Structural basis for retroviral integration into nucleosomes

Daniel P. Maskell1, Ludovic Renault2,3, Erik Serrao4, Paul Lesbats1, Rishi Matadeen5, Stephen Hare6, Dirk Lindemann7, Alan N. Engelman4, Alessandro Costa2 & Peter Cherepanov1,6

Retroviral integration is catalysed by a tetramer of integrase (IN) assembled on viral DNA ends in a stable complex, known as the intasome1–6. How the intasome interfaces with chromosomal DNA, which exists in the form of nucleosomal arrays, is currently unknown. Here we show that the prototype foamy virus (PFV) intasome is proficient at stable capture of nucleosomes as targets for integration. Single-particle cryo-electron microscopy reveals a multivalent intasome–nucleosome interface involving both gyres of nucleosomal DNA and one H2A–H2B heterodimer. While the histone octamer remains intact, the DNA is lifted from the surface of the H2A–H2B heterodimer to allow integration at strongly deviated from that in the structure of free nucleosome 3.5 positions (Fig. 1a and Extended Data Fig. 1b), a property that depended on nucleosome structure (Extended Data Fig. 2d). Lower thermal stability of the selected nucleosomes (Extended Data Fig. 3) suggests enhanced flexibility, which may aid in the conformational adaptation required for intasome binding (see later). The D02 nucleosome afforded isolation of a stable complex with the intasome, which, upon incubation with 5 mM Mg2+, converted into the strand transfer complex with integrated viral DNA ends (Fig. 1c, d). DNA sequencing analysis of the resulting products revealed integration into a single site, offset from the middle of the D02 DNA by 36 bp, indicating that the complexes comprised the dyad-related nucleosomal site dissociated during purification (Extended Data Fig. 1d).

To determine the structure of the 400 kDa intasome–D02 nucleosome complex before strand transfer, we acquired single-particle cryo-electron microscopy (EM) data. The resulting electron density map, calculated to 7.8 Å resolution (Extended Data Fig. 4), allowed unambiguous docking of the intasome2,3 and the nucleosome5,16 crystal structures (Fig. 2a). The intasome contains a homotetramer of IN made of two types of subunits. The inner IN chains provide catalytic function, synapse the viral DNA ends and form the tDNA-binding groove. The function of the outer IN subunits, which attach to the inner subunits via the canonical catalytic core domain (CCD) dimerization interface17, has been unclear. The path of the viral and nucleosomal DNA backbone, the histone octamer and the inner subunits of the IN tetramer are well defined in the electron density map (Figs 2a, 3a and Supplementary Video 1).

The cryo-EM structure reveals an extensive intasome–nucleosome interface, involving three IN subunits, both gyres of the nucleosomal DNA and one H2A–H2B heterodimer (Fig. 2a and Supplementary Video 1). The tDNA-binding groove of the intasome engages nucleosomal DNA above one of the H2A–H2B histone heterodimers, in agreement with the location of the preferred integration site. The path of DNA captured within the tDNA-binding groove of the intasome strongly deviates from that in the structure of free nucleosome (Fig. 2b). Here, DNA is deformed and lifted by ~7 Å from the surface of the H2A–H2B heterodimer to adopt a conformation strikingly matching the sharply bent naked tDNA in crystals of the PFV target capture complex18 (Fig. 2b, Extended Data Fig. 5 and Supplementary Video 1). Notably, the H2A L1-loop directly underlying the integration site is a hotspot for structural variability within the histone core19. Thus, the segment of nucleosomal DNA preferentially targeted by the intasome may be particularly malleable to conformational adaptation. The interface is extended by ancillary contacts on both sides of the tDNA-binding groove. A triad of loops belonging to the intasome may be particularly malleable to conformational adaptation. The interface is extended by ancillary contacts on both sides of the tDNA-binding groove.
inner IN subunits cradle the carboxy-terminal helix of H2B (Fig. 2a). On the other side, the intasome leans against the second gyre of nucleosomal DNA, using a saddle-shaped surface of the CCD–CCD interface (Figs 2a and 3a; secondary interface). These interactions may compensate for the requirement to deform DNA beyond its ground state on the nucleosome. An early study, albeit using undefined integrase–DNA complexes, implicated SHL 3.5 and SHL 1.5 positions, which are separated from the dyad by 3.5 turns of DNA helix or 36 bp. The products include viral DNA joined to nucleosomal DNA fragments (Extended Data Fig. 6d). FITC, fluorescein isothiocyanate. Pull-down of recombinant nucleosomes with biotinylated intasome in the presence of variable NaCl concentrations; where indicated, the intasome was assembled with A188D IN. Bound material, separated in SDS–PAGE gels, was stained to detect proteins (IN, H3, H2A, H2B and H4) and nucleosomal DNA (Nuc. DNA). WT, wild type. c, Isolation of the intasome–D02 nucleosome complex by size–exclusion chromatography. Peak fractions of intasome (red trace), D02 nucleosome (blue) and the complex (green) were separated by SDS–PAGE (inset). d, DNAs from D02 nucleosome, intasome, and the intasome–nucleosome complex, before and after incubation with 5 mM MgCl₂, were separated by PAGE and detected with GelRed.

