

ARTICLE

Received 29 Nov 2015 | Accepted 23 Mar 2016 | Published 25 Apr 2016

DOI: 10.1038/ncomms11409

OPEN

# Retroviral intasomes search for a target DNA by 1D diffusion which rarely results in integration

Nathan D. Jones<sup>1,2</sup>, Miguel A. Lopez Jr<sup>1</sup>, Jeungphill Hanne<sup>1</sup>, Mitchell B. Peake<sup>1</sup>, Jong-Bong Lee<sup>3,4</sup>, Richard Fishel<sup>1,2,5</sup> & Kristine E. Yoder<sup>1</sup>

Retroviruses must integrate their linear viral cDNA into the host genome for a productive infection. Integration is catalysed by the retrovirus-encoded integrase (IN), which forms a tetramer or octamer complex with the viral cDNA long terminal repeat (LTR) ends termed an intasome. IN removes two 3'-nucleotides from both LTR ends and catalyses strand transfer of the recessed 3'-hydroxyls into the target DNA separated by 4–6 bp. Host DNA repair restores the resulting 5'-Flap and single-stranded DNA (ssDNA) gap. Here we have used multiple single molecule imaging tools to determine that the prototype foamy virus (PFV) retroviral intasome searches for an integration site by one-dimensional (1D) rotation-coupled diffusion along DNA. Once a target site is identified, the time between PFV strand transfer events is 470 ms. The majority of PFV intasome search events were non-productive. These observations identify new dynamic IN functions and suggest that target site-selection limits retroviral integration.

<sup>1</sup>Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University Medical Center, Columbus, Ohio 43210, USA. <sup>2</sup>The Interdisciplinary Biophysics Graduate Program, The Ohio State University, Columbus, Ohio 43210, USA. <sup>3</sup>Department of Physics, Pohang University of Science and Technology (POSTECH), Pohang 790-784, Korea. <sup>4</sup>Interdisciplinary Bioscience and Bioengineering, POSTECH, Pohang 790-784, Korea. <sup>5</sup>Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA. Correspondence and requests for materials should be addressed to R.F. (email: rfishel@osu.edu) or to K.E.Y. (email: yoder.176@osu.edu).

Retroviral integration is infrequent compared with the number of viral complementary DNA reverse transcripts found in infected cells<sup>1</sup>. The possibility that integration chemistry might account for this disparity was underscored by studies that suggested significant time separated the two HIV IN-catalysed strand transfer events<sup>2</sup>. These integration kinetics contrast the time between the murine leukaemia virus (MLV) strand transfer and the resolution of the 5'-Flap and 4 bp gap by the host DNA repair machinery that appeared to occur near simultaneously<sup>3</sup>. In addition, high-throughput sequencing of integration sites has suggested that most retroviruses display a subtle sequence preference<sup>4,5</sup>, although whether retroviruses seek out these relatively degenerated nucleotide arrangements is unknown. A preference for distorted DNA<sup>6–8</sup> was highlighted by structural reports that revealed the prototype foamy virus (PFV) intasome bound to a bent target DNA<sup>9,10</sup> and a related cryo-EM reconstruction that indicated an unusual DNA curvature when it was bound to a nucleosome<sup>11</sup>. Importantly, these foundational studies have been unable to fully resolve the animated progressions that ultimately result in retroviral integration.

Here we utilize single molecule imaging analysis to examine the dynamic properties of the retroviral intasome with a target DNA. We determine that the two strand transfer events associated with PFV integration are separated by 470 ms, which is nearly four orders-of-magnitude faster than that reported for HIV<sup>2</sup>. In addition, we observe vigorous target search processes on linear DNA in which the PFV intasome performs one-dimensional (1D) rotation-coupled diffusion while in continuous contact with the DNA backbone. The target search does not require catalytically active PFV IN suggesting additional functional domains within the PFV intasome that control the target search process. Remarkably, integration is rare compared with the number of target search events. These observations are consistent with the conclusion that target-site selection limits retroviral integration.

## Results

**Strand transfer catalysed by PFV intasomes.** The retroviral intasome is the fundamental nucleoprotein complex that catalyses integration into a target DNA<sup>12</sup>. To examine the dynamics of retroviral integration, we reconstituted PFV intasomes that contained recombinant PFV IN tetramer and two Cy5-labelled oligonucleotides bearing the U5 sequence of the viral cDNA LTR (Supplementary Fig. 1). As expected, the purified PFV intasomes preferred supercoiled DNA as a target (Fig. 1a)<sup>13</sup> and were capable of catalysing robust integration at ambient temperature (21 ± 1 °C), where the single molecule imaging instruments operate (Fig. 1b,c).

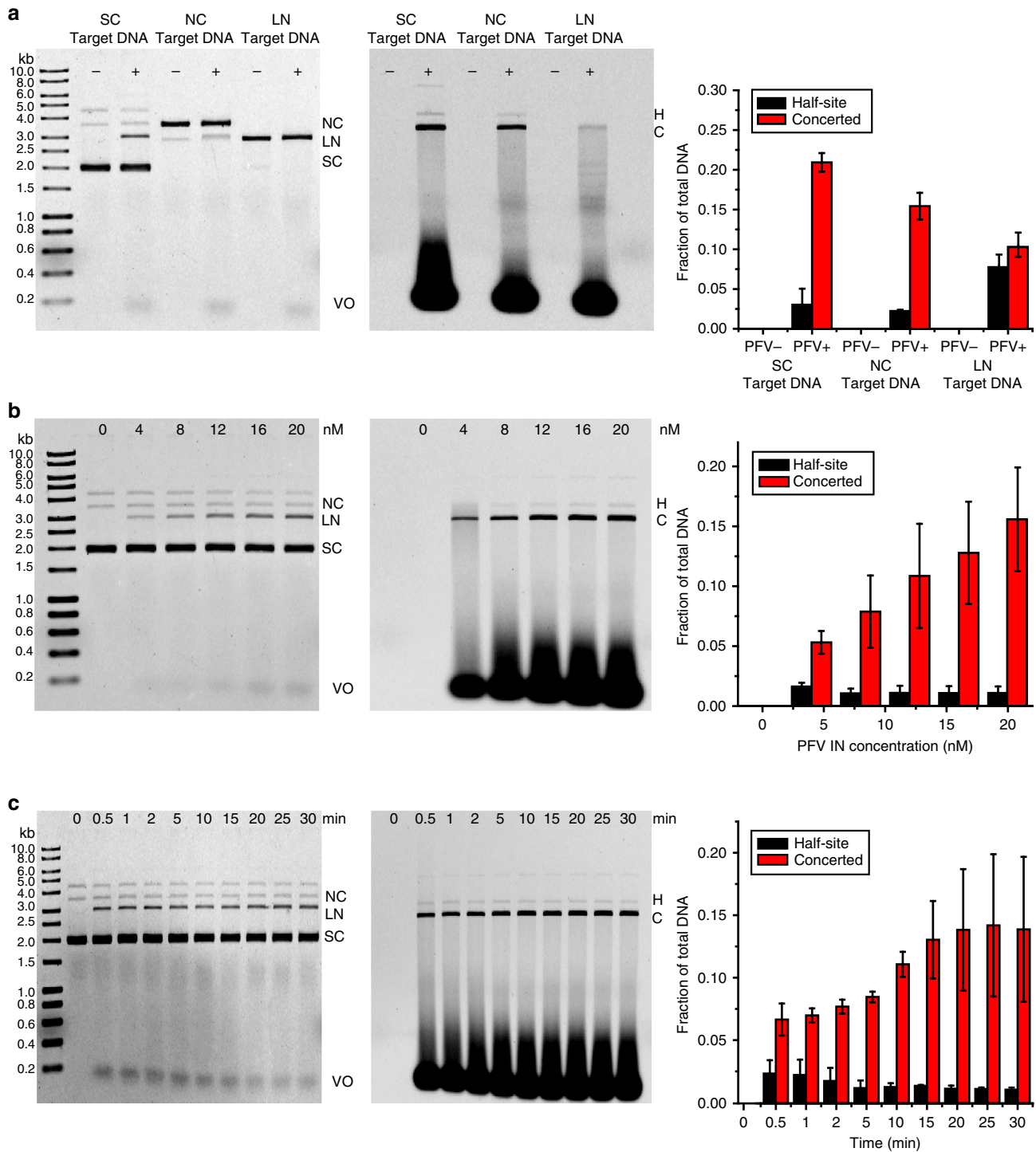
Single molecule magnetic tweezers (smMT) were utilized to determine the kinetic characteristics of two retroviral strand transfer events<sup>14–17</sup>. In this system, one end of a linear target DNA was attached to an engineered flow cell surface via multiple biotin–NeutrAvidin linkages, while the other end of the DNA was attached to a superparamagnetic (SPM) bead via multiple digoxigenin–antidigoxigenin (dig-antidig) linkages (Fig. 2a). These assorted linkages constrain the DNA preventing rotation of the helical strands around their common axis. Controlled positioning of a permanent magnet relative to the SPM bead may be exploited to apply well-defined extension and rotational forces onto this tethered target DNA (Supplementary Fig. 2)<sup>14,17,18</sup>. For example, clockwise rotation of the magnet physically introduces linking number changes ( $-\Delta Lk$ ) in the DNA between the SPM bead and the surface that at low extension forces (0.1–0.5 pN) generates defined negative plectonemic supercoils ( $-\Delta Wr$ ) by the simple relationship

