Article:
Ballandras-Colas, A, Maskell, DP, Serrao, E et al. (2017) A supramolecular assembly mediates lentiviral DNA integration. Science, 355 (6320). pp. 93-95. ISSN: 0036-8075

https://doi.org/10.1126/science.aah7002

Title: Cryo-EM reveals that a supramolecular assembly mediates lentiviral DNA integration

Authors: Allison Ballandras-Colas,[1]† Daniel P. Maskell,[1]†‡ Erik Serrao,[2] Julia Locke,[3] Paolo Swuec,[3] Stefán R. Jónsson,[4] Abhay Kotecha,[5] Nicola J. Cook,[1] Valerie E. Pye,[1] Ian A. Taylor,[6] Valgerdur Andrésdóttir,[4] Alan N. Engelman,[2] Alessandro Costa,[3]* and Peter Cherepanov[1,7]*

Affiliations:

[1]Chromatin Structure and Mobile DNA, The Francis Crick Institute, London, NW1 1AT, UK.

[2]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.

[3]Macromolecular Machines Laboratory, The Francis Crick Institute, London, NW1 1AT, UK.

[4]Institute for Experimental Pathology, University of Iceland, Keldur, 112 Reykjavik, Iceland.

[5]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK.

[6]Macromolecular Structure Laboratory, The Francis Crick Institute, London, NW1 1AT, UK.

[7]Division of Medicine, Imperial College London, W2 1PG, UK.

†These authors contributed equally to this work.

‡Present address: School of Molecular and Cellular Biology, University of Leeds, Leeds, LS2 9JT, UK.

*Corresponding author. Email: alessandro.costa@crick.ac.uk (A.C.); Email: peter.cherepanov@crick.ac.uk (P.C.)

One Sentence Summary: Lentiviral integration is catalyzed by a higher-order multimeric assembly.

Abstract: Retroviral integrase (IN) functions within the intasome nucleoprotein complex to catalyze the insertion of viral DNA into cellular chromatin. The lack of lentiviral intasome structural information has hampered the development of anti-HIV drugs and the understanding of viral resistance. Using cryo-electron microscopy, we now visualize the functional maedi-visna lentivirus intasome at 4.9 Å resolution. The intasome, which comprises a homo-hexadecamer of IN with a tetramer-of-tetramers architecture, harbors eight structurally distinct types of IN protomers including two catalytically competent subunits. The conserved intasomal core, previously observed in simpler retroviral systems, is formed between two IN tetramers, with a pair of C-terminal domains from flanking tetramers completing the synaptic interface. Our results explain how HIV-1 IN, which self-associates into higher order multimers, can form a functional intasome, reconcile the bulk of early HIV-1 IN biochemical and structural data, and provide a lentiviral platform for structure-guided design of HIV-1 IN inhibitors.

Main Text: Integrase (IN) acts on the ends of the linear double stranded viral DNA (vDNA) molecule produced by reverse transcription of the retroviral RNA genome. Initially, IN catalyzes 3'-processing to expose 3' hydroxyl groups attached to invariant CA dinucleotides at the vDNA ends. Following entry into the nuclear compartment IN inserts the processed vDNA 3' termini across the major groove of chromosomal target DNA using the 3' hydroxyls as nucleophiles in the strand transfer reaction. These events take place within the intasome, a stable synaptic complex comprising a multimer of IN assembled on vDNA ends (*1*).

Characterization of prototype foamy virus (PFV, belonging to the spumavirus genus), Rous sarcoma virus (RSV, an α-retrovirus), and mouse mammary tumor virus (MMTV, a β-retrovirus) intasomes illuminated the conserved intasome core (CIC) structure minimally comprising a pair of IN dimers, as in the case of the PFV intasome (*2, 3*), or decorated by flanking IN dimers in RSV (*4*) and MMTV (*5*). The architecture of the lentiviral intasome, the genus that includes HIV-1 and HIV-2 along with highly pathogenic animal viruses, has remained elusive.

