



Repair of Oxidative DNA Base Damage in the Host Genome Influences the HIV Integration Site Sequence Preference

Geoffrey R. Bennett¹, Ryan Peters¹, Xiao-hong Wang³, Jeunghill Hanne¹, Robert W. Sobol^{2,3,4}, Ralf Bundschuh^{5,6}, Richard Fishel^{1,5}, Kristine E. Yoder^{1*}

1 Molecular Virology, Immunology, and Medical Genetics, The Ohio State University Wexner Medical Center and Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, United States of America, **2** Department of Pharmacology & Chemical Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America, **3** University of Pittsburgh Cancer Institute, Hillman Cancer Center, Pittsburgh, Pennsylvania, United States of America, **4** Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania, United States of America, **5** Department of Physics, The Ohio State University, Columbus, Ohio, United States of America, **6** Department of Chemistry and Biochemistry, Division of Hematology, Department of Internal Medicine, Center for RNA Biology, The Ohio State University, Columbus, Ohio, United States of America

Abstract

Host base excision repair (BER) proteins that repair oxidative damage enhance HIV infection. These proteins include the oxidative DNA damage glycosylases 8-oxo-guanine DNA glycosylase (OGG1) and mutY homolog (MYH) as well as DNA polymerase beta (Pol β). While deletion of oxidative BER genes leads to decreased HIV infection and integration efficiency, the mechanism remains unknown. One hypothesis is that BER proteins repair the DNA gapped integration intermediate. An alternative hypothesis considers that the most common oxidative DNA base damages occur on guanines. The subtle consensus sequence preference at HIV integration sites includes multiple G:C base pairs surrounding the points of joining. These observations suggest a role for oxidative BER during integration targeting at the nucleotide level. We examined the hypothesis that BER repairs a gapped integration intermediate by measuring HIV infection efficiency in Pol β null cell lines complemented with active site point mutants of Pol β . A DNA synthesis defective mutant, but not a 5' dRP lyase mutant, rescued HIV infection efficiency to wild type levels; this suggested Pol β DNA synthesis activity is not necessary while 5' dRP lyase activity is required for efficient HIV infection. An alternate hypothesis that BER events in the host genome influence HIV integration site selection was examined by sequencing integration sites in OGG1 and MYH null cells. In the absence of these 8-oxo-guanine specific glycosylases the chromatin elements of HIV integration site selection remain the same as in wild type cells. However, the HIV integration site sequence preference at G:C base pairs is altered at several positions in OGG1 and MYH null cells. Inefficient HIV infection in the absence of oxidative BER proteins does not appear related to repair of the gapped integration intermediate; instead oxidative damage repair may participate in HIV integration site preference at the sequence level.

Citation: Bennett GR, Peters R, Wang X-h, Hanne J, Sobol RW, et al. (2014) Repair of Oxidative DNA Base Damage in the Host Genome Influences the HIV Integration Site Sequence Preference. PLoS ONE 9(7): e103164. doi:10.1371/journal.pone.0103164

Editor: Maria Spies, University of Iowa, United States of America

Received: April 16, 2014; **Accepted:** June 24, 2014; **Published:** July 22, 2014

Copyright: © 2014 Bennett et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All HIV integration site sequence files are available in the NCBI GSS database. NCBI accession numbers for the sequence data described in this manuscript have now been assigned. They are KG960306–KG961230.

Funding: The following National Institutes of Health grants supported this work: National Cancer Institute (www.cancer.gov) grant CA148629 to RWS, and National Institute of Allergy and Infectious Diseases Extramural Activities (www.niaid.nih.gov) grants AI82422 and AI99854 were both awarded to co-investigators KEY and RF. A National Science Foundation Division of Materials Research (http://www.nsf.gov/div/index.jsp?div=DMR) grant DMR-01105458 supported RB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: RWS is currently a member of the PLOS ONE Editorial Board. His service on the Editorial Board has not altered the authors' adherence to the PLOS ONE editorial policies and criteria. RWS is a scientific consultant for Trevigen, Inc. This does not alter the authors' adherence of PLOS ONE policies on sharing data and materials. The other authors declare that no competing interests exist.