$\Delta Lk = \Delta Tw + \Delta Wr$ , where  $Tw$  remains relatively constant at ~10.4 bp per turn<sup>19,20</sup>. In the smMT system, these negative plectonemic DNA supercoils reduce the apparent height of the SPM beads relative to the surface (Fig. 2a; Supplementary Fig. 2b)<sup>21</sup>.

The first strand transfer event during a concerted PFV intasome integration in the smMT system introduces an ssDNA scission that will relax the plectonemic supercoils within the tethered target DNA, returning the SPM bead to its initial position (Fig. 2a). Since the viral U5 cDNA oligonucleotides are not connected, the second strand transfer event will introduce a double-stranded break (DSB) into the target DNA releasing the SPM bead from the image plane (Fig. 2a). A three-dimensional (3D) tracking algorithm was used to follow the SPM bead position<sup>22</sup> that was combined with in-house developed software to analyse  $x$ ,  $y$  and  $z$  coordinates and associated force measure. The strand transfer time ( $\tau_{ST}$ ) was easily visualized and defined as the first detectable SPM bead movement in the  $z$ -direction (first strand transfer) followed by the first detectable  $x$ - and/or  $y$ -movement (second strand transfer) that ultimately released the SPM bead (Fig. 2b; Supplementary Fig. 3a; see Methods for definitions and data analysis). Recombinant PFV intasomes may also perform a single strand transfer event, termed half-site integration (Fig. 1). Half-site integration uniquely introduces an ssDNA scission that only results in  $z$ -direction movement and may be used to determine a supercoil relaxation time ( $\tau_{RE}$ ; see Methods for definitions and data analysis). However, half-site integration events are unlikely to have physiological significance since only concerted joining of the two viral cDNA ends yields a functional provirus *in vivo*.

Both concerted and half-site integration events were observed at multiple DNA extension forces (Fig. 2c,d; Supplementary Fig. 3b,c). The  $\tau_{RE}$  for PFV half-site integration was fit to a normal distribution (0.26 ± 0.10 s, at 0.3 pN) that was virtually identical to the  $\tau_{RE}$  induced by the Nt.BspQ1 nicking endonuclease (0.25 ± 0.05 s, at 0.3 pN; Supplementary Fig. 4). These results are consistent with the conclusion that PFV intasome interactions with the target DNA do not affect DNA supercoil relaxation kinetics. The distribution of concerted integration events fit a single exponential decay consistent with a unique  $\tau_{ST}$  between the first and second strand transfer events (0.47 ± 0.05 s at 0.3 pN). Because we define the start of the first strand transfer as well as the start of half-site integration as the point of the first detectable  $z$ -direction rate change (see Methods for definitions and data analysis), the contribution of  $\tau_{RE}$  to  $\tau_{ST}$  would appear minimal. However, a scenario where supercoil relaxation is required to detect the first strand transfer event could theoretically increase the time between the first and second strand transfer events ( $\tau_{ST}'$ ) for PFV concerted integration to 0.73 s ( $\tau_{ST}' = \tau_{RE} + \tau_{ST}$ ).

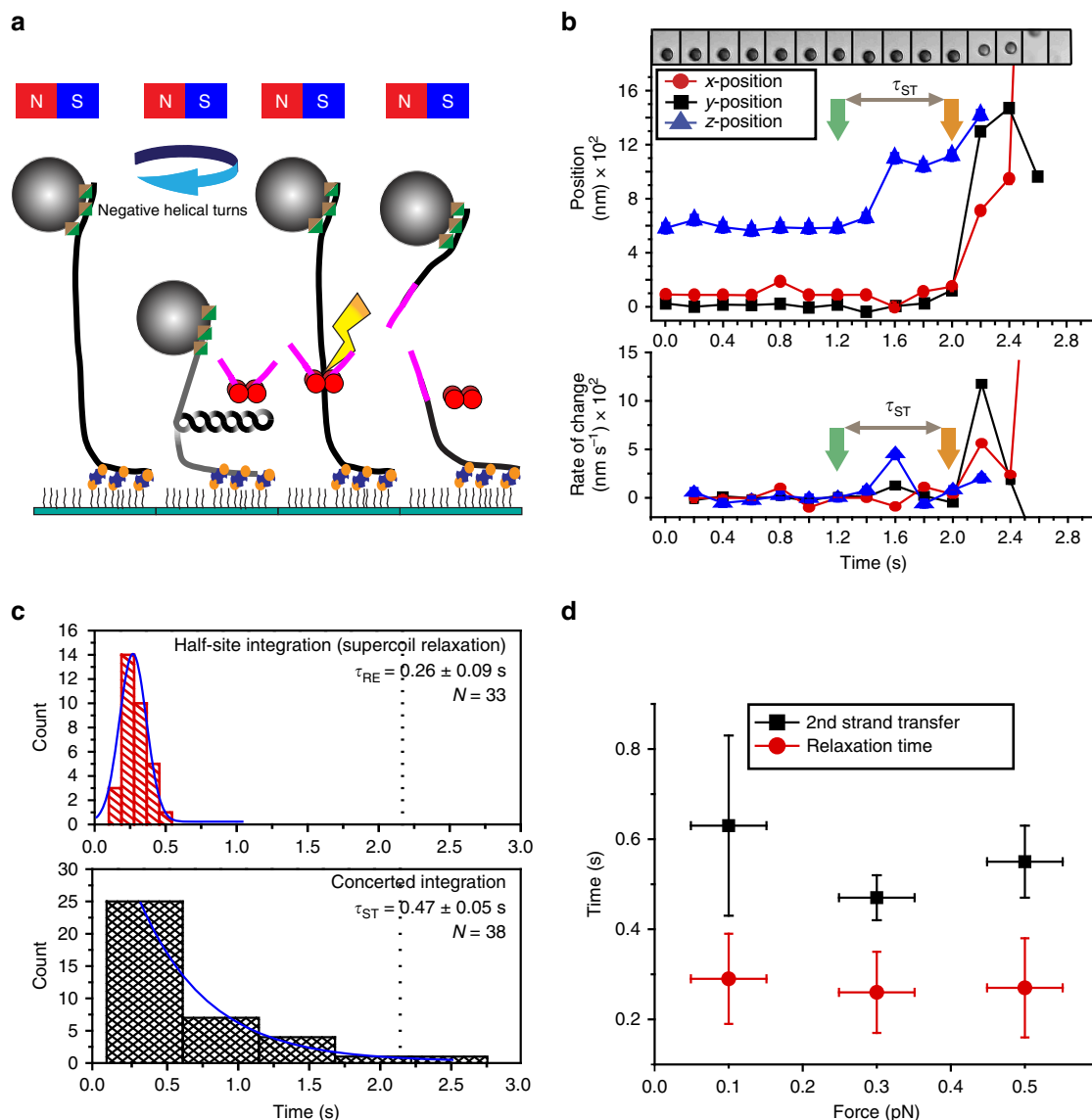
Both  $\tau_{RE}$  and  $\tau_{ST}$  appeared to be independent of DNA extension forces between 0.1 and 0.5 pN (Fig. 2d). There were very few integration events at extension forces above 0.5 pN, where plectonemic supercoils transition to toroidal superhelicity and DNA with altered twist<sup>17,21</sup>. The relative frequency of concerted (52%) and half-site (48%) events appeared similar below 0.5 pN (Supplementary Fig. 5a). Moreover, we found that the distribution of integration times for both concerted and half-site events was identical regardless of whether the reaction was carried out for 30 or 90 min (Supplementary Fig. 5b). These observations imply that PFV intasome-catalysed concerted and half-site integrations are intrinsically unrelated and that long incubation times do not drive half-site events into concerted events. We cannot rule out an effect of the low extension force on PFV integration since the frequency of half-site events appeared ~4-fold greater with the smMT system compared with bulk studies with supercoiled DNA (compare Fig. 1 with Supplementary Fig. 5a). Nevertheless, the



