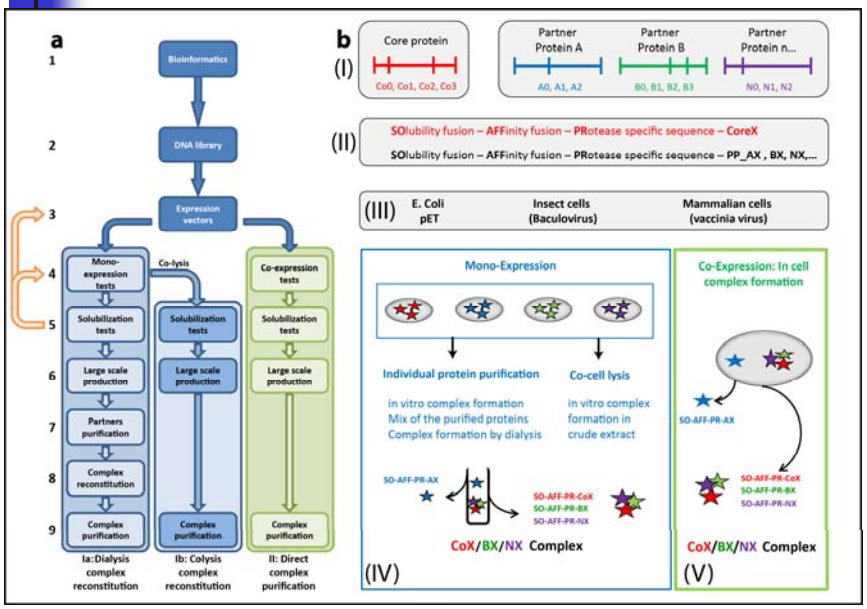
STRUCTURE, FUNCTION AND DYNAMICS

From unstable protein to stable complexes



Levy N, Eiler S, Pradeau-Aubreton K, Maillot B, Stricher F, Ruff M (2016). Nature Communications 7: 10932

From unstable protein to stable complexes

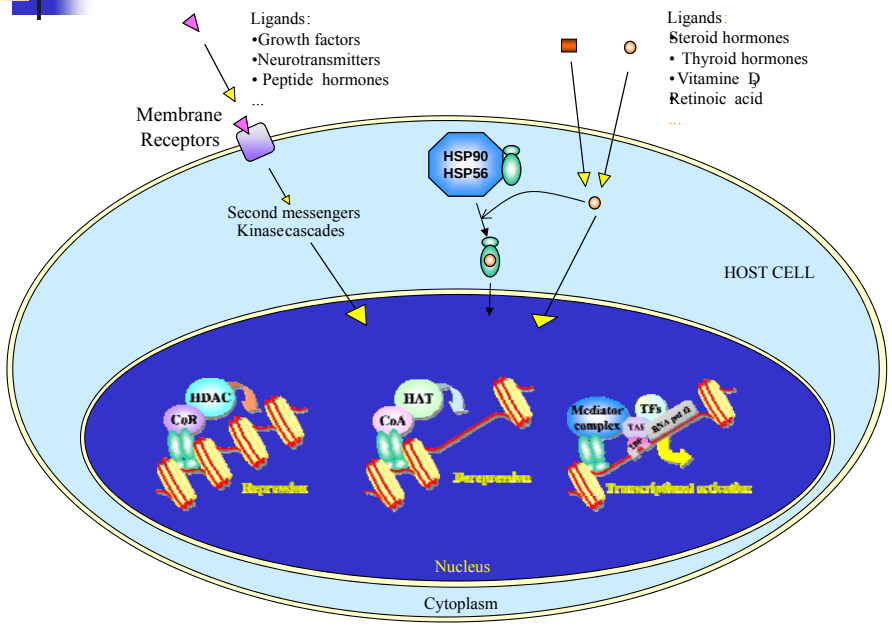


Levy N, Eiler S, Pradeau-Aubreton K, Maillot B, Stricher F, Ruff M (2016). Nature Communications 7: 10932

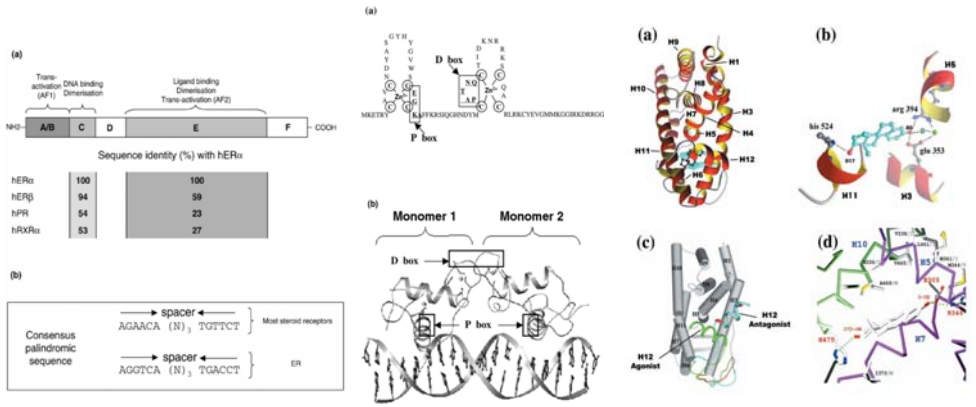
Example 1

Steroid Hormones Nuclear Receptors

Steroid hormone nuclear receptors

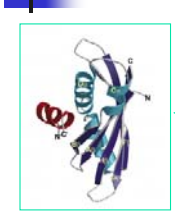


Steroid nuclear receptors organized into domains with IDRs

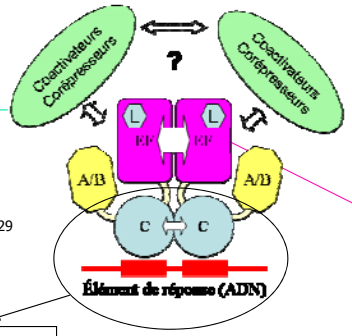


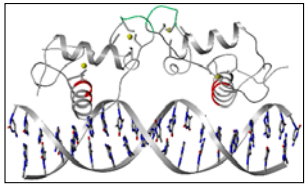
Marc Ruff, Monique Gangloff, Jean Marie Wurtz and Dino Moras. Estrogen receptor transcription and transactivation. Structure–function relationship in DNA- and ligand-binding domains of estrogen receptors. *Breast Cancer Res* 2000, **2**:353–359

Steroid hormone nuclear receptors

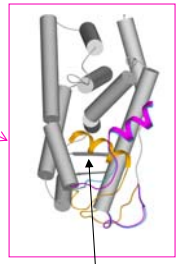


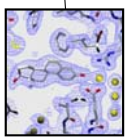
Razeto et al, (2004), J. Mol. Biol. 336, 319-329





Luisi et al, (1991). Nature *352*, 497-505.

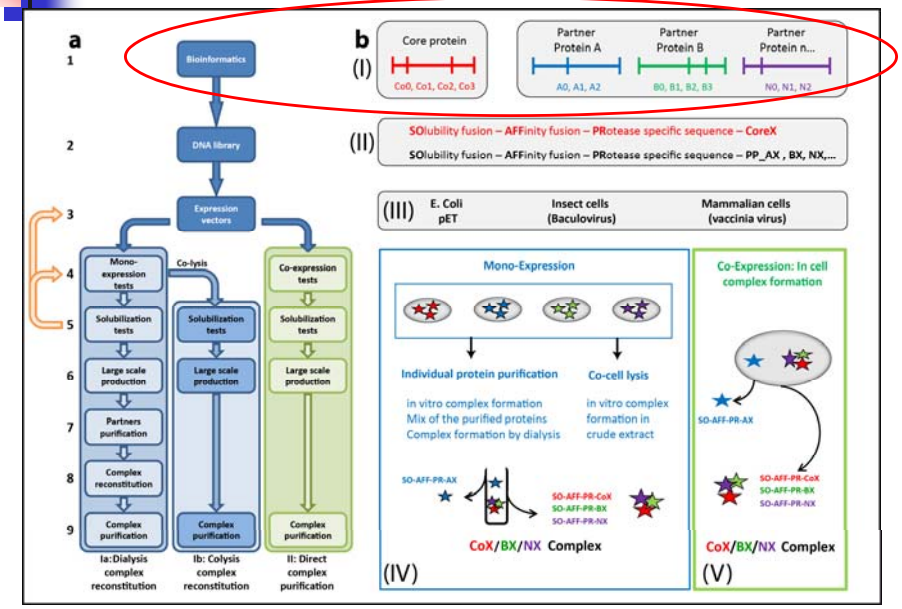




Gangloff et al, (2001), JBC, 276, 15059-15065.

→ Interaction of SNRs with TIF2 coactivator

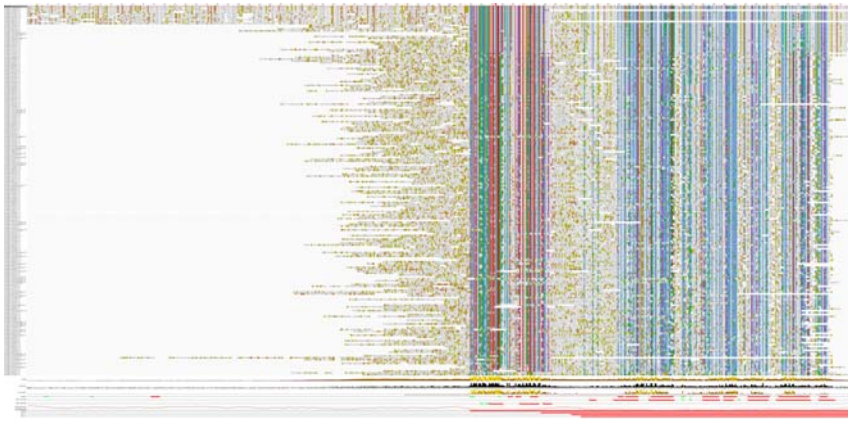
From unstable protein to stable complexes



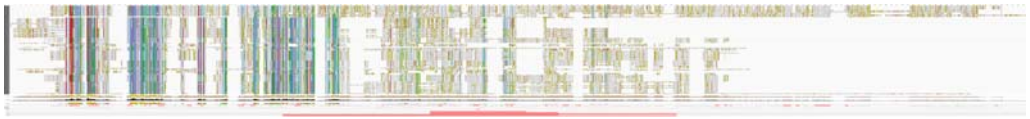
Levy N, Eiler S, Pradeau-Aubreton K, Maillot B, Stricher F, Ruff M (2016). Nature Communications 7: 10932

Bioinformatics analysis of the sequences: design of constructs

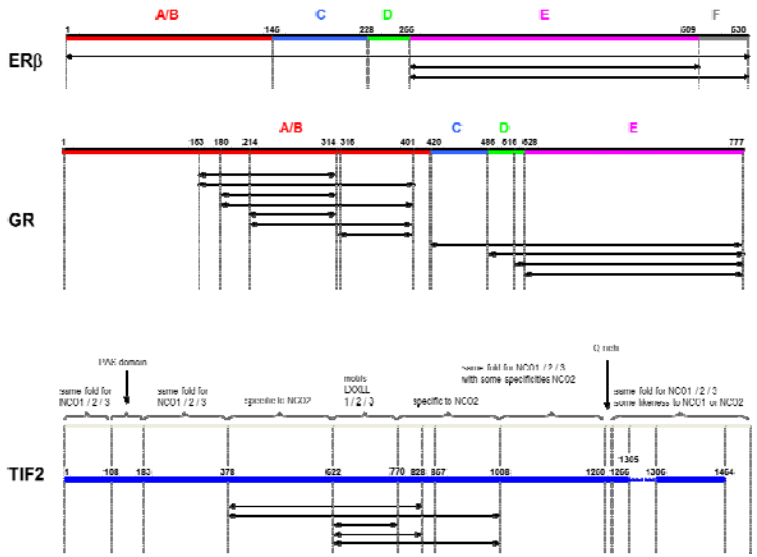
GR multi - alignment



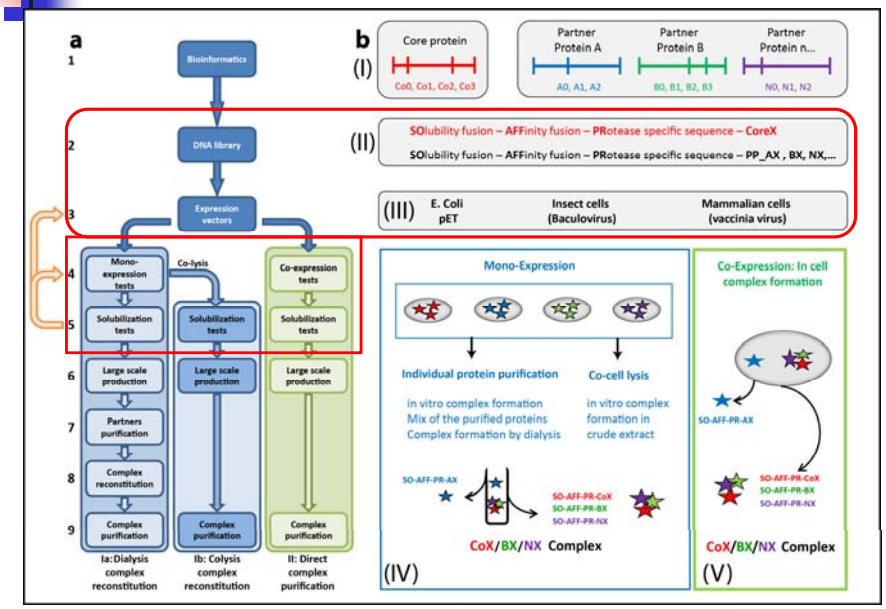
TIF2 multi - alignment



Bioinformatics analysis of the sequences of ERß GR and TIF2 : design of constructs



From unstable protein to stable complexes



Levy N, Eiler S, Pradeau-Aubreton K, Maillot B, Stricher F, Ruff M (2016). Nature Communications 7: 10932

ERβ Expression and solubilisation tests

E. Coli expression

ERβ	HIS		GST		HIS-TRX		HIS-MBP	
	LBS/18°C		LBS/18°C		LBS/18°C		LBS/18°C	
	T	S	T	S	T	S	T	S
FL 1 – 530	++	-	++	-	++	-	++	-
EF 255 – 530	nd	nd	nd	nd	+++	++	nd	nd

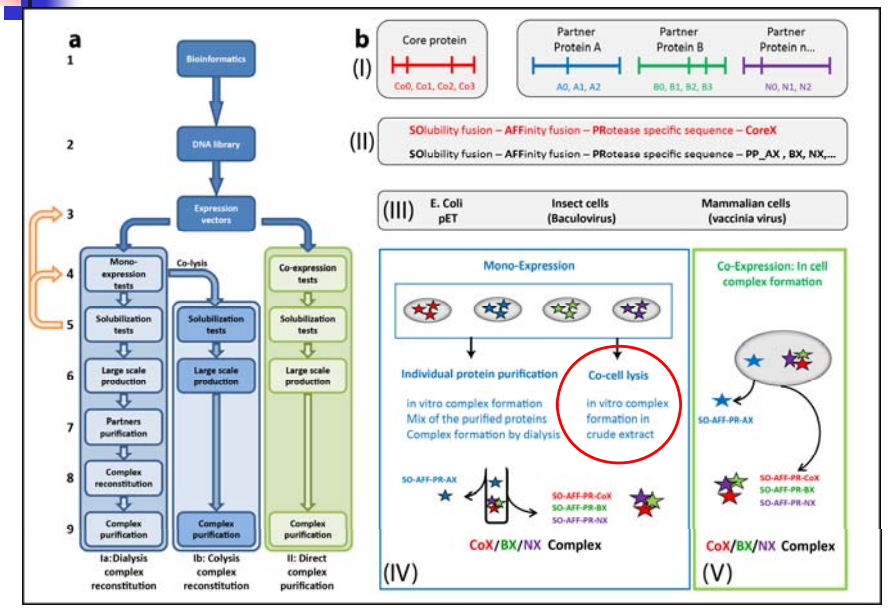
The optimized buffer for protein solubilization is 300 mM KCl, 100 mM Phosphate buffer pH=7.5, 10 mM βmercaptoethanol.