Unfavorable biochemical properties of HIV-1 IN necessitate the use of hyperactive and/or solubilizing mutations (*6-8*), which, by their nature, dramatically change the properties of the protein. Taking a more holistic approach, we sought to identify a lentiviral IN that is amenable for structural studies as a wild type protein. We discovered that the IN from maedi-visna virus (MVV), an ovine lentivirus, displays robust strand transfer activity when supplied with oligonucleotides mimicking the vDNA ends in the presence of the common lentiviral integration host factor LEDGF (*9, 10*) (fig. S1). MVV IN assembled into a functional nucleoprotein complex that could be isolated by size exclusion chromatography (fig. S2A). In the presence of the essential Mg^{2+} cofactor, the purified nucleoprotein complex catalyzed strand transfer activity and could be inhibited by the HIV-1 IN strand transfer inhibitor (INSTI) dolutegravir (*11*) (fig. S2B). Sequence analysis of reaction products ascertained that they were formed by full-site integration – coordinated insertion of pairs of vDNA ends across the major groove in target DNA – leading to short duplications of target DNA sequences (fig. S2C). To confirm that the most commonly observed duplication size – 6 bp – is representative of MVV integration, we sequenced 2,526 unique integration sites in primary sheep cells infected with pathogenic MVV and compared them to *in vitro* integration sites obtained with purified intasomes and deproteinized sheep or plasmid DNA. Aligning the three sets of integration site sequences revealed symmetric and highly similar sequence

preferences that are fully consistent with integration of vDNA ends across 6 bp in target DNA (fig. S3). As expected for a lentivirus (*9*), MVV displayed a strong preference for transcription units, with 70.2% of integration sites found within predicted sheep genes, compared to 43.7% in the *in vitro* generated sample ($p<10^{-150}$).

Inspection of the intasome by negative stain electron microscopy (EM) revealed a flat, two-fold symmetric molecule measuring over 20 nm in the widest dimension (fig. S4), which is much larger than any of the previously characterized retroviral intasomes. To determine its structure, we acquired images of single particles in vitreous ice using a transmission electron microscope equipped with a direct detector. The final structure was refined using a dataset of 94,283 single particles to an overall resolution of 4.94 Å, with local resolution varying from 9 Å in the periphery of the structure to ~4 Å throughout the core region (Fig. 1, figs S5-7). A crystal structure spanning the N-terminal and the catalytic core domains (NTD and CCD) of MVV IN is available (*12*). In addition, we determined two crystal structures spanning the MVV IN C-terminal domain (CTD, table S1). Sixteen MVV IN subunits and two double stranded DNA oligonucleotides representing the synapsed vDNA ends could be unambiguously placed in the electron density map (Fig. 1A, fig. S8, movie S1), consistent with the observed molecular mass of ~0.5 MDa for the complex (fig. S2D).

The intasome represents a tetramer of tetramers, each comprising a pair of imperfectly symmetric IN dimers with CCD-CTD linkers in extended α-helical configurations that is strikingly similar to the HIV-1 IN dimer observed in crystals (*7*) (figs S9, S10A). Although intasome formation required the presence of LEDGF, only traces of the host factor remained after purification by two-stage chromatography (fig. S2A). Consequently, no density could be attributed to LEDGF in the structure. It is possible that the remaining LEDGF molecules are distributed over 16 possible binding positions on the intasome.

The CIC, analogous to those found in other retroviral systems (*2, 4, 5*), is located at the center of the assembly (Figs. 1B, 2, fig. S7C), and each of the four MVV IN tetramers is involved in its formation: core tetramers I and II contribute a CCD dimer each, while flanking tetramers III and IV provide a pair of synaptic CTDs that join the halves of the CIC structure. Approximately 20 bp of each vDNA end are well-defined in the electron density. The vDNA ends pass through the CIC structure approaching each other at an angle of 60°, with their terminal base pairs separated by IN CCD α4 helix (Fig. S10C). Each recessed 3' vDNA end is placed in the active site of a catalytic IN subunit (chains A and I), while the complementary non-transferred strand is threaded between the CCD and the synaptic CTD (Fig. S10C). The catalytic subunits intertwine by exchanging a pair of NTDs, with CCD-NTD linkers crossing the synaptic interface and contacting vDNA minor grooves (Fig. 2, fig. S10B, movie S1).