Next, we introduced amino acid substitutions into the relevant regions of PFV IN and evaluated the ability of the resulting intasomes to engage nucleosomes in vitro. Some substitutions were engineered selectively into the inner or the outer subunits of the intasome using a pair of IN variants—K120E and D273K—that co-assemble into hybrid intasomes through restricted localization to the inner and the outer subunit, respectively (Extended Data Fig. 6a–c). IN residues Gln 137, Lys 168 and Lys 159 are located in the vicinity of the contacts with the second gyre of nucleosomal DNA, while Pro 135, Pro 239 and Thr 240 approach the C-terminal helix of H2B (Fig. 3a). Substitutions at these positions affect the ability of the intasome to engage nucleosomes to various degrees. In particular, K168E in the inner and outer subunits, or the double substitution P135E/T240E in the inner subunits, grossly affected the interaction with recombinant and native human nucleosomes (Fig. 3b, c). The same substitutions reduced the ability of the intasome to integrate into the W601 nucleosome (Fig. 3d), while importantly preserving strand transfer activity into naked plasmid DNA (Extended Data Fig. 6d).

To assess the importance of the observed interactions under conditions of viral infection, we used a PFV vector system with a green fluorescent protein (GFP) reporter (Extended Data Fig. 7a)⁹. Initially, we mapped the genomic positions of 153,447 unique integration events of wild-type vector in human epithelial HT1080 cells. As a reference data set, we determined ∼2.2 million integration sites using purified PFV intasomes and deproteinized human DNA. In agreement with published observations⁹,²¹, PFV disfavoured transcription units: 32% of integration sites were found in RefSeq genes, which is ∼16% below the gene-targeting frequency in the reference data set (P = 10⁻³⁰⁶; Fig. 3e and Extended Data Table 1). Ranking transcription units by their activity in HT1080 cells, we discovered that PFV integration strongly contrasted with local transcriptional activity (P = 10⁻³⁰⁶; Fig. 3e). Furthermore, integration sites accumulated within gene-poor and lamin-A/B-associated genomic regions²², indicating a strong bias towards condensed perinuclear heterochromatin (Fig. 3e and Extended Data Table 1).
PFV vectors incorporating K168E IN or P135E/T240E IN as the inner subunit of the intasome displayed 2–4-fold defects in their ability to stably transduce HT1080 cells, while still strongly outperforming matched controls harbouring catalytically inert IN variants (Extended Data Fig. 7b–d). The residual infectivity allowed the mapping of 14,872 unique integration sites of K168E and P135E/T240E IN, respectively. In accordance with their location outside of the primary interface, the substitutions introduced into the inner subunits in the hybrid control (vector produced using a combination of K120E and D273K/D185N/E221Q pcoP-POL packaging vectors encoding functional inner and catalytically incompetent outer IN subunits, respectively) or P135E/T240E (the substitutions introduced into the inner subunits in the hybrid control background). Asterisks indicate significant departures of the mutants from their respective controls ($\chi^2$ test: *P < 0.01, **P < 10$^{-5}$).

Using the nucleosomal structure as a landing platform may allow the viral machinery to sense epigenetic marks as well as provide a surface for recruitment of chromatin remodellers for disassembly of the post-catalytic strand transfer complex.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 February; accepted 15 April 2015.

Published online 10 June 2015.

1. Li, M., Mizuuchi, M., Burke, T. R. Jr & Cniogile, R. Retroviral DNA integration: reaction pathway and critical intermediates. EMBO J. 25, 1295–1304 (2006).
2. Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cheranopan, P. Retroviral intasome assembly and inhibition of DNA strand transfer. Nature 464, 232–236 (2010).
3. Maertens, G. N., Hare, S. & Cheranopan, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature 468, 326–329 (2010).
4. Montañá, S. P., Pigli, Y. Z. & Rice, P. A. The murine transpososome structure sheds light on DDE recombinase evolution. Nature 491, 413–417 (2012).
5. Lugur, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389, 251–260 (1997).
6. Przyca, P. M., Sil, A. & Varma, H. E. Retroviral integration into minichromosomes in vitro. EMBO J. 11, 291–303 (1992).
7. Pryciak, P. M. & Varumus, H. E. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell 69, 769–780 (1992).

8. Pruss, D., Bushman, F. D. & Wolffe, A. P. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc. Natl Acad. Sci. USA 91, 5913–5917 (1994).

