



HAL
open science

In vitro, in cellulo and structural characterizations of the interaction between the integrase of Porcine Endogenous Retrovirus A/C and proteins of the BET family

Kathy Gallay, Guillaume Blot, Margaux Chahpazoff, Halima Yajjou-Hamalian, Marie-Pierre Confort, Claire de Boisseson, Aurélie Leroux, Catherine Luengo, Francesca Fiorini, Marc Lavigne, et al.

► To cite this version:

Kathy Gallay, Guillaume Blot, Margaux Chahpazoff, Halima Yajjou-Hamalian, Marie-Pierre Confort, et al.. In vitro, in cellulo and structural characterizations of the interaction between the integrase of Porcine Endogenous Retrovirus A/C and proteins of the BET family. *Virology*, 2019, 532, pp.69-81. <10.1016/j.virol.2019.04.002>. <hal-02359028>

HAL Id: hal-02359028

<https://hal.science/hal-02359028v1>

Submitted on 22 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

1 ***In vitro, in cellulo* and structural characterizations of the interaction**
2 **between the Integrase of Porcine Endogenous Retrovirus A/C and proteins**
3 **of the BET family**

4

5 Kathy Gallay ^{1, †}, Guillaume Blot ^{2, †}, Margaux Chahpazoff ^{3, †}, Halima Yajjou-Hamalian ³,
6 Marie-Pierre Confort ¹, Claire De Boisséson ², Aurélie Leroux ², Catherine Luengo ¹,
7 Francesca Fiorini ³, Marc Lavigne ⁴, Yahia Chebloune ⁵, Patrice Gouet ³, Karen Moreau ^{1, 6},
8 Yannick Blanchard ^{2*, +}, Corinne Ronfort ^{1*, +}

9

10 ¹ INRA, Université Lyon 1, UMR754, Viral Infections Compared Pathology. 69007 Lyon,
11 France. Université de Lyon, 69000 Lyon, France. UMSI3444 Biosciences Gerland Lyon
12 Sud, 69007 Lyon, France

13 ² ANSES, Ploufragan/Plouzané Laboratory, Viral Genetics and Bio-Security Unit, Université
14 Européenne de Bretagne, Ploufragan, France

15 ³ Molecular Microbiology and Structural Biochemistry, MMSB-IBCP, UMR 5086 CNRS
16 University of Lyon, 7, passage du Vercors, 69367 Lyon Cedex 07, France

17 ⁴ Pasteur Institute, Virology dept. Institut Cochin- Inserm U1016-CNRS UMR8104-
18 Université Paris Descartes. 27, rue du faubourg Saint Jacques. 75014 Paris, France

19 ⁵ Pathogenesis and Lentiviral Vaccination Laboratory. INRA. Univ Grenoble Alpes. 570 rue
20 de la Chimie, 38400 St Martin d'Hères

21 ⁶ Present address : CIRI, International Center for Infectiology Research, INSERM U1111,
22 Université Lyon 1, Ecole Normale Supérieure de Lyon, CNRS UMR5308, 7 rue Guillaume
23 Paradin, 69008 Lyon, France

24

25 [†]: KG, GB and MC contributed equally

26 ⁺: YB and CR contributed equally

27

28 *Corresponding authors: C. Ronfort (corinne.ronfort@inra.fr) and Y. Blanchard
29 (yannick.blanchard@anses.fr).

1 **Abstract**

2 Retroviral integrase (IN) proteins catalyze the permanent integration of the viral genome into
3 host DNA. They can productively recruit cellular proteins, and the human Bromodomain and
4 Extra-Terminal domain (hBET) proteins have been shown to be co-factors for integration of
5 gamma-retroviruses such as Murine Leukemia Virus (MLV) into human cells. By using two-
6 hybrid, co-immunoprecipitation and *in vitro* interaction assays, we showed that IN of the
7 gamma- Porcine Endogenous Retrovirus-A/C (PERV IN) interacts through its C-terminal
8 domain (CTD) with hBET proteins. We observed that PERV IN interacts with the BRD2,
9 BRD3 and BRD4 proteins *in vitro* and that the BRD2 protein specifically binds and co-
10 localizes with PERV IN protein in the nucleus of cells. We further mapped the interaction
11 sites to the conserved Extra-Terminal (ET) domain of the hBET proteins and to several amino
12 acids of the of the C-terminal tail of the PERV IN CTD. Finally, we determined the first
13 experimental structure of an IN CTD – BET ET complex from small-angle X-ray scattering
14 data (SAXS). We showed that the two factors assemble as two distinct modules linked by a
15 short loop which confers partial flexibility. The SAXS-restrained model is structurally
16 compatible with the binding of the PERV intasome to BRD2. Altogether, these data confirm
17 the important role of host BET proteins in the gamma-retroviruses' targeting site and
18 efficiency of integration.

19

20 **Keywords:** gamma-retrovirus; integration; IN co-factors; BET proteins; SAXS

21

1 **Background**

2 During the retroviral life cycle, integrases (INs) that catalyze the key step of
3 integration for the provirus of the retrotranscribed DNA into the host cell genome can interact
4 with cellular cofactors to target specific integration sites (Kvaratskhelia et al., 2014). Hence,
5 the Human Immunodeficiency Virus type 1 (HIV-1) binds the host Lens Epithelium-Derived
6 Growth Factor (LEDGF/p75) protein (Busschots et al., 2005; Cherepanov et al., 2003;
7 Emiliani et al., 2005; Llano et al., 2006; Maertens et al., 2003) to preferentially target the
8 body of actively transcribed genes (Mitchell et al., 2004; Schroder et al., 2002). More
9 recently, the cleavage and polyadenylation specificity factor 6 protein (CPSF6) complexed
10 with the HIV-1 capsid (CA) protein has been shown to play a dominant role in directing
11 integration within euchromatin (Sowd et al., 2016). Integration of Avian Sarcoma and
12 Leukemia Viruses (ASLV) is more random (Mitchell et al., 2004) and no integration cell
13 cofactor directing ASLV integration has been reported to date even though the ASLV IN-
14 binding factor FACT (FAcilitates Chromatin Transcription) might be a good candidate
15 (Winans et al., 2017). Also, it remains to be seen whether the delta-retroviral IN-binding
16 serine/threonine protein phosphatase 2A (PP2A) protein plays a role in directing integration
17 of delta-retroviruses (Maertens, 2016). Foamy virus (FV) integration seems to also be
18 random, but FVs tend to target promoter-closed regions when integration into a transcribed
19 region has occurred (Nowrouzi et al., 2006).

20 Concerning gamma-retroviruses, the Murine Leukemia Virus (MLV) favors
21 integration near the promoters, enhancers, CpG islands and DNase I hypersensitive sites (De
22 Ravin et al., 2014; LaFave et al., 2014; Mitchell et al., 2004; Wu et al., 2003). Numerous
23 groups (De Rijck et al., 2013; Gupta et al., 2013; Sharma et al., 2013) have confirmed earlier
24 preliminary data (Studamire and Goff, 2008) indicating a potential binding of MLV IN with
25 the Bromodomain-containing protein Brd2, and they have shown that Brd2, 3 and 4 target the
26 integration of MLV in transcription start sites. These three ubiquitously expressed proteins of
27 the Bromodomain and Extra-Terminal (BET) family specifically bind MLV IN and co-
28 localize in the cell nucleus. Furthermore, the integration site distribution of MLV corresponds
29 to the chromatin-binding profile of BET proteins (De Rijck et al., 2013). Brd4 was reported to
30 interact with other gamma-retroviral INs, such as those of Feline Leukemia Virus (FeLV)
31 (Gupta et al., 2013) and avian Reticulo-Endotheliosis Virus (REV) (Serrao et al., 2015). In
32 addition, a recent paper has pinpointed an important role of the MLV CA protein in

1 facilitating the targeting of the pre-integration complex to nucleosomes via a mechanism that
2 could be similar to the one of the HIV-1 CA–CPSF6 complex (Wanaguru et al., 2018).

3 Members of the BET protein family contain tandem bromodomains (BD-I and BD-II)
4 and a conserved Extra-Terminal domain (ET) (Thorpe et al., 1996). The BET family contains
5 chromatin-binding factors that have been implicated in the control of cell cycle progression,
6 transcription and DNA replication (Belkina and Denis, 2012; Wu and Chiang, 2007). These
7 proteins bind to the acetylated tails of histones H3 and H4 via their bromodomains. The extra-
8 terminal motif ET in the C-terminal region of BET proteins was shown to interact with
9 numerous cellular — *e.g.* transcription factors and histone modification enzymes — and viral
10 factors. Interestingly, BRD2 was shown to interact with Kaposi’s Sarcoma-Associated
11 Herpesvirus (KSHV) and the latency-associated nuclear antigen 1 (LANA 1), a protein
12 required for the replication of the episomal genome in dividing cells (Viejo-Borbolla et al.,
13 2005). The C-terminal part of the long isoform of BRD4 can bind the bovine papilloma virus
14 type 1 E2 protein (Jang et al., 2009).

15 Retroviral INs are composed of three conserved domains — *i.e.* N-terminal domain
16 (NTD), catalytic core domain (CCD) and C-terminal domain (CTD) — joined by flexible
17 linkers (see Fig. 1A) which affect crystallization trials. For this reason structural information
18 was obtained only for one- or two-domain IN fragments for a decade (reviewed by (Andrake
19 and Skalka, 2015)). The NTD was characterized with an all- α fold, the CCD with an RNase-H
20 fold and the CTD with an SH3 fold. Since 2010, crystal and cryo-electron microscopy
21 structures of intasomes (*i.e.* the entire IN complex with bound viral DNA ends) have been
22 published for spumavirus, lentivirus, alpha-retrovirus and beta-retrovirus genera (Ballandras-
23 Colas et al., 2016; Ballandras-Colas et al., 2017; Hare et al., 2010; Passos et al., 2017; Yin et
24 al., 2016). Concerning gamma-retroviruses, the isolated NTD and CTD structures of MLV IN
25 are known (Aiyer et al., 2014; Guan et al., 2017), and we have recently published a model of
26 the Porcine Endogenous Virus A/C (PERV-A/C) intasome (Demange et al., 2015). BET
27 proteins also display a conserved modular architecture and the isolated structures of BD-I,
28 BD-II and ET have been determined, each with an all- α fold ((Lin et al., 2008); for a review,
29 see (Sanchez and Zhou, 2009)). The ET binding motif (EBM) of MLV IN was located within
30 the C-terminal tail of the CTD (Aiyer et al., 2014), as observed in the nuclear magnetic
31 resonance (NMR) structure of MLV IN EBM complexed with BRD4 ET (Crowe et al., 2016).
32 The latter study revealed that the 17-long flexible EBM segment (MLV IN aa 389–405) folds
33 into a β -hairpin when it binds the ET domain (Crowe et al., 2016).

1 We have previously shown that the integration of the gamma-retrovirus PERV-A/C
2 (hereafter referred to as PERV) in human cells is not random and is strongly favored at
3 transcriptional start sites and CpG islands (Moalic et al., 2006; Moalic et al., 2009). These
4 results revealed similarities between PERV and MLV integration site selection, suggesting
5 that gamma-retroviruses have a common pattern of integration and that their cellular
6 integration target cofactors might be similar. However, this assumption remains to be
7 demonstrated. In this study, we show *in vitro* and *in cellulo* that PERV IN interacts strongly
8 with BET proteins like MLV IN, in contrast with the behavior exhibited by ASLV and HIV-1
9 INs. We further perform small-angle X-ray Scattering (SAXS) experiments on the PERV IN
10 CTD – BRD2 ET complex and generate a 3D model that corroborates our interaction results.

11

12 **Materials and Methods**

13 *Plasmids*

14 *Plasmids for the two-hybrids assay.* PERV, MLV, HIV-1 and RAV-1 INs — as well
15 as BRD2 encoding sequences — were amplified by PCR from the recombinant PERV-A/C
16 clone AY570980 (supplied by Takeuchi, Y.), the pHit60 (supplied by S. P. Goff), pET15b-
17 IN-HIV-1 (Cellier et al., 2013), pET30a-IN-RAV-1 (Moreau et al., 2004), GST-
18 Brd2/RingG3B, GST-Brd4/HUNK, GST-Brd43/ORF XRingG3B (supplied by T. Schulz),
19 respectively, using appropriate primers (Tables S1 and S2). The resulting fragments were
20 cloned into either the pGBKT7 plasmid (for INs) or the pGAD-T7 plasmid (for BRD2, 3, 4)
21 using the Matchmaker Gold Yeast Two Hybrid System (*Clontech*). Deletion constructs were
22 generated for PERV IN (pGBKT7-PERV IN (1–98), pGBKT7-PERV IN (99–272), pGBKT7-
23 PERV IN (273–405), pGBKT7-PERV IN (1–272), pGBKT7-PERV IN (99–405), pGBKT7-
24 PERV IN (99–394), pGBKT7-PERV IN (99–383), pGBKT7-PERV IN (372–405) and for
25 MLV IN (pGBKT7-MLV IN (1–276), pGBKT7-MLV IN (277–408) by PCR fusion of DNA
26 fragments encoding the corresponding amino acid (aa) residues from either the PERV or the
27 MLV INs, respectively, with the Gal4-DNA binding domain. Point mutations (pGBKT7-
28 PERV INmut (P387A/L388A), pGBKT7-PERV INmut (K389A/L390A), pGBKT7-PERV
29 INmut (R391A/L392A), pGBKT7-PERV INmut (H393A/R394A) of the full-length PERV IN
30 were generated by PCR-directed mutagenesis with the appropriate primers. The
31 oligonucleotides used are listed in Table S1. All constructs were verified by DNA sequencing.

1 ***Cloning of the BRD2-encoding cDNA from HEK293T cells into the pCMV-Myc***
2 ***plasmid.*** The full-length BRD2 coding sequences were isolated from HEK293T cells. RNA
3 from HEK293T cells was isolated with the RNeasy kit (*Qiagen*) and converted to
4 complementary DNA (cDNA) by reverse transcription using the M-MLV RT (*Promega*) with
5 an oligo(dT) primer. The BRD2 encoding cDNA was then amplified by two independent
6 polymerase chain reactions (PCRs): (i) the first with the hF1 (position 1–22) and hR1
7 (position 1217–1196) primers, and (ii) the second with hF2 (position 1144–1162) and hR2
8 (position 2406–2379) primers (Table S3). PCR amplifications were performed with 35
9 amplification cycles (30 s at 95°C, 30 s at 60°C and 105 s at 72°C) using the Pfu DNA
10 polymerase (*Promega*). The two overlapping PCR fragments were then denatured (10 min at
11 95°C) and hybridized at room temperature for 30 minutes followed by an elongation step of
12 15 min at 72°C using the Pfu polymerase to generate the full-length cDNA of the BRD2 gene.
13 The BRD2 cDNA was then inserted into the mammalian expression vector pCMV-Myc
14 (*Invitrogen*) according to the Gateway technology (*Invitrogen*). This plasmid (named pCMV-
15 Myc-BRD2) contains the hCMV enhancer-promoter and an N-terminal Myc tag allowing the
16 expression of the Myc-BRD2 protein. Sequencing was performed to control for the integrity
17 of the cDNA.