A layer of CTDs bridges the flanking and core tetramers of the intasome (fig. S8). Four pairs of CTDs belonging to the inner- and outermost IN chains of each lobe stack to form dimers nearly identical to those observed in MVV IN crystals and formed by the isolated HIV-1 CTD in solution (*13*) (figs S11). The β1-β2 loops of the CTD dimers from tetramers I and II insert into minor grooves of vDNA (fig. S10D) close to the end engaged by the active site of the opposing main tetramer (Fig. 1A). The interactions made across the synaptic interface likely ensure that a stable intasome forms only when the enzyme engages both vDNA ends. The NTDs from the inner core IN and one of the flanking IN lobes interact with the vDNA backbone (fig. S10D), forming an interface previously observed in crystals of the HIV-1 IN NTD-CCD construct (*14*) (fig. S12).

Due to the role of DNA in synaptic interface formation, retroviral INs and closely related DNA transposases tend to assemble into functional multimers only in the presence of their DNA substrates (*2, 4, 5, 15, 16*). The intasome would thus be expected to contain a

multiple of the minimal multimeric species found in solution, and this conjecture holds true for characterized intasomes containing tetramers of monomers (PFV) or dimers (RSV and MMTV) (*2, 4, 5*). In contrast, HIV-1 IN forms higher-order multimers in the absence of vDNA (*12, 17-19*), and crosslinked HIV-1 IN tetramers are functional *in vitro* (*20*). Similarly, MVV IN also forms tetramers and higher-order multimers in solution (fig. S2E). Thus, our structure explains how lentiviral INs, which are highly prone to self-associate, combine into the CIC structure. In lieu of the remarkable differences between intasome structures it would be of interest to compare quantitative proteomes of retroviral genera, although the number of IN molecules carried by the virus is unlikely to be limiting (*21, 22*).

The structural basis for α- and β- retroviral intasomes to comprise more than the minimal IN dimer-of-dimers architecture is relatively short IN CCD-CTD linkers (*4, 5*), which prohibit the CTD from the core subunits to insert into the synaptic interface. In HIV-1 and MVV IN, the CCD-CTD linkers assume α-helical conformations (*7*) (figs S9B, S11B), which likewise make it impossible for core tetramer subunits to provide the synaptic CTDs. Strikingly, although the linker region is the least conserved among lentiviral INs, it is invariably predicted to form an extended α helix (fig. S9C), arguing for conservation of the higher-order state of IN within lentiviral intasomes. The high stoichiometry of IN within the lentiviral intasome may help explain the notoriously pleiotropic phenotypes of HIV-1 IN mutant viruses (*23*). Because the 2-fold symmetric assembly contains eight structurally distinct IN subunits, each IN residue could play as many as eight distinct functions. The CTD plays the most functionally diverse roles within the intasome, contributing to intra- and inter-tetramer interactions, as well as DNA binding.

To visualize how the lentiviral intasome engages target DNA, we determined a cryo-EM structure of the MVV strand transfer complex to 8.6 Å resolution (Fig. 3A, fig. S13). In agreement with the analogous PFV and RSV structures (*3, 4*), target DNA binds between the

halves of the CIC structure. The synaptic CTDs insert their β1-β2 loops into expanded major grooves, which contributes to target DNA bending (Fig. 3B). Inspection of the surface potential distribution on the target DNA side of the complex highlighted several patches of positive charge, each corresponding to the cleft at the IN CCD dimerization interface (*6*) (Fig. 3B), which was recently implicated in non-catalytic interactions with nucleosomal DNA in the PFV system (*24*). Lentiviral integration is exquisitely selective towards highly active and gene-rich genomic loci, a property that is explained, at least in part, by the direct interaction between IN and chromatin-associated LEDGF (*9*). The MVV intasome structure seems to be compatible with binding as many as 16 molecules of the host factor (fig. S14). The ability to form such super-multivalent interactions may facilitate the viral integration machinery to locate chromatin highly enriched in LEDGF and possibly other marks associated with transcriptional activity. The MVV intasome system described here should be applicable to studies of HIV-1 INSTIs (fig. S2B). Moreover, the complexity of the lentiviral intasome, presenting multiple IN-IN interfaces, may be exploitable in anti-HIV/AIDS drug development.