9. Müller, H. P. & Varumus, H. E. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. 13, 4704–4714 (1994).

10. Wang, G. P., Ciuffi, A., Leipzig, J., Berry, C. C. & Bushman, F. D. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res. 17, 1186–1194 (2007).

11. Roth, S. L., Malani, N. & Bushman, F. D. Gammaretroviral integration into nucleosomal target DNA in vivo. J. Virol. 85, 7393–7401 (2011).

12. Baller, J. A., Gao, J., Stamenova, R., Curcio, M. J. & Voytas, D. F. A nucleosomal surface defines an integration hotspot for the Saccharomyces cerevisiae Ty1 retrotransposon. Genome Res. 22, 704–713 (2012).

13. Hare, S., Stamnen, R., Cucito, M. J. & Voytas, D. F. A nucleosomal surface defines an integration hotspot for the Saccharomyces cerevisiae Ty1 retrotransposon. Genome Res. 22, 3020–3028 (2012).

14. Yin, Z., Lapkowski, M., Yang, W. & Craigie, R. Assembly of prototype foamy virus strand transfer complexes on product DNA bypassing catalysis of integration. Protein Sci. 21, 1849–1857 (2012).

15. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. 276, 19–42 (1998).

16. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 A resolution. J. Mol. Biol. 319, 1097–1113 (2002).

17. Dyda, F. et al. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidytransferases. Science 266, 1381–1386 (1994).

18. Shayan, A. K., Landsman, D. & Panchenko, A. R. Nucleosome adaptability conferred by sequence and structural variations in histone H2A–H2B dimers. Curr. Opin. Struct. Biol. 32, 48–57 (2015).

19. Mullers, E. et al. Novel functions of prototype foamy virus Gag glycine-arginine-rich boxes in reverse transcription and particle morphogenesis. J. Virol. 85, 1452–1463 (2011).

20. Trobridge, G. D. et al. Foamy virus vector integration sites in normal human cells. Proc. Natl Acad. Sci. USA 103, 1498–1503 (2006).

21. Nowrouzi, A. et al. Genome-wide mapping of foamy virus vector integrations into a human cell line. J. Gen. Virol. 87, 1339–1347 (2006).

22. Guelen, L. et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453, 948–951 (2008).

23. Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science 328, 1161–1164 (2010).

24. Schwabish, M. A. & Struhl, K. Ast1 mediates histone eviction and deposition during elongation by RNA polymerase II. Mol. Cell 22, 415–422 (2006).

25. Gupta, K. et al. Solution conformations of prototype foamy virus integrase and its stable synaptic complex with US viral DNA. Structure 20, 1918–1928 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the European Union FP7 HIVINNOV consortium grant 305137 (to P.C.), the US National Institute of General Medical Sciences P50 grant GM082251-06 (to P.C.) and the US National Institutes of Health R01 grant AI070042-08 (to A.N.E.). Data collection was in part funded by the Netherlands Centre for Electron Nanoscopy (NeCEN) by grants from the Netherlands Organisatie voor Wetenschappelijk Onderzoek (project 175.010.2009.001) and by the European Union’s Regional Development Fund through ‘Kansen voor West’ (project 21Z014). We would like to thank L. Collinson, R. Carzaniga and Kirsty MacLellan-Gibson for EM access, R. Horton-Harpin for provision of HeLa cell pellets and assistance with tissue culture. We also thank F. Santoni, N. Sweeney and all our colleagues for helpful discussions.

Author Contributions D.P.M. analysed interactions of the PFV intasome and nucleosomes, discovered conditions to produce the stable intasome–nucleosome complex and prepared it for EM; L.R., D.P.M. and A.C. performed all EM work with the exception of cryo-EM grid preparation and screening, which was performed by L.R.; R.M. collected cryo-EM data; L.R. and A.C. determined the structure. S.H. designed the co-dependent K120E–D273K PFV IN pair; D.L. designed and provided wild-type PFV vector constructs; P.C. cloned PFV vector mutants; P.C. and P.L. carried out PFV infections; E.S. and A.N.E. developed the protocol and carried out sequencing of PFV integration sites; P.C. analysed integration site distributions.