18 ***Construction of the pPERV IN CTD – BRD2 (667–808) complex expression***
19 ***plasmid.*** An original plasmid construction was built to obtain the protein complex formed by
20 PERV IN CTD (aa 309–405) and BRD2 (aa 667–808), cloned from the pCI-Flag-PERV IN
21 and the pCMV-Myc-BRD2 plasmids respectively. This construction contains three genes
22 encoding: i) TEV protease, ii) PERV IN CTD with an N-terminal Strep-tag II and iii) BRD2
23 (667–808) with an N-terminal hexahistidine (His) tag (see Fig. 6A). At first, each gene was
24 amplified by PCR using the pRK793b, pGEM-T Easy and pPROEX-Hta plasmids,
25 respectively. The whole insert containing these three genes was generated by PCR. In
26 addition, the genes were separated from each other by TEV protease cleavage sites introduced
27 by the primers. The insert was cloned using an In-Fusion[®] HD-Cloning kit (*Clontech*) and
28 amplified by PCR in the pET28a vector.

29 The constructed expression plasmid was also used to generate one single and one
30 double alanine mutant affecting the complexed IN CTD (*i.e.* W379A and R391A/L392A), and
31 to generate the shorter pPERV IN CTD – BRD2 (667–735) complex expression plasmid also
32 termed pPERV IN CTD – BRD2 ET. All plasmids were confirmed by DNA sequencing.

1 **Construction of the pCMV-Myc-BRD2-, BRD3-, BRD4-Cter plasmids.** The pCMV-
2 Myc-BRD2 plasmid was used for the amplification of the BRD2 C-terminal fragments
3 (BRD2-Cter) by PCR using hF3 (position 1789–1807) and hR3 (position 2406–2389) primers
4 (Table S3). The BRD3 C-terminal fragment (BRD3-Cter) was amplified from pGVB3403
5 using the hF4 (position 1705–1722) and hR4 (position 2181–2161) primers, and the BRD4 C-
6 terminal fragment (BRD4-Cter) was amplified from pGVB3115 using hF5 (position 1819–
7 1836) and hR5 (position hF5-hR5) primers (Table S3). PCRs were performed with 35
8 amplification cycles (30 s at 95°C, 45 s 65°C and 160 s at 72°C) using the Pfu DNA
9 polymerase (*Promega*). The resulting fragment was inserted into the pCMV-Myc vector using
10 the Gateway technology as above and controlled by sequencing.

11 **Construction of the pCI-Flag-PERV IN plasmids.** The PERV IN coding sequence
12 was amplified from plasmid A14/220 containing the whole genome of PERV (Genbank ref.
13 AY570980) using attB11SP/PERV and attB21ASP/PERV IN primers (Table S3). PCRs were
14 performed with 35 amplification cycles (30 s at 95°C, 30 s 60°C and 120 s at 72°C) using the
15 Pfu DNA polymerase (*Promega*). The resulting fragment was inserted into the pDONR223
16 and pCI-Flag plasmids using the Gateway technology as above.

17 The pCI-Flag-PERV IN CTD was obtained by PCR amplification of the CTD of
18 PERV IN (aa 273–405) from pCI-Flag PERV IN by PCR amplification using F and R
19 primers. The resulting fragment was inserted into the pCI-Flag plasmid using the Gateway
20 technology. The pCI-Flag PERV IN CTD (P387A/L388A) plasmid containing a CTD
21 mutated was obtained by PCR-directed mutagenesis with the appropriate primers.

22 *The two-hybrid assay*

23 The yeast reporter strain Y2H-Gold containing the *HIS3* and *ADE2* Gal4-inducible
24 genes were co-transformed and plated on a selective medium lacking tryptophan and leucine.
25 Y2H-Gold double transformants were patched on the same media and then analyzed for
26 histidine and adenine auxotrophy by replica plating on selective medium lacking tryptophan,
27 leucine, histidine and adenine (-His / -Ade).

28 *Cells and immunoassays*

29 **Cells.** Human embryonic kidney HEK293T cells were cultured in Dulbecco's
30 modified eagle medium (PAA), 10% fetal bovine serum (PAA) and 1% penicillin-
31 streptomycin (Gibco) at 37°C in 5% CO₂.

1 **Confocal immunofluorescence microscopy.** For immunofluorescence microscopy,
2 HEK293T cells were grown in LabTek II chamber slides (*BD Falcon*) for 24 hours and
3 transfected with 800 ng of plasmid using Lipofectamine 2000 (*Invitrogen*). Forty-eight hours
4 later, the transfected HEK293T cells were fixed with 1% paraformaldehyde in phosphate-
5 buffered saline (PBS) overnight at 4°C, washed with PBS and then incubated in a
6 blocking/permeabilization buffer (BP) (10% FBS and 0.1% saponin in PBS) for 10 min at
7 4°C. After incubation, the cells were washed twice with the BP buffer and incubated with the
8 primary antibodies. The Myc-BRD2 fusion protein and the Flag-PERV IN fusion protein were
9 detected using a mix of polyclonal rabbit antibody directed against the Myc tag (*Sigma-*
10 *Aldrich*, catalog no. C3956, dilution 1/500) and a monoclonal mouse antibody directed against
11 the Flag tag (*Sigma-Aldrich*, catalog no. F3165; dilution 1/500) in the BP buffer for one hour
12 at 4°C. After incubation, the cells were washed twice with the BP buffer and incubated with
13 the alexa-488 conjugated goat anti-rabbit antibody (*Invitrogen*, catalog no. A11008, dilution
14 1/500) or the alexa-633 goat anti-mouse antibody (*Invitrogen*, catalog no. A21053; dilution
15 1/500), respectively. The cells were then mounted in the Vectashield Mounting Medium
16 (*vector Laboratories*) containing 1.5 µg/ml of 4',6-diamidino-2-phenylindole (DAPI) to stain
17 the nuclear DNA. Images were acquired with a Confocal Spectral SP5 (*Leica*) controlled by
18 LAS AF software (*Leica Microsystems CMS GmbH*) with a 63x oil immersion lens.

19 **Co-Immuno Precipitation (Co-IP) assays.** HEK293T cells were plated at a
20 confluence of 60-70% in 6-well dishes. The following day, cells were co-transfected with 2
21 µg of each of two plasmids: (i) the pCI-Flag, the pCI-Flag-PERV IN CTD or the pCI-Flag-
22 PERV IN CTD (P387A/L388A) plasmids and (ii) the pCMV-Myc-BRD2-Cter, pCMV-Myc-
23 BRD3-Cter or pCMV-Myc-BRD4-Cter encoding plasmids. Transfection was carried out
24 using the Lipofectamine 2000 transfection reagent (*Invitrogen*). Forty-eight hours later,
25 transfected cells were trypsinized, pelleted and resuspended in the IP buffer (50 mM Tris, pH
26 8, 150 mM NaCl, 1% Nonidet-P40, 0.5% sodium deoxycholate) supplemented with a
27 complete EDTA-free protease inhibitor cocktail (*Roche*). The resuspended cells were then
28 disrupted by sonication and incubated on ice for 30 min before being centrifugated for 10 min
29 at 4°C at 12,000 g. The protein concentration of the supernatant (cell lysate) was determined
30 by the Bradford assay. For the IP reaction, 20 µl of pure proteome Protein G Magnetic Beads
31 (*Millipore*) were washed and incubated for 15 min in the IP buffer with 4 µg of a polyclonal
32 rabbit anti-Myc (*Sigma-Aldrich*, catalog no. C3956) on a rotator at room temperature. The
33 magnetic Myc-beads were then washed three times with 500 µl of the IP buffer before the

1 addition of 500 µg of the cell lysate to 600 µl of the magnetic Myc-beads in the IP buffer. The
2 mixture was then incubated overnight at 4°C on a rotator. The next day, the magnetic beads
3 were washed three times with 500 µl of IP buffer and then boiled in 60 µl of Laemmli buffer
4 containing 5% β-mercaptoethanol. The protein mixture was analyzed by SDS-PAGE and
5 immuno-detected by Western blotting.

6 **Immunodetection.** For immunoblotting, 20 µg of the whole cell extracts or 20 µl of
7 the immuno-precipitated proteins were separated on a 10% SDS-PAGE. The proteins were
8 transferred to a nitrocellulose membrane and detected either (i) with a mouse anti-Myc
9 antibody (*Sigma-Aldrich*, catalog no. M4439, dilution 1/5000) and an anti-mouse IgG-
10 Horseradish Peroxidase antibody (*Sigma-Aldrich*, catalog no. A5906, dilution 1/10,000) or (ii)
11 with an anti-Flag M2 Horseradish Peroxidase (HRP) antibody (*Sigma-Aldrich*, catalog no.
12 A8592, dilution 1/10,000). The Western blots were developed using the ECL
13 chemiluminescent HRP substrate (*Thermo Scientific*).

14 *Protein purification*

15 **Purification of the PERV IN CTD – BRD2 ET complex.** After transformation, the
16 protein complex was expressed in T7 Express competent *E. coli* in LB broth by induction
17 with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37°C for 3 hours. Bacteria were
18 harvested by centrifugation at 6,000 g for 10 min and stored at -20°C. The thawed bacterial
19 pellet was resuspended in buffer A (20 mM Tris HCl, pH 8.0, 0.5 M NaCl, 25 mM imidazole)
20 in the presence of lysozyme (0.2 mg/ml, *Sigma*) and complete ULTRA EDTA-free protease
21 inhibitor cocktail 1X (*Roche*). Following sonication for 8 min at 40 mA on ice, the lysate was
22 clarified by centrifugation at 14,000 g for 45 min and filtered through a 0.45 µm filter (*VWR*
23 *International*). The supernatant containing soluble His-tagged BRD2 ET, Strep-tagged II
24 PERV IN CTD and the resulting double-tagged complex was loaded on a Ni²⁺ charged 5 ml
25 Protino[®] Ni-NTA (*Macherey Nagel*) equilibrated with buffer A using the ÄKTA purifier
26 system (*GE Healthcare*). The column was extensively washed with buffer B (20 mM Tris
27 HCl, pH 8.0, 1 M NaCl, 25 mM imidazole). The His-tagged BRD2 ET and the double-tagged
28 complex were eluted by a linear gradient of imidazole (0.025–1 M). Eluted fractions were
29 diluted 1:4 with buffer C (20 mM Tris HCl, pH 8, 0.5 M NaCl) and immediately loaded on a
30 5 ml Strep-Tactin affinity column (Strep-Tactin Superflow[®] cartridge, *IBA*) equilibrated with
31 buffer W (100 mM Tris HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA). The double-tagged
32 complex was eluted with a desthiobiotin-containing buffer (100 mM Tris HCl, pH 8.0,
33 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin). After 15% SDS-PAGE analysis,

1 purified protein complex aliquots were flash frozen in liquid N₂ and stored at -80°C. Each
2 partner of the complex was identified by western blot and MALDI-TOF (Voyager DE-PRO,
3 *AB Sciex*).

4 ***Purification of the PERV IN CTD – BRD2 (667–806) complex.*** A similar protocol
5 was used to purify the complex and the two related mutants (*i.e.* W379A, and R391A/L392A
6 mutated on IN).

7 *Structural studies*

8 ***Determination of the oligomeric state of the complex.*** PERV IN CTD – BRD2 ET
9 was concentrated at 14 mg/ml using a Vivaspin-PES-10K centrifugal concentrator (*Sartorius*)
10 and loaded on Superdex 200 10/300 GL (*GE Healthcare*) equilibrated with buffer C for
11 analysis by Size Exclusion Chromatography (SEC). The main peak fraction was concentrated
12 to 10 mg/ml and analysed by Dynamic Light Scattering (DLS) with a Zetasizer Nano S ZEN
13 1600 instrument (*Malvern Instruments*). Each measure is the mean of 10 runs repeated three
14 times. A cross-linking experiment with glutaraldehyde was performed by mixing 10 µg of
15 purified protein complex, 20 µl of 20 mM Tris HCl, pH 8 and 5 µl of 0.1% glutaraldehyde.
16 After 20 minutes of incubation at room temperature, the cross-linking reaction was stopped by
17 adding 5 µl of 200 mM glycine. The resulting protein sample was loaded onto a 15% SDS
18 PAGE. A similar protocol was used for PERV IN CTD – BRD2 (667–808).

19 ***SAXS study of the complex.*** SAXS data were collected at the European Synchrotron
20 Radiation Facility (ESRF) beamline BM29, Grenoble, France. Sample and buffer were
21 centrifuged at 14,000 g for 10 min before acquisition to remove aggregates. The complex, at a
22 concentration of 15 mg/ml, was injected in SEC coupled mode using an analytical Superdex
23 S200 increase 5/150 GL column (*GE Healthcare*) equilibrated in buffer C. The temperature
24 was maintained at 4°C. Processing and analysis of SEC-SAXS data were performed with the
25 programs CHROMIXS and SAS-DATA of the ATSAS suite (Franke et al., 2017). The radius
26 of gyration (R_g) was calculated from both the Guinier plot and the pair-distribution function
27 $P(r)$. The maximum inter-atomic distance of particle (D_{max}) was calculated from $P(r)$ analysis.
28 *Ab initio* bead models were generated using DAMMIN. After superposition and averaging,
29 the filtered SAXS envelope was displayed with CHIMERA (Pettersen et al., 2004).

30 A sequence search performed with SWISS-MODEL (Waterhouse et al., 2018) showed
31 that the C-terminal fragment 736–808 of BRD2 has no homologous sequence of known

1 structure in the Protein Data Bank (PDB). In consequence, the generation of a SAXS-
2 restrained atomic model was possible for PERV IN CTD – BRD2 ET (667-735) only.