REFERENCES AND NOTES

1. P. Lesbats, A. N. Engelman, P. Cherepanov, Retroviral DNA Integration. *Chem. Rev.*, DOI: 10.1021/acs.chemrev.1026b00125 (2016).
2. S. Hare, S. S. Gupta, E. Valkov, A. Engelman, P. Cherepanov, Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232-236 (2010).
3. G. N. Maertens, S. Hare, P. Cherepanov, The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326-329 (2010).
4. Z. Yin, K. Shi, S. Banerjee, K. K. Pandey, S. Bera, D. P. Grandgenett, H. Aihara, Crystal structure of the Rous sarcoma virus intasome. *Nature* **530**, 362-366 (2016).
5. A. Ballandras-Colas, M. Brown, N. J. Cook, T. G. Dewdney, B. Demeler, P. Cherepanov, D. Lyumkis, A. N. Engelman, Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* **530**, 358-361 (2016).
6. F. Dyda, A. B. Hickman, T. M. Jenkins, A. Engelman, R. Craigie, D. R. Davies, Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981-1986 (1994).

7. J. C. Chen, J. Krucinski, L. J. Miercke, J. S. Finer-Moore, A. H. Tang, A. D. Leavitt, R. M. Stroud, Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8233-8238 (2000).
8. M. Li, K. A. Jurado, S. Lin, A. Engelman, R. Craigie, Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLoS One* **9**, e105078 (2014).
9. R. Craigie, F. D. Bushman, Host Factors in Retroviral Integration and the Selection of Integration Target Sites. *Microbiol. Spectr.* **2**, DOI: 10.1128/microbiolspec.MDNA1123-0026-2014 (2014).
10. P. Cherepanov, LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res.* **35**, 113-124 (2007).
11. B. A. Johns, T. Kawasuji, J. G. Weatherhead, T. Taishi, D. P. Temelkoff, H. Yoshida, T. Akiyama, Y. Taoda, H. Murai, R. Kiyama, M. Fuji, N. Tanimoto, J. Jeffrey, S. A. Foster, T. Yoshinaga, T. Seki, M. Kobayashi, A. Sato, M. N. Johnson, E. P. Garvey, T. Fujiwara, Carbamoyl pyridone HIV-1 integrase inhibitors 3. A diastereomeric approach to chiral nonracemic tricyclic ring systems and the discovery of dolutegravir (S/GSK1349572) and (S/GSK1265744). *J. Med. Chem.* **56**, 5901-5916 (2013).
12. S. Hare, F. Di Nunzio, A. Labeja, J. Wang, A. Engelman, P. Cherepanov, Structural basis for functional tetramerization of lentiviral integrase. *PLoS Pathog.* **5**, e1000515 (2009).
13. P. J. Lodi, J. A. Ernst, J. Kuszewski, A. B. Hickman, A. Engelman, R. Craigie, G. M. Clore, A. M. Gronenborn, Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* **34**, 9826-9833 (1995).
14. J. Y. Wang, H. Ling, W. Yang, R. Craigie, Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.* **20**, 7333-7343 (2001).
15. D. R. Davies, I. Y. Goryshin, W. S. Reznikoff, I. Rayment, Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**, 77-85 (2000).
16. S. P. Montano, Y. Z. Pigli, P. A. Rice, The mu transpososome structure sheds light on DDE recombinase evolution. *Nature* **491**, 413-417 (2012).
17. P. Cherepanov, G. Maertens, P. Proost, B. Devreese, J. Van Beeumen, Y. Engelborghs, E. De Clercq, Z. Debyser, HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* **278**, 372-381 (2003).
18. C. J. McKee, J. J. Kessl, N. Shkriabai, M. J. Dar, A. Engelman, M. Kvaratskhelia, Dynamic modulation of HIV-1 integrase structure and function by cellular lens epithelium-derived growth factor (LEDGF) protein. *J. Biol. Chem.* **283**, 31802-31812 (2008).
19. S. P. Lee, J. Xiao, J. R. Knutson, M. S. Lewis, M. K. Han, Zn2+ promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. *Biochemistry* **36**, 173-180 (1997).
20. A. Faure, C. Calmels, C. Desjobert, M. Castroviejo, A. Caumont-Sarcos, L. Tarrago-Litvak, S. Litvak, V. Parissi, HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res.* **33**, 977-986 (2005).
21. T. Jacks, M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, H. E. Varmus, Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **331**, 280-283 (1988).