Author Information The cryo-EM electron density map has been deposited in the Electron Microscopy Data Bank under accession number EMD-2992. Integration sites have been deposited in the NCBI Gene Expression Omnibus under accession number GSE67730. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C. (alessandro.costi@crick.ac.uk) or P.C. (peter.chepepanov@crick.ac.uk).
PFV intasome assembly. PFV IN was expressed in bacteria and purified as previously described. Synthetic DNA oligonucleotides used for intasome assembly were purified using high-performance liquid chromatography (HPLC) (Midland Certified). For pull-downs, activity assays and EM studies, PFV intasomes were assembled with stabilized processed U5 DNA (ref. 27) obtained after autoclaving oligonucleotides 5′-TGCGAAATTCCATGACA (transferred strand) and 5′-ATTGCTCATGAAAATGGTGACGCA. For experiments involving biotin pull-downs or in-gel fluorescent detection of strand transfer products, the transferred strand oligonucleotide was synthesized with 5′ triethylene glycol biotin or fluorescein, respectively. To allow compatibility with the linker-mediated PCR protocols, a longer donor DNA construct spanning 47 bp of processed PFV U5 end and made by annealing oligonucleotides 5′-GATGA TCTCCTTTGATAATCAATATACAAAAATCTTGCAGA and 5′-ATTGCTCATGAAAATGGTGACG were used in a model in vitro genetic integration sites.

The intasomes were assembled according to published procedures27, by dialysing 120 μM PFV IN and 50 μM donor DNA duplex combined in 500 mM NaCl and 50 mM BisTris propane-HCl, pH 7.45, against excess low-salt buffer containing 200 mM NaCl, 20 mM BisTris propane-HCl, pH 7.45, 4 mM dithiothreitol (DTT) and 25 μM ZnCl2 for 16 h at 18 °C. For hybrid intasome assembly, 60 μM of each IN was produced (R120E and D227R) was used. The intasomes were purified by size-exclusion chromatography through a 10/300 GL Superdex-200HR column (GE Healthcare) in 20 mM BisTris propane-HCl, pH 7.45, 320 mM NaCl and kept on ice for immediate use.

Preparation of native human mononucleosomes. Native human mononucleosomes were prepared according to established protocols28 with minor modifications. HeLa cells were lysed in 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT, 0.5 mM phenylmethylsulfonylfluoride (PMSF), 10 mM HEPES-NaOH, pH 7.9, using a 40 μl Dounce homogenizer (Wheaton). Nuclei, harvested by gentle centrifugation, were digested with 0.5 U ml−1 micrococcal nuclease (Sigma-Aldrich) in 4 volumes of 0.34 M sucrose, 3 mM CaCl2, 60 mM KCl, 0.5 mM PMSF and 50 mM Tris-HCl, pH 7.5, for 10 min at 37 °C. The reaction was stopped with 50 mM EDTA, and the nucleosomes, extracted by addition of 0.5 M NaCl, were dialysed overnight at 4 °C against 650 mM NaCl, 2 mM EDTA, 1 mM β-mercaptoethanol and 20 mM HEPES, pH 7.4. Nucleosomes were isolated by size-exclusion chromatography through a Superdex-200 column operated in 650 mM NaCl, 25 mM Tris-HCl, pH 7.5, supplemented with 2.5% w/v PolyG. Nucleosomes were concentrated by precipitation with ethanol, re-dissolved in 1 mM EDTA and used for nucleosome assembly were generated by PCR using in-house produced Pfu polymerase (GE Healthcare) and stored on ice.

PFV intasome and 200 μM BisTris propane, pH 7.45, for 15 min at 37 °C. The reaction was stopped by addition of 0.5% SDS and 25 mM EDTA. DNA products, deproteinized by digestion with 30 μg proteinase K at 37 °C for 1 h and ethanol precipitation, were separated in 4–12% TBE PAGE gels. Fluorescein-labelled DNA was detected using a Typhoon TRIO fluorescence scanner (GE Healthcare); non-labelled DNA was visualized by staining with GelRed. Strand transfer assays using naked supercoiled plasmid target DNA was done according to published procedures2.

Preparation of intasome-nucleosome complex for EM. The complex, assembled by incubating 200 μM PFV intasome and 200 μM D02 nucleosome in 300 μl of 320 mM NaCl, 20 mM BisTris propane, pH 7.45, for 20 min at room temperature, was purified by size-exclusion chromatography over a Superdex-200HR 10/300 column in 320 mM NaCl, 25 mM BisTris propane-HCl, pH 7.45. Fractions containing the complex were immediately used to prepare EM grids.