**Figure 1 | Biochemical analysis of wild type PFV intasome integration.** Bulk integration studies into the pMP2 target DNA were performed at ambient temperature ( $21 \pm 1^\circ\text{C}$ ). (Left panels) Ethidium bromide stained gel; (middle panels) Cy5 viral donor emission; (right panels) Quantitative analysis of band intensities. H, half-site integration product; C, concerted integration product. From supercoiled (SC), nicked circular (NC) and linear (LN) target DNAs. For the LN DNA target the half-site products co-migrate with linear target (marked by C), while concerted products migrate as a smear between LN target and viral oligonucleotide DNA (VO). **(a)** Integration of purified Cy5-labelled wild type PFV IN intasome (20 nM) with various target DNA (50 ng) conformations. **(b)** Titration of wild type PFV IN intasome (nM at top) into supercoiled target DNA (50 ng). Note that the fraction of concerted integration product increases with increasing concentration of PFV intasome, while the fraction half-site remained relatively unchanged with increasing PFV intasome. **(c)** Time course (min at top) of wild type PFV IN intasome (20 nM) integration into supercoiled target DNA (50 ng). Error bars indicate s.d. from two independent experiments.

sub-second time between the two PFV intasome-catalysed strand transfer events strongly suggests that integration catalysis is unlikely to limit concerted retroviral integration.

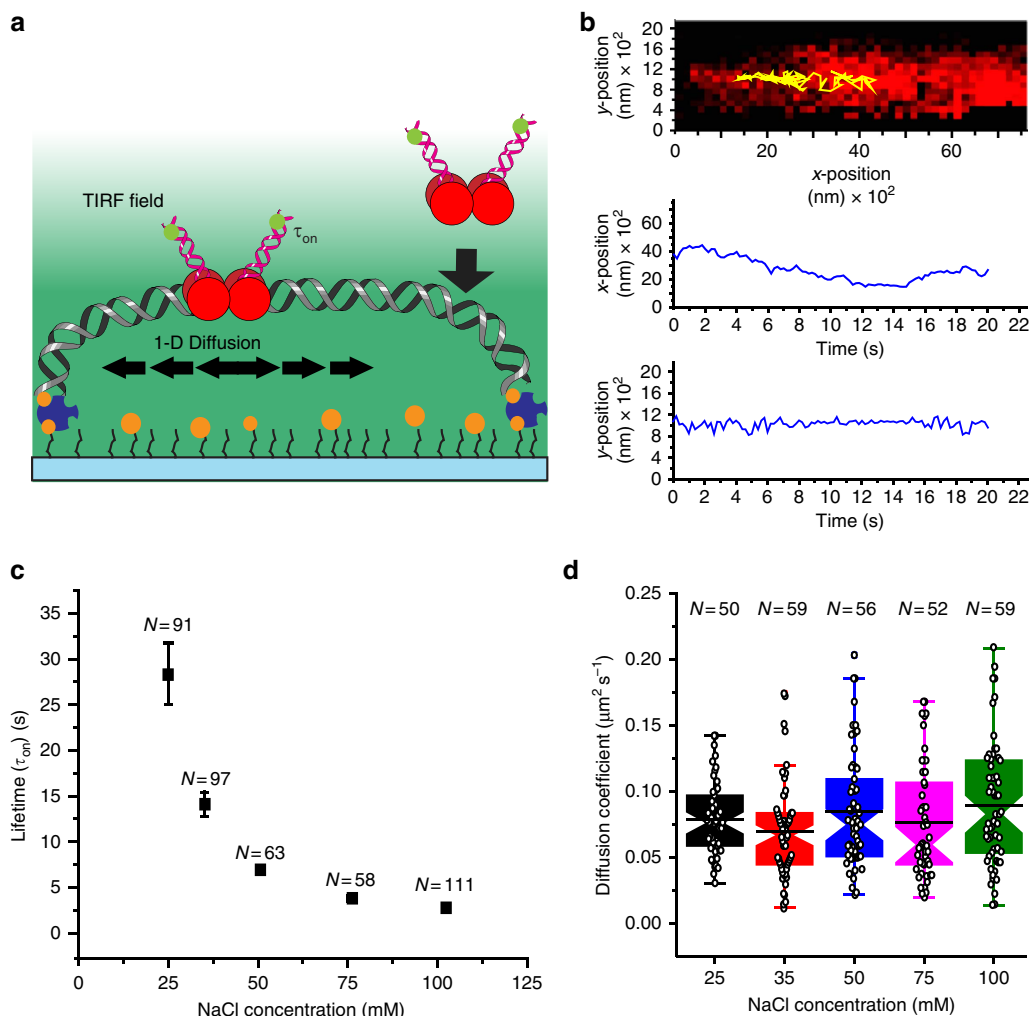
**PFV intasomes search the target DNA by 1D diffusion.** The target search before retroviral integration may occur by 3D collision alone, or 3D collision followed by 1D diffusion along the