* Email: kristine.yoder@osumc.edu

Introduction

Retroviruses are defined by two enzymatic activities, reverse transcription and integration [1]. The viral enzyme reverse transcriptase copies the viral genomic RNA to a linear double stranded cDNA. The cDNA is part of a pre-integration complex (PIC) that is poorly understood. The PIC travels to the nucleus where the viral protein integrase mediates the covalent joining of the viral cDNA to the host chromosome. The two ends of the viral cDNA are joined to the host DNA 4–6 base pairs apart, depending

on the retrovirus. In the case of HIV, 5 base pairs separate the points of joining, the distance across one major groove (Figure 1A). The process of integration yields an integration intermediate comprised of the viral cDNA flanked by 5 base gaps of host sequence and 5' dinucleotide flaps of viral sequence. Host enzymes are assumed to repair this integration intermediate, but identities of the specific repair proteins remain unknown [2].

The host protein PSIP1/LEDGF p75 is the major host co-factor for HIV integration [3–7]. This protein specifically binds HIV integrase with an integrase-binding domain (IBD) and binds to

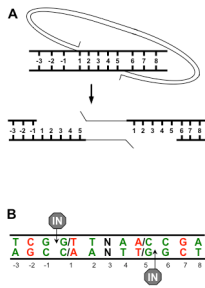


Figure 1. Retroviral integration. (A) Viral cDNA is depicted by a thin line and host target DNA is indicated by a thick line. Base pairs in the host target DNA are numbered. The HIV LTR ends are covalently joined to the target DNA 5 base pairs apart. The intervening host DNA denatures yielding an integration intermediate with two 5 base pair gaps. (B) The sequence preference observed at HIV integration sites. The numbering is identical to (A) and the points of joining are indicated by "IN". Base pairs in green are favored and base pairs in red are disfavored at HIV integration sites.
doi:10.1371/journal.pone.0103164.g001

chromatin via a PWWP domain and two AT hook motifs [8–16]. Studies of *LEDGF* depletion and deletion cells have shown that *LEDGF* is required for efficient integration to chromatin *in vivo* [4,5]. *LEDGF* also enhances integration by recombinant integrase *in vitro* [17–19].

Integration sites from multiple retroviruses have been sequenced and mapped in relation to chromatin features such as transcription units, promoters, CpG islands, and local G:C content [5,20–23]. Retroviruses have distinct integration site preferences. For example, HIV preferentially integrates to transcription units but does not favor promoters while MLV favors integration near promoter sequences. The sequence at the site of integration, including the 4–6 base pairs between and up to three base pairs flanking the points of joining, show a palindrome of preferred and disfavored nucleotides (Figure 1B, [24,25]). Sequences from cytoplasmic HIV PICs integrated to a naked DNA target showed that the central 5 base pairs retained the sequence preference as well as positions –3 and 6 [25]. This comparison suggests that many base pairs at the integration site are recognized by integrase, but additional host factors may also influence integration at some flanking nucleotides. While deletion of *LEDGF* leads to significant differences of integration targeting to genomic features *in vivo*, it had no effect on the palindromic sequence preference [5,26].

The short patch oxidative base excision repair (BER) pathway has been implicated in the integration of HIV [27,28]. The BER pathway recognizes damage on individual bases, including deamination, alkylation, and oxidation [29–31]. Glycosylases recognize specific damages of DNA bases. The damaged base is removed by a glycosylase leaving an abasic site. The essential enzyme apurinic/apyridinic endonuclease (APEX1) recognizes the abasic site and cleaves the DNA backbone at the 5' side of the lesion to generate a 3' hydroxyl and a 5' deoxyribose phosphate (5'dRP) flap. In the short patch BER pathway, DNA polymerase β (Pol β) synthesizes one base and also removes the 5'dRP lesion. Repair is completed when the remaining nick is ligated by a heterodimer of Lig3 and XRCC1 or via LigI [32]. Proteins throughout the pathway from oxidative base damage recognition through the final ligation step were identified in an siRNA screen for DNA repair factors affecting HIV infection [27]. Reduced expression of oxidative BER genes led to a decrease of HIV infection efficiency. The absence of BER proteins was characterized to reduce HIV integration [27,28].

The mechanism of BER mediated effects on HIV integration has not been shown. One formal possibility is repair of the gapped integration intermediate by BER. We have further explored the effects of BER proteins on HIV integration with well-defined BER mutant cell lines. The previously observed effects of the BER pathway on HIV integration do not appear related to gap repair of an integration intermediate. Instead, oxidative glycosylases with specific base recognition may affect the sequence preference of HIV integration. This suggests that integration may be favored at sites of oxidative DNA repair intermediates.