Negative-stain single-particle analysis. Negative-stain EM grids were prepared as follows. Carbon was evaporated onto freshly cleaned mica using a Q150TE coater (Quorum Technologies) and baked for 2 h at 50 °C before floating onto 400-mesh copper grids (Agar Scientific). Dry continuous carbon grids were glow discharged for 30 s at 45 mA using a 100 μm glow discharger (Electron Microscopy Sciences). A 4 μl drop of sample was applied onto the glow-discharged grid immediately after elution from the gel filtration column. The grid was sequentially laid on top of four distinct 75 μl drops of a 2% (w/v) uranyl formate solution, stirred for 10 s, before blotting to dryness and being stored at room temperature before imaging. Negative stain grids were scanned on a T2 Spirit Lab6 microscope (FEI) and data were collected on a JEOL-2100 LaB6 electron microscope (JEOL) operated at 200 kV. Images were recorded at a nominal magnification of ×50,000 on a Ultrascan 4K × 4K CCD camera (Gatan), resulting in a 2.2 Å pixel size at the specimen level. A total of 135 micrographs were collected with a 1 to 2 μm defocus and using 20 e− per Angström². Contrast transfer function (CTF) estimation was performed with CTFFind3d (ref. 39) and micrographs were phase-flipped using BSoFT (ref. 31). Reference-free three-dimensional averages were calculated using routine MSA/MRA IMAGIC protocols29. The initial three-dimensional model was determined starting from a sphere, and further refined with multi-model projection-matching approaches using libraries from the EMAN2 and SPARX packages12,34 (Extended Data Fig. 9).

Cryo-EM structure determination. A 4 μl sample drop was applied to plasma-cleaned C-Flat grids (400 mesh CF-1/1, Electron Microscopy Sciences). After 1 min incubation, grids were double side blotted for 3.8 s in a CP3 cryo-plunger (Gatan) at 80% humidity and plunged frozen into −172 °C liquefied ethane. Cryo-grids were screened for ice quality using a 914 side-entry cryo-holder (Gatan) on a JEOL-2100 LaB6 operated at 200 kV and equipped with an UltraScan 4K × 4K CCD camera (Gatan). Cryo-grids were stored in liquid nitrogen and dry-shipped to the Netherlands Centre for Electron Nanoscopy (NeCEN). At NeCEN, cryo-grids were loaded into a Titan Krios electron microscope (FEI) for automated data collection with the EPU software (FEI) over a period of 3 days. Images were recorded at a nominal magnification of ×59,000 on a Falcon direct electron detector. One-thousand four-hundred and seventy-nine micrographs were recorded using a 1.5/3.5 μm defocus range with an electron dose of 40 e− per Å². A first cryo-EM map was calculated starting from a subset of 16,702 particles manually picked from 157 micrographs using Xmipp software25. Multimodel refinement protocols described earlier were employed to solve a 19.7 Å resolution cryo-structure (based on 6,057 particles), using the 60 Å low-pass filtered negative stain volume as a starting model. A complete, 33-micrograph refinement was performed after automatic processing on the best 932 micrographs, selected after inspection of the power spectrum. Contrast transfer function was estimated using CTFFIND3d. All further processing was performed within the RELION 1.2 environment36.
resulting in a cleaned 193,569-particle data set. A three-dimensional classification was subsequently performed using three classes and starting from the initial cryo-EM map; 83,887 particles belonging to the best three-dimensional class were selected and subjected to one further round of three-dimensional classification. This approach yielded an improved volume calculated from 53,887 particles. These particles were subsequently separated into 187 groups, on the basis of their refined intensity scale-factor, and used in a final three-dimensional refinement using the selected, 60 Å low-pass filtered three-dimensional class as a starting model. The density map was corrected for the modulation transfer function (MTF) of the Falcon detector and sharpened by applying a $-818 \AA^2$ factor, estimated automatically with the RELION post-processing function (Extended Data Fig. 4g). Final resolution after post-processing was 7.84 Å, according to the 0.143 cut-off criterion (Extended Data Fig. 4d). The handedness of the map was verified by inspection of the duplex DNA density. UCSF Chimera was used for automated rigid-body docking and generating figures and videos.35

**PFV vector infections and integration site analysis.** Single-cycle viruses were produced by co-transfection of HEK293T cells (Cell Services, London Research Institute) with pMD9 (GFP reporter PFV vector) and codon-optimized foamy virus gag (PFV GAG, POL and ENV packaging constructs, as previously described19,38). Hybrid PFV vectors and matched controls were pseudotyped with the less fusogenic envelope of the type B envelope, using MiSeq Illumina platform at the Dana-Farber Cancer Institute Molecular and lamina-associated domains (GSE22428)38, were analysed using BEDTools software suite45. Nucleotide sequence logos were generated using WebLogo software (http://weblogo.threeplusone.com/)49.