TIF2 Expression and solubilisation tests

TIF2	HIS		GST		MBP	
	LBS/18°C		LBS/18°C		LBS/18°C	
	T	S	T	S	T	S
378-828	+	+/-	+	+	+++	++
378-1008	nd	nd	nd	nd	+++	++
623-772	+	+	+	+	+++	+++
623-828	+	+/-	+	+	+++	++
623-1008	+	+/-	+	+	+++	++

Solubilization buffer: 50mM Phosphate Na/K pH 7.5, 50 mM NaCl

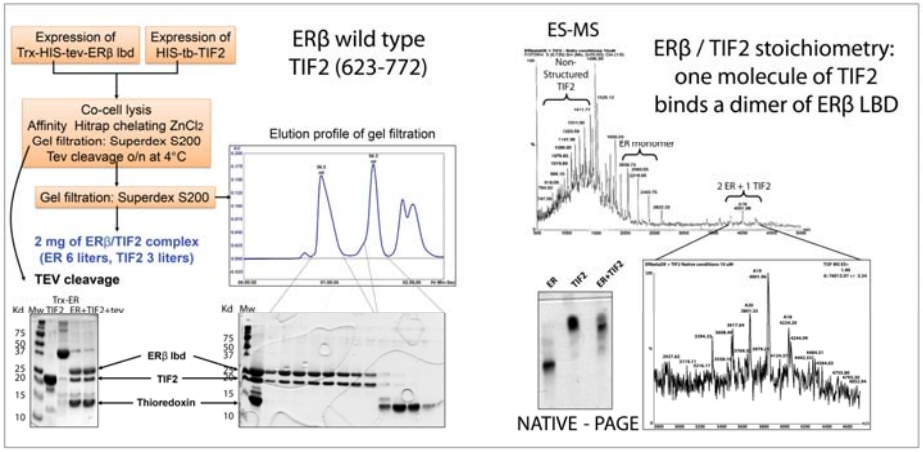
From unstable protein to stable complexes



Levy N, Eiler S, Pradeau-Aubreton K, Maillot B, Stricher F, Ruff M (2016). Nature Communications 7: 10932

ERβ / TIF2 complex reconstitution

Co-cell lysis



TIF2 binds ERβ with an induce folding mechanism

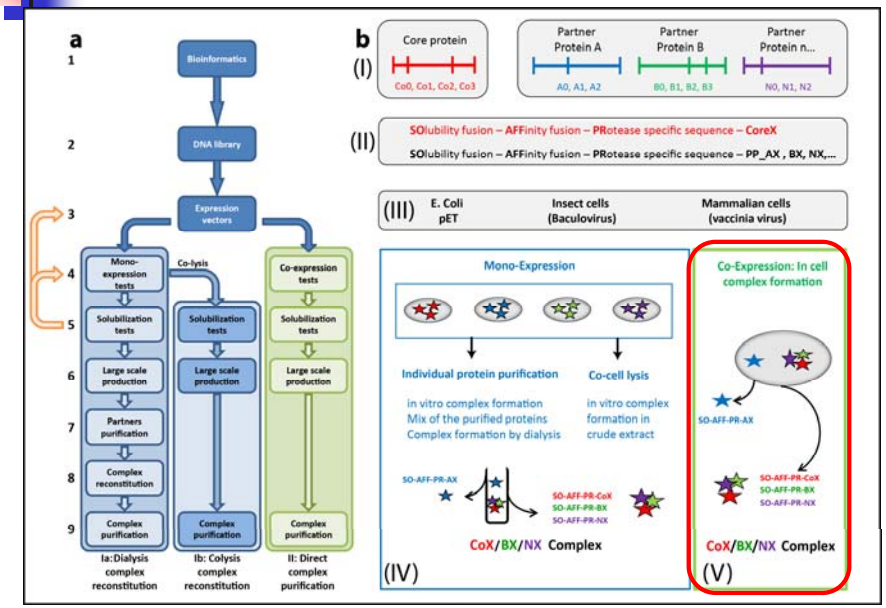
GR Expression and solubilisation tests

GR (*E. coli*)	HIS				FLAG				HIS-NUS			
	LBS/37°C		LBS/18°C		LBS/37°C		LBS/18°C		LBS/37°C		LBS/18°C	
	T	S	T	S	T	S	T	S	T	S	T	S
CDE 420 – 777	+/-	-	-	-	+	-	-	-	++	-	++	+
DE 486 – 777	++	-	-	-	++	-	+	-	++	-	++	++
E 516 – 777	+/-	-	+/-	-	++	-	+	+/-	++	-	++	+
E 524 – 777	-	-	-	-	+	-	++	-	++	-	++	+

GR (insect)	HIS		GST	
	T	S	T	S
CDE 420 – 777	+	-	++	+/-
DE 486 – 777	nd	nd	++	+/-
E 516 – 777	+	-	++	+/-
E 524 – 777	+	-	nd	nd

Solubilization buffer: 250 mM NaCl, 50 mM Tris pH7.5, 10 μM dexamethasone

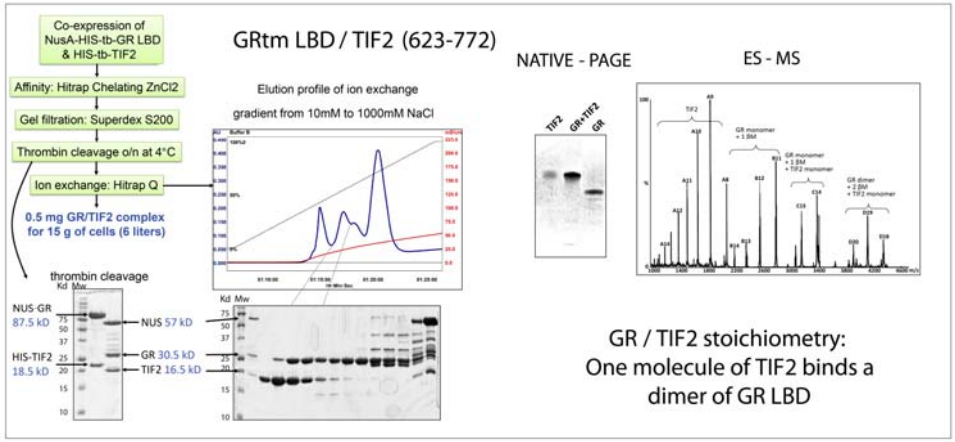
From unstable protein to stable complexes



Levy N, Eiler S, Pradeau-Aubreton K, Maillot B, Stricher F, Ruff M (2016). Nature Communications 7: 10932

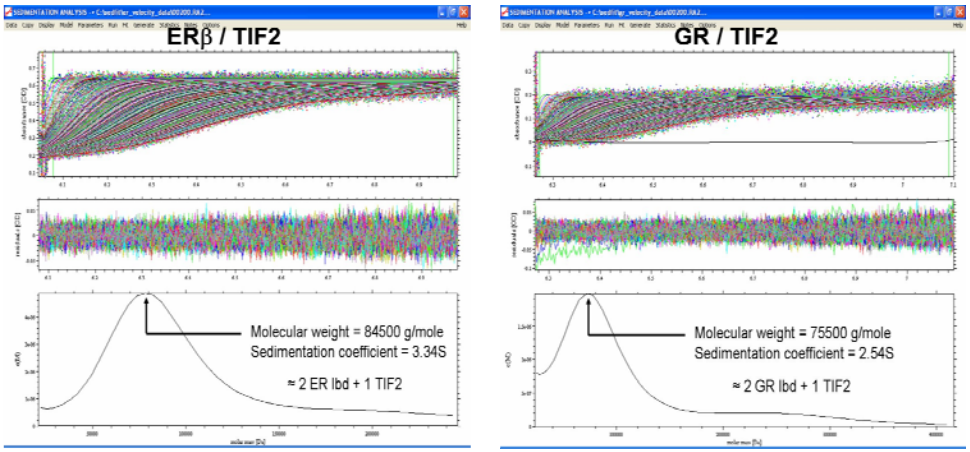
GRtm / TIF2 complex reconstitution

Co-expression

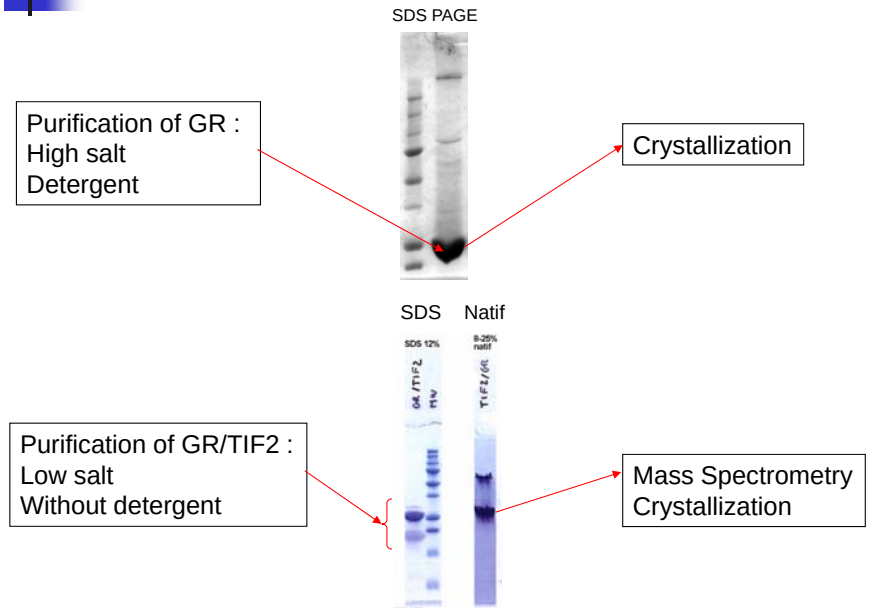


TIF2 binds GR with an induce folding mechanism

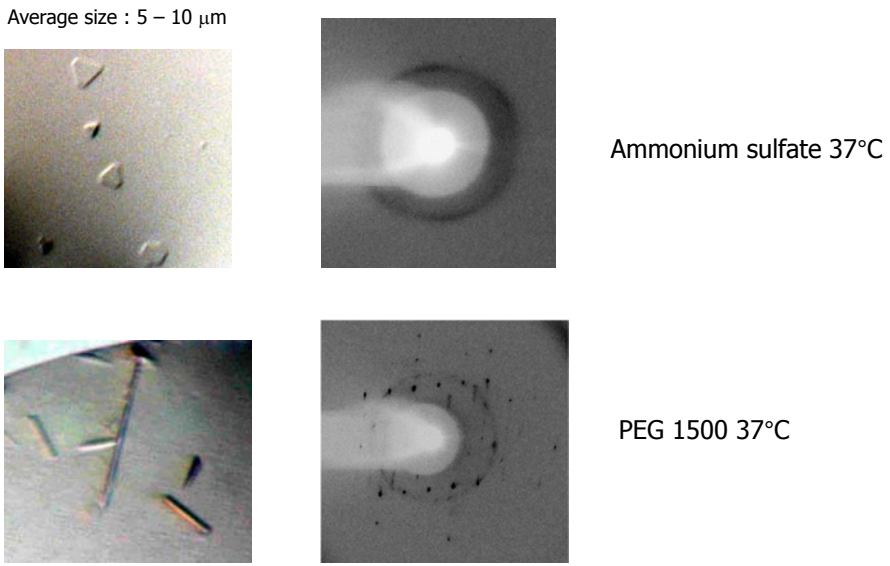
Analytical Ultracentrifugation analysis



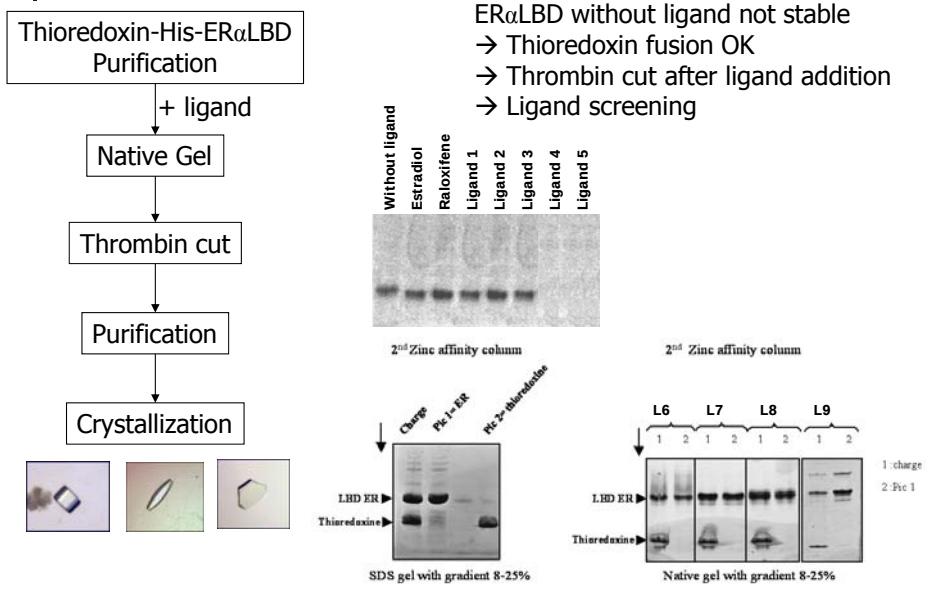
Purification of GR LBD and GR LBD/TIF2



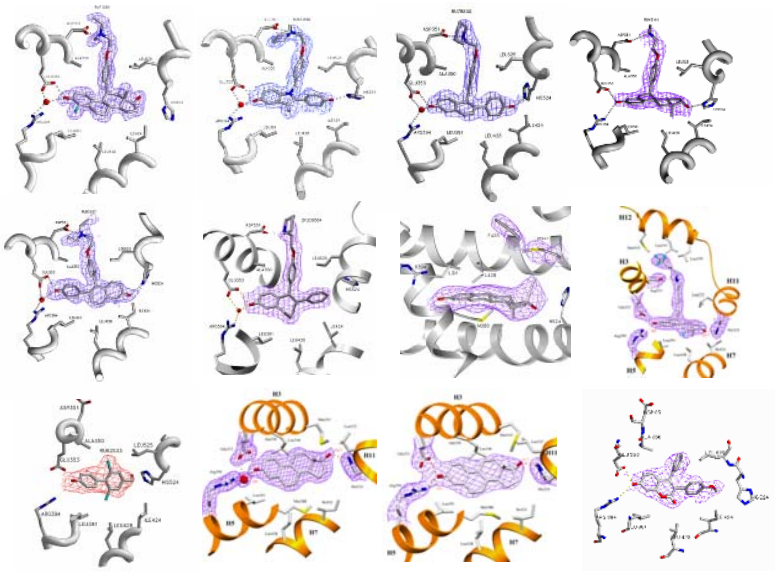
Crystallization of GR LBD + Dex



Drug design: structure with ligands (Ligand screening)



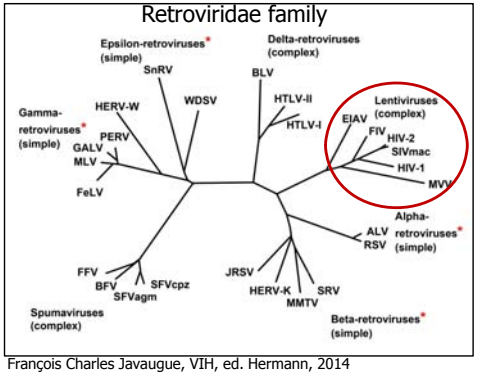
ER alpha LBD: ligand screening

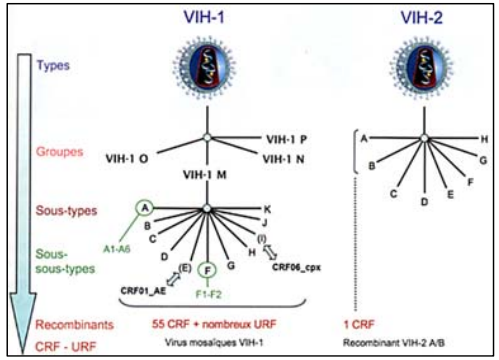


Example 2

HIV-1 Pre-integration complexes

HIV phylogeny





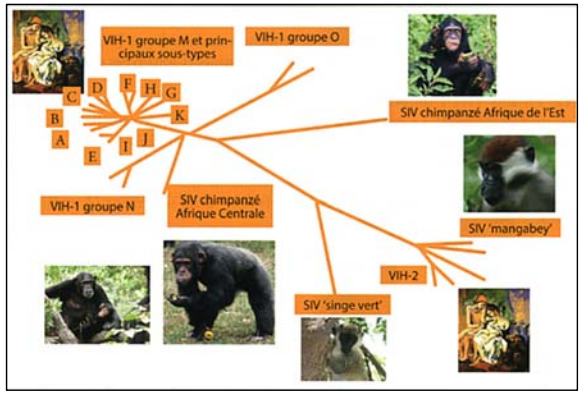
François Charles Javaugue, VIH, ed. Hermann, 2014

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)