3 First, the NMR structure of MLV IN EBM in complex with BRD4 ET (PDB code
4 2n3k) was taken as a guide to model with the program SCULPTOR (Bunkoczi and Read,
5 2011) the structure of PERV IN EBM (aa 378–394) in complex with BRD2 ET. This model
6 of the bound C-terminal tail was linked with an arbitrary relative position and orientation with
7 the structure of the remaining CTD (aa 309–377) extracted from our model of the PERV
8 intasome (Demange et al., 2015). The resulting homology model was fitted against
9 experimental SAXS data with the program DADIMODO (Evrard et al., 2011) using a two
10 domain rigid-body strategy with the loop 376–380 connecting the IN CTD core to the bound
11 EBM segment rendered as flexible. The goodness of fit was estimated by χ^2 calculation with
12 CRY SOL (Franke et al., 2017). The resulting SAXS-restrained model was positioned with
13 CHIMERA into the SAXS envelope for observation.

14

15 **Results**

16 *The human BRD2 protein interacts specifically with PERV and MLV INs.*

17 We first analyzed the interactions between different retroviral INs and BRD2 in the
18 yeast two-hybrid assay (Fig. 1B). We tested the INs of the gamma-retroviruses PERV and
19 MLV, the lentivirus HIV-1 and the alpha-retrovirus Rous Associated Virus type 1 (RAV-1)
20 (Fig. 1B, left table). The full-length PERV IN (aa 1–405), MLV IN (aa 1–408), HIV-1 IN (aa
21 1–288), and RAV-1 IN (aa 1–286) were fused to the Gal4-DNA binding domain (Gal4-BD).
22 The full length ORF of BRD2 (836 residues long) was fused to the Gal4 activation domain
23 (Gal4-AD). The yeast strain was analyzed for histidine and adenine auxotrophy (-His/-Ade)
24 acquired by the interaction of the fused proteins (Fig. 1B, left table). As expected, no growth
25 was observed if PERV and MLV INs or BRD2 were not fused to their respective Gal4-BD
26 and -AD genes. In accordance with others (Gupta et al., 2013), we detected no interaction
27 between RAV-1 or HIV-1 INs and BRD2. We confirmed that MLV IN interacts with BRD2
28 and showed that PERV IN gave a positive interaction (Fig. 1B).

29 *PERV IN and BRD2 co-localize in the nucleus of cells*

30 To further investigate the interaction between BRD2 and PERV IN, we used
31 immunofluorescent microscopy to examine the localization of the full-length Flag-PERV IN

1 and the Myc-BRD2 proteins in HEK 293T cells. In the first experiment, the pCMV-Flag-
2 PERV IN and the pCMV-Myc-BRD2 plasmids were individually transfected into HEK 293T
3 cells. The correct expression of each protein was checked by western blot analysis (data not
4 shown). The cellular distribution of each fusion protein was analyzed using confocal
5 fluorescent microscopy 48 hours later (Fig. 2). More than 30 positive cells were observed in
6 each sample. In HEK 293T cells, exogenously expressed Flag-PERV IN was mostly detected
7 in the nucleus (either diffused (top images) or as punctuate nuclear foci (bottom images)) with
8 a minority expression in cytoplasm (Fig. 2A). In HEK 293T cells, exogenously expressed
9 Myc-BRD2 showed diffuse nuclear localization (main pattern) or both a nuclear and
10 cytoplasmic localization (Fig. 2B).

11 In the second experiment, the pCMV-Flag-PERV IN and the pCMV-Myc-BRD2
12 plasmids were co-transfected into HEK 293T cells. Several patterns of co-localization (in
13 yellow) were observed (Fig. 2C): (i) at the internal periphery of the nucleus (top images) and
14 (ii) diffuse and punctuate in the nucleus (middle images). We also observed a chromosomal
15 co-localization in a dividing cell (bottom images). Although a χ^2 statistical test did not reveal
16 any change in the localization of the proteins — whether they were expressed alone or
17 together (data not shown) — these co-localization studies provide further evidence supporting
18 an interaction between PERV IN and BRD2 in cytoplasm and nucleus of cells.

19 *The BRD2 binding site maps within the C-terminal tail of PERV IN CTD*

20 The retroviral IN protein contains three evolutionarily conserved and functional
21 domains interconnected with flexible linkers: the N-terminal domain (NTD), the catalytic core
22 domain (CCD) and the C-terminal DNA binding domain (CTD) (Fig. 1A), all necessary for
23 the biological activities of the IN protein. The predicted structure of PERV IN also comprises
24 a singular 46-residue long N-terminal extension domain (NED) (Demange et al., 2015). This
25 NED structure is also observed in MLV and prototype foamy virus (PFV) IN.

26 In order to map the region of PERV IN involved in the interaction with BRD2, we
27 constructed a total of five IN truncation mutants fused with Gal4-DB: *i.e.* the one-domain
28 fragments NTD (aa 1–98), CCD (aa 99–272) and CTD (aa 273–405) and the two-domain
29 fragments NTD+CCD (aa 1–272) and CCD+CTD (aa 99–405) (see Fig. 1A). These five
30 fragments were tested in the yeast two-hybrid system against the Gal4-AD-BRD2 protein
31 (Fig. 3A). BRD2 interacts with PERV IN (99–405) and IN (273–405) but not with IN (1–98),
32 IN (99–272) and IN (1–272). Therefore, only constructs containing PERV IN CTD can bind
33 efficiently to BRD2 and the NTD and CCD are dispensable. Thus, our data showed that the

1 CTD alone (aa 273–405) is necessary and sufficient for the interaction with BRD2. Similar
2 deletion mutants were constructed with MLV IN and tested using the yeast two-hybrid system
3 against Gal4-AD-BRD2 (supplementary Fig. S1). In accordance with data from other studies
4 (De Rijck et al., 2013; Gupta et al., 2013; Sharma et al., 2013), we observe that the CTD of
5 MLV IN is sufficient for the interaction with BRD2. Altogether, these results demonstrate that
6 INs from gamma-retroviruses interact specifically with BRD2 through their CTD domains.

7 To characterize more precisely the peptide sequence of IN required for the interaction,
8 we used two CCD+CTD constructs deleted from the C-terminal sequences of the CTD, PERV
9 IN (99–394) and IN (99–372) (Fig. 3B). The truncation of the 11 C-terminal residues of the
10 CTD (PERV IN (99–394)) did not disturb the interaction. By contrast, removal of the 33 C-
11 terminal residues of the CTD in PERV IN (99–372) resulted in a loss of interaction. This
12 deletion study indicated that several residues from the segment 372–394 are necessary to
13 mediate the interaction with BRD2 (Fig. 3B). We also tested positively the short IN fragment
14 372–405 carrying this interaction sequence (Fig. 3B) showing that the C-terminal tail of
15 PERV IN (residues 372–405), alone, is sufficient for interaction with BRD2.

16 We then focused on the most C-terminal motif of CTD encompassing residues 378–
17 394 in PERV IN. This segment corresponds to the MLV IN EBM (aa 389–405) which is
18 critical for the interaction with BRD4 (Crowe et al., 2016). The primary sequence of this
19 region is highly conserved among the gamma-retroviral INs (Fig. 4A) and not between
20 retroviral genera (supplementary Fig. S2). To assess the importance of these conserved
21 residues, we performed a two-hybrid assay using five different mutants: the W379A single
22 mutant (mutant containing an Alanine in place of the W379) and the P387A/L388A,
23 K389A/L390A, R391A/L392A and H393A/R394A double mutants (Fig. 4B left panel). All of
24 these mutants affect the interaction between PERV IN and BRD2 (Fig. 4B right panel).
25 Notably the W379A single mutant disrupts the interaction in accordance with what was
26 previously shown for the equivalent MLV W390A IN mutant and BRD4 (El Ashkar et al.,
27 2014). This means that not only W379 but at least one residue in each double mutant (387–
28 388, 389–390, 391–392 and 393–394) is important for the formation of the PERV IN – BRD2
29 complex.

30 *PERV IN interacts with the ET domain of different BRD proteins in the two-hybrid assay and*
31 *in HEK293T cells*

32 BET proteins such as BRD2 (Fig. 5A), BRD3 and BRD4 contain two conserved
33 tandem bromodomains and a conserved ET domain. The ET from BRD4 is known to be

1 involved in the interaction with MLV IN (Crowe et al., 2016; Gupta et al., 2013). We first
2 tested the interaction of PERV IN with ET from different BET proteins using the two-hybrid
3 system (Fig. 5B). A positive signal was detected with all constructs. The BRD4 ET fragment
4 alone (aa 607–671), was sufficient for this interaction whereas the use of smaller fragments of
5 ET of BRD2 or BRD4 did not show any interaction (data not shown).

6 We also investigated the interaction of the PERV IN CTD and the hBET C-terminal
7 domain in HEK 293T cells. PERV IN CTD and hBET ET-containing C-terminal region
8 (BRD2-Cter, BRD3-Cter and BRD4-Cter) were tagged with the Flag and Myc tags,
9 respectively. The Flag-PERV IN CTD was tested for its ability to co-immuno-precipitate the
10 Myc-hBET-Cter after transfection in HEK 293T cells (Fig. 5C). As a control, we used a
11 PERV IN CTD (P387A/L388A) mutant containing two alanines in place of P387 and L388 in
12 the interacting domain of IN. The Flag-PERV IN-CTD and the Flag-PERV IN-CTD
13 (P387A/L388A) co-immunoprecipitation with Myc-hBET-Cter were assessed in the Myc-
14 immunoprecipitated fractions using a Flag antibody (Fig. 5C). Data showed that the Flag-
15 PERV IN-CTD was co-immunoprecipitated with Myc-BRD2-Cter, Myc-BRD3-Cter and
16 Myc-BRD4-Cter (panel D, lanes 2, 5 and 8). The mutation of two residues in the Flag-PERV
17 IN-CTD (P387A/L388A) diminished but did not fully abolish the interaction with the Myc-
18 BRD2-Cter and Myc-BRD4-Cter (panel D, lanes 3 and 9, respectively). By contrast, the
19 interaction of Myc-BRD3-Cter with the Flag-PERV IN-CTD (P387A/L388A) was completely
20 lost (panel D, lane 6).

21 *Co-purification of the two proteins as a complex*

22 With the aim of strengthening our data and obtaining novel structural information
23 about the IN/Brd2 complex, we overexpressed and purified recombinant PERV IN CTD and
24 the BRD2 C-terminal fragment from a prokaryotic expression system and performed *in vitro*
25 interaction assays.

26 To date, structural knowledge on the C-terminal domain of BET proteins is limited to
27 the ET (Crowe et al., 2016; Sanchez and Zhou, 2009). A prediction of intrinsic disorder by the
28 GLOBPLOT server (Linding et al., 2003) shows that the C-terminal region of BRD2 (aa 667–
29 836) that contains the ET (aa 676–738) is mostly ordered with the exception of the C-terminal
30 extremity which is rich in serine residues (Supplementary Fig. S3). Since this long flexible
31 tail can prevent crystallization of the protein and — in doing so — the use of X-ray
32 crystallography to determine the atomic structure of the complex, we worked on the BRD2
33 fragment without the C-terminal tail. We produced independently PERV IN CTD (aa 309–

1 405) and the BRD2 fragment (aa 667–808) to form the complex, but an aggregation
2 phenomenon was observed when the two partners were mixed together (data not shown). To
3 overcome this problem, we used a new strategy of protein co-expression inspired by Levy and
4 co-workers that have already expressed HIV-1 IN and LEDGF/p75 recombinant proteins from
5 a sophisticated polycistronic construction in a mammalian cells expression system (Levy et
6 al., 2016). We optimized the same kind of construction for a bacterial expression system (Fig.
7 6A). A Strep-tagged-PERV IN fragment (aa 309–405) and a His-tagged-BRD2 fragment (aa
8 667–808) were cloned in frame with Tobacco Etch Virus (TEV) protease and interspaced
9 with TEV digestion sites to form a unique, cleavable, polyprotein. During overexpression, the
10 TEV protease separates the Strep-PERV IN CTD and the His-BRD2 (667–808) of the
11 polyprotein leading to formation of the complex. The complex was then purified by two
12 successive affinity steps with nickel and Strep-Tactin column (Fig. 6B top left panel). The
13 two partners were identified by western blot using anti-His tag and Anti-Strep tag antibodies
14 (Fig. 6B bottom left panels) and mass spectrometry from gel bands (data not shown). When
15 using a wild type Strep-PERV IN CTD and the His-BRD2 (667–808), we were able to co-
16 purify both proteins in the elution of the nickel column (lane 2) and also in the elution sample
17 of the strep column (lane 3). This indicated that both partners, His-BRD2 (667–808) and
18 Strep-PERV IN CTD were indeed co-purified as a complex form and not as two isolated
19 proteins.

20 In order to assess the interaction of the previously described IN mutants with BRD2,
21 the PERV IN CTD was modified within this construction to obtain the PERV IN CTD-
22 W379A and PERV IN CTD- R391A/L392A mutants. Using these constructions, both proteins
23 were efficiently co-expressed (Fig. 6B, lanes 4 and 7). However, after passing on the nickel
24 column, although a small proportion of the strep-PERV IN CTD mutants were retained (lanes
25 5 and 8), the His-BRD2 (667–808) fragment was not retained on the Strep column (lanes 6
26 and 9). This observation strongly supports the interaction of BRD2 with PERV IN CTD and
27 assesses the critical role of each conserved residue of the IN EBM region.

28 We also performed the co-purification of the shorter PERV IN CTD – BRD2 complex,
29 hereafter referred as to PERV IN CTD – BRD2 ET, by using the TEV expression system. We
30 confirmed the interaction of this new two-fragment complex *in vitro*, as described below.

31 *The complex is heterodimeric in solution and assembles as distinct modules*

32 We next investigated the oligomeric state of the complex by Size Exclusion
33 Chromatography (SEC) and cross-linking approaches.

1 In more details for PERV IN CTD – BRD2 ET, the SEC experiment indicated one
2 main peak at an elution volume of 16.5 ml and a minor peak at an elution volume of 17.5 ml,
3 corresponding to ~23 kDa (heterodimer) and ~13 kDa (IN CTD), respectively (Supplementary
4 Figs. S4A and S4B). This demonstrated that the two-fragment complex is mainly
5 heterodimeric at 14 mg/ml in 20 mM Tris HCl pH 8.0, 500 mM NaCl. Fractions from the
6 main peak were pooled, concentrated and sample homogeneity was assessed by diffuse light
7 scattering, resulting in a satisfactory polydispersity index of 0.2.