**Figure 2 | Magnetic tweezers resolves the kinetics of PFV intasome strand transfer.** (a) Experimental analysis of PFV intasome strand transfer kinetics. DNA bound to an SPM bead and the surface of the flow cell (left) is supercoiled (left centre) by clockwise rotation of ferromagnets under low extension force. PFV intasome interaction with the supercoiled target DNA results in integration where the first strand transfer event creates a strand scission that releases the supercoils (right centre). The second strand transfer event creates a double strand break (DSB) that results in SPM bead dissociation from the target DNA tethered to the surface (right). (b) (top panel) A representative SPM bead image during a concerted integration event. (middle panel) A 3D (x,y,z) trace of the SPM bead position. (bottom panel) The rate of SPM bead position change during a concerted integration event. The first strand transfer (green arrow; at 1.2 s) alters the SPM bead height in the z-direction as a result of DNA relaxation. The second event (orange arrow; at 2.0 s) alters both the x- and y-direction, while the z-direction could not be clearly resolved because the SPM bead leaves the field of view (see Supplementary Fig. 3a for additional examples). The definition of strand transfer time ( $\tau_{ST}$ ) is shown. (c) The kinetics of DNA relaxation ( $\tau_{RE}$ ) as a result of one strand transfer event during half-site integration (top panel) and the kinetics of two strand transfer ( $\tau_{ST}$ ) events as a result of concerted integration. See Methods for the analytical definitions of  $\tau_{RE}$  and  $\tau_{ST}$ . The integration analysis shown was performed at a fixed applied force ( $F_D = 0.3$  pN) on the supercoiled DNA (-10 turns). (d) The distribution of relaxation times ( $\tau_{RE}$ ) and the times between the two strand transfer events ( $\tau_{ST}$ ) at various force values (Supplementary Fig. 3b). X-axis error bars indicate s.e. of force measurements. Y-axis error bars indicate s.e. and s.d. of  $\tau_{RE}$  and  $\tau_{ST}$ , respectively. The frame rate for smMT was 100–200 ms.

DNA<sup>23</sup>. 1D diffusion is a well-recognized process that significantly enhances the search for sequences, lesions and structures on DNA<sup>24</sup>. We utilized single molecule total internal reflection fluorescence (smTIRF) microscopy to visualize individual PFV intasomes associated with a target DNA<sup>25–27</sup>. A truncated 24 kb  $\lambda$ -based target DNA containing bilateral 5'-biotins was attached at one end to a passivated flow cell surface via PEG-biotin-NeutrAvidin and the DNA gently stretched by defined hydrodynamic flow force to ultimately link the other end via PEG-biotin-NeutrAvidin (Fig. 3a, see Methods)<sup>25–27</sup>.

PFV intasomes were reconstituted with Cy3-labelled U5 viral cDNA (Cy3-PFV; Supplementary Table 1). The substitution of Cy3- for Cy5-labelled U5 viral cDNA enhanced photostability and lifetime in the smTIRF system. These Cy3-PFV intasomes were then introduced into the flow cell, the flow was stopped and the DNA continuously imaged for up to 30 min. We observed single and multiple PFV intasomes that travelled along the target DNA by 1D diffusion (Fig. 3b; Supplementary Movie 1). Individual PFV intasome particles were tracked and the resulting histogram fit to a single exponential decay to determine the lifetime



**Figure 3 | PFV intasomes search a target DNA by 1D diffusion.** (a) Experimental system for PFV intasome interaction with a target DNA using smTIRF imaging. The total internal reflection fluorescence (TIRF) field extends 200–250 nm into the flow cell (teal background). A target DNA is stretched by hydrodynamic flow and bound at both ends by the biotin–NeutrAvidin over the surface. The PFV intasomes were reconstituted with Cy3-labelled viral U5 cDNA. When the Cy3-labelled PFV intasome interacts with the target DNA inside the TIRF field, the fluorophore is excited allowing visual tracking of particles. (b) Representative trace of PFV intasome diffusion (yellow) on a target DNA (red) at a 300-ms frame rate. (c) Distribution of lifetimes ( $\pm$  s.e.) for wild type PFV intasomes at different NaCl salt concentrations (Supplementary Fig. 6). (d) Box plots of the diffusion coefficients distribution for wild type PFV intasomes at different NaCl salt concentrations at a 100-ms frame rate (Supplementary Fig. 6). Box plots show the mean (indentation), median (black line), upper and lower quartile (box ends), and outliers (whiskers).

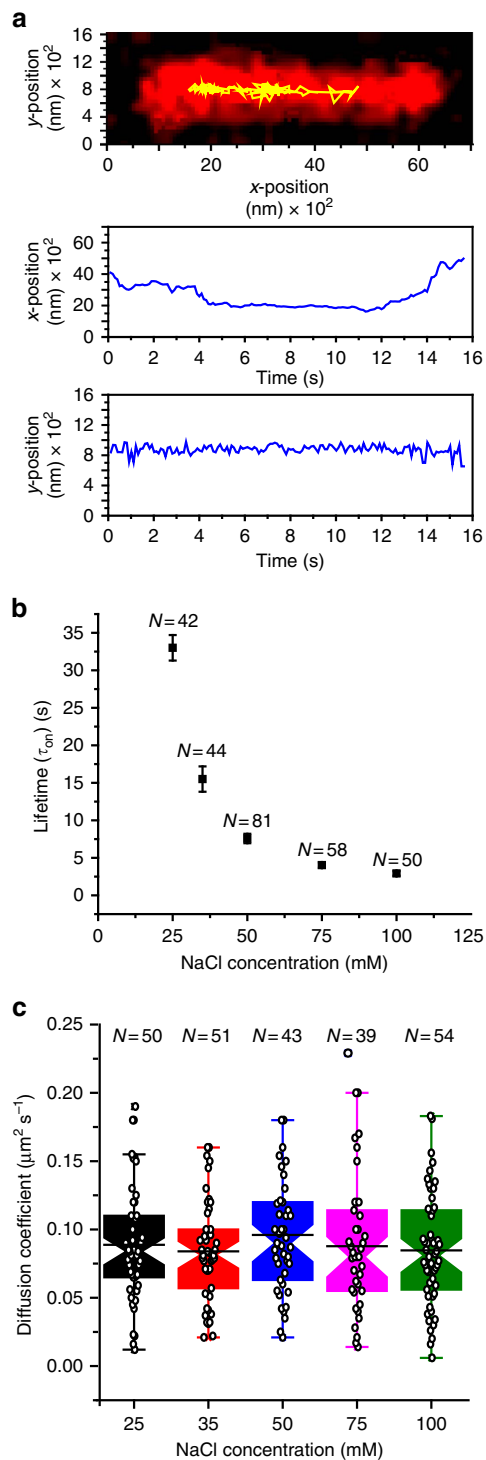
( $\tau_{on} = 2.1 \pm 0.1$  s at 100 mM NaCl,  $N = 111$ ; Supplementary Fig. 6). While too short to accurately determine above 100 mM, the lifetime showed an inverse correlation with NaCl concentration (Fig. 3c; Supplementary Fig. 6). These observations are consistent with the significant ionic interaction(s) between the PFV intasome and the target DNA.

The  $x$ - and  $y$ -positions of Cy3-PFV were followed on the DNA to determine the mean squared displacement (MSD) and a diffusion coefficient ( $D = 0.082 \pm 0.005 \mu\text{m}^2 \text{s}^{-1}$  at 100 mM NaCl,  $N = 59$ ; Supplementary Fig. 6; see Methods). We found that the MSD in the  $y$ -direction ( $\langle \Delta y^2 \rangle$ ) resulting from tethered DNA Brownian motions contributed  $0.7 \pm 0.2\%$  (s.d.; with a maximum of 2%) to the diffusion coefficient determined from the MSD in the  $x$ -direction alone ( $\langle \Delta x^2 \rangle$ ; Supplementary Fig. 7; see Methods). Together with the lifetime, these observations suggest that the PFV intasome may scan  $\sim 1,600$  bp of target DNA before disassociation.

Two fundamentally different 1D search mechanisms have been described, sliding and hopping, that engender different ionic

interactions with the DNA<sup>23,28</sup>. Diffusion by a sliding mechanism conserves the protein–DNA ionic interactions and is expected to be constant and independent of ionic conditions<sup>23</sup>. In contrast, diffusion by a hopping mechanism must alter protein–DNA ionic interactions with each step and is expected to increase with ionic strength<sup>23</sup>. We found that diffusion of the PFV intasome was constant and independent of ionic strength ( $D_{avg} = 0.079 \pm 0.006 \mu\text{m}^2 \text{s}^{-1}$ ; Fig. 3d; Supplementary Fig. 6). While our studies cannot rule out secondary DNA interactions such as inter-segmental transfer<sup>29</sup>, these results are consistent with the conclusion that the PFV intasome diffuses by a sliding mechanism along the DNA.