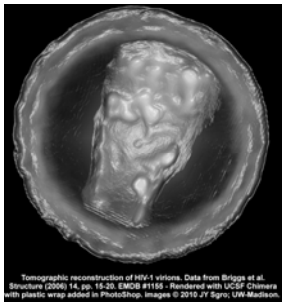
Origin of HIV pandemics

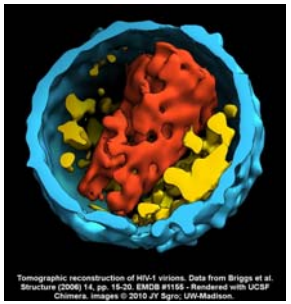


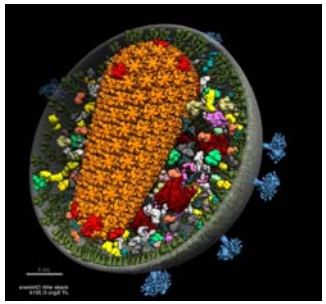
François Charles Javaugue, VIH, ed. Hermann, 2014

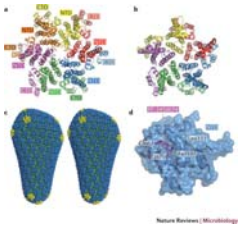
HIV emerged from SIV in late 19th / early 20th century through zoonosis between apes or monkeys and men:
HIV-1 group M : chimpanzee (Central Africa : Cameroon, Equatorial Guinea, Gabon…)
HIV-1 group P : gorilla (Cameroon)
HIV-2 : sooty mangabey monkey (West Africa : Sierra Leone, Liberia, Ivory Coast…)

HIV-1 virion structure

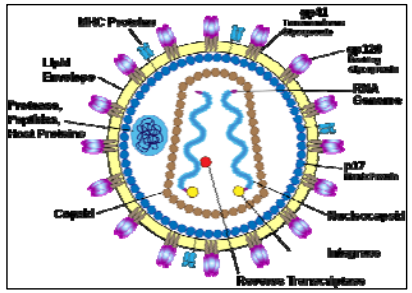


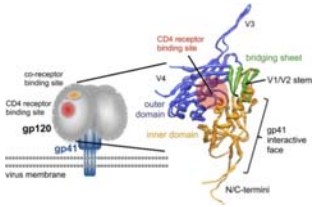






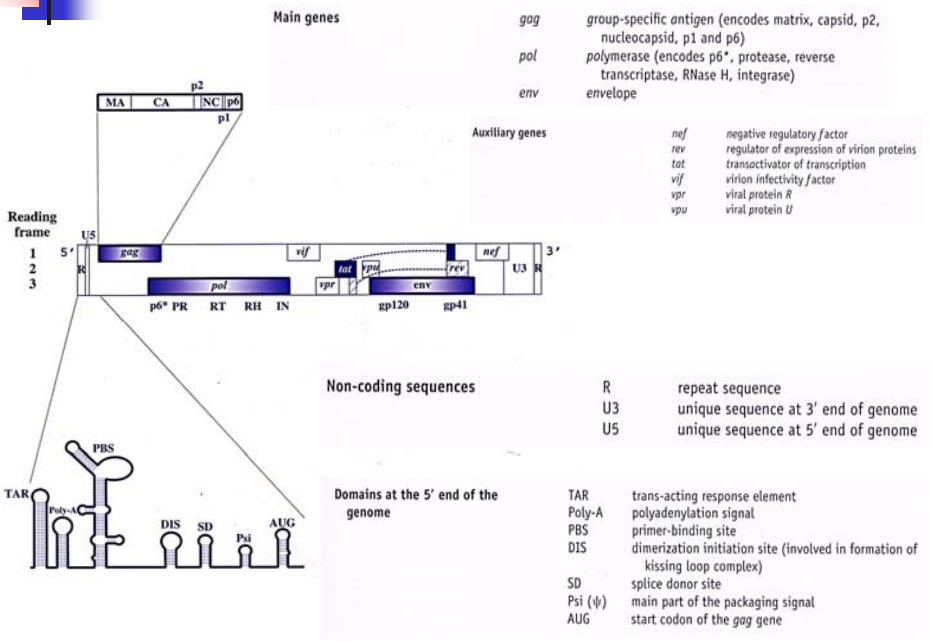
Alan Engelman and Peter Cherepanov, Nature Review in Microbiology, 2012, 10, 279-289



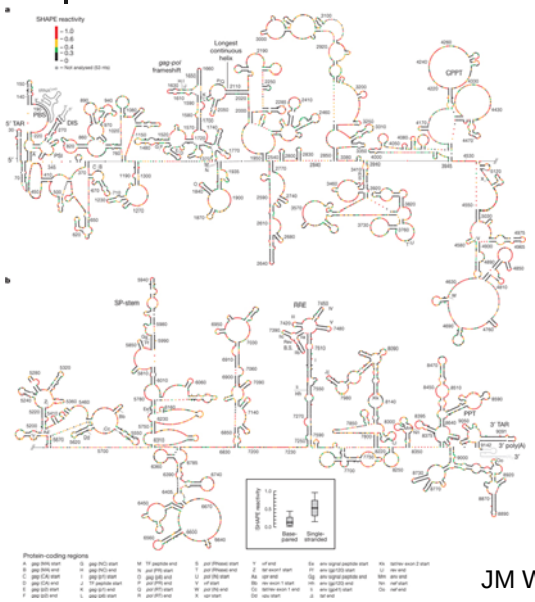


Miklos Guttman et al., J Virol. Aug 2012; 86(16): 8750–8764

HIV-1 genome organization



Structure of the HIV-1 NL4-3 genome

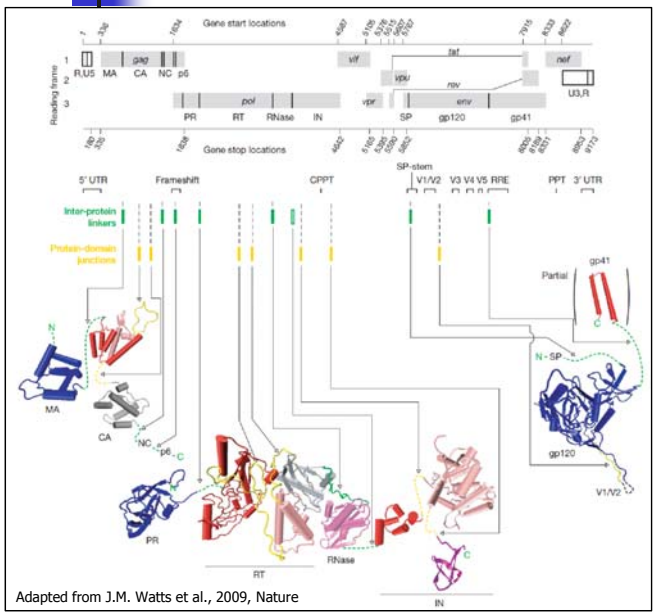


The 5' (**a**) and 3' (**b**) genome halves are shown. Nucleotides are coloured by their absolute SHAPE reactivities. Every nucleotide is shown explicitly as a sphere; base pairing is indicated by adjacent parallel orientation of the spheres.

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



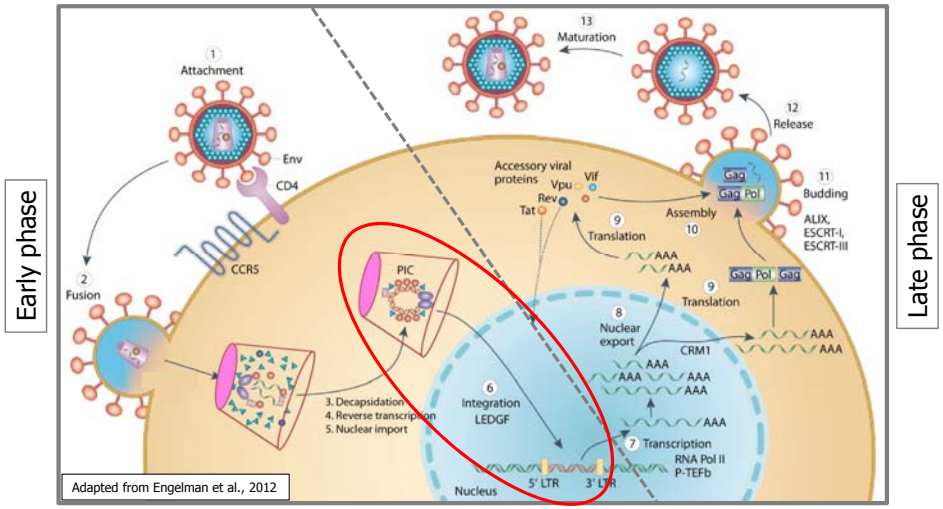
Adapted from J.M. Watts et al., 2009, Nature

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)

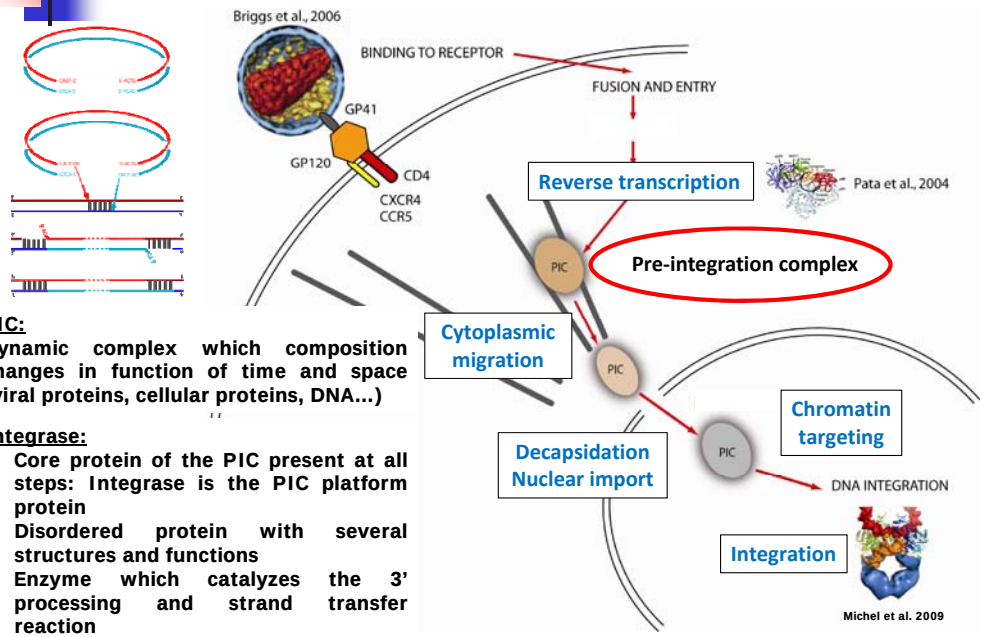
Accessory genes :
Vif, Vpr, Vpu, Nef
(regulation of synthesis and processing viral RNA and other functions)

Schematic diagram of HIV replication cycle



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

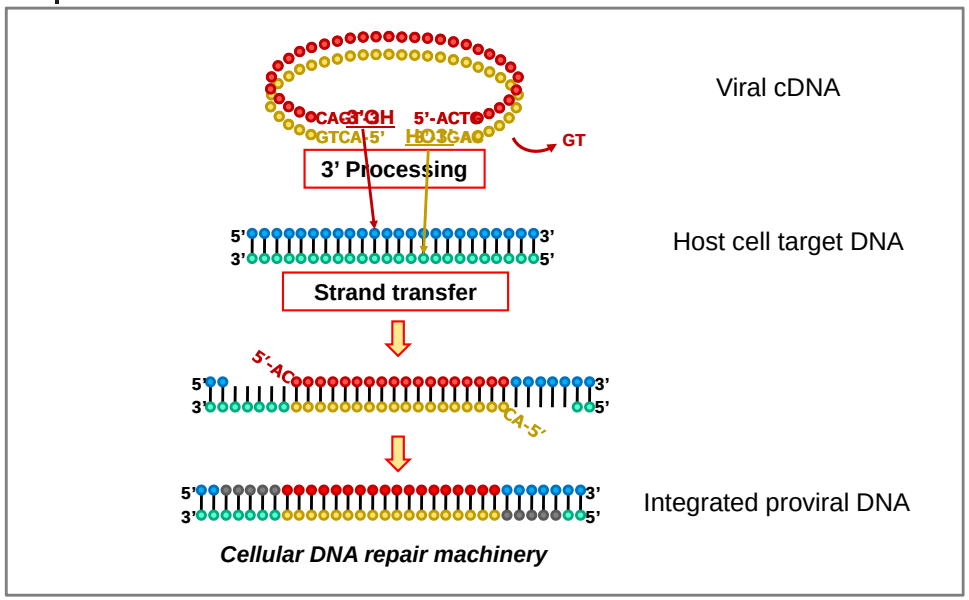
HIV-1 pre-integration complex



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



The HIV-1 viral DNA integration

Decapsidation
Reverse transcription

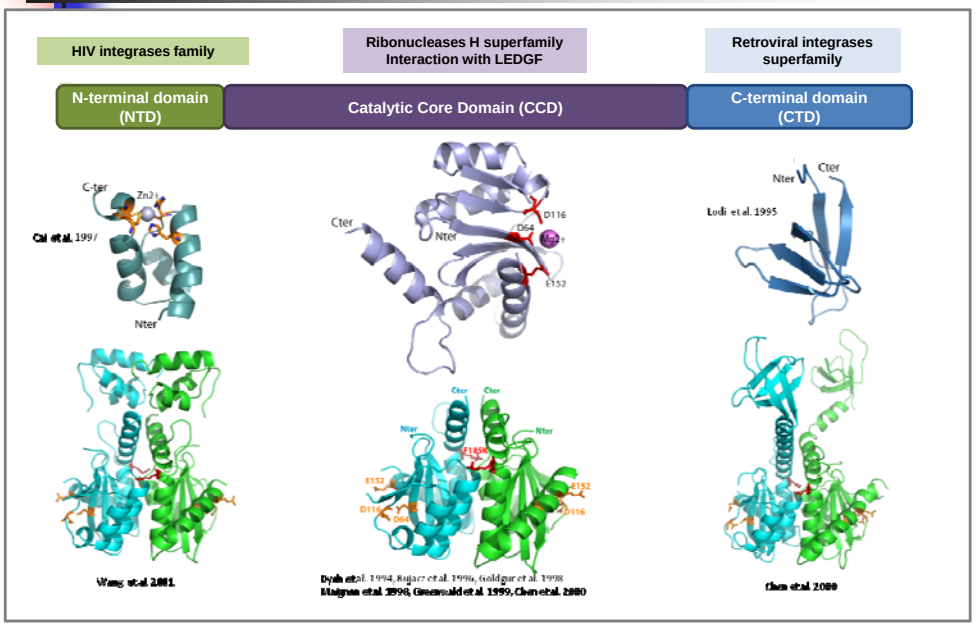
Nuclear import

Continuous and dynamic process

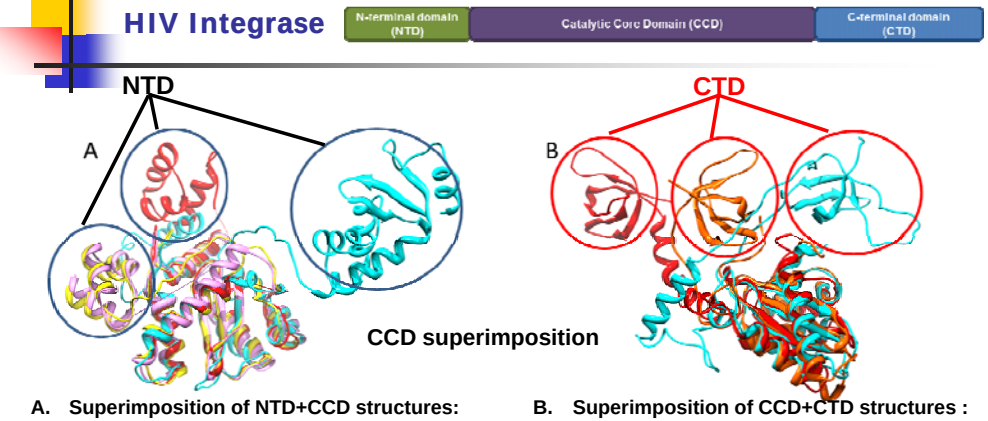
Integration

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

Structural domains of HIV-1 integrase



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

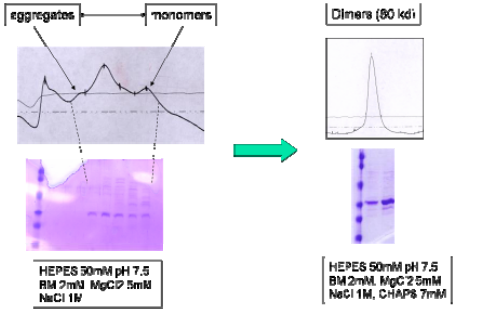
Isolate stable IN complexes (VBP1, TRN-SR2, LEDGF, INI1....)