22. J. A. Briggs, M. N. Simon, I. Gross, H. G. Krausslich, S. D. Fuller, V. M. Vogt, M. C. Johnson, The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol.* **11**, 672-675 (2004).
23. A. Engelman, In vivo analysis of retroviral integrase structure and function. *Adv. Virus Res.* **52**, 411-426 (1999).
24. D. P. Maskell, L. Renault, E. Serrao, P. Lesbats, R. Matadeen, S. Hare, D. Lindemann, A. N. Engelman, A. Costa, P. Cherepanov, Structural basis for retroviral integration into nucleosomes. *Nature* **523**, 366-369 (2015).

Acknowledgements: We thank G. Schoehn for help with preliminary cryo-EM, the staff of the Diamond beamlines I04 and I04-1 for assistance with X-ray data collection, P. Afonine for generous advice on real-space refinement, L. Collinson, R. Carzaniga, T. Pape, P. Walker and A. Purkiss for EM, X-ray crystallography and software support, L. Heck for expert assistance with the computer cluster, and G. Maertens for comments on the manuscript. The cryo-EM maps, pseudo-atomic models and the X-ray structures were deposited with the PDB and the EM databank with accession codes EMD-4138, EMD-4139, PDB-5M0Q, PDB-5M0R, PDB-5LLJ, and PDB-5T3A. This work was supported by NIH grant GM082251 (PC and ANE), The Francis Crick Institute (PC), The Wellcome Trust (AK), and Icelandic Research Fund (VA, SRJ). The Division of Structural Biology Particle Imaging Center EM Facility at University of Oxford was founded by The Wellcome Trust JIF award 060208/Z/00/Z and is supported by equipment grant 093305/Z/10/Z.

FIGURE LEGENDS

Fig. 1: Cryo-EM reconstruction of the MVV intasome. A. Fitted intasome model color-coded to highlight IN subunits including 12 NTDs, 16 CCDs, and 14 CTDs. Molecules of vDNA in dark grey are surrounded by core tetramers I and II (colored in green, light green, sky blue, and blue), and flanking tetramers III and IV (red, yellow, pink, and purple). **B-C.**

Views of the map in two alternative orientations. The CIC structure is highlighted with a black outline in panel B.

Fig. 2: CIC structure in MVV and previously characterized synaptic complexes. The CIC in each structure is shown in color with the remainder in grey; yellow CTDs indicate domains donated by flanking IN subunits.

Fig. 3: Target DNA binding and surface electrostatic potential distribution. A. Cryo-EM reconstruction of the MVV strand transfer complex at 8.6 Å resolution viewed in two orientations. Protein, vDNA, and target DNA are shown in white, dark grey and bordeaux, respectively. **B.** Pseudo-atomic model of the STC with the protein portion of the structure in space-fill mode and colored by charge.

Supplementary Materials:

Materials and Methods

Figs S1 - S14

Tables S1, S2

Movie S1

References

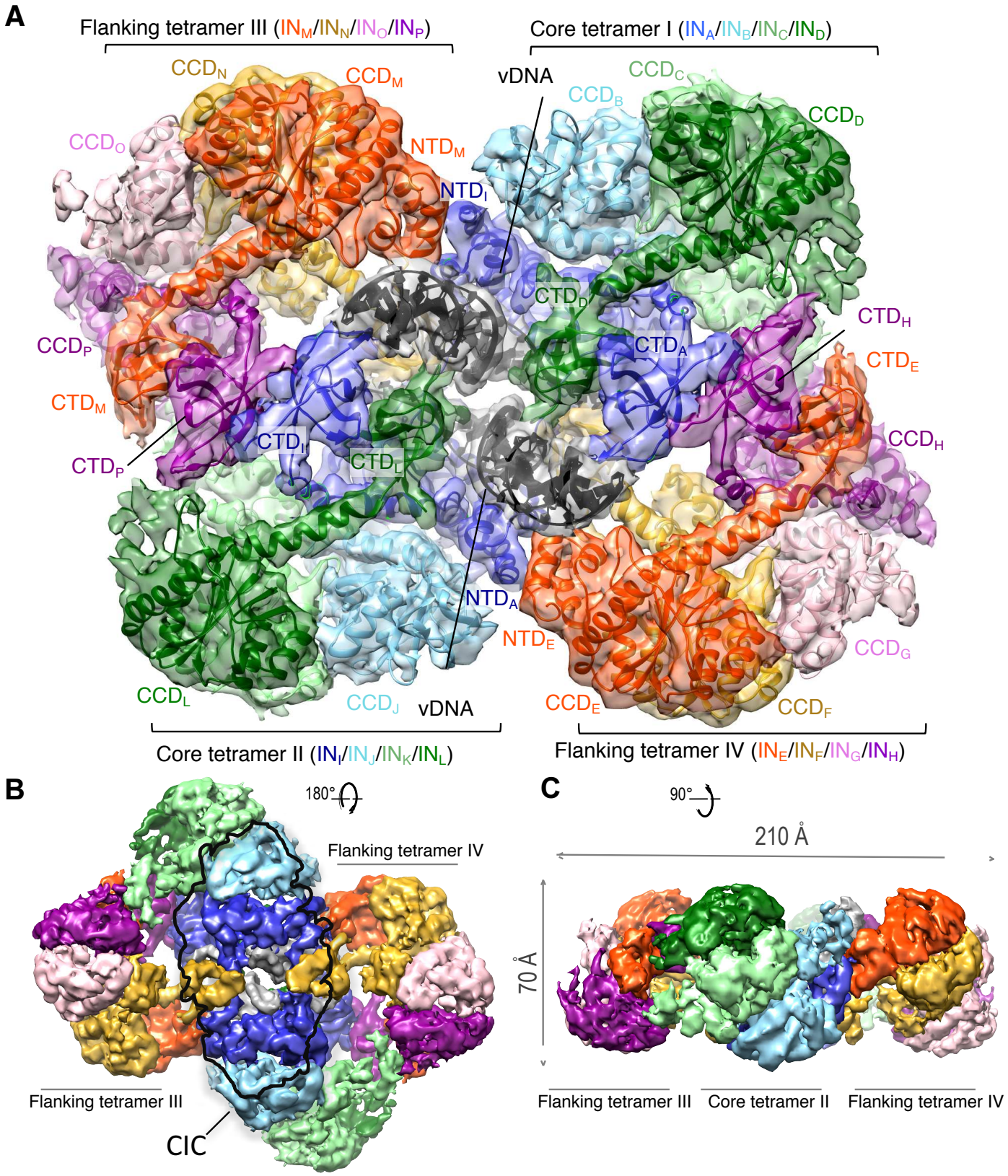
A
Flanking tetramer III (INM/INN/INO/INP)
Core tetramer I (INA/INB/INC/IND)
CCDN
CCDM
CCDO
NTDM
NTDI
vDNA
CCDC
CCDB
CCDD
CTDD
CTDH
CTDA
CTDE
CCDP
CTDM
CTDI
CTDL
CTDP
CCDH
NTDA
NTDE
CCDG
CCDL
CCDJ
vDNA
CCDE
CCDF
Core tetramer II (INI/INJ/INK/INL)
Flanking tetramer IV (INE/INF/ING/INH)
B
180°
Flanking tetramer IV
Flanking tetramer III
CIC
C
90°
210 Å
70 Å
Flanking tetramer III
Core tetramer II
Flanking tetramer IV