8 The heterodimeric state of PERV IN CTD – BRD2 ET was further confirmed by a
9 chemically cross-linking assay using glutaraldehyde (Supp. Fig. 4C). Indeed, in the presence
10 of glutaraldehyde, the covalently bound PERV IN CTD and BRD2 ET migrate with a higher
11 molecular weight of about 23 kDa (shown on lane 1). The heterodimeric state of the cross-
12 linked sample was confirmed by mass spectrometry from gel bands (data not shown). The
13 heterodimeric state of the cross-linked PERV IN CTD – BRD2 (667-808) was checked with a
14 similar protocol (data not shown).

15 Crystallization trials were performed on the two purified complexes but without
16 success. The two samples were analyzed by SAXS using a synchrotron radiation source and
17 the collected HPLC-SAXS data were processed (Table 1 and Fig. 7 for PERV-A/C IN CTD -
18 BRD2 ET). The calculated Kratky plots yielded bell-shaped curves at low q values but did not
19 converge at high q (Fig. 7B), indicating partial flexibility which might have hindered our
20 crystallization attempts. The linearity of the Guinier plots were indicative of monodisperse
21 solutions and similar R_g values were obtained from Guinier and $P(r)$ analyses (Table 1 and
22 Fig. 7C). *Ab initio* modeling led to elongated shape envelopes with an estimated resolution of
23 3 nm (Fig. 7D).

24 PERV IN CTD and BRD2 ET share significant sequence identities with MLV IN CTD
25 and BRD4 ET (60% and 86%, respectively) whose experimental structures are known (Aiyer
26 et al., 2014; Crowe et al., 2016). An atomic model of PERV IN CTD – BRD2 ET was
27 generated by homology modeling, assuming that the bound peptide tail of the CTD (aa 378–
28 394) adopts a β -hairpin fold like the bound MLV IN EBM (Crowe et al., 2016). This model
29 was refined against related SAXS data using a two-domain rigid body strategy with a flexible
30 CTD loop (aa 373–377) in order to experimentally position the SH3 core domain of the IN
31 CTD (aa 309–372) relative to the bound EBM segment. The resulting SAXS-restrained model
32 had an excellent dependence to the experimental data with a χ^2 of 1 (Fig. 7A), and it exhibited
33 two distinct structural blocks connected by the five-residue loop 373–377 (Fig. 7D). Such

1 modular structure gives flexibility to the heterodimer and further explains why the EBM
2 segment is sufficient and essential for IN - BET assembly.

3

4 **Discussion**

5 Herein, we show by using two-hybrid experiments (Figs 1, 3–5), co-localization into
6 cells (Fig. 2), co-purification and co-immunoprecipitation (Fig. 6), and SAXS (Fig. 7) that the
7 gamma-retroviral PERV A/C IN interacts with the hBET proteins BRD2, BRD3 and BRD4 as
8 does MLV IN. The interaction with hBET proteins seems specific to gamma-retroviral INs
9 since BRD2 does not interact with either RAV-1 (alpha retrovirus) or HIV-1 (lentivirus) INs
10 (Fig. 1). Altogether, these results show that interaction between BET proteins and INs is a
11 common feature of gamma-retroviruses and is specific to this retroviral genus. This specificity
12 may explain the similarity of integration site distribution between PERV and MLV (Moalic et
13 al., 2006; Moalic et al., 2009), and the difference with other retroviruses from other genera.

14 The CTD of PERV IN is responsible for the interaction with hBET proteins *in vitro*
15 and *in cellulo* (Figs. 3–7). We also observed by two-hybrid and *in vitro* interaction assays that
16 mutations in a 17-residue long peptide (aa 378–394), which pertains to the PERV IN CTD C-
17 terminal peptide tail, abolish the interaction with BRD2. This segment is equivalent in
18 sequence position with the BET-interaction motif EBM characterized in MLV IN (Crowe et
19 al., 2016; Larue et al., 2014) (Fig. 4A), thereby confirming that gamma-retroviral INs share a
20 common mechanism of selection of BET proteins. We also observed that the W379 PERV IN
21 residue in the CCD is essential for interaction with BRD2 in accordance with other authors
22 that established that the corresponding residue W390 from the MLV IN CTD is a critical
23 hotspot for interaction with the BRD4 protein (El Ashkar et al., 2014). We also showed that,
24 besides W379, at least one residue in each double mutant of the PERV IN EBM 387–394
25 stretch is also essential for the interaction. Others reported that residues (E266, L268 and
26 Y269 of MLV IN CCD) — located outside the MLV EBM motif — are also important for the
27 interaction with BET proteins (Gupta et al., 2013), but substitution of these residues in the
28 context of the full-length MLV IN did not abolish the interaction with Brd4 (El Ashkar et al.,
29 2014). Finally, PERV IN CTD was shown to interact with the conserved BET ET domain
30 (Fig. 5) as does MLV IN (Aiyer et al., 2014; Crowe et al., 2016).

31 The interaction with BET protein was shown to be important for the nuclear
32 localization of MLV IN. Indeed, when expressed separately, MLV IN and BET proteins show

1 a diffuse cytoplasmic and nuclear localization, respectively. When co-expressed with BET
2 proteins, IN MLV changes localization from cytoplasmic to nuclear (Gupta et al., 2013). By
3 contrast, we found that the exogenously expressed full-length Flag-PERV IN and Myc-BRD2
4 proteins were detected mainly in the nucleus of 293T cells when expressed alone, while only
5 a slight diffuse cytoplasmic localization was observed for both proteins. We did not observe
6 any change in the localization of the Flag-PERV IN when co-expressed with the Myc-BRD2
7 protein. However, we clearly detected a co-localization of the two protein domains in the
8 nucleus and cytoplasm of cells when expressed together, which confirms the interaction of
9 both proteins *in cellulo* and strengthens the role of the BET protein to direct PERV IN
10 integration (Fig. 2).

11 All of these results are in accordance with results obtained by others who showed that
12 BET proteins interact with the MLV IN protein and that both proteins co-localize into the
13 nucleus of the cells (De Rijck et al., 2013; Gupta et al., 2013). Interaction between the two
14 partners has been shown to involve the CTD of IN and the ET domain of BET proteins (Gupta
15 et al., 2013; Sharma et al., 2013). Moreover, disruption of the MLV IN – BET interaction by
16 using specific BET inhibitors, blocks MLV replication at the integration step and — when
17 integration occurs — retargets MLV integration away from transcription start sites (De Rijck
18 et al., 2013; Gupta et al., 2013; Sharma et al., 2013).

19 We combined SAXS and homology modeling to generate the first experimentally
20 restrained atomic model of a PERV IN CTD – BET ET complex. This structural study reveals
21 that PERV IN CTD and BRD2 ET assemble as heterodimers to form a modular structure (Fig.
22 7D). The first part of the C-terminal tail of IN bridges the SH3 heart of the CTD to an
23 intermolecular domain that contains the PERV IN EBM and the BET ET.

24 Keeping in mind that the estimated resolution of the SAXS-derived envelope is 3 nm
25 and that atomic details are inferred by homology-modeling, the heterodimeric assembly is
26 likely stabilized by an intermolecular β -sheet with two EBM β -strands and one ET β -strand as
27 observed in the NMR structure of bound MLV IN EBM (Crowe et al., 2016). which shares a
28 strong sequence identity with PERV IN EBM – BRD2 ET.

29 The SAXS-restrained model can also be used with caution (at the atomic level,
30 structural features must be conserved in bound MLV IN EBM) to propose structural
31 explanations to our interaction results: (i) the single mutant W379A is disruptive and W379 is
32 modeled as a buried residue making numerous intramolecular interactions, suggesting that
33 this tryptophan is essential for the stability of the folded tail. Moreover, this residue is strictly

1 conserved in gamma-retroviral INs and known as a critical hotspot for interaction; (ii) the
2 disruptive double mutant P387A/L388A may alter the turn connecting the two β -strands of
3 the modelled bound EBM (Fig. 3) and hinder the induced folding process; (iii) the disruptive
4 mutants K389A/L390A and R391A/L392A (Figs. 4 and 6) cancel the -BHBH- motif of the
5 conserved gamma-retroviral stretch, where -B- and -H- are basic and hydrophobic residues.
6 Thus, the modelled intermolecular interactions between the basic side-chains and conserved
7 acidic residues of the neighboring ET β -strand could be essential to the stability of the
8 heterodimer. Of interest, an amphipathic -BHBH- binding motif has been observed in other
9 viral and cellular partners of BET proteins—such as the transcriptional regulator NSD3—
10 where residues equivalent in position to the pair R391/L392 are essential for ET recognition
11 (Zhang et al., 2016); (iv) the disruptive double mutant H393A/R394A affects the C-terminal
12 arginine of EBM, which is strictly conserved among gamma-retroviral INs and contacts the
13 critical tryptophan W379 in the model.

14 Our SAXS data also reveal that the PERV IN CTD – BRD2 ET complex possesses an
15 inherent flexibility due to a short loop connecting the bound IN EBM segment and the IN
16 CTD core. This flexibility is also observed in the bigger PERV IN CTD – BRD2 (667–808)
17 complex and might be advantageous in the context of the full-length proteins for a dynamic
18 selection of the integration sites. Moreover, the recent publications of numerous intasome
19 structures (Ballandras-Colas et al., 2016; Ballandras-Colas et al., 2017; Hare et al., 2010;
20 Passos et al., 2017; Yin et al., 2016) allowed us to gain information on the formation of the
21 PERV IN – BRD2 assembly.

22 The oligomerization state of the known intasome structure differs between retroviral
23 genera and seems to be related to the length of the IN CCD-CTD linker (for review, see
24 (Engelman and Cherepanov, 2017)). Gamma-retroviral intasomes should be tetrameric, as
25 observed in the PFV intasome (Engelman and Cherepanov, 2017). Our previously published
26 model of the PERV intasome respects this architecture (Demange et al., 2015) with two
27 symmetrical inner and outer monomers of INs. The two inner CTDs are positioned between
28 their corresponding inner NTDs and CCDs (Fig. 8), whereas the outer CTDs are not modeled
29 because they were too flexible to be observed in the PFV intasome. A superposition with our
30 SAXS-restrained model of the PERV IN CTD – BRD2 ET complex shows that the inner
31 CTDs are well accessible to the ET domain of the BET cofactor (Fig. 8). This accessibility
32 remains unchanged if we consider the intasome to be bound with target DNA as observed for
33 PFV (Maertens et al., 2010). Altogether, these data suggest that, before integration, the

1 accessible, flexible C-terminal tail of the inner IN CTD can hook the ET domain of the BET
2 cofactor. The induced β -folding of the IN EBM segment locks the two partners together,
3 while the short loop preceding IN EBM gives partial flexibility to the complex. This structural
4 combination participates in the efficient targeting of the bound vDNA to the promoters of
5 active genes.

6

7 **Conclusion**

8 Retroviral vectors are important tools and have been widely used to deliver DNA for
9 therapeutic purposes. However, retroviral integration can be associated with insertional
10 oncogene activation, especially for gamma-retroviral vectors due to their strong bias of
11 integration into transcription start sites. Eliminating the interaction of gamma-retroviral IN
12 with hBET family members has implications for the safety of human gene therapy and gene
13 delivery when using gamma-retroviral vectors. Besides, it has been shown that the PERV
14 recombinant virus can infect human cells and represents a potential major risk of zoonotic
15 disease in the case of xenotransplantation of pig organs to humans. We have previously
16 shown that Raltegravir (Demange et al., 2015), as well as AZT (Denner, 2017), are potent
17 inhibitors of PERV replication and could be useful for the treatment of PERV infections.
18 Eliminating the interaction of PERV IN with hBET proteins using inhibitors such as JQ1
19 would be another way to control this zoonotic risk

20 .

1 **Declarations**

2 **Ethics approval and consent to participate**

3 Not applicable

4 **Consent for publication**

5 Not applicable

6 **Availability of data and material**

7 The datasets used and/or analyzed during the current study are available from the
8 corresponding authors on reasonable request.

9 **Competing interests**

10 The authors declare that they have no competing interests.

11 **Funding**

12 This work was funded by the French Institutes INRA (Institut National de la Recherche
13 Agronomique), ANSES (Agence Nationale sécurité sanitaire de l'alimentation, de
14 l'environnement et du travail) and CNRS (Centre National de la Recherche Scientifique).

15 MC is supported by a PhD grant from the Rhône-Alpes-Auvergne region (N°18003930 ARC
16 Health 2016).

17 **Authors' contributions**

18 CR, YB, KM and PG designed the experiments, analyzed data and drafted the manuscript.
19 KG, GB, MC, HY, and FF made substantial contributions to conception, design, acquisition
20 and analysis of data and drafted the manuscript. MPC, CDB, AL, CL performed the
21 experiments. ML and YC discussed the conception and analysis of some experiments. All
22 authors read and approved the final manuscript.

23 **Acknowledgments**

24 We are grateful to the Production Sciences Facility (PSF) and to the Plateau Technique
25 Imagerie/Microscopie (PLATIM) platforms of UMS3444 BioSciences Gerland Lyon Sud. We
26 thank Lise-Marie Donnio ¹, Romain Appourchaux ¹, Anne-Sophie Guyomard ², Séverine
27 Calvo ² and Margot François ⁶ for technical assistance.

28

29

1 **Legends to figures**

2

3 **Figure 1. Interaction of PERV IN with the BRD2 protein**

4 *(A) Domain organization of the full-length gamma-retroviral PERV IN protein.* The IN
5 proteins are made up of three domains: (i) the N-terminal domain (NTD) which is involved in
6 the binding of the viral DNA and zinc, (ii) the catalytic core domain (CCD) which contains
7 the invariant catalytic triad D,D-35-E (indicated with asterisks), (iii) the C-terminal domain
8 (CTD) involved in non-specific DNA binding. The PERV also contains an N-terminal
9 extension domain (NED).

10 *(B) In the two-hybrid assay, INs from PERV and MLV gamma-retroviruses interact with*
11 *BRD2, while INs from HIV-1, a lentivirus, and from RAV-1, an alpha-retrovirus, do not.*

12 The full-length ORFs of PERV, MLV, HIV-1 and RAV-1 INs were fused to the Gal4-DNA
13 binding domain (Gal4-BD). The BRD2 ORF was fused to the Gal4-activation domain
14 (GAL4-AD). The yeast reporter strain Y2H-Gold producing the hybrid proteins indicated was
15 analyzed for histidine and adenine auxotrophy (-His / -Ade) acquired by the interaction of the
16 fused proteins.