Theoretical modelling of PFV intasome diffusion based on comparable observations with other DNA binding proteins<sup>30,31</sup> and using the crystal structure to determine intasome dimensions<sup>9</sup>, we calculated a simple sliding diffusion coefficient of  $\sim 45 \mu\text{m}^2 \text{s}^{-1}$  (Supplementary Note 1)<sup>32</sup>. This is at least two orders-of-magnitude faster than our measured diffusion coefficient. We estimated an average free-energy barrier ( $\epsilon$ ) of



**Figure 4 | The PFV IN(D128N) intasome search on target DNA.** (a) Representative trace of PFV IN(D128N) intasome diffusion (yellow) on a target DNA (red) at a 200-ms frame rate. (b) Distribution of lifetimes ( $\pm$  s.e.) for PFV IN(D128N) intasomes at different NaCl concentrations at a 200-ms frame rate (Supplementary Fig. 8). (c) Box plots of the diffusion coefficients for PFV IN(D128N) intasomes at different NaCl concentrations at a 100-ms frame rate (Supplementary Fig. 8). Box plots show the mean (indentation), median (black line), upper and lower quartile (box ends), and outliers (whiskers).

$1.1 \pm 0.2 k_B T$  for rotational sliding of the PFV intasome along DNA (Supplementary Note 1)<sup>30,31</sup>. This is significantly less than the  $2 k_B T$  energy barrier for biologically relevant sliding with

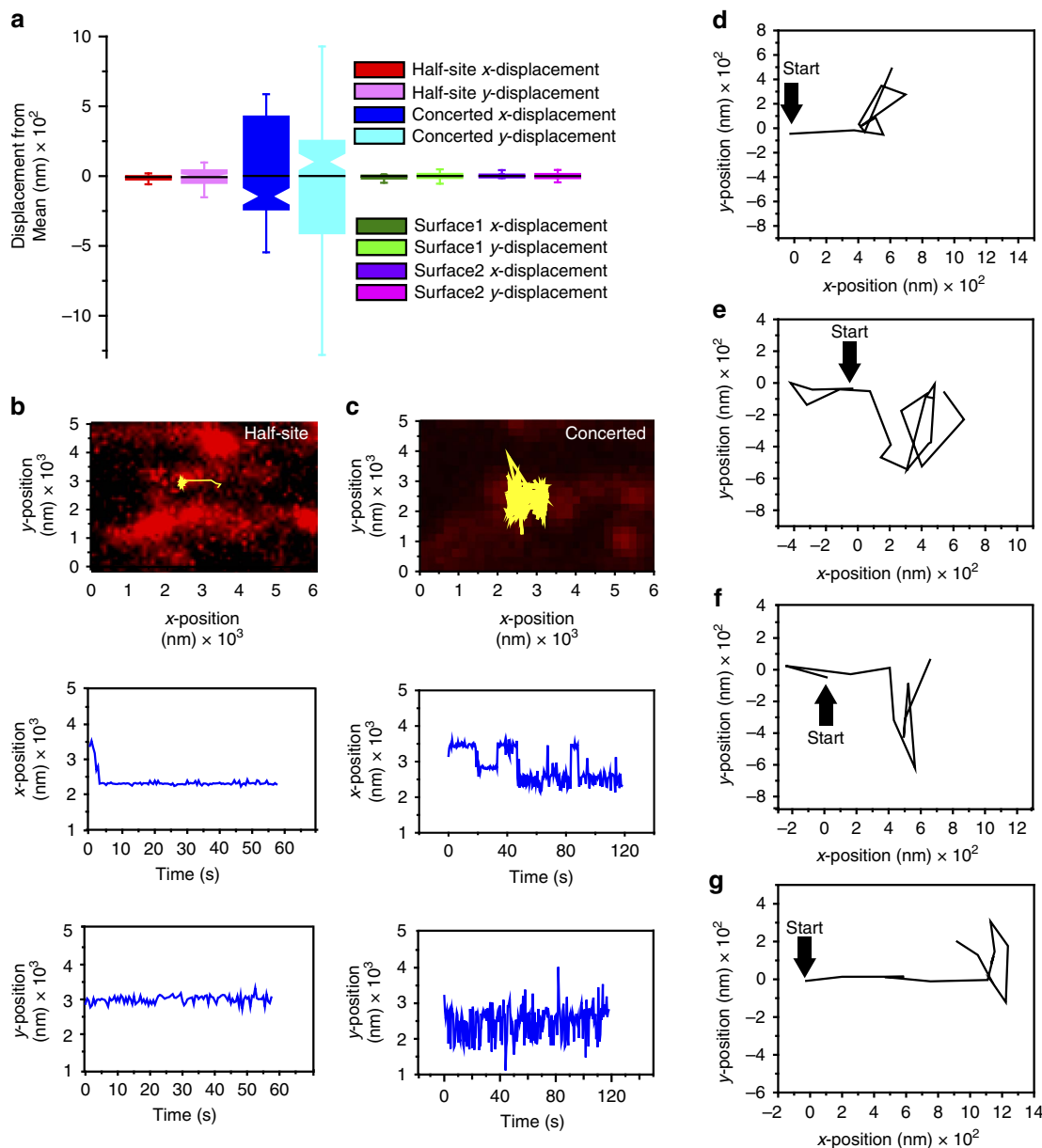
rotation on DNA<sup>33</sup>. Taken as a whole, this evaluation is consistent with the hypothesis that PFV intasomes search a target DNA by 1D rotation-coupled translational diffusion along the DNA.

We prepared catalytically inactive intasomes using PFV IN(D128N) and Cy3-labelled viral U5 cDNA. A similar smTIRF analysis found that the lifetime ( $2.9 \pm 0.3$  s), diffusion coefficient ( $D_{\text{avg}} = 0.082 \pm 0.031 \mu\text{m}^2 \text{s}^{-1}$ ) and diffusion properties were similar to the wild-type PFV intasomes (Fig. 4; Supplementary Fig. 8). These results are consistent with the conclusion that altering the PFV intasome catalytic site does not affect the target DNA search process. Taken together, these observations suggest that residues and/or domains outside the catalytically active core are required to accomplish the 1D diffusion target search process.

**Integration is rare compared with the target search.** Remarkably, only four potential concerted and three potential half-site integration events were observed in 812 recorded target DNA search events (Fig. 5). Integration events could be distinguished from surface-bound fluorophores by their unusual  $x$ - and  $y$ -direction motions (Fig. 5a). For example, probable half-site integration events clearly overlaid a Sytox-stained DNA and were associated with a Cy3-PFV intasome that initially underwent  $x$ -direction diffusion, which was followed by localized  $x$ - and  $y$ -direction movement attributable to Brownian motion of the Cy3-labelled viral U5 cDNA attached to a target DNA tethered to the surface at both ends (Fig. 5a,b; Supplementary Movie 2). In contrast, none of the potential concerted integration events were correlated with significant Sytox DNA staining since the broken target DNA is expected to coil near the tether site<sup>34</sup> (Fig. 5c). Instead, a short  $x$ -direction diffusion of the Cy3-PFV intasome was followed by a relatively dramatic  $x$ -direction movement that we interpret reflects the broken DNA contracting to one tether site, with the attached Cy3-labelled viral U5 DNA ultimately displaying  $x$ - and  $y$ -direction Brownian motions consistent with random coiled DNA tethered at one end (Fig. 5d–g; Supplementary Movie 3; note that the tracked particle in Fig. 5c,d and Supplementary Movie 3 are the same event). There may be two contributors to the reduced frequency of integration events: (1) the necessary use of a linear DNA that is intrinsically a less-efficient target substrate (Fig. 1) (ref. 2) and/or (2) the resulting broken integration product consisting of coiled target DNA might extended out of the TIRF field making them invisible in the smTIRF system. To further consider this latter possibility, we calculate that 2–4 kb of coiled DNA from either tethered end (20–30% of the total target DNA) may remain continuously in the 200–250 nm TIRF field generated by our smTIRF system<sup>35</sup>. This implies that we may be missing 70–80% of the concerted integration events, which could increase the total to between 15 and 20 out of the 812 recorded target DNA search events. However, even including obscured integrations it is clear that  $>97\%$  of the PFV intasome search events appear to be nonproductive. It is important to note that even though our recorded events likely reflect half-site and concerted integrations, they are not statistically significant to draw mechanistic conclusions.