Materials and Methods

Cell lines

Matched wild type and *Pol β* null murine embryonic fibroblasts were derived from littermates and have been previously described [33]. The *Pol β* null cell line was stably transfected with an empty vector (Pol β –/– Neo) or the same vector expressing the wild type *Pol β* cDNA (PolB–/– +PolB), a polymerase defective *Pol β* (*D256A*) mutant cDNA (PolB–/– +PolB (*pol*-)), a lyase defective *Pol β* (*K35,68,72A*) mutant cDNA (PolB–/– +PolB (*lyase*-)), or a completely inactive *Pol β* (*K35,68,72A, D256A*) mutant cDNA (PolB–/– +PolB (*pol-lyase*-)) [33]. Nuclear extracts were used to determine the protein expression of the endogenous and transgenic Pol β proteins in established stable cell lines. Nuclear extracts were prepared with the NucBuster protein extraction kit and protein concentration was determined as described previously. 20–30 μ g nuclear extract proteins were loaded on a precast 4–12% NuPAGE Tris-glycine gel, run 2–3 h at 100 volts. The gel was transferred to a 0.45 μ m nitrocellulose membrane (Bio-Rad) at 0.2 mA for 2–3 hours. The membrane was blotted with anti-Pol β (1:5000, Clone 18S) or anti-PCNA (1:2000, SC-56, Santa Cruz Biotechnology), followed by the secondary antibody Immun-Star goat anti-mouse HRP conjugate (Bio-Rad). After washing, the membrane was illuminated with Immun-Star HRP peroxide buffer with luminol/enhancer (Bio-Rad).

Matched wild type, *OGG1* null, and *MYH* null murine embryonic fibroblasts were also derived from littermates and have been previously described [28,34]. All cells were grown in DMEM supplemented with 10% fetal bovine serum, GlutaMAX glutamine, and Penicillin-Streptomycin (Gibco, Life Technologies). The media of transfected *Pol β* null cell lines was supplemented with 600 μ g/ml geneticin. Cells were grown at 37°C with 10% CO₂.

HIV vector particles

HIV lentiviral vectors expressing GFP were generated by triple transfection of human fibroblast 293T cells with a packaging construct Δ R9, a viral genomic RNA plasmid p156RRLsinPPTCMVGFPPRE, and a plasmid expressing the envelope gene *VSV-G* [35,36]. Supernatants containing vector particles were sterile filtered to remove producer cells and treated with DNaseI to digest producer plasmids. Target cells were plated in 6 well dishes at a density to achieve 2×10^5 cells 24 hours after plating. Cell density was verified by counting with trypan blue (Gibco, Life Technologies). Vector particles were added to target cell media in the presence of 10 μ g/ml DEAE-Dextran (Sigma Aldrich). The MOI of wild type Pol β cells in Figure 2 was 0.4 and in Figure 3 was 0.3. Cells were analyzed for infection efficiency 72 hours post infection by fixing with 4% paraformaldehyde (Sigma Aldrich) and scanning for GFP expression with a FACScalibur (BD Biosciences). Flow cytometry data was analyzed with FlowJo software. Cells were exposed to varying concentrations of H₂O₂ (Sigma Aldrich) in PBS for one hour at 37°C prior to infection.

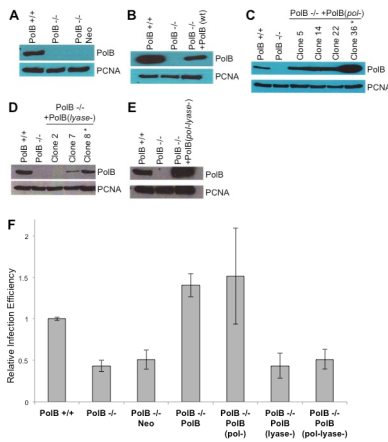


Figure 2. Polβ polymerase activity is not required for efficient HIV infection. Murine embryonic fibroblasts with a deletion of the Polβ cDNA (Polβ^{-/-}) were complemented with (A) an empty vector (Neo), (B) the wild type cDNA (Polβ), (C) a polymerase defective point mutant gene (Polβ (pol-)), (D) a lyase deficient mutant cDNA (Polβ (lyase-)), or (E) an enzymatically dead mutant (Polβ (pol-lyase-)). (F) These cell lines and a matched wild type cell line were infected with an HIV based vector expressing GFP following integration. Cells were analyzed by flow cytometry for GFP expression, an indicator of successful infection efficiency. Results are from three independent experiments of duplicates and are expressed relative to wild type cells. Error bars indicate the standard deviations. doi:10.1371/journal.pone.0103164.g002

Viability was measured by trypan blue exclusion at 72 hours post infection.