17

18 **Figure 2. Intracellular co-localization of the full-length Myc-BRD2 and Flag-IN PERV** 19 **proteins in HEK 293T cells**

20 HEK 293T cells were transfected with the pCMV-Flag-PERV IN plasmid or/and with the
21 pCMV-Myc-BRD2 plasmids. 48 h after transfection, cells were fixed and confocal laser
22 scanning microscopy was used to study expression of the fluorescent fusion proteins. Cells
23 were also colored with DAPI to show DNA.

24 In **(A)**, cells were transfected with the pCMV-Flag-PERV IN plasmid. The DNA (in blue) and
25 the Flag (in red) signals and their overlays are shown in the left, middle, and right images of
26 each lane, respectively. Flag-PERV IN was localized to both the nucleus and—to a lesser
27 extent—the cytoplasm.

28 In **(B)**, cells were transfected with the pCMV-Myc-BRD2 plasmid. Myc (in green) and DNA
29 (in blue) signals and their overlays are shown in the top left, bottom left and right-hand
30 images of the panel, respectively. Expression of Myc-BRD2 is mainly nuclear (top and
31 bottom images) or both nuclear and cytoplasmic (bottom image).

1 In (C), 293T cells were co-transfected with both the pCMV-Flag-PERV IN and the pCMV-
2 Myc-BRD2 plasmids. The Flag-PERV IN (in red), Myc-BRD2 (in green) and DNA (in blue)
3 signals are shown on the left of the images. A Pearson coefficient (in brackets) greater than
4 0.75 indicates a good co-localization of the two proteins.

5

6 **Figure 3. The CTD domain of PERV IN (and more specifically, residues 372–405)**
7 **interacts with BRD2 in the two-hybrid assay**

8 (A) *The PERV IN CTD is involved in the interaction with BRD2 protein.* The NTD (aa 1–
9 98), CCD (aa 99–272) or CTD (aa 273–405) isolated domains of PERV IN as well as
10 combined domains of PERV IN (NTD+CTD (aa 1–272) and CCD+CTD (aa 99–405)) were
11 fused to the Gal4-BD domain. Interaction with BRD2 was evaluated as described in Figure 1.

12 (B) *Refinement of the PERV IN CTD interaction domain.* PERV IN fragments 99–394, 99–
13 372 and 372–405 were fused to the Gal4-BD domain. Interaction with BRD2 was evaluated
14 as described in Fig. 1B.

15

16 **Figure 4. Alanine replacements of W379 or amino acid pairs from residues 387–394 of**
17 **PERV IN abolish interaction with BRD2**

18 (A) *Multiple sequence alignment of the CTD of gamma-retroviral INs* from PERV, MLV,
19 FeLV, Gibbon ape Leukemia Virus (GaLV) and REV. Identical and well-conserved residues
20 (> 70% similarity) have a red and a yellow background, respectively. The 17-residue long
21 EBM segment is marked below sequence blocks with green stars, and the residues in PERV
22 IN we have mutated into Alanine (A) are written. Secondary structure elements are shown
23 above sequence blocks and colored in blue for the NMR structure of MLV IN CTD (PDB
24 code 2m9u) and colored in green for the NMR structure of MLV IN EBM in complex with
25 BRD4 ET (PDB code 2n3k). Secondary structure elements are shown above sequence blocks
26 (helices with squiggles, strands with arrows, turns with T letters). The figure was rendered
27 with ESPript.

28 (B) *Alanine scanning* of W379 (single mutation) and of residues in the well-conserved
29 stretch 387–394 (double mutations). The ability of the PERV IN mutants to interact with
30 BRD2 was analyzed in the two-hybrid assay as described in Fig. 1B.

31

1 **Figure 5. The C-terminal domains of BRD2, BRD3 and BRD4 proteins interact with**
2 **PERV IN in the two-hybrid and *in cellulo* assays**

3 **(A) Domain organization of the BRD2 proteins.** The hBET proteins contain several
4 functional domains: (i) the bromodomains I and II (BD1 and BD2) bind acetylated marks on
5 histones H3 and H4, (ii) the Extra-Terminal domain (ET) interacts with various viral and
6 cellular proteins. Adapted from (Weidner-Glunde et al., 2010).

7 **(B) Two-hybrid assay using the full-length PERV IN and the C-terminal domains of hBET**
8 **proteins.** The C-terminal domains of BRD2, BRD3 and BRD4 proteins encompassing the ET
9 domain, or the BRD4 ET domain alone were fused to the Gal4-activation domain (GAL4-
10 AD). The ability of the full-length PERV IN to interact with these domains was analyzed in
11 the two-hybrid assay as described in Fig. 1.

12 **(C) *In cellulo* assay: co-immunoprecipitation.** HEK 293T cells were co-transfected: (i) with
13 the pCMV-Myc-BRD2-Cter (lanes 1, 2 and 3), the pCMV-Myc-BRD3-Cter (lanes 4, 5 and 6)
14 or the pCMV-Myc-BRD4-Cter (7, 8 and 9) and (ii) with the pCI-Flag-PERV IN-CTD (lanes
15 2, 5 and 8) or the pCI-Flag-PERV IN-CTD (P387A/L388A) (lanes 3, 6 and 9). Forty-eight
16 hours later, cell lysates were collected. Expression of the Myc-BRD2-Cter, Myc-BRD3-Cter
17 and Myc-BRD4-Cter proteins and of the Flag-PERV IN CTD or Flag-PERV IN-CTD
18 (P387A/L388A) proteins was verified in the total cell lysates with anti-Myc and anti-Flag
19 antibodies, respectively (panels A and B). Myc-BRD2-Cter, Myc-BRD3-Cter and Myc-
20 BRD4-Cter proteins were then immuno-precipitated with the anti-Myc antibody and revealed
21 with the Myc antibody (panel C). Co-immunoprecipitation of the PERV IN-CTD and PERV
22 IN- CTD (P387A/L388A) proteins was assessed by a western blot analysis using an anti-Flag
23 antibody (panel D).

24

25 **Figure 6. Analysis of PERV IN CTD interaction with hBET C-terminal domains *in vitro***

26 **(A)** The pPERV IN CTD-BRD2 (667–808) plasmid is a polycistronic construction for the co-
27 expression and co-purification of PERV IN CTD and BRD2 (667–808). This construction
28 expresses a polyprotein containing the TEV protease and the Strep-PERV IN CTD (309–405)
29 fragment with the His-BRD2 (667–808) fragment, both placed after a TEV digestion site
30 (TEV DS). During overexpression, the TEV protease cleaves the TEV sites, if any, leaving
31 the formation of a complex between PERV IN CTD and BRD2 (667–808) fragments.

1 **(B)** SDS-PAGE gel illustrating the co-expression and co-purification of the wild-type (WT)
2 Strep-PERV IN CTD, or of the Strep-PERV IN CTD- W379A and - R391A/L392A mutants
3 with His-BRD2 (667–808). Two sequential affinity purifications were performed on nickel
4 and Strep-Trap columns. Non-eluted samples (Lanes 1, 4 and 7; called “Input”) and samples
5 eluted from the nickel column (lanes 2, 5 and 8; called flow-through “FT”) after eluted from
6 the Strep-Trap column (lanes 3, 6 and 9; called “E”) were loaded on an SDS-PAGE gel;
7 proteins were either revealed by Coomassie staining (upper panels) or were visualized by
8 western blot using anti-His and anti-Strep-tag II antibodies targeting His-BRD2 (667–808)
9 and Strep-PERV IN CTD, respectively (lower panels).

10

11 **Figure 7. SAXS analysis of PERV IN CTD – BRD2 ET.**

12 **(A)** Experimental diffusion curve in blue and CRY SOL curve in red calculated from the
13 atomic model.

14 **(B)** Dimensionless Kratky plot with a plateau in the higher q region reflecting flexibility.

15 **(C)** Distance distribution function $P(r)$.

16 **(D)** *Ab initio* SAXS envelope and ribbon representation of the SAXS-restrained model with
17 the CTD of PERV IN in blue, the EBM of IN in green and the bound ET domain of BRD2 in
18 red. The figures A, B and C were rendered with ATSAS, and the figure D was rendered with
19 CHIMERA.

20

21 **Figure 8. Ribbon representation of the modeled PERV IN intasome superposed with the**
22 **SAXS-restrained model of PERV IN CTD – BRD2 ET.** The two symmetric inner IN
23 monomers are painted cyan and violet, whereas the two outer INs are blue and yellow. The
24 two bound DNA ends are in grey. The PERV IN EBM segments and the bound BRD2 ET of
25 the superposed SAXS-restrained model are colored in green and red, respectively. The inner
26 CTD of the intasome and the CTD of the heterodimeric complex were superimposed with
27 COOT. The figure was rendered with CHIMERA.

28

1 **Supplementary figures**

2

3 **Figure S1: MLV IN domains involved in the MLV IN / BRD2 interaction**

4 Mapping of the MLV IN domain involved in the interaction with BRD2 protein. The numbers
5 1–408 indicate the positions of residues in the MLV IN protein. The DNA fragments
6 encoding for residues 1–408, 1–276 and 277–408 of the MLV IN were fused to the Gal4-BD
7 binding domain. Binding to BRD2 (fused to Gal4-AD) was determined by assessing
8 auxotrophic growth in media lacking histidine and adenine (-His / -Ade), as in Fig. 1.

9

10 **Figure S2: Multiple sequence alignment of the CTD INs from different retroviral genera**

11 with PERV and MLV (gamma-retrovirus), PFV (spumavirus), HIV-1 (lentivirus) and RAV-1
12 (alpha-retrovirus). Identical and well-conserved residues (>70% similarity) have a red and a
13 yellow background, respectively. The 17-residue long EBM segment of gamma-retroviral INs
14 is marked below sequence blocks with green stars. The figure was rendered with ESPript.

15

16 **Figure S3: GlobPlot (Waterhouse et al., 2018) with prediction of globularity and**

17 **disorder for the C-terminal domain of BRD2 from residue 667 to 836.** The sequence of
18 the fragment (aa 667-836) is written with the predicted disordered zones colored in blue and
19 the potential globular domains colored in green.

20

21 **Figure S4: Expression, purification and cross-linking of PERV IN CTD (309–405) –**

22 **BRD2 ET (667–735) complex.**

23 (A) UV absorption chromatogram of size exclusion chromatography (SEC) performed on the
24 PERV IN CTD (312–408) / BRD2 ET (667–735) co-purified complex.

25 (B) SDS-PAGE gel illustrating the co-expression and co-purification of Strep-tag-PERV-IN
26 CTD with His-tag BRD2 ET. Two sequential affinity purifications were performed: Nickel
27 and Strep-Trap columns. Pellet (lane 1), total fraction (lane 2), samples from Nickel column
28 flow through (lane 3); elution (lane 4) and elution from Strep-Trap column (lanes 5, 6 and 7)
29 were loaded on the gel. The purity of the fractions was evaluated by Coomassie staining.

1 (C) SDS-PAGE of the chemically cross-linked PERV IN CTD – BRD2 ET complex with on
2 lane 1, 10 µg of sample incubated with 0.1% of glutaraldehyde for 20 min at room
3 temperature, and on lane 2, the control sample without glutaraldehyde.

4

5

6 **References**

- 7 Aiyer, S., Swapna, G.V., Malani, N., Aramini, J.M., Schneider, W.M., Plumb, M.R.,
8 Ghanem, M., Larue, R.C., Sharma, A., Studamire, B., Kvaratskhelia, M.,
9 Bushman, F.D., Montelione, G.T., Roth, M.J., 2014. Altering murine leukemia
10 virus integration through disruption of the integrase and BET protein family
11 interaction. *Nucleic Acids Res* 42, 5917-5928.
- 12 Andrade, M.D., Skalka, A.M., 2015. Retroviral Integrase: Then and Now. *Annu Rev*
13 *Virol* 2, 241-264.
- 14 Ballandras-Colas, A., Brown, M., Cook, N.J., Dewdney, T.G., Demeler, B.,
15 Cherepanov, P., Lyumkis, D., Engelman, A.N., 2016. Cryo-EM reveals a novel
16 octameric integrase structure for betaretroviral intasome function. *Nature* 530,
17 358-361.
- 18 Ballandras-Colas, A., Maskell, D.P., Serrao, E., Locke, J., Swuec, P., Jonsson, S.R.,
19 Kotecha, A., Cook, N.J., Pye, V.E., Taylor, I.A., Andresdottir, V., Engelman,
20 A.N., Costa, A., Cherepanov, P., 2017. A supramolecular assembly mediates
21 lentiviral DNA integration. *Science* 355, 93-95.
- 22 Belkina, A.C., Denis, G.V., 2012. BET domain co-regulators in obesity, inflammation
23 and cancer. *Nat Rev Cancer* 12, 465-477.
- 24 Bunkoczi, G., Read, R.J., 2011. Improvement of molecular-replacement models with
25 Sculptor. *Acta Crystallogr D Biol Crystallogr* 67, 303-312.
- 26 Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ,
27 F., Debyser, Z., 2005. The interaction of LEDGF/p75 with integrase is
28 lentivirus-specific and promotes DNA binding. *J Biol Chem* 280, 17841-17847.
- 29 Cellier, C., Moreau, K., Gallay, K., Ballandras, A., Gouet, P., Ronfort, C., 2013. In
30 vitro functional analyses of the human immunodeficiency virus type 1 (HIV-1)
31 integrase mutants give new insights into the intasome assembly. *Virology* 439,
32 97-104.
- 33 Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J.,
34 Engelborghs, Y., De Clercq, E., Debyser, Z., 2003. HIV-1 integrase forms
35 stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol*
36 *Chem* 278, 372-381.
- 37 Crowe, B.L., Larue, R.C., Yuan, C., Hess, S., Kvaratskhelia, M., Foster, M.P., 2016.
38 Structure of the Brd4 ET domain bound to a C-terminal motif from gamma-
39 retroviral integrases reveals a conserved mechanism of interaction. *Proc Natl*
40 *Acad Sci U S A* 113, 2086-2091.
- 41 De Ravin, S.S., Su, L., Theobald, N., Choi, U., Macpherson, J.L., Poidinger, M.,
42 Symonds, G., Pond, S.M., Ferris, A.L., Hughes, S.H., Malech, H.L., Wu, X.,
43 2014. Enhancers are major targets for murine leukemia virus vector
44 integration. *J Virol* 88, 4504-4513.
- 45 De Rijck, J., de Kogel, C., Demeulemeester, J., Vets, S., El Ashkar, S., Malani, N.,
46 Bushman, F.D., Landuyt, B., Husson, S.J., Busschots, K., Gijssbers, R.,

1 Debyser, Z., 2013. The BET family of proteins targets moloney murine
2 leukemia virus integration near transcription start sites. *Cell Rep* 5, 886-894.