Isolate protein with PTMs: Production in eukaryotic cells Specific modifications

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	+	+-	+++	+++	++



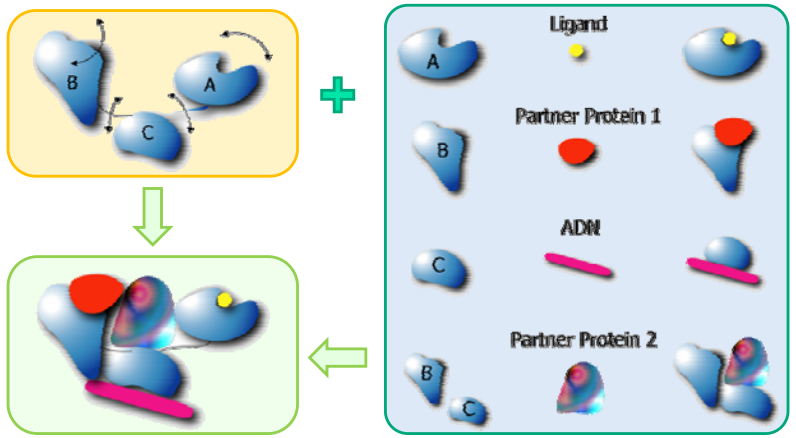
Affinity Chromatography → Gel Filtration (HEPES 50 mM pH 7.5, NaCl 1M, 1 mM Mg++, 7mM CHAPS) → HIC (Ammonium Sulfate, HEPES 50 mM pH 7.5, NaCl 1M, 1 mM Mg++, 7mM CHAPS)

No crystals

No crystals

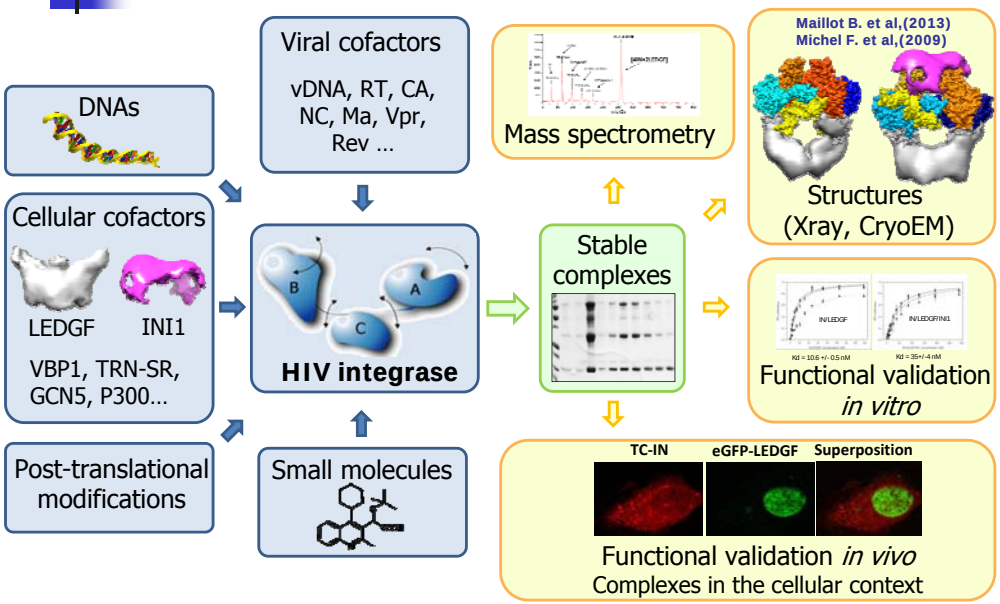
Strategy

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

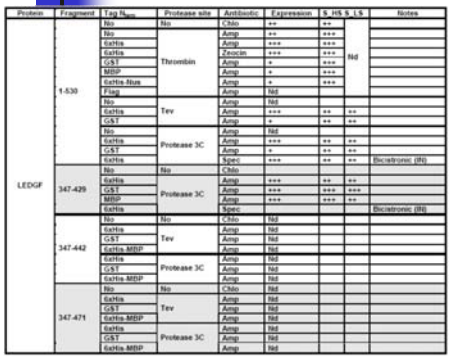


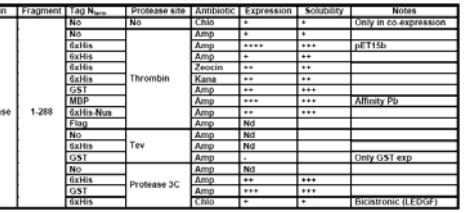
- Proteins domains, mutants, solubility fusions
- Stabilization by the interaction with partner proteins, ligands, DNA.

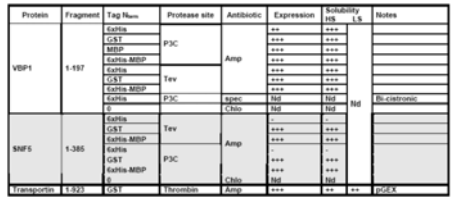
HIV Integrase structural and functional studies

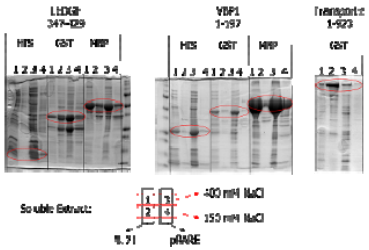


Cloning, expression and solubility tests (E. Coli)







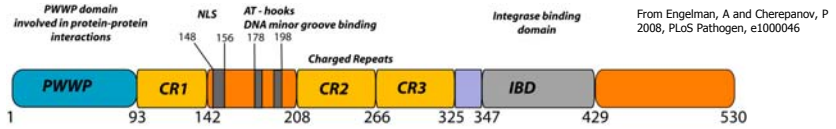


	37°		25°		18°	
	ET	EB	ET	EB	ET	EB
LB	+	-	+	+	+	+
LB/sucrose	+	-	+	+	++	++

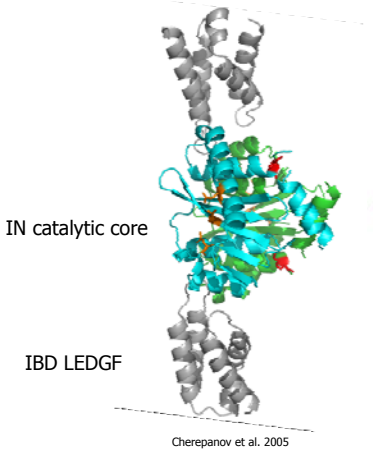
	-	Glycerol 10%	Chaps 10mM	Triton 0.1%	Sucrose 20%
50mM NaCl	+/-	+/-	++	+/-	+/-
1M NaCl	+	++	+++	+++	++

HIV-1 IN / LEDGF complex

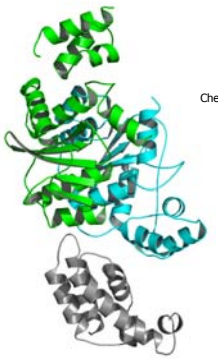
HIV-1 IN interact with LEDGF



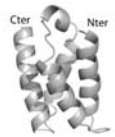
From Engelman, A and Cherepanov, P 2008, PLoS Pathogen, e1000046



Cherepanov et al. 2005



Hare et al. 2009

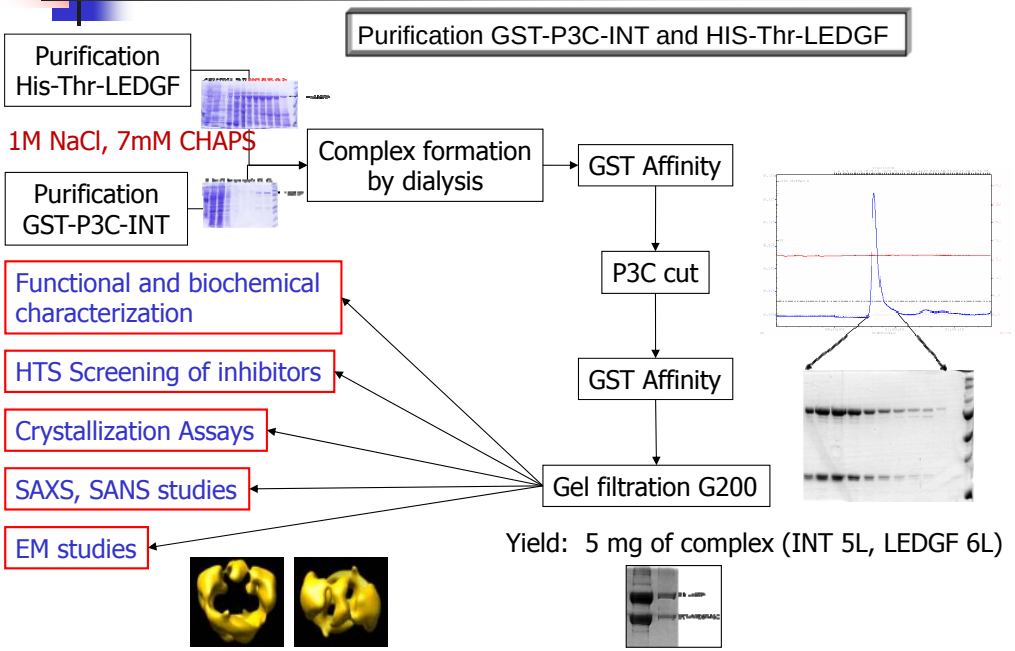


Cherepanov et al. 2005

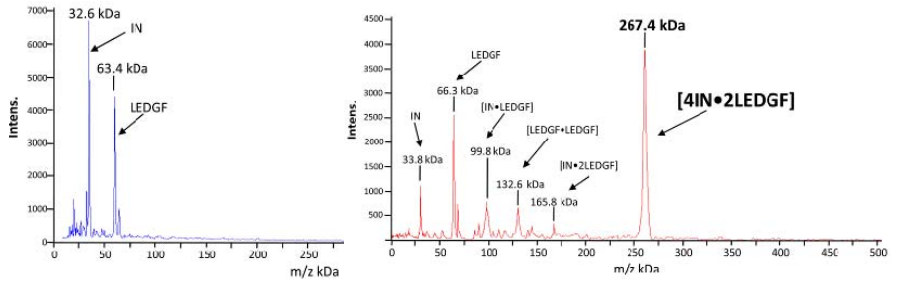
LEDGF/P75

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



IN/LEDGF : characterization by High Mass MALDI-TOFF

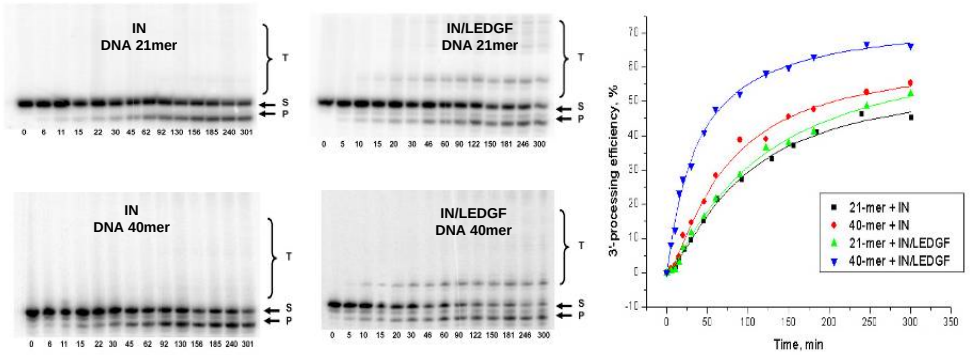


Absence of cross linking agents

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

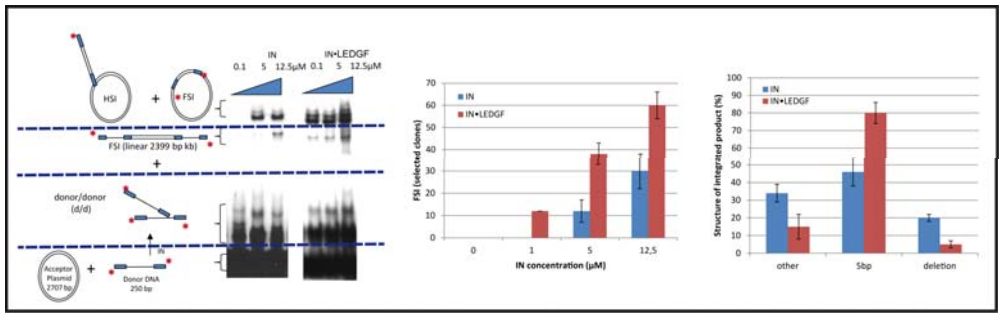


DNA: 21 or 40-mer duplex that mimic the HIV-1 U5 viral DNA end

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

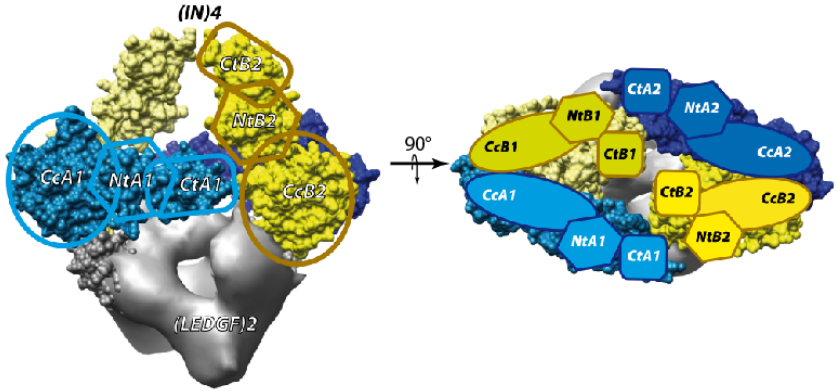
LEGDF stimulates the 3′ processing activity of IN with a 40-mer DNA.
In the presence of LEGDF, the strand transfer efficiency is strongly enhanced for both the 21- and the 40-mer DNA.

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

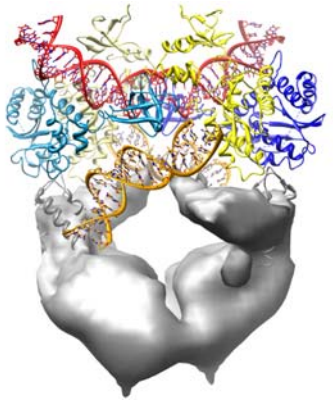


IN-LEDGF complex contains 4 IN molecules (A1, A2, B1 and B2) organized in two IN dimers

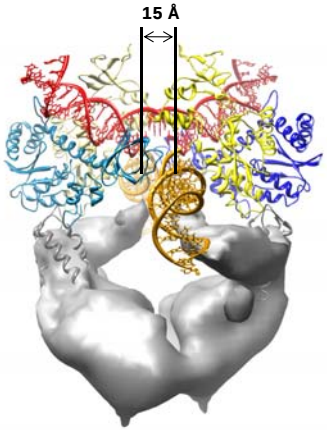
Each IN monomer within the IN dimer has a different conformation → the IN dimer is asymmetric

Each IN molecule has a distinct function within the dimer.

IN/LEDGF/DNA EM structure



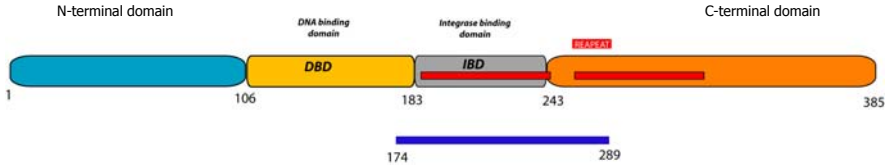
3' Processing



Integration

HIV-1 IN / LEDGF / INI1 complex

IN interact with INI1/SNF5

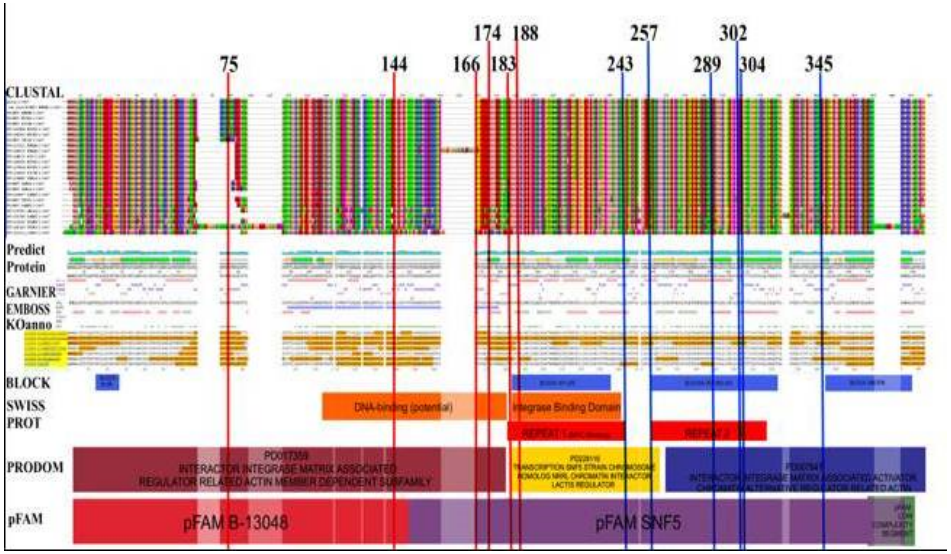


SNF5/Ini1, a subunit of the SWI/SNF chromatin remodeling complex, is the first cofactor identified to interact with IN.