26. Valkov, E. et al. Functional and structural characterization of the integrase from the protozoan foamy virus. Nucleic Acids Res. 37, 243–255 (2009).
27. Hare, S. et al. Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. Proc. Natl Acad. Sci. USA 107, 20057–20062 (2010).
28. Schnitzler, G. R. Isolation of histones and nucleosome cores from mammalian cells. Curr. Protoc. Mol. Biol. Chapter 21, Unit 21 (2001).
29. Dyer, P. N. et al. Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol. 375, 23–44 (2004).
30. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 142, 334–347 (2003).
31. Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. J. Struct. Biol. 157, 3–18 (2007).
32. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. J. Struct. Biol. 116, 17–24 (1996).
33. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 46–47 (2007).
34. Hohn, M. et al. SPARX, a new environment for Cryo-EM image processing. J. Struct. Biol. 157, 47–55 (2007).
35. de la Rosa-Trevin, J. M. et al. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J. Struct. Biol. 184, 321–328 (2013).
36. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530 (2012).
37. Petersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612 (2004).
38. Heinikel, M. et al. Improved primate foamy virus vectors and packaging constructs. J. Virol. 76, 3774–3783 (2002).
39. Stimmagel, K. et al. Differential pH-dependent cellular uptake pathways among foamy viruses elucidated using dual-colored fluorescent particles. Retrovirology 9, 71 (2012).
40. Stange, A., Lutzenegger, D., Reh, J., Weisshenkorn, W. & Lindemann, D. Subviral particle release determinants of prototype foamy virus. J. Virol. 82, 9858–9869 (2008).
41. Rasheed, S., Nelson-Rees, W. A., Toth, E. M., Arnstein, P. & Gardner, M. B. Characterization of a newly derived human sarcoma cell line (HT-1080). Cancer Res. 33, 1027–1033 (1974).
42. Matrejek, K. A. et al. Host and viral determinants for MoxB restriction of HIV-1 infection. Retrovirology 11, 90 (2014).
43. Chatterjee, A. G. et al. Serial number tagging reveals a prominent sequence preference of retrotransposon integration. Nucleic Acids Res. 42, 8449–8464 (2014).
44. Schroeder, A. R. et al. HIV-1 integration in the human genome activates active genes and local hotspots. Cell 110, 521–529 (2002).
45. Kent, W. J. BLAT—the BLAT-like alignment tool. Genome Res. 12, 656–664 (2002).
46. Deyle, D. R. et al. A genome-wide map of adeno-associated virus-mediated human gene targeting. Nature Struct. Mol. Biol. 21, 969–975 (2014).
47. Meuleman, W. et al. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. Genome Res. 23, 270–280 (2013).
48. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
49. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190 (2004).
Extended Data Figure 1 | PFV integration into recombinant mononucleosomes. a, W601, D02, F02 and H04 nucleosome core particles (left) and W601 nucleosome with 30 bp tails mimicking linker DNA (W601 L30; right) were separated by native PAGE and detected by staining with ethidium bromide. b, Major products of PFV integration into a nucleosome core particle. Concerted integration of intasomal oligonucleotides (blue lines) into discontinuous tDNA (black lines) produces pairs of strand transfer products containing viral DNA mimics joined to tDNA fragments via 4 bp gaps. nt, nucleotides. c, PFV integration into nucleosome core particles (W601) and extended nucleosomes (W601 L30). Fluorescein-labelled intasomal DNA and reaction products were separated by PAGE and detected by fluorescence scanning. Migration positions of the strand transfer products obtained with W601 L30 nucleosome shift relative to those with the W601 core particle by ~30 bp. Thus, linker DNA does not appear to influence integration. d, Positions of integration events on D02 nucleosomal DNA before (left) or after (right) purification of the complex. The histograms show relative frequencies of integration events along the D02 DNA fragment into the top (blue bars) or bottom (pink bars) strands. The inset shows the nucleotide sequence at the preferred integration site; arrowheads indicate precise positions of the major integration events into the top and bottom strands of D02 DNA.
Extended Data Figure 2 | Pull-down of native nucleosomes and naked DNA by biotinylated PFV intasome. a, Mono–nucleosomes prepared by micrococcal nuclease digestion of HeLa cell chromatin were incubated with biotinylated intasomes under conditions of indicated ionic strength (190–340 mM NaCl). The intasomes used were wild type (WT), A188D or a hybrid intasome lacking the NTDs and CTDs on the outer subunits (indicated as DNTD/DCTD; see main text and Extended Data Fig. 6a–c for details of hybrid intasome design). The intasome–nucleosome complexes were isolated on streptavidin agarose and separated by SDS–PAGE. Proteins and nucleosomal DNA were detected by staining with Coomassie blue and GelRed, respectively. Two leftmost lanes contained 50% and 25% of input nucleosomes, as indicated. Migration positions of protein sizes standards (kDa) are shown to the left of the gel. b, Isolation of HeLa nucleosomes preferentially binding to the PFV intasome. Biotinylated wild-type or A188D intasomes were incubated with tenfold excess HeLa nucleosomes in the presence of 290 mM NaCl. Nucleosomal DNA recovered with wild-type intasome was cloned into a bacterial vector; the histogram depicts distribution of nucleosomal insert sizes obtained in this experiment. The inset shows separation of deproteinized nucleosomal DNA from 10% of input nucleosome material and from the fractions recovered with wild-type and A188D intasomes. c, Nucleotide sequences of three human DNA fragments (H04, F02 and D02) recovered with the intasome and used to assemble recombinant nucleosomes in this work. d, Naked W601 D02, F02 or H04 DNA was incubated with biotinylated wild-type or A188D intasomes in the presence of 190 or 240 mM NaCl, as indicated; DNA fractions recovered after pull-down on streptavidin beads were separated by PAGE and detected by staining with GelRed.