3 Demange, A., Yajjou-Hamalian, H., Gallay, K., Luengo, C., Beven, V., Leroux, A.,
4 Confort, M.P., Al Andary, E., Gouet, P., Moreau, K., Ronfort, C., Blanchard, Y.,
5 2015. Porcine endogenous retrovirus-A/C: biochemical properties of its
6 integrase and susceptibility to raltegravir. *J Gen Virol* 96, 3124-3130.

7 Denner, J., 2017. Can Antiretroviral Drugs Be Used to Treat Porcine Endogenous
8 Retrovirus (PERV) Infection after Xenotransplantation? *Viruses* 9.

9 El Ashkar, S., De Rijck, J., Demeulemeester, J., Vets, S., Madlala, P., Cermakova,
10 K., Debyser, Z., Gijssbers, R., 2014. BET-independent MLV-based Vectors
11 Target Away From Promoters and Regulatory Elements. *Mol Ther Nucleic
12 Acids* 3, e179.

13 Emiliani, S., Mousnier, A., Busschots, K., Maroun, M., Van Maele, B., Tempe, D.,
14 Vandekerckhove, L., Moisant, F., Ben-Slama, L., Witvrouw, M., Christ, F.,
15 Rain, J.C., Dargemont, C., Debyser, Z., Benarous, R., 2005. Integrase
16 mutants defective for interaction with LEDGF/p75 are impaired in chromosome
17 tethering and HIV-1 replication. *J Biol Chem* 280, 25517-25523.

18 Engelman, A.N., Cherepanov, P., 2017. Retroviral intasomes arising. *Curr Opin
19 Struct Biol* 47, 23-29.

20 Evrard, G., Mareuil, F., Bontems, F., Sizun, C., Perez, J., 2011. DADIMODO: A
21 program for refining the structure of multidomain proteins and complexes
22 against small-angle scattering data and NMR-derived restraints. *Applied
23 Crystallography* 44, 1264-1271. .

24 Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens,
25 H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., Svergun,
26 D.I., 2017. ATSAS 2.8: a comprehensive data analysis suite for small-angle
27 scattering from macromolecular solutions. *J Appl Crystallogr* 50, 1212-1225.

28 Guan, R., Aiyer, S., Cote, M.L., Xiao, R., Jiang, M., Acton, T.B., Roth, M.J.,
29 Montelione, G.T., 2017. X-ray crystal structure of the N-terminal region of
30 Moloney murine leukemia virus integrase and its implications for viral DNA
31 recognition. *Proteins* 85, 647-656.

32 Gupta, S.S., Maetzig, T., Maertens, G.N., Sharif, A., Rothe, M., Weidner-Glunde, M.,
33 Galla, M., Schambach, A., Cherepanov, P., Schulz, T.F., 2013. Bromo- and
34 extraterminal domain chromatin regulators serve as cofactors for murine
35 leukemia virus integration. *J Virol* 87, 12721-12736.

36 Hare, S., Gupta, S.S., Valkov, E., Engelman, A., Cherepanov, P., 2010. Retroviral
37 intasome assembly and inhibition of DNA strand transfer. *Nature* 464, 232-
38 236.

39 Jang, M.K., Kwon, D., McBride, A.A., 2009. Papillomavirus E2 proteins and the host
40 BRD4 protein associate with transcriptionally active cellular chromatin. *J Virol*
41 83, 2592-2600.

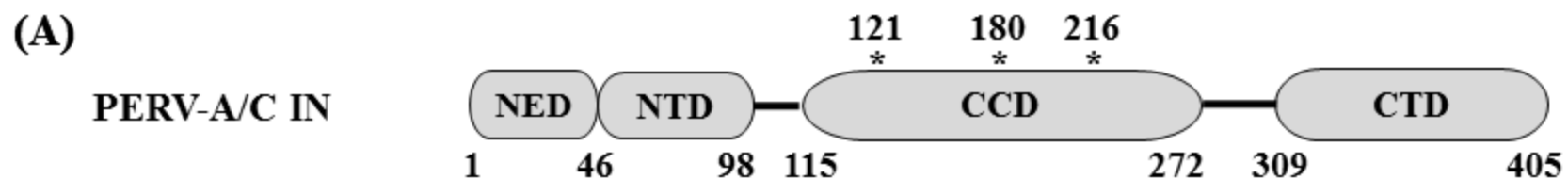
42 Kvaratskhelia, M., Sharma, A., Larue, R.C., Serrao, E., Engelman, A., 2014.
43 Molecular mechanisms of retroviral integration site selection. *Nucleic Acids
44 Res* 42, 10209-10225.

45 LaFave, M.C., Varshney, G.K., Gildea, D.E., Wolfsberg, T.G., Baxevanis, A.D.,
46 Burgess, S.M., 2014. MLV integration site selection is driven by strong
47 enhancers and active promoters. *Nucleic Acids Res* 42, 4257-4269.

48 Larue, R.C., Plumb, M.R., Crowe, B.L., Shkriabai, N., Sharma, A., DiFiore, J., Malani,
49 N., Aiyer, S.S., Roth, M.J., Bushman, F.D., Foster, M.P., Kvaratskhelia, M.,

- 1 2014. Bimodal high-affinity association of Brd4 with murine leukemia virus
2 integrase and mononucleosomes. *Nucleic Acids Res* 42, 4868-4881.
- 3 Levy, N., Eiler, S., Pradeau-Aubretton, K., Maillot, B., Stricher, F., Ruff, M., 2016.
4 Production of unstable proteins through the formation of stable core
5 complexes. *Nat Commun* 7, 10932.
- 6 Lin, Y.J., Umehara, T., Inoue, M., Saito, K., Kigawa, T., Jang, M.K., Ozato, K.,
7 Yokoyama, S., Padmanabhan, B., Guntert, P., 2008. Solution structure of the
8 extraterminal domain of the bromodomain-containing protein BRD4. *Protein*
9 *Sci* 17, 2174-2179.
- 10 Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B., 2003.
11 Protein disorder prediction: implications for structural proteomics. *Structure* 11,
12 1453-1459.
- 13 Llano, M., Saenz, D.T., Meehan, A., Wongthida, P., Peretz, M., Walker, W.H., Teo,
14 W., Poeschla, E.M., 2006. An essential role for LEDGF/p75 in HIV integration.
15 *Science* 314, 461-464.
- 16 Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser,
17 Z., Engelborghs, Y., 2003. LEDGF/p75 is essential for nuclear and
18 chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem* 278,
19 33528-33539.
- 20 Maertens, G.N., 2016. B'-protein phosphatase 2A is a functional binding partner of
21 delta-retroviral integrase. *Nucleic Acids Res* 44, 364-376.
- 22 Maertens, G.N., Hare, S., Cherepanov, P., 2010. The mechanism of retroviral
23 integration from X-ray structures of its key intermediates. *Nature* 468, 326-329.
- 24 Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker,
25 J.R., Bushman, F.D., 2004. Retroviral DNA integration: ASLV, HIV, and MLV
26 show distinct target site preferences. *PLoS Biol* 2, E234.
- 27 Moalic, Y., Blanchard, Y., Felix, H., Jestin, A., 2006. Porcine endogenous retrovirus
28 integration sites in the human genome: features in common with those of
29 murine leukemia virus. *J Virol* 80, 10980-10988.
- 30 Moalic, Y., Felix, H., Takeuchi, Y., Jestin, A., Blanchard, Y., 2009. Genome areas
31 with high gene density and CpG island neighborhood strongly attract porcine
32 endogenous retrovirus for integration and favor the formation of hot spots. *J*
33 *Virol* 83, 1920-1929.
- 34 Moreau, K., Faure, C., Violot, S., Gouet, P., Verdier, G., Ronfort, C., 2004. Mutational
35 analyses of the core domain of Avian Leukemia and Sarcoma Viruses
36 integrase: critical residues for concerted integration and multimerization.
37 *Virology* 318, 566-581.
- 38 Nowrouzi, A., Dittrich, M., Klanke, C., Heinkelein, M., Rammling, M., Dandekar, T.,
39 von Kalle, C., Rethwilm, A., 2006. Genome-wide mapping of foamy virus
40 vector integrations into a human cell line. *J Gen Virol* 87, 1339-1347.
- 41 Passos, D.O., Li, M., Yang, R., Rebensburg, S.V., Ghirlando, R., Jeon, Y., Shkriabai,
42 N., Kvaratskhelia, M., Craigie, R., Lyumkis, D., 2017. Cryo-EM structures and
43 atomic model of the HIV-1 strand transfer complex intasome. *Science* 355, 89-
44 92.
- 45 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng,
46 E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory
47 research and analysis. *J Comput Chem* 25, 1605-1612.
- 48 Sanchez, R., Zhou, M.M., 2009. The role of human bromodomains in chromatin
49 biology and gene transcription. *Curr Opin Drug Discov Devel* 12, 659-665.

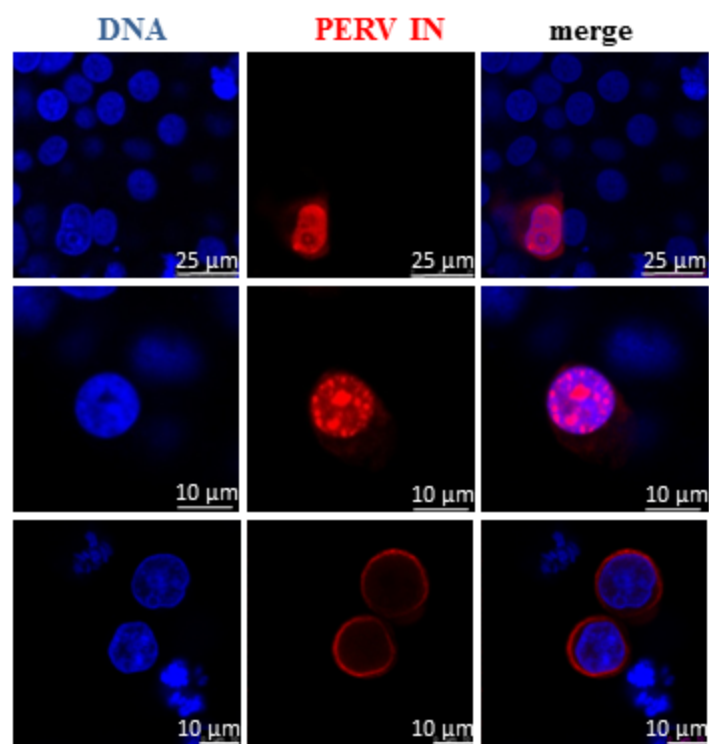
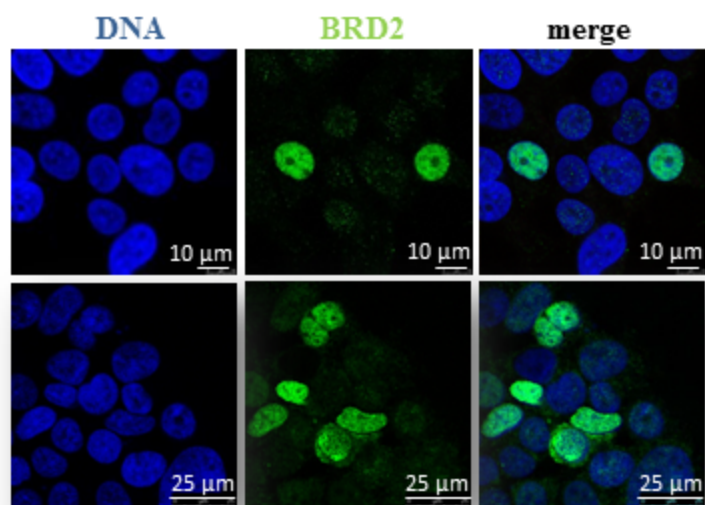
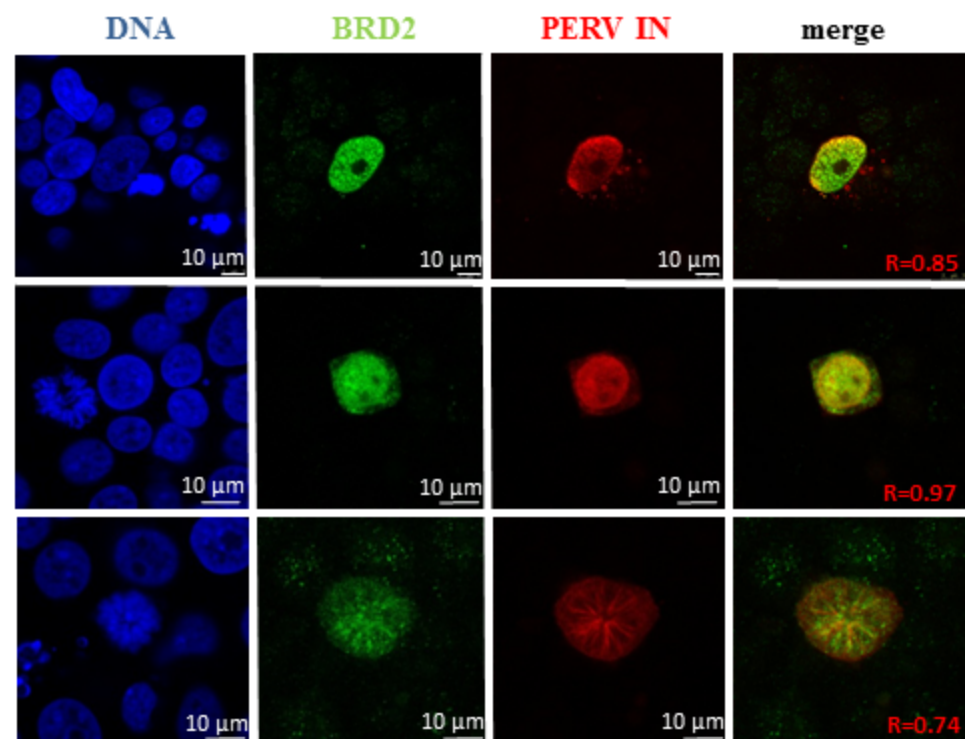
- 1 Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., Bushman, F., 2002. HIV-1
2 integration in the human genome favors active genes and local hotspots. *Cell*
3 110, 521-529.
- 4 Serrao, E., Ballandras-Colas, A., Cherepanov, P., Maertens, G.N., Engelman, A.N.,
5 2015. Key determinants of target DNA recognition by retroviral intasomes.
6 *Retrovirology* 12, 39.
- 7 Sharma, A., Larue, R.C., Plumb, M.R., Malani, N., Male, F., Slaughter, A., Kessler, J.J.,
8 Shkriabai, N., Coward, E., Aiyer, S.S., Green, P.L., Wu, L., Roth, M.J.,
9 Bushman, F.D., Kvaratskhelia, M., 2013. BET proteins promote efficient
10 murine leukemia virus integration at transcription start sites. *Proc Natl Acad*
11 *Sci U S A* 110, 12036-12041.
- 12 Sowd, G.A., Serrao, E., Wang, H., Wang, W., Fadel, H.J., Poeschla, E.M., Engelman,
13 A.N., 2016. A critical role for alternative polyadenylation factor CPSF6 in
14 targeting HIV-1 integration to transcriptionally active chromatin. *Proc Natl Acad*
15 *Sci U S A* 113, E1054-1063.
- 16 Studamire, B., Goff, S.P., 2008. Host proteins interacting with the Moloney murine
17 leukemia virus integrase: multiple transcriptional regulators and chromatin
18 binding factors. *Retrovirology* 5, 48.
- 19 Thorpe, K.L., Abdulla, S., Kaufman, J., Trowsdale, J., Beck, S., 1996. Phylogeny and
20 structure of the RING3 gene. *Immunogenetics* 44, 391-396.
- 21 Viejo-Borbolla, A., Ottinger, M., Bruning, E., Burger, A., Konig, R., Kati, E., Sheldon,
22 J.A., Schulz, T.F., 2005. Brd2/RING3 interacts with a chromatin-binding
23 domain in the Kaposi's Sarcoma-associated herpesvirus latency-associated
24 nuclear antigen 1 (LANA-1) that is required for multiple functions of LANA-1. *J*
25 *Virol* 79, 13618-13629.
- 26 Wanaguru, M., Barry, D.J., Benton, D.J., O'Reilly, N.J., Bishop, K.N., 2018. Murine
27 leukemia virus p12 tethers the capsid-containing pre-integration complex to
28 chromatin by binding directly to host nucleosomes in mitosis. *PLoS Pathog* 14,
29 e1007117.
- 30 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R.,
31 Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T.,
32 2018. SWISS-MODEL: homology modelling of protein structures and
33 complexes. *Nucleic Acids Res* 46, W296-W303.
- 34 Weidner-Glunde, M., Ottinger, M., Schulz, T.F., 2010. WHAT do viruses BET on?
35 *Front Biosci (Landmark Ed)* 15, 537-549.
- 36 Winans, S., Larue, R.C., Abraham, C.M., Shkriabai, N., Skopp, A., Winkler, D.,
37 Kvaratskhelia, M., Beemon, K.L., 2017. The FACT Complex Promotes Avian
38 Leukosis Virus DNA Integration. *J Virol* 91(7). pii: e00082-17.
- 39 Wu, S.Y., Chiang, C.M., 2007. The double bromodomain-containing chromatin
40 adaptor Brd4 and transcriptional regulation. *J Biol Chem* 282, 13141-13145.
- 41 Wu, X., Li, Y., Crise, B., Burgess, S.M., 2003. Transcription start regions in the
42 human genome are favored targets for MLV integration. *Science* 300, 1749-
43 1751.
- 44 Yin, Z., Shi, K., Banerjee, S., Pandey, K.K., Bera, S., Grandgenett, D.P., Aihara, H.,
45 2016. Crystal structure of the Rous sarcoma virus intasome. *Nature* 530, 362-
46 366.
- 47 Zhang, Q., Zeng, L., Shen, C., Ju, Y., Konuma, T., Zhao, C., Vakoc, C.R., Zhou,
48 M.M., 2016. Structural Mechanism of Transcriptional Regulator NSD3
49 Recognition by the ET Domain of BRD4. *Structure* 24, 1201-1208.