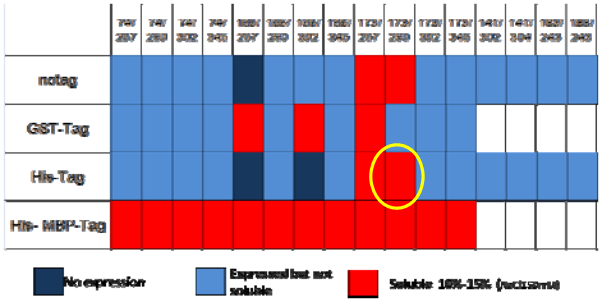
SNF5/Ini1 is one of the core subunits of the ATP-dependent chromatin remodeling complex SWI/SNF that regulates expression of numerous eukaryotic genes by altering DNA/histone interactions

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

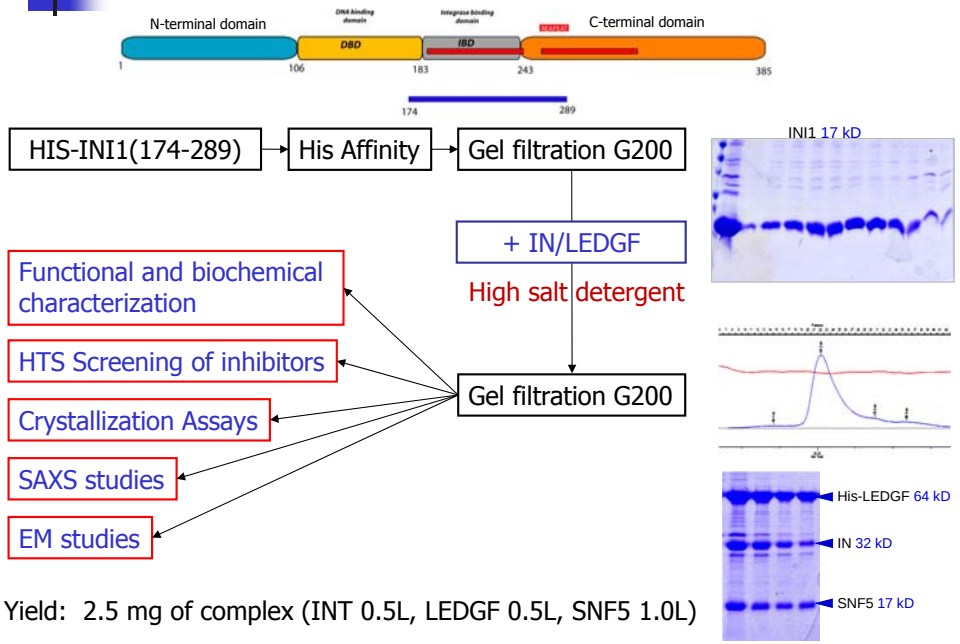


INI1: Cloning, expression and solubility tests (E. Coli)



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

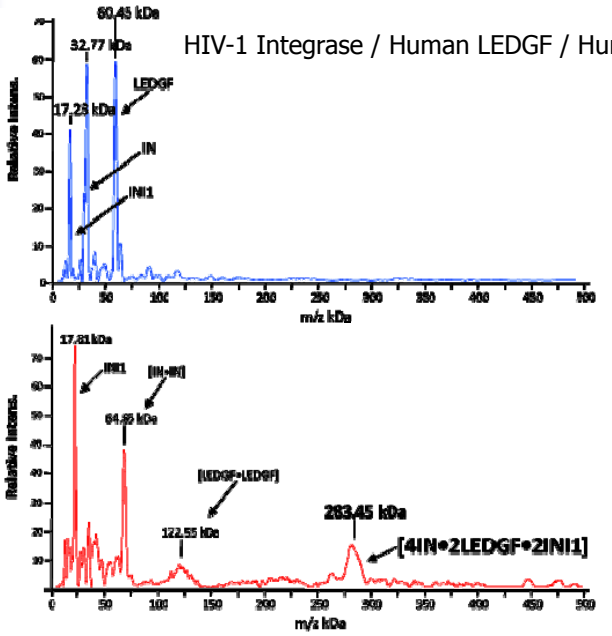
IN/LEDGF/INI1 : complex formation and purification



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

Protein complexes analysis : High Mass MALDI-ToF

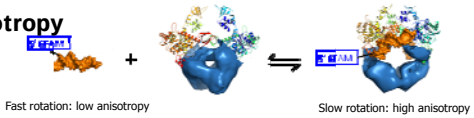
HIV-1 Integrase / Human LEDGF / Human INI1 (173-290)

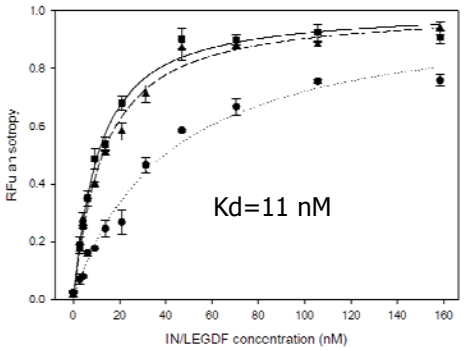


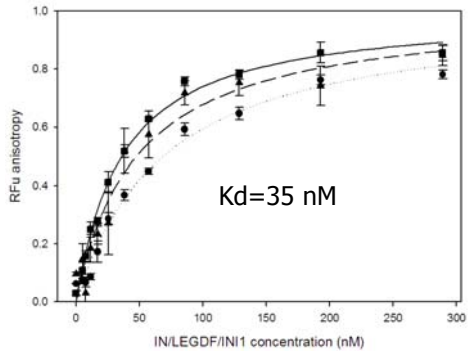
IN/LEDGF and IN/LEDGF/INI1: Functional characterization

Viral DNA Binding assay by fluorescence anisotropy

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

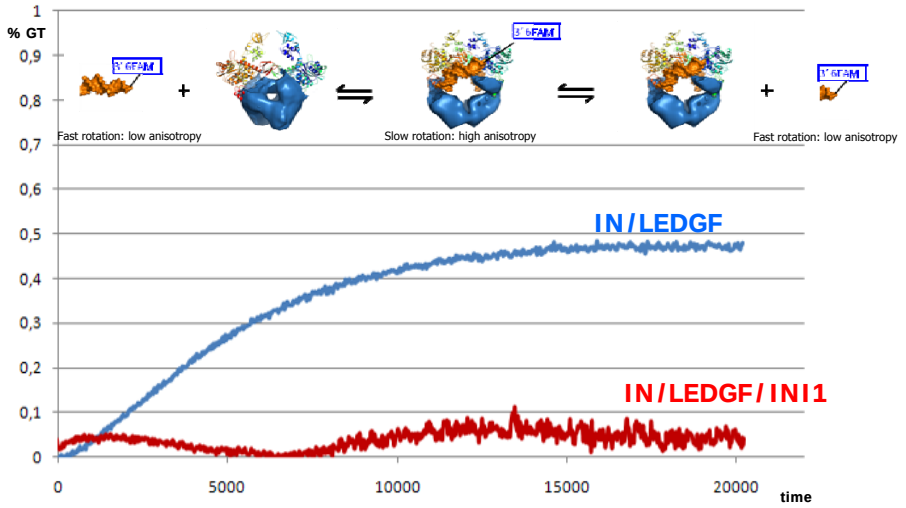






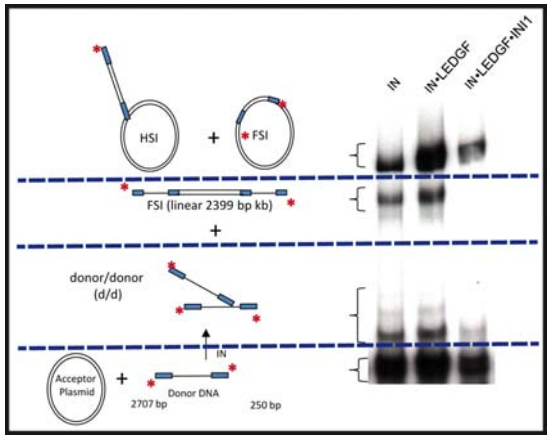
IN/LEDGF and IN/LEDGF/INI1: Functional characterization

The 3' Processing assay by fluorescence anisotropy



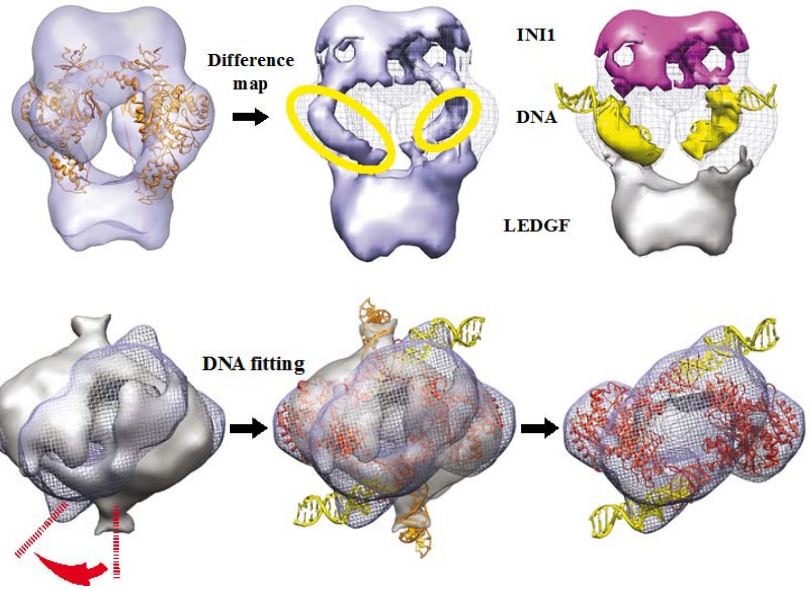
INI1 inhibits the *3' processing* activity of IN/LEDGF

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

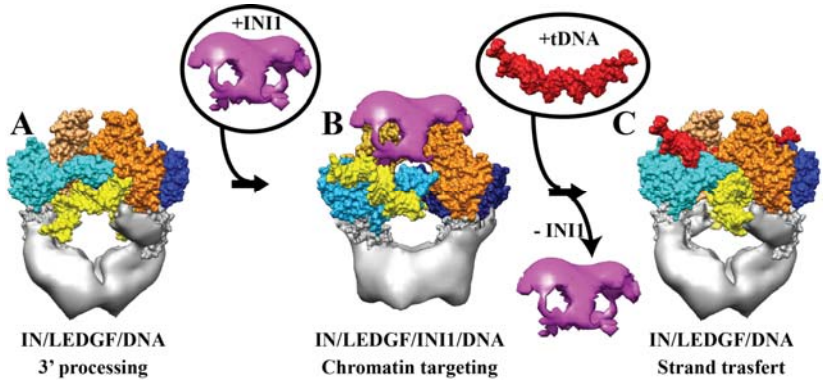


In the presence of INI1, integration occurs with reduced kinetics compared to IN alone or to the IN/LEDGF complex with strongly reduced by-products formation

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



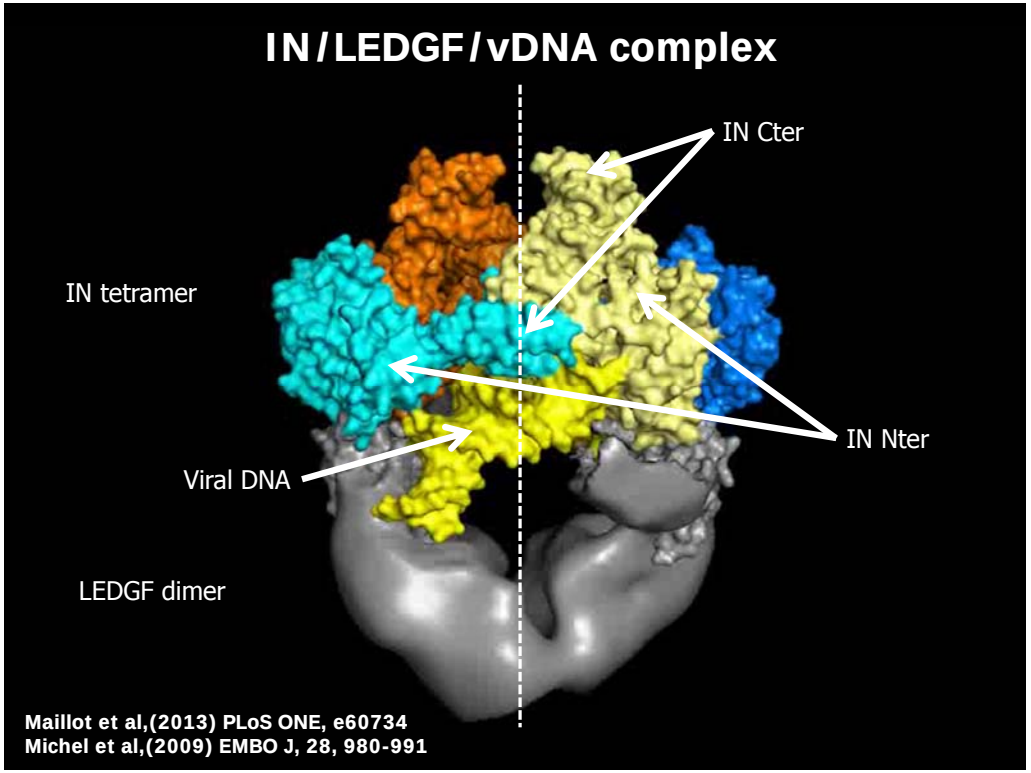
Function of INI1 in HIV-1 infection

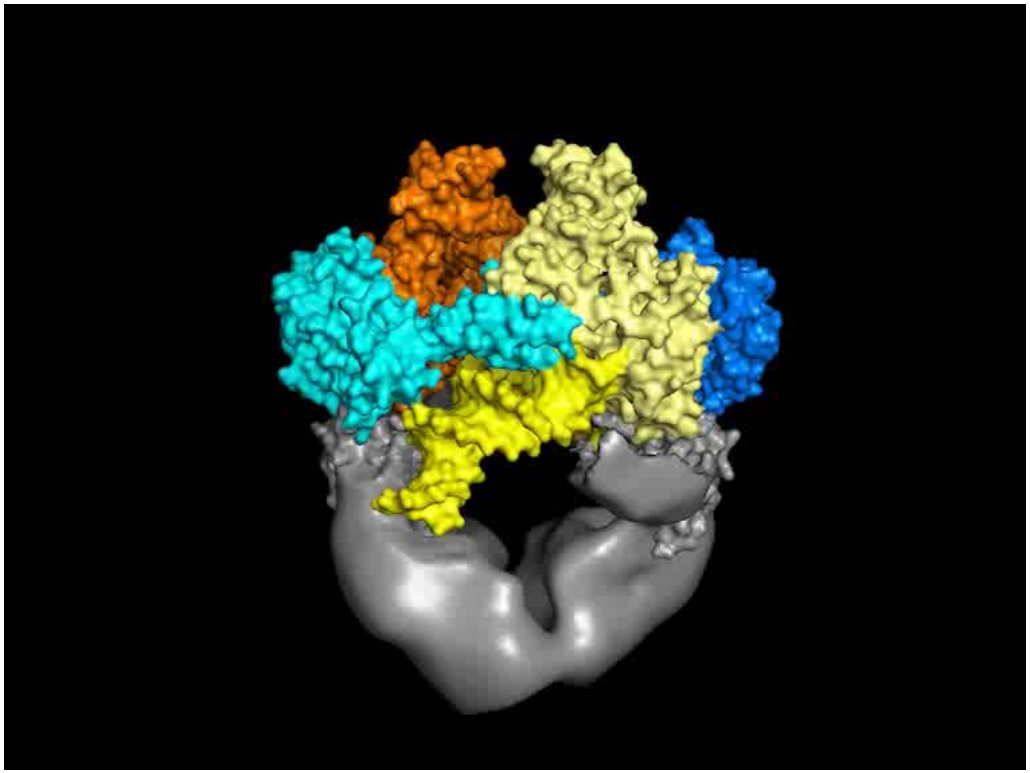


Benoit Maillot, Nicolas Lévy, Sylvia Eiler, Corinne Crucifix, et al., (2013), Structural and functional role of INI1 and LEDGF in the HIV-1 preintegration complex, PlosOne, In Press