## Discussion

To our knowledge, these studies represent the first single molecule imaging analysis of retroviral integration. The results suggest that the two PFV IN-catalysed strand transfer events are discernible but also nearly simultaneous, an observation that significantly contrasts previous bulk HIV-1 integration analysis<sup>2</sup>. In addition, we found that concerted and half-site PFV integration events are intrinsically unrelated, while previous



**Figure 5 | Imaging PFV intasome integration events into linear target DNA using smTIRF.** (a) Box plots of the frame-by-frame x- and y-displacement from the mean for (left-to-right): half-site integration example (**b**); the concerted integration example (**c**); and two random particles that were stuck to the flow cell surface (surface 1 and surface 2). Box plots show the mean (indentation), median (black line), upper and lower quartile (box ends), and outliers (whiskers). **(b)** (top) Representative trace of a wild-type PFV intasome (yellow) half-site integration event on DNA (red) at a 500-ms frame rate. (middle) x-position and (bottom) y-position traces over time. Note reduced x-position and increased y-position indicative of a half-site integration event where the Cy3-labelled viral U5 cDNA donor is linked to a target DNA molecule attached at both ends and undergoing Brownian movement (see **a**). **(c)** (top) Representative trace of a wild type PFV intasome (yellow) concerted integration event on DNA (red) at a 500-ms frame rate. (middle) x-position and (bottom) y-position traces over time. Note large x- and y-position Brownian movement indicative of a broken DNA attached at one end following a concerted integration event (see **a**). **(d-g)** Initial frame-by-frame traces of the four observed concerted integration events. **(d)** The initial tracking frames for the integration event shown in **c**. All tracking traces (500 ms frame rate) show an initial sliding diffusion for 1-2 s that was followed by a large movement in the x-direction (97-236 nm); ultimately resulting in a particle that randomly moved in both the x- and y-direction consistent with a broken DNA containing a Cy3 label.

HIV studies suggested that half-site events progress to concerted events<sup>2</sup>. These results appear to reflect previously unrecognized differences between the PFV and HIV integration processes.

There are ~100 predicted PFV integrase preferred-sequences in the  $\lambda$  target DNA used in the smTIRF analysis<sup>13,36</sup>. Based on the numbers of observable target molecules in the smTIRF system, we calculate that PFV intasomes should have encountered a preferred target sequence ~100,000 times during our

observation period. The exceedingly low number of integration events suggests that a preferred sequence search on a target DNA by retroviral intasomes is at the very least inefficient. Another possibility is that PFV intasomes search for DNA conformations or structures<sup>6-8</sup>. Consistent with this hypothesis we find that supercoiled DNA is a preferred target<sup>21</sup>. However, it should be noted that there is only a 2-5-fold increase in the PFV intasome integration frequency into supercoiled DNA compared with

linear DNA (Fig. 1). These observations suggest that even for a preferred target only 10–15% of search events will be productive. Taken as a whole, we conclude that target site selection limits PFV integration. Finally, the retroviral search mechanics on chromatin is likely to be significantly different since rotation-coupled 1D diffusion is physically limited and host factors may tether the intasome to a nucleosome target<sup>37,38</sup>.

## Methods

**Chemicals and enzymes.** All chemicals were ultrapure and purchased from Sigma Aldrich (St Louis, MO) or Amresco (Cleveland, OH). Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs (NEB, Beverly, MA) or ThermoFisher Scientific (Waltham, MA). Recombinant PFV IN and catalytically inactive PFV IN(D128N) containing N-terminal His<sub>6</sub>-tag (HHHHHH), thrombin protease site (LVPRGS), HRV3C protease site (LEVLFQGP) and spacers (MGSS-HHHHHH-SSG-LVPRGS-HS-LEVLFQGP-G-PFV IN coding sequence minus PFV N-terminal Met) was purified by a modified method<sup>12</sup>. Briefly, the IN protein was overexpressed in *Escherichia coli* Rosetta (DE3) that were lysed by sonication and soluble material separated from cellular debris by ultracentrifugation. The cell lysate in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl was subjected to Ni-NTA chromatography, IN protein fractions pooled and the majority of the N-terminal tag removed with in-house purified HRV3C protease (leaving GP-G-PFV IN coding sequence). The protease-treated protein was diluted fourfold with 50 mM Tris-HCl (pH 7.5), 5 mM DTT, 0.1 mM EDTA (Buffer A) loaded onto a Heparin column, washed with Buffer A plus 200 mM NaCl and stepped off with Buffer A plus 1 M NaCl. Highly concentrated protein samples were dialysed into storage buffer containing 50 mM Tris pH 7.5, 500 mM NaCl, 5 mM DTT and 10% glycerol, snap frozen in LN<sub>2</sub> and were stable at  $-80^{\circ}\text{C}$  for several months.

Intasomes were assembled by combining highly concentrated PFV IN (120  $\mu\text{M}$  final) with annealed oligomers mimicking pre-processed PFV U5 viral cDNA ends (50  $\mu\text{M}$  final) in 50 mM Bis-Tris Propane-HCl 7.45, 500 mM NaCl and then dialysed overnight at  $18^{\circ}\text{C}$  against 20 mM Bis-Tris Propane-HCl 7.45, 200 mM NaCl, 2 mM DTT, 25  $\mu\text{M}$  ZnCl<sub>2</sub>. Assembled intasome complexes were separated from aggregates and non-intasome material by Superose 12 10/30 GL size exclusion chromatography in 20 mM Bis-Tris Propane-HCl (pH 7.45), 320 mM NaCl, 50  $\mu\text{M}$  ZnCl<sub>2</sub>, 10 mM DTT and 10% glycerol (Supplementary Fig. 1). Fractions were snap frozen and stable at  $-80^{\circ}\text{C}$  for several months.

**Bulk PFV IN strand transfer analysis.** PFV IN intasome catalysed strand transfer into supercoiled, nicked or linear pMP2 target DNA (50 ng) was examined over time at ambient temperature ( $21 \pm 1^{\circ}\text{C}$ ). The 15- $\mu\text{l}$  reactions included PFV intasome (20 nM) in Buffer A (10 mM Bis-Tris Propane-HCl (pH 7.45), 110 mM NaCl, 5 mM MgSO<sub>4</sub>, 4  $\mu\text{M}$  ZnCl<sub>2</sub> and 10 mM DTT). Following incubation, the reactions were stopped with the addition of 0.45% SDS and 900  $\mu\text{g ml}^{-1}$  proteinase K. Products were separated on a 1% agarose gel, stained with ethidium bromide (0.1  $\mu\text{g ml}^{-1}$ ), scanned using a Typhoon 9410 variable mode fluorescent imager (GE Healthcare Life Sciences) and analysed with ImageQuant TL software.