Figure 1

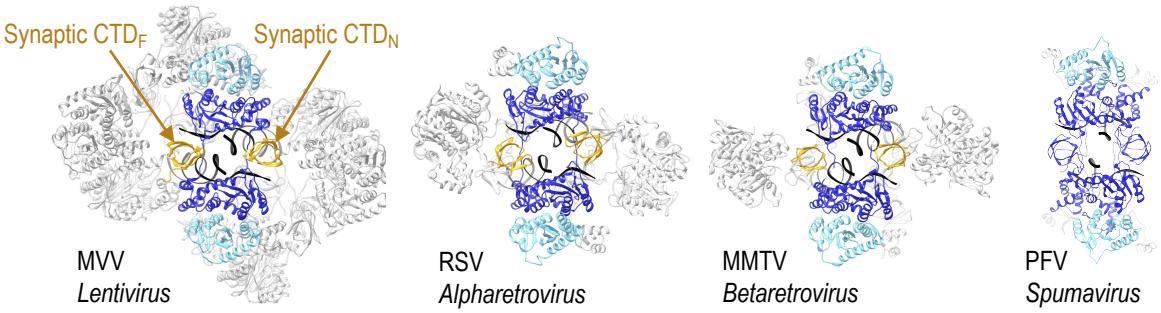
Synaptic CTD$_F$
Synaptic CTD$_N$
MVV
Lentivirus
RSV
Alpharetrovirus
MMTV
Betaretrovirus
PFV
Spumavirus

Figure 2

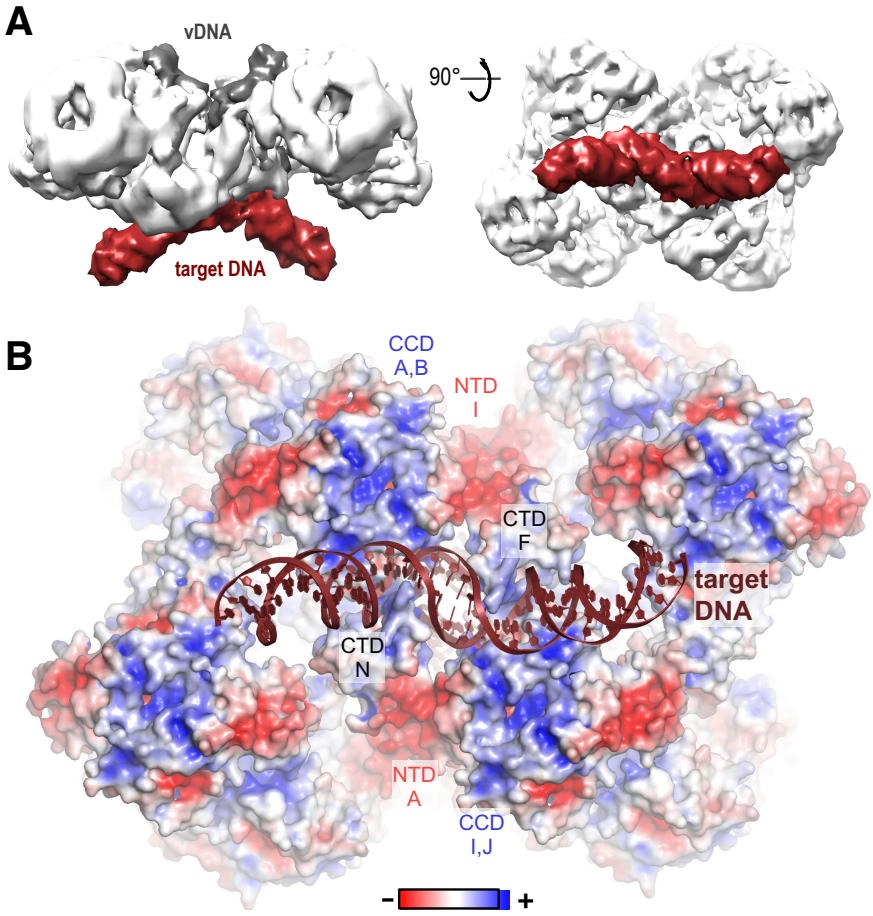

Figure 3 (1column)