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



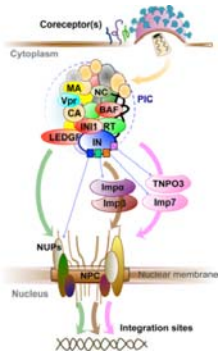


HIV-1 IN cytoplasm – nuclear trafficking

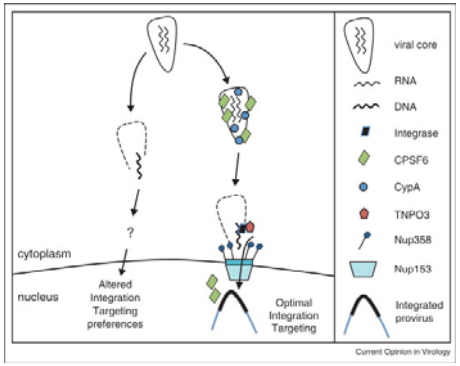
PIC nuclear import:
IN / LEDGF / TNPO3
IN / LEDGF_IBD / CA

IN nuclear export:
HIV-1 IN / TNPO3 / VBP1 complex

PIC Nuclear trafficking



PIC nuclear import from Marx and Alian, 2015



PIC nuclear import from Hilditch and Towers, 2014

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

We have characterized two stables complexes involved in PIC nuclear import:
IN/CA/LEDGF-IBD and IN/TNPO3/LEDGF
And one stable complex involved in IN nuclear export:
IN/TNPO3/VBP1

IN nuclear export

- VBP1: IN degradation through the ubiquitin – proteasome pathway
- Interaction with the Cul2 / VHL ubiquitin ligase complex
- Integration / transcription transition of the viral replication cycle
- VBP1 / IN direct and specific interaction shown by Y2H screen

IN Nuclear export: IN/TNPO3/VBP1 : complex formation and purification: E. Coli

High salt, detergent

Purification His-Thr-IN

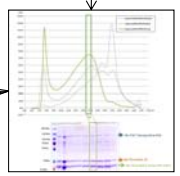
Purification HIS-Strep-VBP1

Purification HIS-P3C-TNPO3

Complex formation by dialysis

Strep Affinity

Gel filtration G200



Functional and biochemical characterization

HTS Screening of inhibitors

Crystallization Assays

SAXS, SANS studies

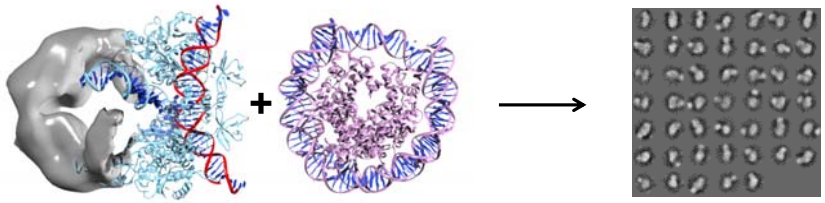
EM studies

Cryo-EM analysis

Work in progress

Other PIC complexes..... In progress

IN interacts with the nucleosome



IN/LEDGF/DNA complex

Nucleosome

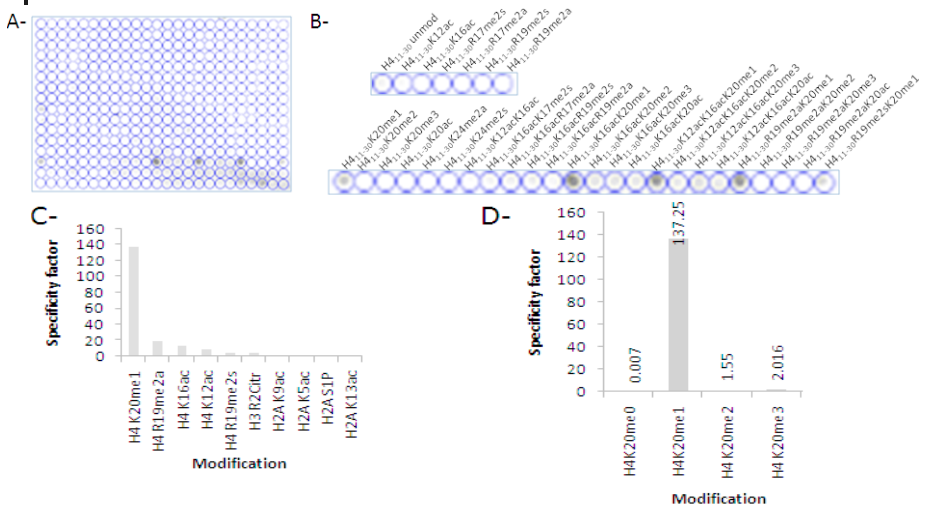
Preliminary Negative stained EM

Structure and function of HIV Integrase domains

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

1QMC, NMR structure

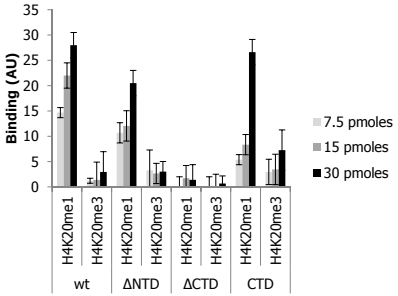
Histone peptide array: H4K20Me1 interacts with HIV-1 IN



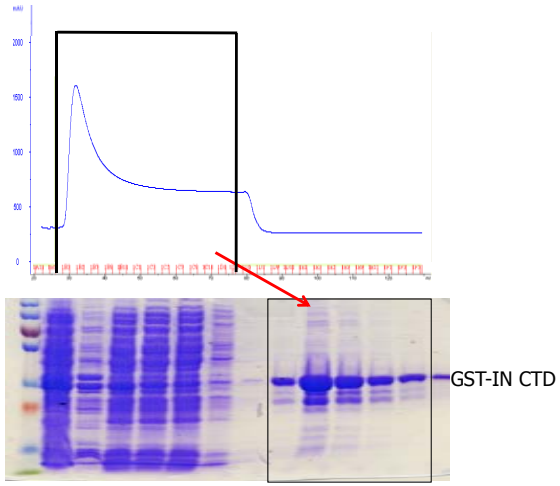
Matysiak et al., submitted

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

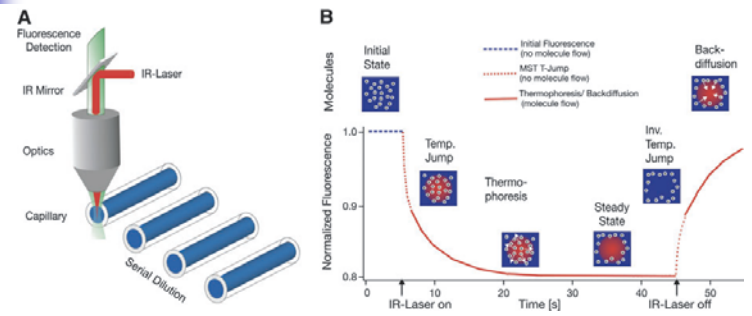


Far dot blot assay



Protein purified using GSTrap column on the AKTA system.

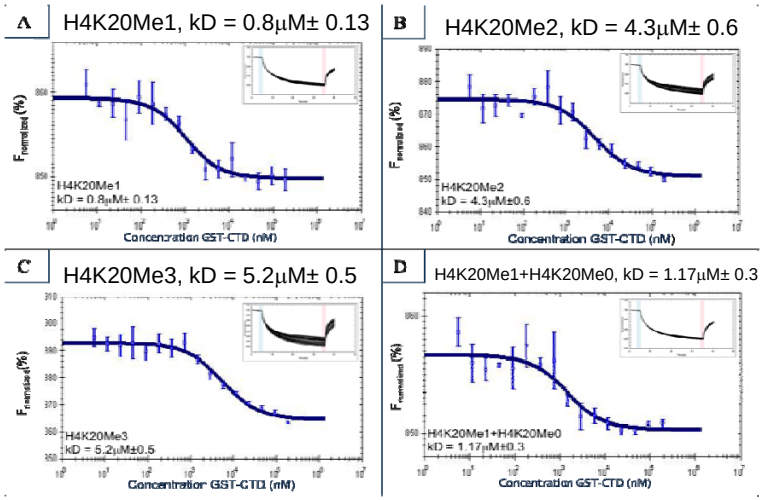
Microscale thermophoresis for Kd measurement



(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.

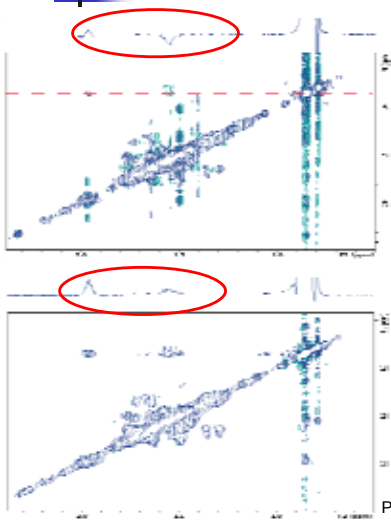
Microscale Thermophoresis confirm a specific interaction between H4K20Me1 and GST-CTD HIV-1 Integrase

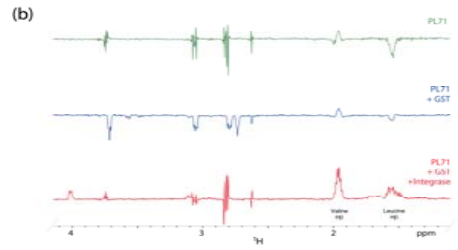


Matysiak et al., submitted

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.

GST-P3C -IN 221-288: binding assay





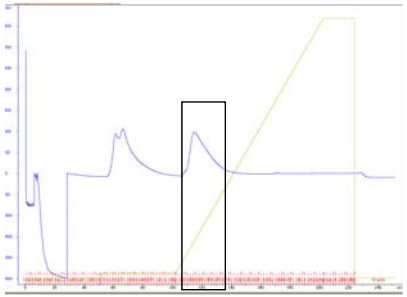
Interaction is specific to GST-CTD as there is no change in NOE signal in the sample with GST only. Change is observed on Leucine and Valine residues

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

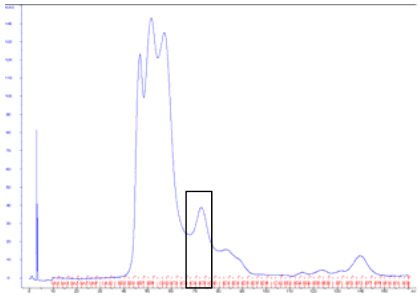
NMR: HIS-P3C-IN 221-288

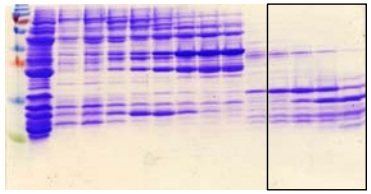
HIS-P3C-IN 221-288: Cultured in N15 labeled minimal medium. Induced with 0.5mM IPTG for 4 hours.

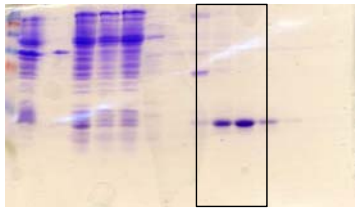
HIS-TRAP 25mM HEPES pH7, 2mM bME, 1M NaCl as lysis buffer, and 25mM HEPES pH7, 2mM bME, 1M NaCl, 500mM Imidazole



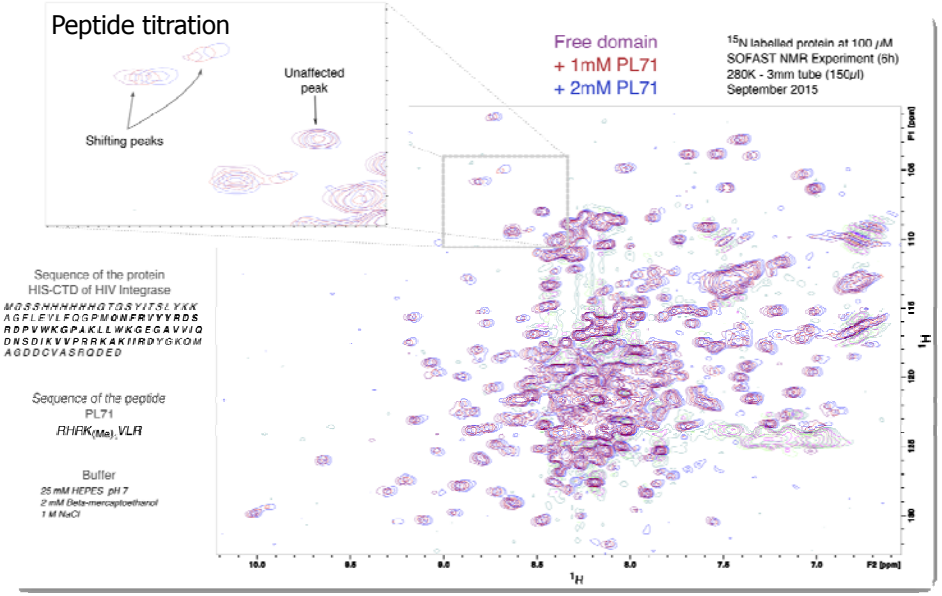
GF S75 16/60 (25mM HEPES pH7, 2mM bME, 1M NaCl)





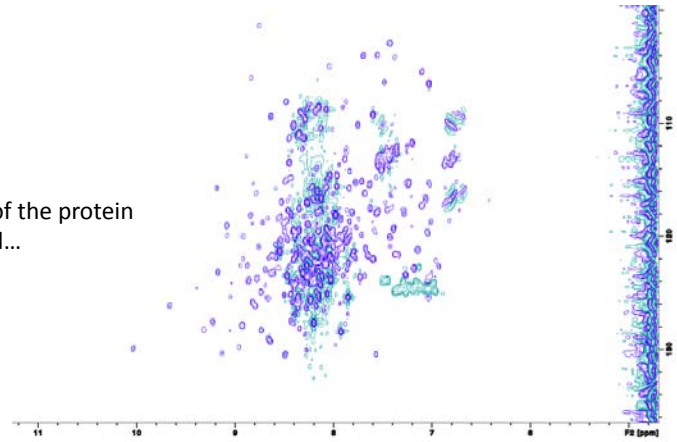


HIS-P3C-IN 221-288: HSQC spectra



HIS-P3C-IN 221-288 HSQC spectra

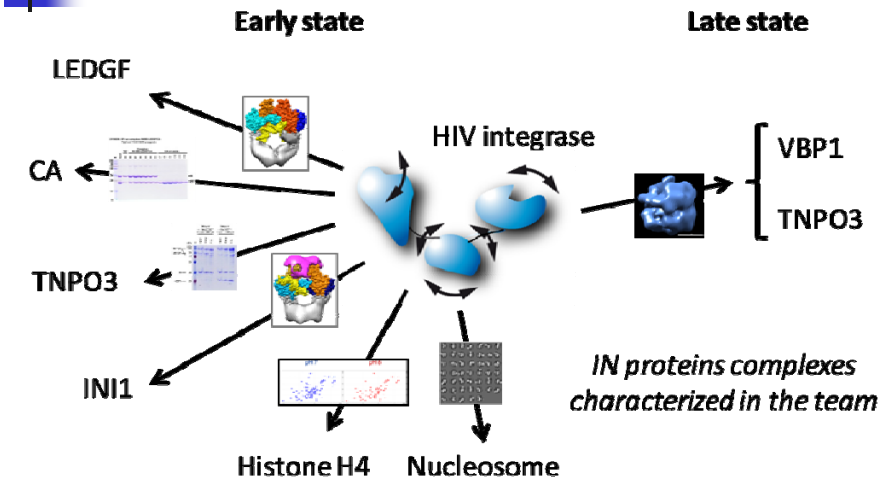
A section of the protein is unfolded…



HIS-P3C-IN 221-288 :

M G S S H H H H H H G T G S Y I T S L Y K K
A G F L E V L F Q G P M ^{221}Q N F R V Y Y R
D S R D P V **W** K G P A K L L **W** K G E G A V
V I Q D N S D I K V V P R R K A K I I R D$_{270}$
Y G K Q M A G D D C V A S R Q D E D$_{288}$