**Magnetic tweezers.** Flow cells were engineered with glass cover slides that were affixed with double-sided tape to an aluminium foundation that maximized SPM bead imaging. Prior to attachment, the glass slides were treated with (3-aminopropyl) triethoxysilane followed by a 1:100 mixture of biotin-PEG SVA to mPEG-SVA (Invitrogen). Plasmid pET-29a was digested with EcoRI and SphI yielding a 4,967 bp linear fragment. Biotin and digoxigenin labelled 1 kb linkers were generated by PCR and included a 1:10 ratio of biotin- or dig-labelled dUTP to dTTP in the reaction. The 1-kb PCR products were digested with EcoRI and SphI, ligated to the pET29a fragment with T4 DNA ligase and the linear 7 kb DNA ligation fragment was excised from a low-melt agarose gel stained with ethidium bromide and purified using a QIAquick Gel Extraction kit (Qiagen), precipitated with isopropanol and washed with ethanol to fully remove any remnant ethidium bromide prior to aqueous resuspension. NeutrAvidin (500  $\mu\text{M}$ , Invitrogen) was injected in the flow cell at a rate of 8  $\mu\text{l min}^{-1}$ , followed by the 7-kb ligation fragment in the presence of T4 DNA ligase to seal any remnant strand scissions. Tosylactivated M-280 SPM Dynabeads (ThermoFisher Scientific) were coated with anti-digoxigenin antibodies and injected into the flow cell containing the 7-kb DNA ligation fragment that was bound to the surface via a biotin-NeutrAvidin linkage; at 8  $\mu\text{l min}^{-1}$  while agitating the system. The bound DNA was washed extensively with Buffer A to remove T4 DNA ligase and free SPM beads prior to PFV integration analysis. Force extension and introduction of supercoils used two 1 cm<sup>3</sup> rare earth magnets (Neodymium, Magcraft). For PFV integration analysis, the DNA was wound clockwise with 10 complete turns of the magnets to induce negative supercoils. The SPM beads were imaged using a 530-nm LED lamp (Thorlabs), a  $\times 40$  Olympus oil immersion objective and images collected on a 1,024  $\times$  1,024 pixel charge coupled device camera (Grasshopper Express 1.0 MP Mono FireWire 1394b) at a frame rate of 100–200 ms for at least 1,800 s.

SPM bead analysis was performed with the 3D bead tracking software Video Spot Tracker (CISMM at UNC-CH). The time establishing SPM bead movement in any one of the three-dimensions ( $x,y,z$ ) was determined as the point of departure

from zero of the first-derivative of the position versus time curve. The start of a concerted PFV intasome strand transfer event that also includes initial aspects of the supercoil DNA relaxation time was defined as the point of departure from zero of the first-derivative of  $z$ -position changes versus time curve, while the second strand transfer event was defined as point of departure from zero of the first-derivative of subsequent  $x$ - or  $y$ -position changes versus time curve. The strand transfer time ( $\tau_{\text{ST}}$ ) is defined as the time between the first strand transfer and the second strand transfer events and were fit to a single exponential decay to determine the mean lifetime with the standard error (s.e.) of curve fitting. Half-site integration events uniquely alter the  $z$ -position allowing a clear determination of the supercoil relaxation time ( $\tau_{\text{RE}}$ ), which was defined as the time between the point of departure from zero of the first-derivative of  $z$ -position change versus time curve to the point that the first-derivative departs from zero again. Distributions for  $\tau_{\text{RE}}$  and  $\tau_{\text{ST}}$  events were binned using: [number of bins =  $(\sqrt{N} - 1) \rightarrow$  round up]; and [Size of bins =  $\frac{(\text{Max} - \text{Min})}{(\text{Number of bins} - 0.5)}$ ], where: Max = maximum value of events; Min = minimum value of events; Bin Start = Min; Bin End = (bin size  $\times$  number of bins)<sup>39,40</sup>. We used 100 ms as the bin minimum for  $\tau_{\text{RE}}$  because the data was collected with both 100 and 200 ms frame rates. The  $\tau_{\text{RE}}$  events were fit to a Gaussian distribution to determine the mean and s.d. from the mean. All fittings were done with Origin software (OriginLabs, Northampton, MA).

**Single molecule tracking on a target DNA.** A  $\lambda$ -based target DNA was prepared by digestion with either AvrII (New England Biolabs) or isoschizomer XmaI (ThermoFisher Scientific) yielding 24.3 and 24.1 kb fragments with cosL-AvrII and AvrII-cosR ends, respectively. Annealed AvrII and cosL linker DNAs were first added to the  $\lambda$  fragments with T4 DNA ligase (Supplementary Table 1). Excess linker DNAs were removed by Amicon Ultra-4 centrifugal filters (100 kDa cutoff; Amicon) followed by the addition of the annealed cosR linker DNA in a new T4 DNA ligase reaction. The resulting 24 kb DNA products were excised from a low-melt agarose gel stained with ethidium bromide, treated with  $\beta$ -agarase (NEB), precipitated with isopropanol and washed with ethanol to fully remove any remnant ethidium bromide prior to aqueous resuspension. Engineered flow cells manufactured with quartz slides were treated with (3-aminopropyl) triethoxysilane followed by a 1:300 mixture of biotin-PEG-SVA:mPEG-SVA (Invitrogen)<sup>41</sup>. Neutravidin (500  $\mu\text{M}$ , Invitrogen) was injected in the flow cell prior to the addition of DNA. The  $\lambda$ -based target DNA was introduced into the flow cell at 100  $\mu\text{l min}^{-1}$ , where stochastic attachment of the 5'-biotin ends to the NeutrAvidin-coated surface under these laminar flow conditions results in fully stretched DNAs linked to the surface at both ends<sup>41</sup>. Purified PFV intasome (0.5 nM) containing Cy3-labelled viral U5 cDNA oligonucleotides was introduced at 10  $\mu\text{l min}^{-1}$ , the flow stopped and the Cy3-PFV particles associated with the target DNA were excited with a 532-nm DPSS laser (crystal laser, 50 mW). The Cy3-PFV emission was detected in a prism-type TIRF microscope system (Olympus IX-71, water-type  $\times 60$  objective, N/A = 1.2; extended with a  $\times 1.6$  magnifier) and recorded on an electron-multiplying charge-coupled device camera (EMCCD, Princeton Instruments, ProEM 512 excelon)<sup>41,42</sup>.

Lifetime and diffusion tracking measures with the Cy3-labelled wild type PFV intasomes were followed for 1,650 s at 300 ms and 230 s at 100 ms frame rate, respectively. Lifetime and diffusion tracking measures with the Cy3-labelled PFV IN(D128N) intasomes were followed for 230 s at 200 and 100 ms frame rate, respectively. To detect integration events wild-type PFV intasome were followed for 3,600 s at 500 ms frame rate. Following tracking measurements, Sytox orange (100 nM) was introduced into the flow cell to visualize the  $\lambda$  DNA. These static images were overlaid onto the tracking movies to identify PFV intasomes that were clearly associated with the target DNA. Particles were tracked using DiaTrack software (Sydney, Australia). Lifetime measures were binned (see above), fit to a single exponential decay and the time  $\pm$  s.e. resolved. The  $x$ - and  $y$ -direction tracking information was obtained (DiaTrack) and plotted with Origin software. MSD was determined for single particles at every imaged time point as the square of the diffusion distance in the  $x$ -direction compared with the preceding time point ( $<\Delta x^2>$ ) (ref. 43). The MSD versus time was displayed for all particle tracks and a linear fit was applied to the first 10% of the data with five frames being the minimum fit requirements to obtain  $<\Delta x^2>/t$  (ref. 43). The diffusion coefficient ( $D$ ) was calculated using the equation:  $<\Delta x^2> = n \times D \times t$ , where  $n$  is the dimensionality constant equal to 2 for linear diffusion<sup>43</sup>. Diffusion coefficients were binned (see above) and fit to a Gaussian distribution with the mean and s.d. from the mean displayed. Combining  $x$ - and  $y$ -displacement into the MSD calculation ( $<\Delta x^2> + <\Delta y^2> = <\Delta r^2>$ ) and using  $<\Delta r^2>$  to calculate a diffusion coefficient resulted in an average a  $0.7 \pm 0.2\%$  (s.d.; with a maximum of 2%) difference between the diffusion coefficient calculated using  $<\Delta x^2>$  alone for MSD. Diffusion coefficients were compiled into box and whisker plots. All fittings and box plots were done with Origin software (OriginLabs).