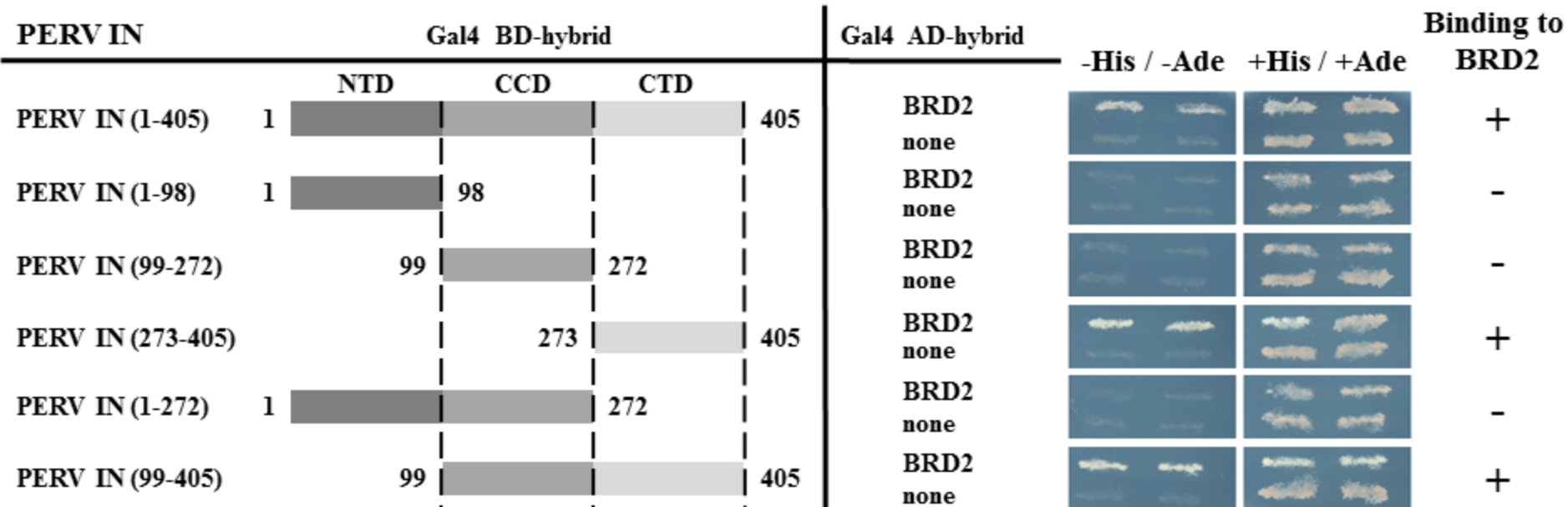
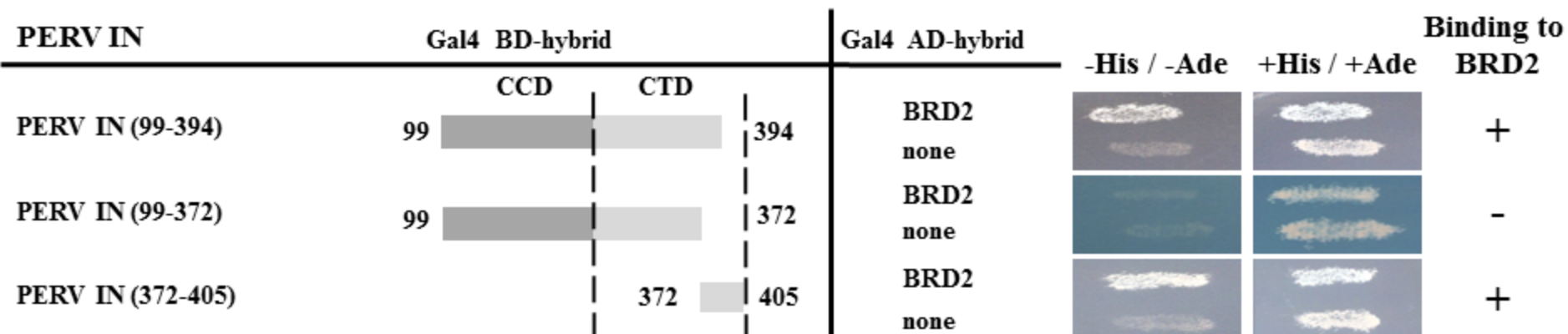


(B)

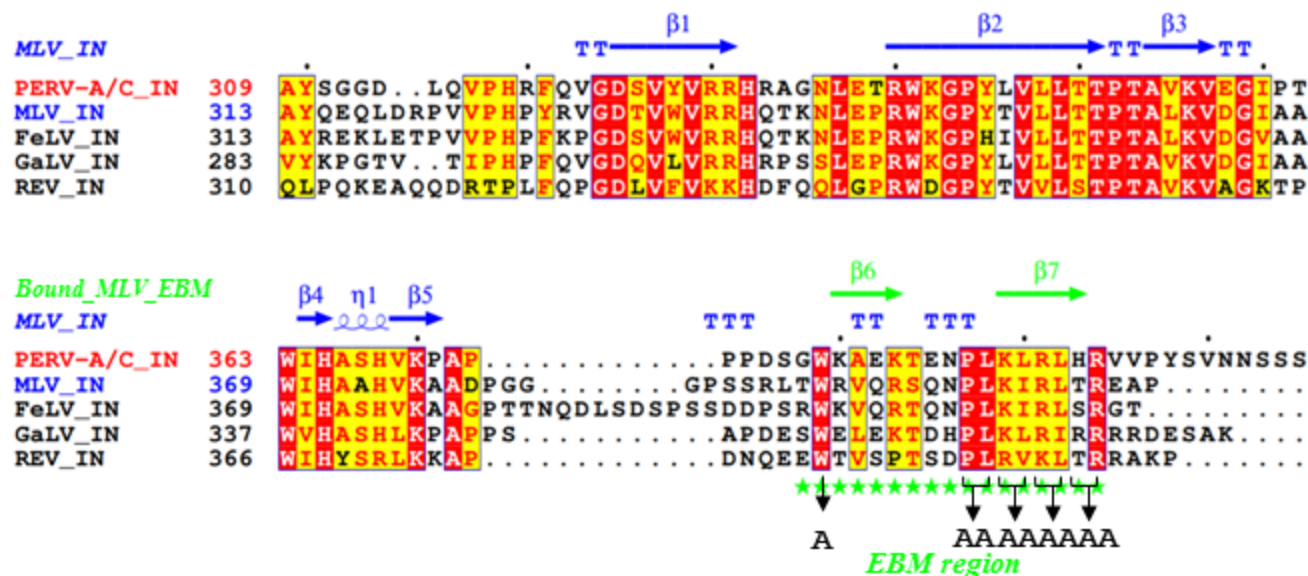
Gal4 BD-hybrid	Gal4 AD-hybrid	-His / -Ade	+His / +Ade	Binding to BRD2
PERV IN (1-405)	BRD2			+
	none			
MLV IN (1-408)	BRD2			+
	none			
HIV-1 IN (1-288)	BRD2			-
	none			
RAV-1 IN (1-286)	BRD2			-
	none			
none	BRD2			-

Figure 1

(A)**(B)****(C)****Figure 2**

(A)**(B)****Figure 3**

(A)



(B)

PERV IN	Gal4 BD-hybrid				Gal4 AD-hybrid		Binding to BRD2	
	NTD	CCD	CTD		-His / -Ade	+His / +Ade		
PERV IN	1		W PLKLRLHR	405	BRD2			+
PERV IN (w379A)	1		A PLKLRLHR	405	BRD2			-
PERV IN (P387A/L388A)	1		W AAKLRLHR	405	BRD2			-
PERV IN (K389A/L390A)	1		W PLAARLHR	405	BRD2			-
PERV IN (R391A/L392A)	1		W PLKLAHR	405	BRD2			-
PERV IN (H393A/R394A)	1		W PLKLRLAA	405	BRD2			-
					None			

Figure 4

(A)



(B)

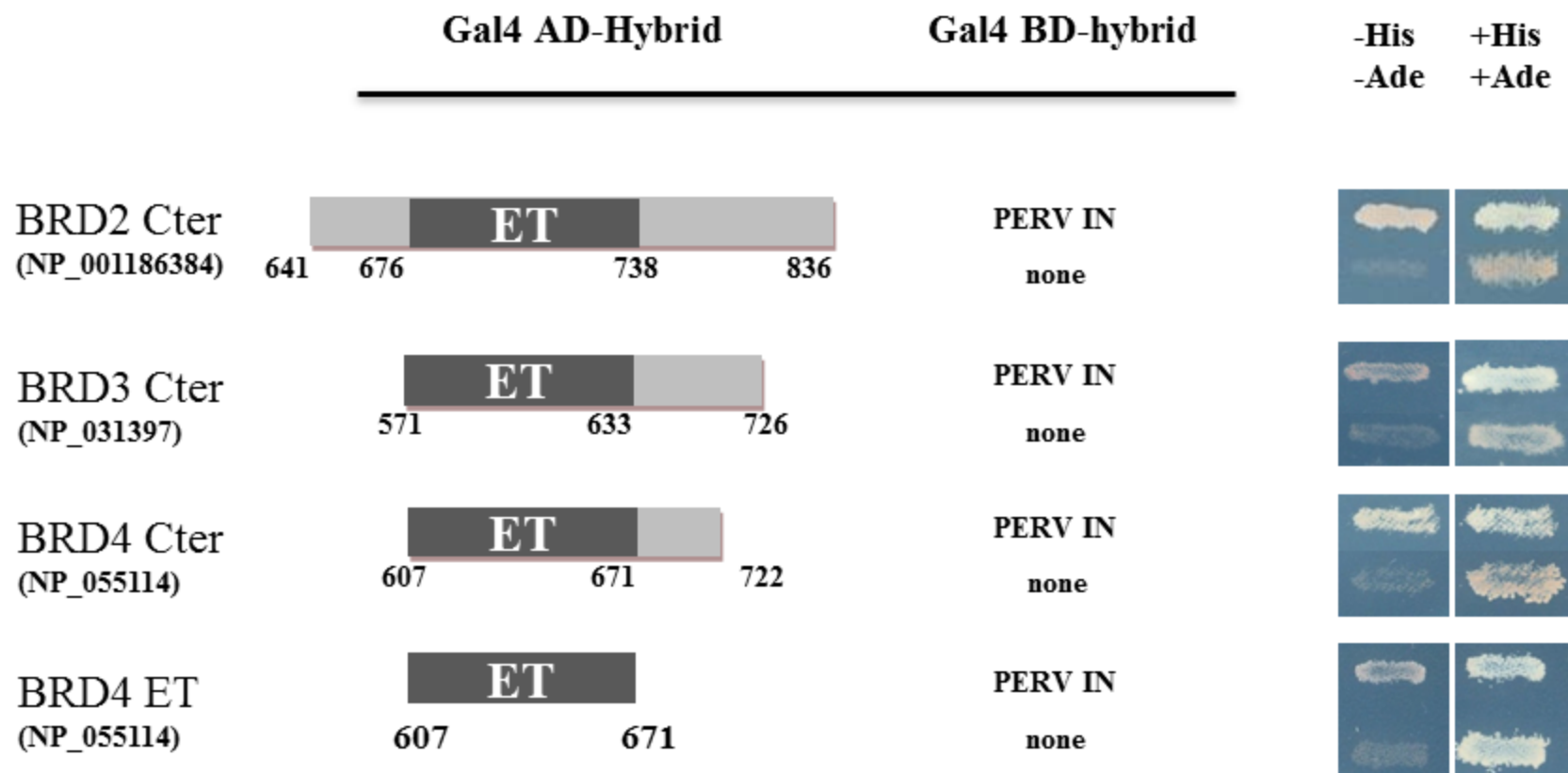


Figure 5

(C)