Integration site sequencing

Cells were infected at a 0.8 multiplicity of infection and grown for 7 days. Genomic DNA was harvested by DNeasy (Qiagen) and prepared for sequencing as described [37,38]. All sequences had perfect homology to the terminal 18 bp of HIV U3 and mouse sequences were at least 20 bp with >98% identity. Sequences were mapped to the mouse genome using BLAT [39]. The number of individual integration sites analyzed included 240 from untreated wild type cells (GenBank Accession numbers KG960523–KG960762), 105 from wild type cells treated with 10 μM H₂O₂ (GenBank Accession numbers KG960306–KG960410), 112 from wild type cells treated with 30 μM H₂O₂ (GenBank Accession numbers KG960411–KG960522), 245 from *OGG1* null cells (GenBank Accession numbers KG960986–KG961230), and 223 from *MYH* null cells (GenBank Accession numbers KG960763–KG960985). The frequency of G or C in these murine cells is 0.205 and the frequency of A or T is 0.295. The observed frequencies in integration sites are expressed as the difference from random frequencies 0.205 or 0.295. Statistical significance of deviations of base frequency was determined by a two-tailed binomial test in Microsoft Excel. Statistical analysis of integration sites compared to genomic elements was by Fisher's exact test using GraphPad Prism version 4.0 for Macintosh (GraphPad Software).

Results

Polβ 5'dRP lyase activity, but not polymerase activity, is necessary for efficient HIV infection

An HIV integration intermediate may be repaired *in vitro* by any combination of a polymerase, flap endonuclease, and ligase [2]. The host DNA repair polymerase Polβ could repair a model

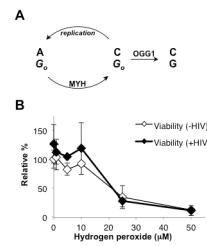


Figure 3. Oxidative damage during HIV infection. (A) OGG1 and MYH both recognize 8-oxo-dG damage (shown as G_o). OGG1 repairs 8-oxo-dG to G. Replication of 8-oxo-dG results in an 8-oxo-dG:A mispair, which is recognized by MYH. The MYH glycosylase initiates repair of the mispaired A to C yielding an 8-oxo-dG:C base pair. The product of the MYH repair reaction must still be repaired by OGG1. (B) Wild type cells were infected with an HIV based vector expressing GFP. Cells were treated with increasing concentrations of H₂O₂ immediately prior to infection. Viability in the absence (open diamonds) or presence of HIV (closed diamonds) was measured by trypan blue exclusion. Error bars indicate the standard deviations from three independent experiments. doi:10.1371/journal.pone.0103164.g003

HIV integration intermediate *in vitro* by synthesizing nascent DNA at the gap. Single deletion of oxidative BER genes *OGG1*, *MYH*, or *Polβ* reduces HIV infection to approximately 40% of wild type cells [28]. *Polβ* null murine embryonic fibroblasts were complemented with an empty vector or the wild type *Polβ* cDNA (Figure 2A, B) [33]. The cells were infected with an HIV based retroviral vector expressing GFP following successful integration [36]. Infection efficiency was measured by flow cytometry of GFP positive cells. Complementation with the wild type *Polβ* cDNA, but not the empty vector, completely rescued infection efficiency (Figure 2F, *Polβ*^{-/-} compared to *Polβ*^{-/-} complemented with an empty expression vector p=0.91, *Polβ*^{-/-} compared to *Polβ*^{-/-} complemented with a wild type *Polβ* transgene p<0.0001).