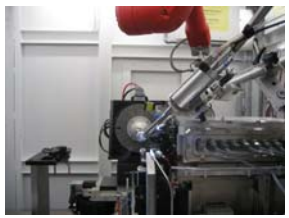
Characterized complexes



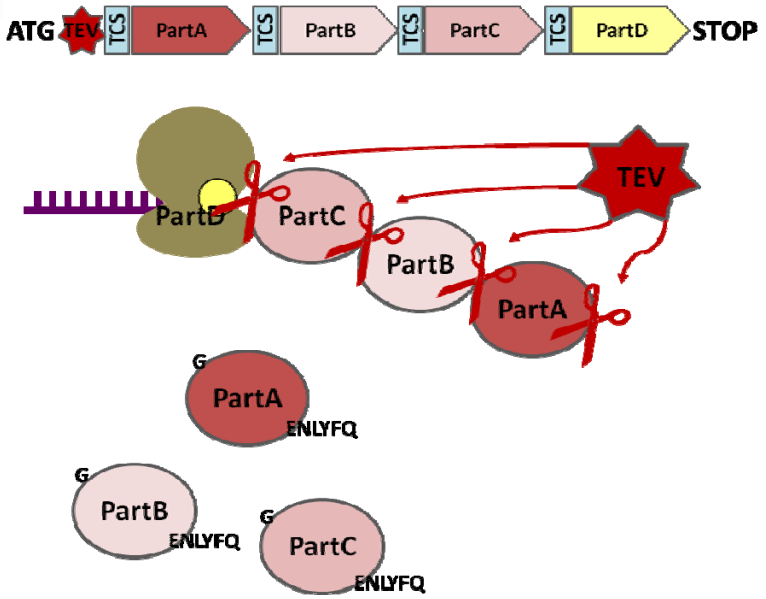
Complexes suitable for low resolution EM studies but not for high resolution and not for crystallization : too many flexible / unstructured parts

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, nanobodies, …)
  - 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)



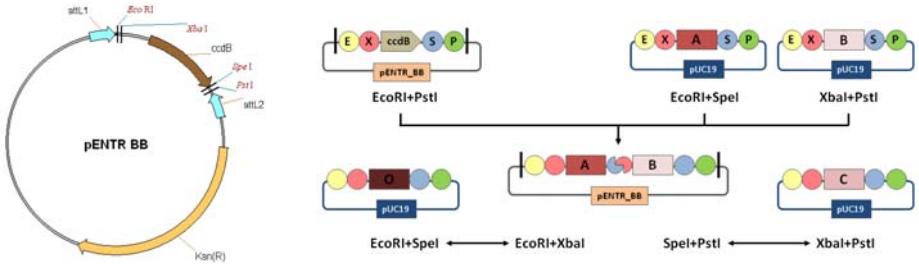
Protein complexes expression in E. Coli, Insect and mammalian cells



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.



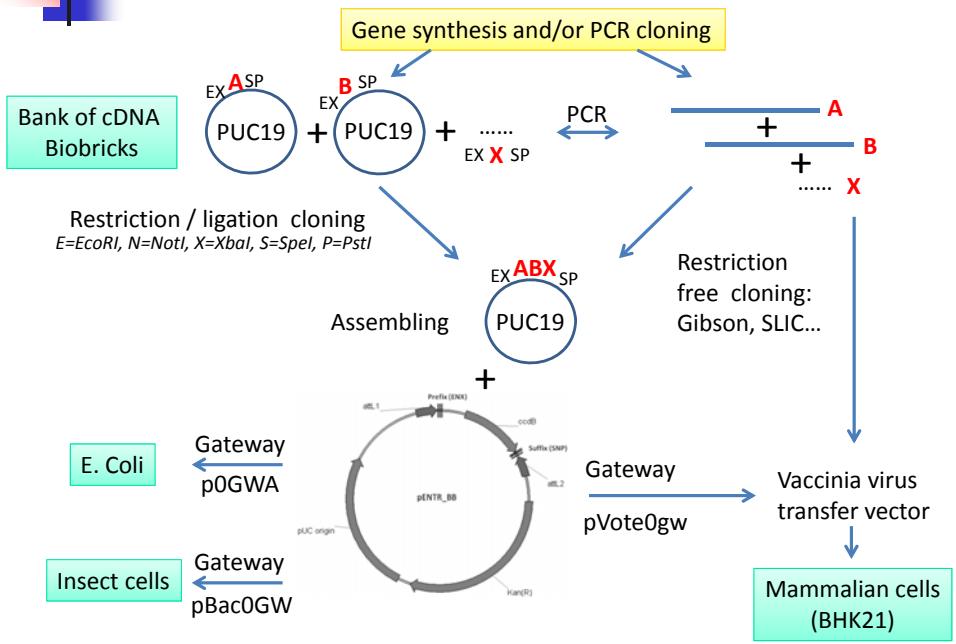
Assembly vector: New pENTR

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

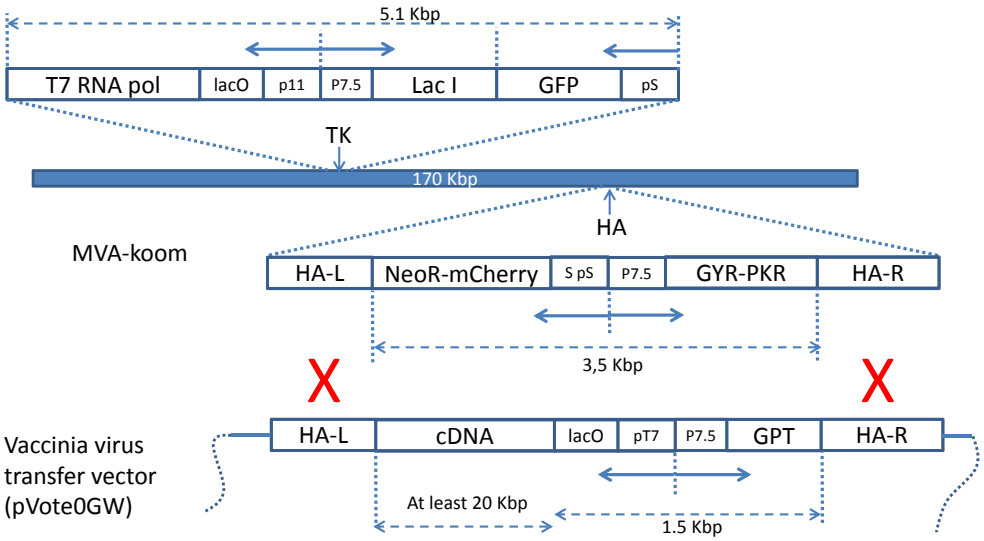


Biobrick collection for protein expression in E. Coli, Insect and mammalian cells

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)

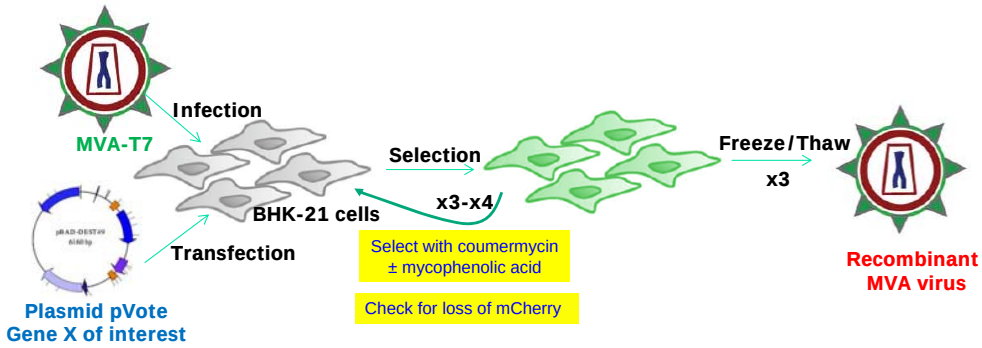
- **Vaccinia virus**: poxvirus family
  - dsDNA virus (≈ 200 kb) encoding its own transcription and replication machinery
  - viral multiplication in the cell cytoplasm: no RNA splicing
  - viral infection diverts the cellular machinery in its favour
  - at least 25 kb of foreign DNA
- **MVA**: Modified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

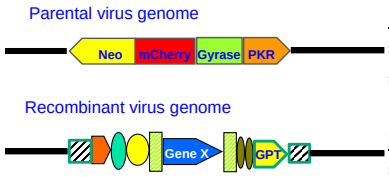


Before recombination: GFP(+), mCherry(+), NeoR(+), GYR-PKR(+) (Coumermycin Sensible), GPT(-) (Mycophenolic acid sensible)

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

Recombinant MVA selection

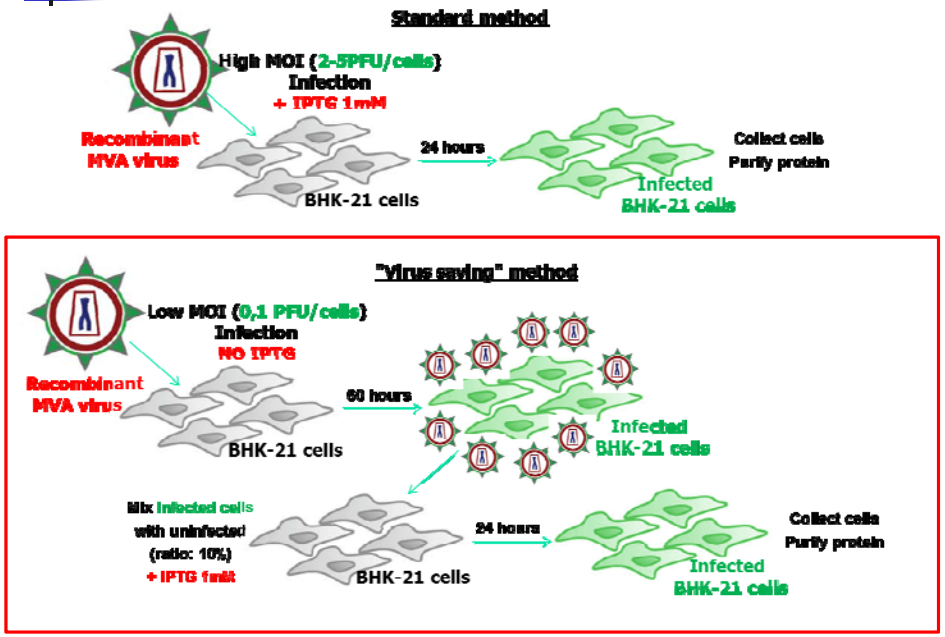




Negative selection: (monitored by mCherry loss) **+ coumermycin**
=> Gyrase dimerisation => active PKR => eif2α phosphorylation
=> translation inhibition => cell death

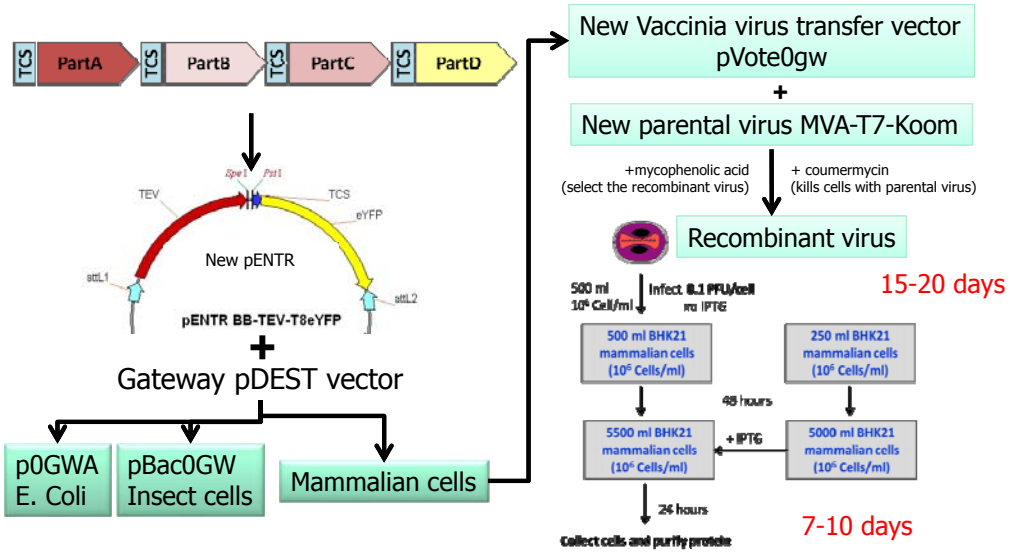
Positive selection: **+ mycophenolic acid**
=> GPT integration selection

MVA Infection methods

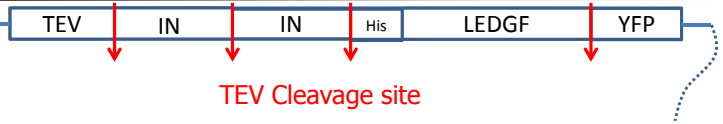


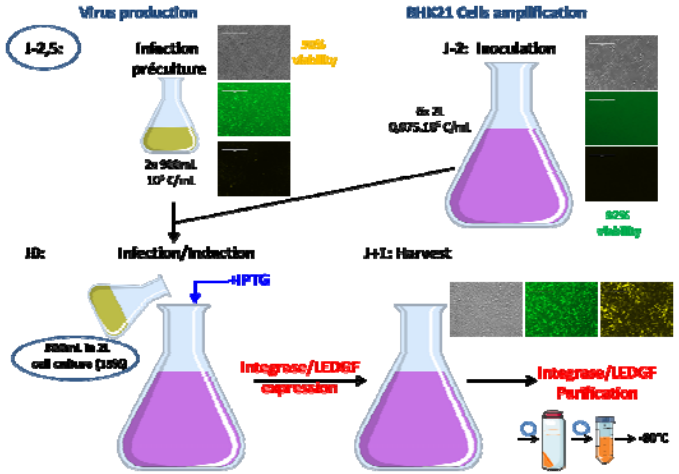
Mammalian cell protein complexes production

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

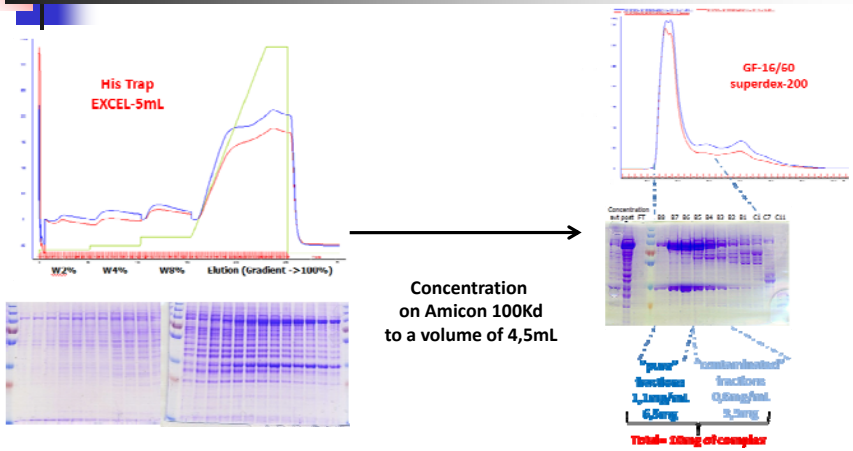


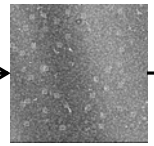
Large scale production of Integrase/LEDGF complex in mammalian cells