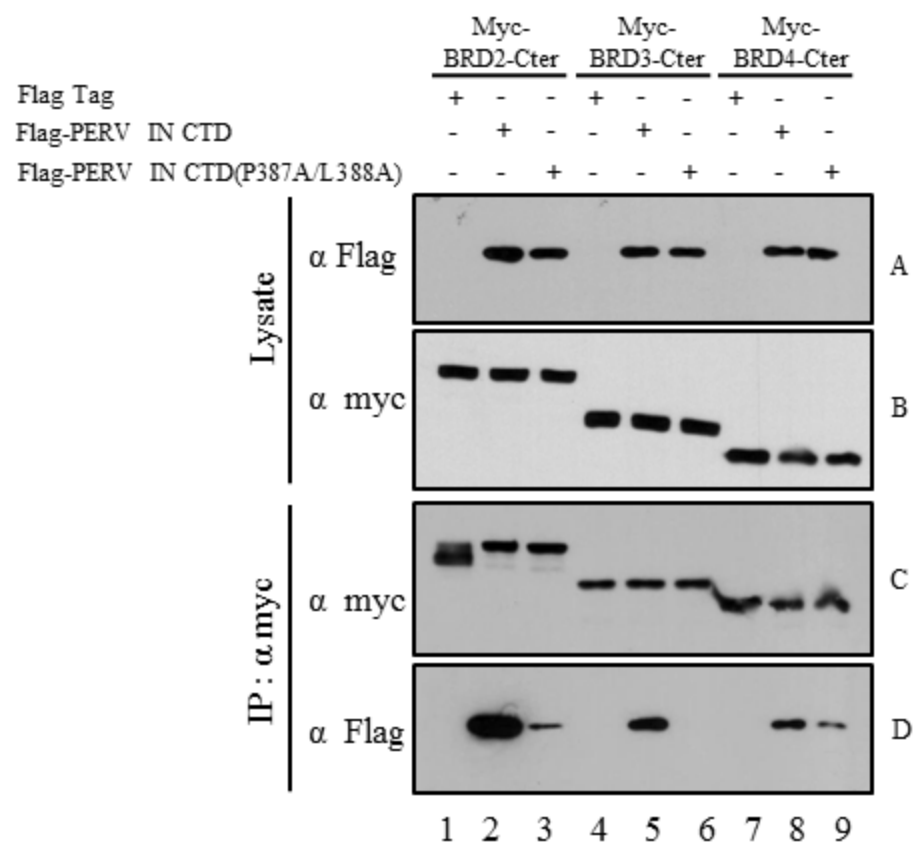


Figure 5

(A)**(B)**

PERV IN CTD –BRD2 (667-808) co-purification

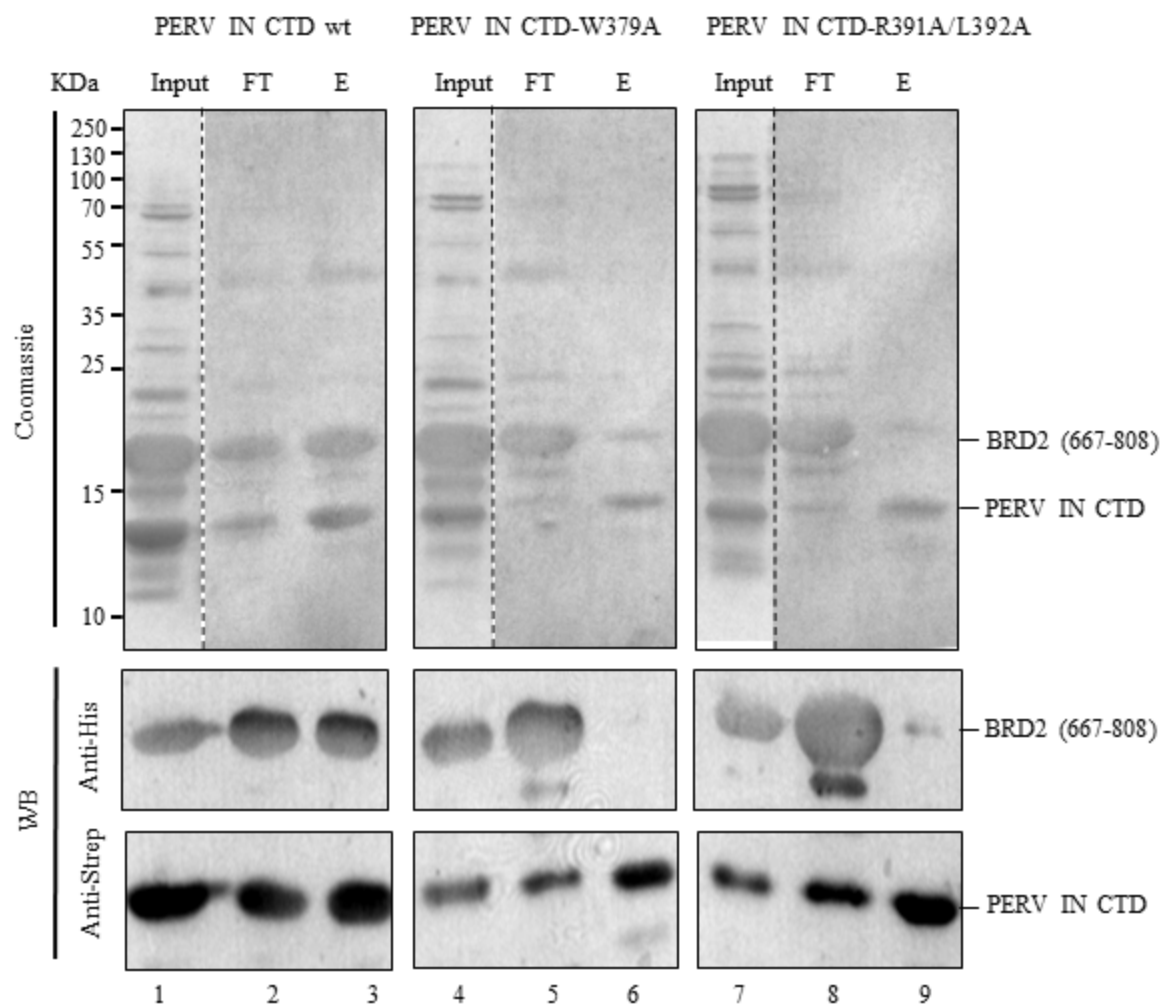


Figure 6

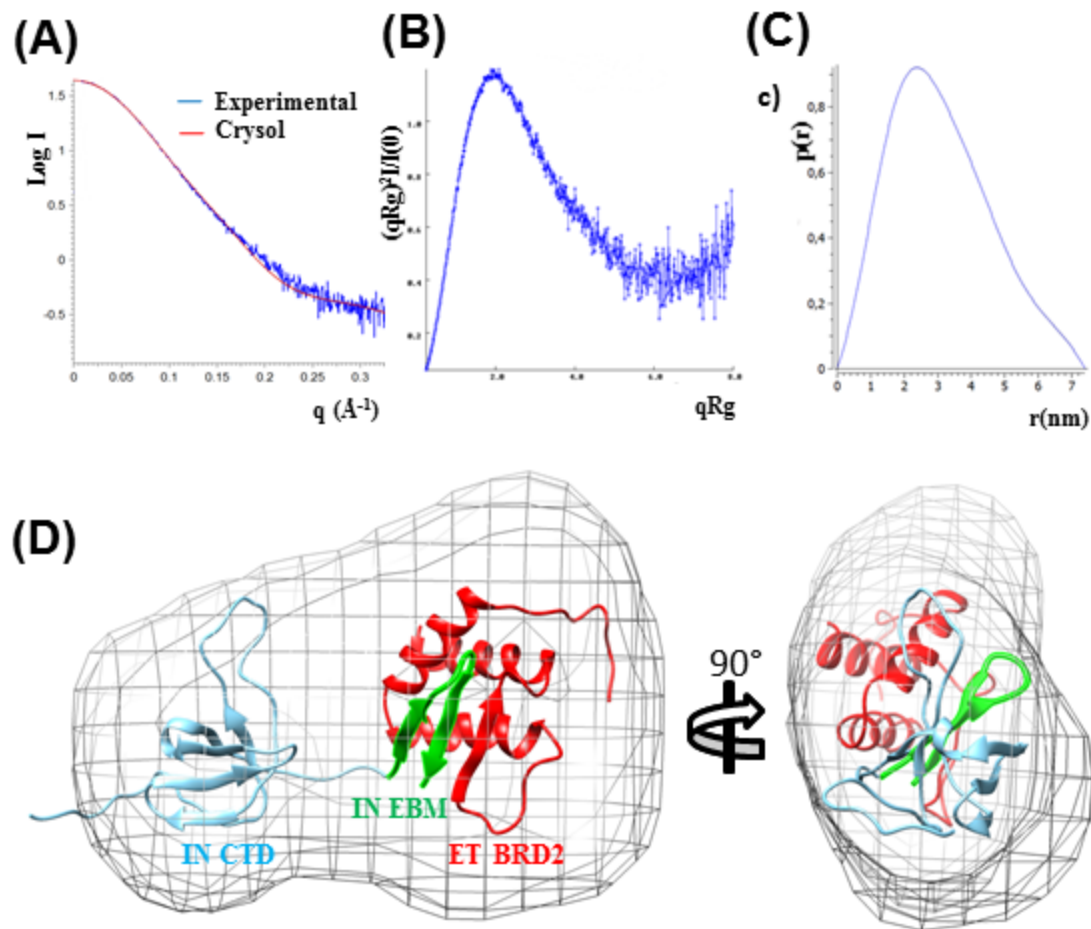


Figure 7

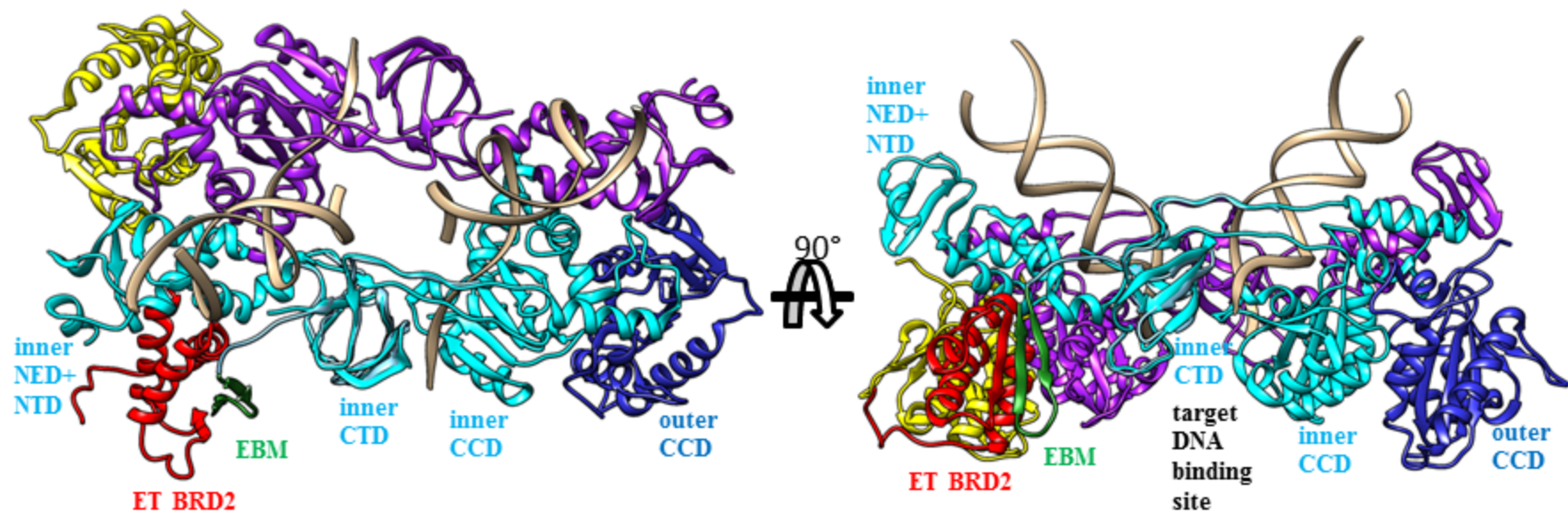


Figure 8

Table 1. Statistics of SAXS data collection and refinement from the two-fragment complexes

<i>Data collection</i>		
Analyzed fragment	IN CTD - ET	IN CTD – BRD2 (667–808)
Instrument	BM29 ESRF	BM29 ESRF
q-range (nm ⁻¹)	0–5	0–5
Exposure time (s)/nb frames	1/10	1/10
Concentration (mg/ml)	4	4
Temperature (°C)		
<i>Data processing</i>		
Primary data reduction	EDNA	EDNA
Data processing	CHROMIXS	CHROMIXS
q-range (nm ⁻¹)	0.1–3.2	0.1–2.2
R _g (nm) [from Guinier]	2.4 ± .07	3.1 ± .1
R _g (nm) [from P(r)]	2.4	3.1
D _{max} (nm)	7.4	11
<i>Ab initio</i> analysis	DAMMIF	DAMMIF
Number of models	50	50
Validation and averaging	DAMAVER	DAMAVER
	0.87 ± .06	0.74 ± .07
NSD	DADIMODO	N.A.
Rigid-body modeling	1.0	N.A.
Model χ^2		