Extended Data Figure 3 | Thermal denaturation of recombinant nucleosomes. Derivative melt profiles of recombinant nucleosomes used in this study. The table in the inset shows experimentally determined melting temperatures.
Extended Data Figure 4 | Overview of the cryo-EM data. a, Representative micrograph of frozen hydrated intasome–nucleosome complex. b, Two-dimensional class averages (phase-flipped only; box size 26 nm). c, Euler angle distribution of all particles included in the final three-dimensional reconstruction. Sphere size relates to particle number. d, Gold standard Fourier-shell correlation and resolution using the 0.143 criterion. e, Three-dimensional volume of the intasome–nucleosome complex refined with RELION. f, Match between reference-free two-dimensional class averages and three-dimensional re-projections of the cryo-EM structure. Two-dimensional class averages of fully contrast transfer function (CTF)-corrected particles are matched with the re-projections of the refined three-dimensional structure before map sharpening (post-processing); 30–6 Å band-pass filter imposed. g, Overview of the three-dimensional classification and structure refinement. The initial negative stain structure was used for one round of structure refinement using a smaller cryo data set. The resulting map was used as a starting model for one round of three-dimensional classification (three classes) on a complete cryo data set. Particles from the two most populated three-dimensional classes were merged and used for one further round of three-dimensional classification (six classes). Each three-dimensional class was refined independently; the most populated three-dimensional class comprising 53,887 particles refined to 7.8 Å resolution.
Extended Data Figure 5 | Nucleosomal DNA plasticity. Overview of the intasome–nucleosome complex structure (left) and a magnified stereo view of nucleosomal DNA engaged within tDNA binding cleft of the intasome (right). DNA conformations as in free nucleosomes (Protein Data Bank accession 1KX5) and as tDNA in complex with the PFV intasome (3OS1) are shown in light and dark grey, respectively; the arrowhead shows approximate direction of the DNA deformation. Asterisks indicate nucleosomal DNA ends.
Extended Data Figure 6 | Hybrid intasomes: structure-based design and validation in vitro. a, Views on the environment of Lys 120 and Asp 273 PFV IN residues within the intasome structure. Protein is shown as cartoons with side chains of selected amino acid residues shown as sticks; the cartoons and carbon atoms of the inner and outer IN chains are shown in green and light orange, respectively. Lys 120 of the outer and Asp 273 of the inner IN subunit are involved in a network of interactions; by contrast, Lys 120 of the inner and Asp 273 of the outer IN subunit are solvent-exposed. Consequently, IN mutants harbouring substitutions of Lys 120 or Asp 273 can only have a role in the inner or outer intasomal subunits, respectively. b, PFV IN mutants K120E and D273K are co-dependent for intasome assembly. Products of intasome assembly using wild-type (WT), D273K, K120E PFV IN or an equimolar mixture of D273K and K120E INs were separated by size-exclusion chromatography. Elution positions of the intasome, IN and free DNA are indicated. The assembly was successful with wild-type IN or with a mixture of the two IN mutants, but not with either of the IN variants separately. c, Validation of the hybrid intasome design. Left, possible types of strand transfer products obtained by reacting the intasome with circular DNA target (pGEM, black lines). Full-site integration (strand transfer involving both intasomal DNAs, dark blue lines) results in a linear concerted product, which may be targeted by further strand transfer events. Half-site integration (strand transfer involving a single intasomal DNA end) results in a circular branched single-end product. Right, strand transfer assays using mutant intasomes and circular pGEM DNA target. The intasomes were assembled using wild-type IN or a mixture of K120E and D273K mutants, as indicated on top of the gel. IN variants marked with a cross additionally incorporated the E221Q amino acid substitution that disables the enzyme active site. Reaction products were separated by agarose gel electrophoresis. Intasomes were used at indicated concentrations; the leftmost lane contained a mock sample, which received no intasome. Migration positions of the reaction products, intasomal DNA and unreacted supercoiled (s.c.) pGEM are indicated to the right of the gel. As predicted, the strand transfer function of the hybrid intasome strictly requires the active site from the K120E (inner) IN subunit, but not the D273K (outer) subunit. d, Strand transfer activity of mutant intasomes on naked plasmid DNA. Mutations indicated in orange or green were restricted to the outer or inner subunits of the hybrid PFV intasome, respectively.
Extended Data Figure 7 | Infectivity of the mutant PFV vectors.