## References

- Butler, S. L., Johnson, E. P. & Bushman, F. D. Human immunodeficiency virus cDNA metabolism: notable stability of two-long terminal repeat circles. *J. Virol.* **76**, 3739–3747 (2002).
- Li, M., Mizuchi, M., Burke, Jr T. R. & Craigie, R. Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J.* **25**, 1295–1304 (2006).



3. Roe, T., Chow, S. A. & Brown, P. O. 3'-end processing and kinetics of 5'-end joining during retroviral integration *in vivo*. *J. Virol.* **71**, 1334–1340 (1997).
4. Holman, A. G. & Coffin, J. M. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc. Natl Acad. Sci. USA* **102**, 6103–6107 (2005).
5. Wu, X., Li, Y., Crise, B., Burgess, S. M. & Munroe, D. J. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.* **79**, 5211–5214 (2005).
6. Muller, H. P. & Varmus, H. E. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.* **13**, 4704–4714 (1994).
7. 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).
8. Pruss, D., Reeves, R., Bushman, F. D. & Wolffe, A. P. The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J. Biol. Chem.* **269**, 25031–25041 (1994).
9. Gupta, K. *et al.* Solution conformations of prototype foamy virus integrase and its stable synaptic complex with U5 viral DNA. *Structure* **20**, 1918–1928 (2012).
10. Maertens, G. N., Hare, S. & Cherepanov, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329 (2010).
11. Maskell, D. P. *et al.* Structural basis for retroviral integration into nucleosomes. *Nature* **523**, 366–369 (2015).
12. Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cherepanov, P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236 (2010).
13. Valkov, E. *et al.* Functional and structural characterization of the integrase from the prototype foamy virus. *Nucleic Acids Res.* **37**, 243–255 (2009).
14. De Vlaminck, I., Henighan, T., van Loenhout, M. T. J., Burnham, D. R. & Dekker, C. Magnetic forces and DNA mechanics in multiplexed magnetic tweezers. *PLoS ONE* **7**, e41432 (2012).
15. Lionnet, T. *et al.* Magnetic trap construction. *Cold Spring Harb. Protoc.* **2012**, 133–138 (2012).
16. Lipfert, J., Hao, X. & Dekker, N. H. Quantitative modeling and optimization of magnetic tweezers. *Biophys. J.* **96**, 5040–5049 (2009).
17. Strick, T. R., Allemand, J. F., Bensimon, D. & Croquette, V. Behavior of supercoiled DNA. *Biophys. J.* **74**, 2016–2028 (1998).
18. Gosse, C. & Croquette, V. Magnetic tweezers: micromanipulation and force measurement at the molecular level. *Biophys. J.* **82**, 3314–3329 (2002).
19. Selvin, P. R. *et al.* Torsional rigidity of positively and negatively supercoiled DNA. *Science* **255**, 82–85 (1992).
20. Crick, F. H. Linking numbers and nucleosomes. *Proc. Natl Acad. Sci. USA* **73**, 2639–2643 (1976).
21. van Loenhout, M. T. J., de Grunt, M. V. & Dekker, C. Dynamics of DNA Supercoils. *Science* **338**, 94–97, (2012).
22. Cribb, J. *et al.* Measuring local microrheological properties of human mucus with magnetically driven microbeads. *Biophys. J.* **88**, 505a–505a (2005).
23. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
24. Berg, O. G. & von Hippel, P. H. Diffusion-controlled macromolecular interactions. [Review] [96 refs]. *Annu. Rev. Biophys. Chem.* **14**, 131–160 (1985).
25. Cho, W.-K. *et al.* ATP alters the diffusion mechanics of MutS on mismatched DNA. *Structure* **20**, 1264–1274 (2012).
26. Jeong, C. *et al.* MutS switches between two fundamentally distinct clamps during mismatch repair. *Nat. Struct. Mol. Biol.* **18**, 379–385 (2011).
27. Lee, J. B. *et al.* Single-molecule views of MutS on mismatched DNA. *DNA Repair (Amst.)* **20**, 82–93 (2014).
28. Gowers, D. M., Wilson, G. G. & Halford, S. E. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl Acad. Sci. USA* **102**, 15883–15888 (2005).
29. Halford, S. E. & Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).
30. Bagchi, B., Blainey, P. C. & Xie, X. S. Diffusion constant of a nonspecifically bound protein undergoing curvilinear motion along DNA. *J. Phys. Chem. B* **112**, 6282–6284 (2008).
31. Blainey, P. C. *et al.* Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.* **16**, 1224–1229 (2009).
32. Einstein, A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik (in German)* **326**, 756–780 (1905).
33. Slutsky, M. & Mirny, L. A. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035 (2004).
34. van Oijen, A. M. *et al.* Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science* **301**, 1235–1238 (2003).
35. Glattig, G., Winkler, R. G. & Reineker, P. Freely jointed chain with variable segment number and length. *Colloid Polym. Sci.* **273**, 32–37 (1995).
36. Serrao, E. *et al.* Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: implications for the mechanism of target DNA binding. *Nucleic Acids Res.* **42**, 5164–5176 (2014).
37. Cherepanov, P. *et al.* HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* **278**, 372–381 (2003).
38. Shun, M. C. *et al.* LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev.* **21**, 1767–1778 (2007).
39. Scott, D. W. Optimal and data-based histograms. *Biometrika* **66**, 605–610 (1979).
40. Sturges, H. A. The choice of class intervals. *J. Am. Stat. Assoc.* **21**, 65–66 (1926).
41. Cho, W. K. *et al.* ATP alters the diffusion mechanics of MutS on mismatched DNA. *Structure* **20**, 1264–1274 (2012).
42. Yardimci, H., Loveland, A. B., van Oijen, A. M. & Walter, J. C. Single-molecule analysis of DNA replication in *Xenopus* egg extracts. *Methods* **57**, 179–186 (2012).
43. Qian, H., Sheetz, M. P. & Elson, E. L. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophys. J.* **60**, 910–921 (1991).

## Acknowledgements

We thank Jiaquan Liu and Randi Mackler for helpful discussions; and Michael Poirier for early advice. This work was supported by NRF Grant No. 2011-0013901 (J.-B.L.) and NIH grants AI099854 (K.E.Y.) and GM080176 (R.F.).

## Author contributions

N.D.J., J.-B.L., R.F. and K.E.Y. designed the studies; N.D.J., M.A.L.Jr and J.H. performed the single molecule studies; M.A.L.Jr and M.B.P. performed bulk biochemical analysis; N.D.J., J.-B.L., R.F. and K.E.Y. analysed the data; N.D.J., J.-B.L., R.F. and K.E.Y. wrote the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Jones, N. D. *et al.* Retroviral intasomes search for a target DNA by 1D diffusion which rarely results in integration. *Nat. Commun.* **7**:11409 doi: 10.1038/ncomms11409 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>