Large scale production of Integrase/LEDGF complex in mammalian cells

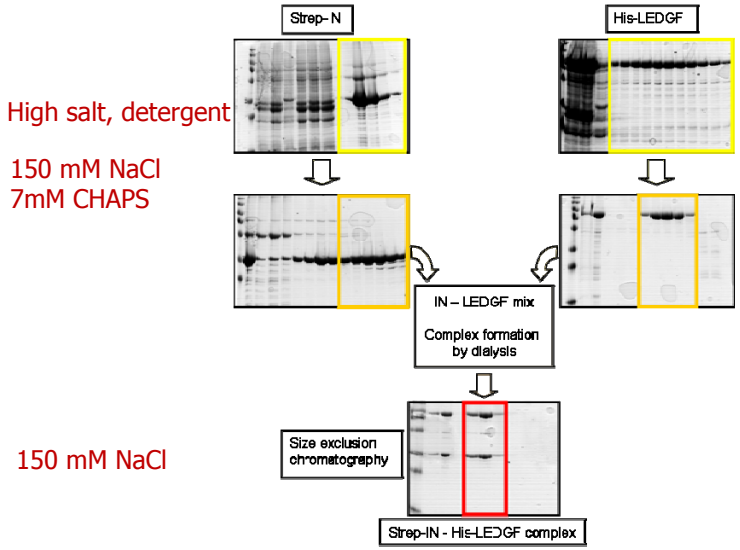


(IN / LEDGF)mam / DNA complex →  → In progress

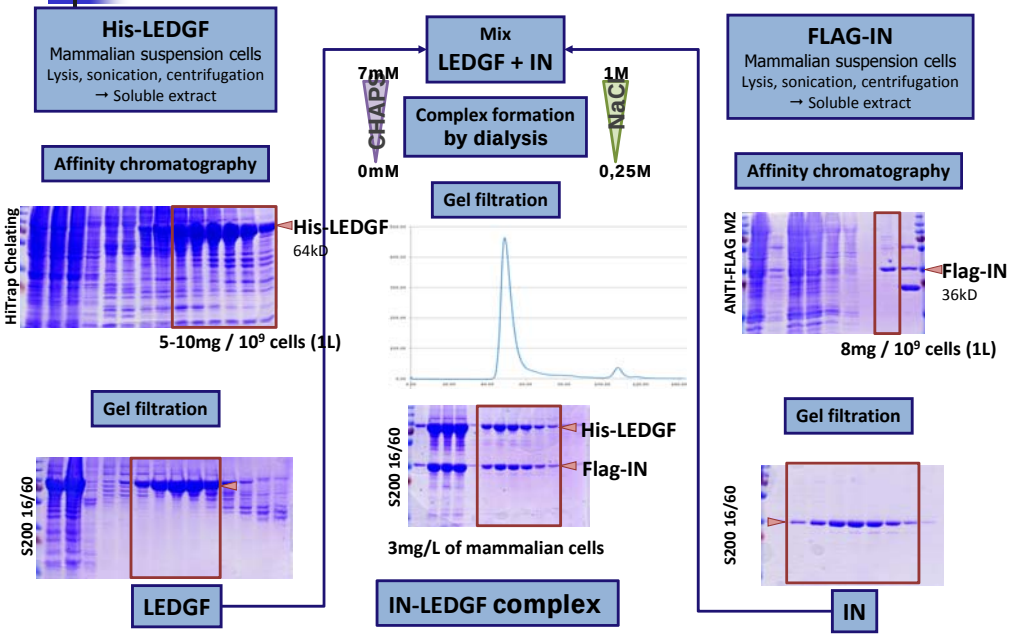
Comparison of HIV-1 IN produced in E. Coli, Insect and mammalian cells

IN/LEDGF : complex formation and purification in Insect cells

Expression vector: Baculovirus



IN/LEDGF complex production in mammalian cells: mono expression



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

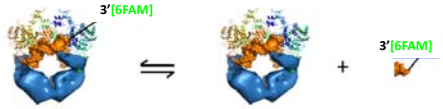
Solubility analysis

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

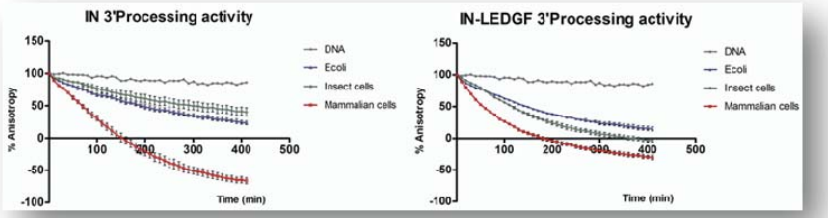
3' processing IN activity by fluorescence anisotropy measurements



Slow rotation
High anisotropy

Fast rotation
Low anisotropy

```
5'-GACTACGGTTCAAGTCAGCGTGTGGAAAATCTCTAGCAGT-3'[6FAM]
3'-CTGATGCCAAGTTCAGTCGCACACCTTTTAGAGATCCTCA-5'
```

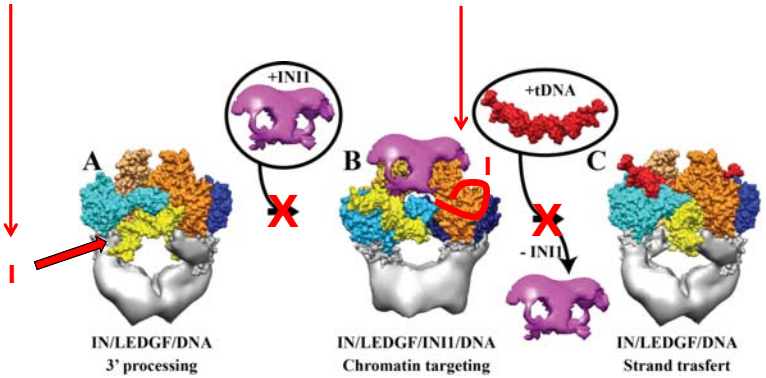


Increase of the 3' processing activity for IN produced in mammalian cells

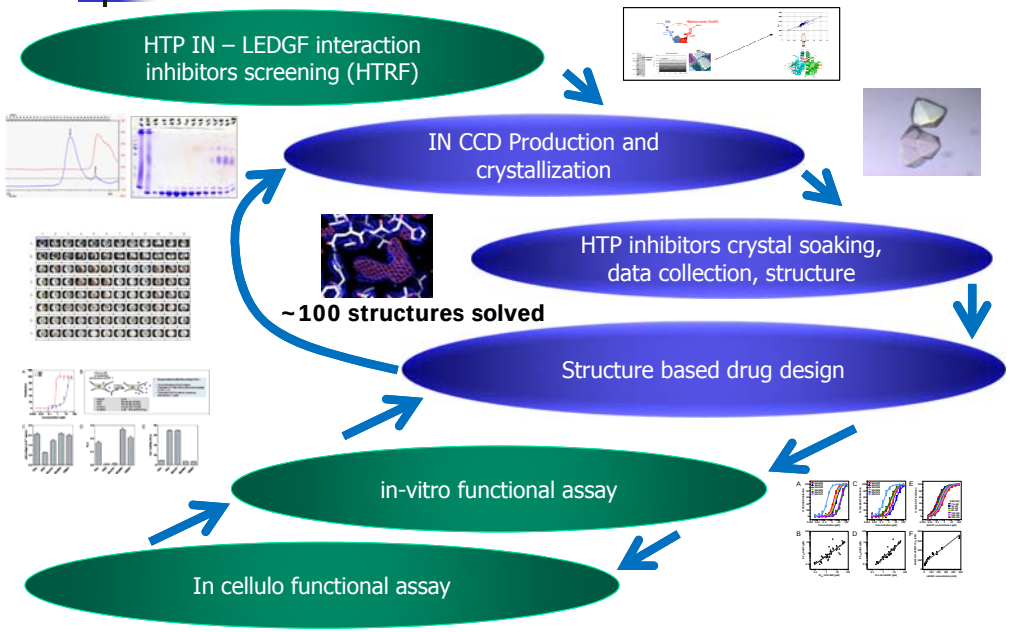
DRUG-DESIGN

Protein – Protein interaction and allosteric inhibitors

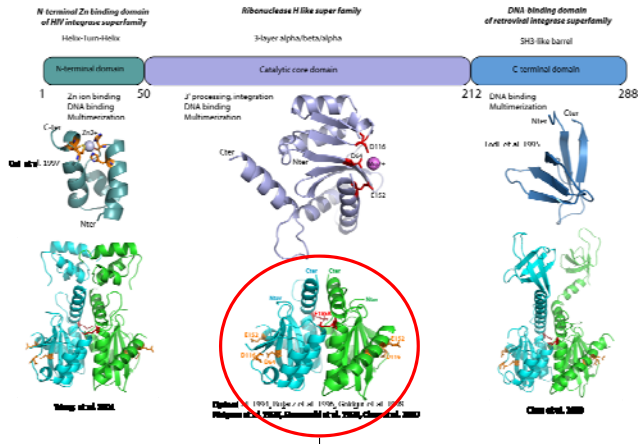
IN/LEDGF interaction inhibitors **IN/LEDGF allosteric inhibitors**



Protein – Protein interaction and allosteric inhibitors



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

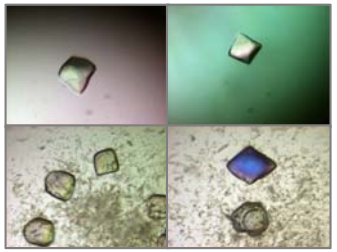
↓

Size exclusion chromatography
Superdex 200

↓

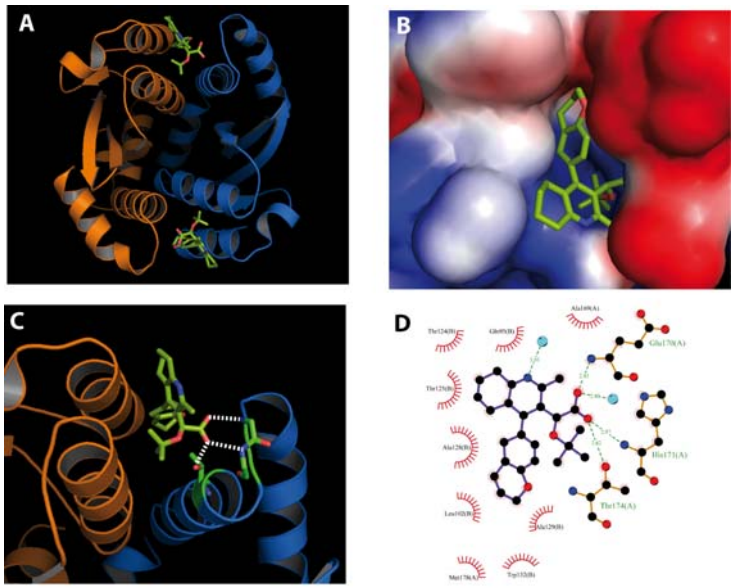
Crystallization
1.26 M AmSO4
100 mM NaCacodylate pH 6.5
Hanging drops
@ 22-23 °C

22-23°C
Diffraction
~2Å

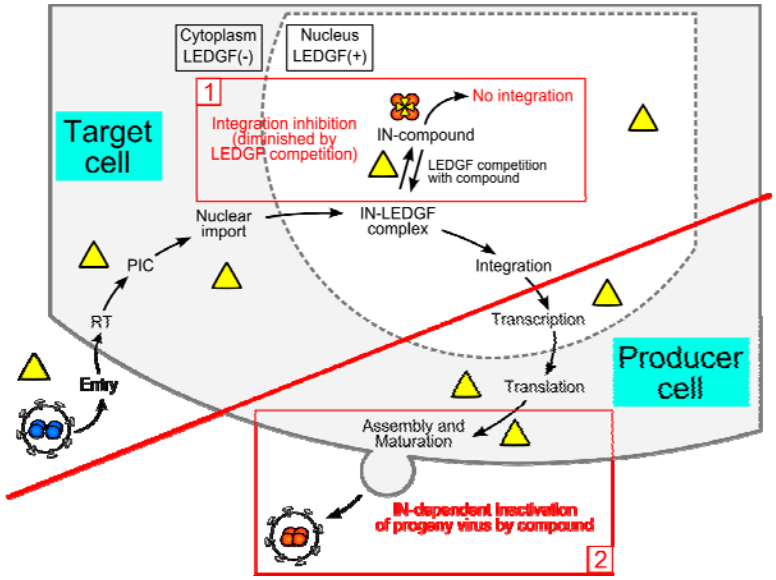


24-25°C
No diffraction !

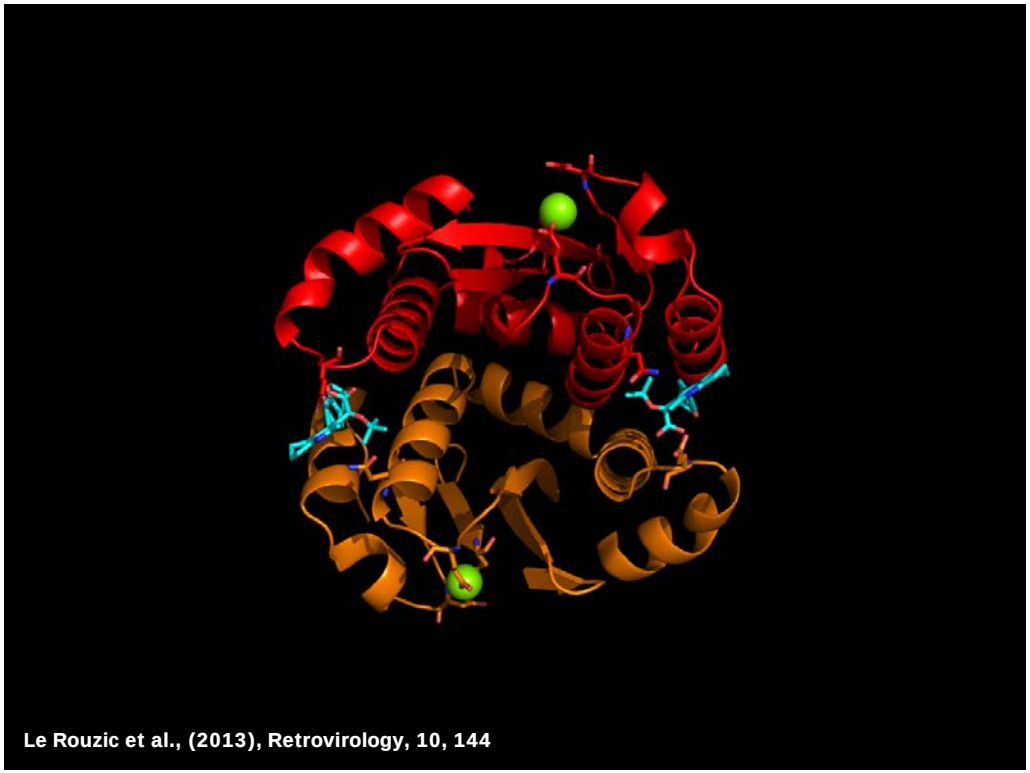
HIV Integrase CCD – ligand structure



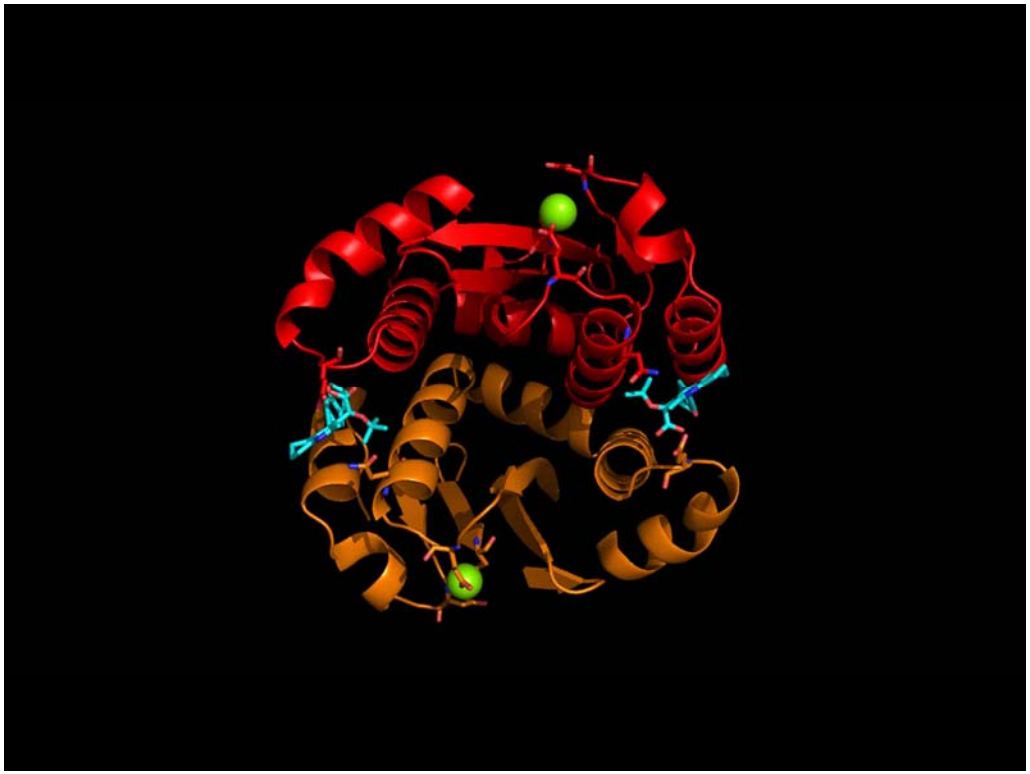
IN – LEDGF interaction and IN allosteric inhibitors



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



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



Acknowledgments

Sylvia Eiler
Nicolas Levy
Julien Batisse
Oyindamola Oladosu
Karine Pradeau
Benoit Maillot
Robert Drillien
Valérie Lamour
Arnaud Vanden Broeck
Christophe Lotz

Fabrice Michel
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