a, Schematic of the experiments. PFV vectors were produced in 293T cells transfected with DNA constructs encoding PFV GAG, POL and ENV, plus a GFP reporter transfer vector (pMD9). The virus, concentrated by centrifugation, was applied onto target HT1080 cells. Five days post-infection, the cells were analysed by FACS and/or used for isolation of genomic DNA and integration site sequencing. IN mutations were introduced into the packaging construct encoding POL (pcP-POL).

b, Validation of the hybrid intasome design in viral culture conditions. PFV GFP virus was produced using wild-type (WT), K120E, D273K POL packaging construct or a mixture of K120E and D273K mutants. The variants marked with a cross additionally contained a double point mutation inactivating the IN active site (D185N/E221Q). The graph and the western blot show mean relative infectivity and GAG contents (pr71 and p68) of the resultant viruses, respectively. All infectivity experiments were done at least in triplicate, with two or more independent virus preparations; error bars represent standard deviations. The K120E and D273K IN mutants are co-dependent for production of infectious PFV vector, and the functional active site of the K120E IN component is essential for production of infectious hybrid virus.

c, Relative infectivity of the PFV vectors harbouring wild-type, K168E, or active-site-dead D185N/E221Q (indicated with a cross) IN.

d, Relative infectivities of hybrid viruses produced using D273K/D185N/E221Q (indicated as D273K with a cross) and K120E, K120E/D185N/E221Q (cross), K120E/P135E, K120E/T240E, and K120E/P135E/T240E. The western blots below the graphs show GAG (pr71 and p68) contents of the respective PFV vector preps.
Extended Data Figure 8 | Local nucleotide sequence biases at PFV integration sites. Nucleotide sequence preferences at PFV integration sites in cellula (wild type (WT), K168E, hybrid control and P135E/T240E) or in vitro displayed in the form of sequence logos. The heights of the logos correspond to the maximum information content at each position (maximum information content being 2 bits per base). Position 0 corresponds to the target nucleotide joined to the processed U5 PFV end.
Extended Data Figure 9 | Negative-stain EM analysis. a, Representative micrograph. b, Reference-free class averages. c, Three-dimensional electron density map of the intasome–nucleosome complex with a docked intasome structure. Note that DNA density is not recovered with negative-stain EM.
Extended Data Table 1 | Integration site preferences of PFV vector mutants

| Dataset                      | Total sites | Intragenic | +/- 5 kb TSS | +/- 5 kb Lamin | Intergenic (≥40 kb) | Gene density (Mbp) |
|------------------------------|-------------|------------|--------------|---------------|--------------------|-------------------|
|                              |             | Number     | %           | Number        | %                  | Number            | %          | Number | %    | Number | %    | Number | %    | Number | %    |
| In vitro*                   | 2,212,496   | 1,050,686  | 47.5        | 213,822       | 9.7               | 880,481           | 39.8        | 713,115 | 32.2 | 10.3   |
| WT                          | 153,447     | 49,023     | 32.0        | 14,646        | 9.5               | 99,307           | 64.7        | 73,367  | 47.8 | 6.7    |
| K168E                       | 14,872      | 5,244      | 35.3        | 1,410        | 9.5               | 8,987            | 60.4        | 6,572   | 44.2 | 7.4    |
| Hybrid control††            | 41,243      | 12,862     | 31.2        | 3,447        | 8.4               | 28,251           | 68.5        | 20,320  | 49.3 | 6.3    |
| P135E/T240E††               | 10,148      | 3,830      | 37.7        | 801          | 7.9               | 5,701            | 56.2        | 4,361   | 43.0 | 7.6    |

*Number and fraction of integration sites within RefSeq genes.
†Integration sites within 5 kb of a RefSeq transcription start site.
‡Integration sites within 5 kb of Lamin-A/C signal in HT1080 cells.
§Integration sites within extended intergenic regions (>40 kb).
Average number of genes within 500 kb of an integration site.
*Percentage of sites within a feature; values for the mutant and paired control, which show statistically significant differences (P < 0.01) are shown in bold type. P-values calculated for equivalence with wild type or hybrid control using Chi-squared (for intragenic, ±5 kb transcription start site, ±5 kb lamin, and intergenic) or Mann–Whitney–Wilcoxon (for gene density) test are given in brackets. Bonferroni correction for multiple comparisons applied to the P-values.
#Reference data set containing integration sites obtained using purified PFV intasome and deproteinized human genomic DNA.
**Hybrid PFV vector harbouring K120E and D273K/D185N/E221Q IN mutants.
††Hybrid PFV vector harbouring K120E/P135E/T240E and D273K/D185N/E221Q IN mutants.

©2015 Macmillan Publishers Limited. All rights reserved