Polβ has two distinct enzymatic activities which may be distinguished by mutations at two separate active sites [33]. In addition to polymerase activity, the lyase activity of Polβ cleaves a 5'dRP flap that occurs during BER. DNA synthesis activity is abolished by the mutation Polβ(D256A); a triple mutation Polβ(K35A/K68A/K72A) abrogates the 5'dRP lyase activity [33]. *Polβ* null cells were complemented with mutant transgenes affecting only the polymerase or the 5'dRP lyase active site (Figure 2C, D). These cells were infected with an HIV based retroviral vector expressing GFP and assayed by flow cytometry. Complementation with a polymerase defective, 5'dRP lyase active *Polβ* transgene was able to rescue HIV infection efficiency to wild type levels (Figure 2F, *Polβ*^{-/-} compared to *Polβ*^{-/-} complemented with a *Polβ*(D256A) transgene p=0.01). However, a polymerase active, 5'dRP lyase defective transgene was unable to rescue HIV infection efficiency (Figure 2F, *Polβ*^{-/-} compared to *Polβ*^{-/-} complemented with a *Polβ*(K35A/K68A/K72A) transgene p=0.21). This data suggests that the polymerase activity of Polβ is not necessary for HIV integration and does not support a model for BER repair of a gapped integration intermediate. The data does indicate that the 5'dRP lyase activity of Polβ is necessary for efficient integration. A combination mutant protein defective for both 5'dRP lyase and polymerase activities was also unable to rescue the HIV infection phenotype (Figure 2E, F, *Polβ*^{-/-} compared to *Polβ*^{-/-} complemented with a *Polβ*(D256A, K35A/K68A/K72A) transgene p=0.14). The presence of enzymatically inactive Polβ protein is not sufficient to support HIV integration.

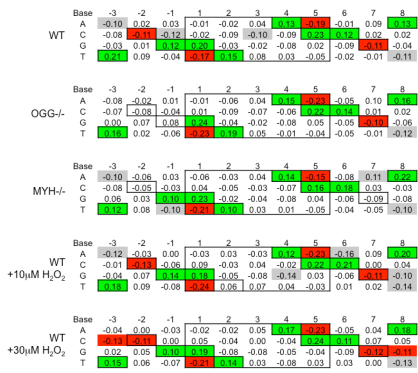


Figure 4. Effects of 8-oxo-dG specific glycosylases or hydrogen peroxide on HIV integration site sequence preference. Wild type, *OGG1* null, *MYH* null, and wild type cells treated with 10 μM or 30 μM H_2O_2 , were infected with an HIV based retroviral vector at 0.8 MOI. 10 μM H_2O_2 is less than the 50% lethal dose for the wild type cells and 30 μM H_2O_2 is greater. After 7 days, genomic DNA was purified. HIV integration sites were subcloned, sequenced, and mapped to the mouse genome. The random frequency of G or C in the mouse genome is 0.205 and A or T is 0.295. The differences in observed base frequencies relative to the random frequencies are shown. Base numbering relative to the HIV points of joining is as in Figure 1. Boxes indicate the known HIV integration base preferences for wild type cells. Green, red, and gray highlights indicate a statistically significant difference of >0.10 from random frequency with a p value of <0.005 . Green highlights are positive differences and red highlights are negative previously published palindromic preferred bases. Deletion of the *OGG1* gene leads to a loss of HIV sequence preference at positions -2 and -1 . Deletion of the *MYH* gene also leads to loss of the HIV sequence preference at positions -2 and 7 . While treatment of wild type cells with 10 μM H_2O_2 did not dramatically alter the sequence preference at integration sites, treatment with 30 μM H_2O_2 led to the disfavor of C at position -3 and disfavor of G at position 8 , highlighted in red.

doi:10.1371/journal.pone.0103164.g004

BER proteins and oxidative damage influence HIV integration site sequence preference

If the oxidative BER pathway is not repairing the integration intermediate, it must enhance integration by an alternative mechanism. The most common form of oxidative DNA base damage is 8-oxo-guanine (8-oxo-dG) [29]. The sequence preference at HIV integration sites includes multiple G:C base pairs (Figure 1B). This observation suggested that sites of oxidative DNA base repair might influence HIV integration; the sequence preference at integration sites could be 8-oxo-dG:C base pairs, which are not detectable by sequencing integration sites. Of the oxidative BER proteins, *OGG1* and *MYH* have specificity for only 8-oxo-dG [30]. *OGG1* directly recognizes 8-oxo-dG and initiates repair by removing the damaged base (Figure 3A). During replication 8-oxo-dG mispairs with A, resulting in an 8-oxo-dG:A mismatch. *MYH* specifically recognizes and binds both bases in this mismatch, but does not remove the damaged base. Instead *MYH* excises the mispaired A and leaves the damaged 8-oxo-dG, allowing a subsequent round of repair by *OGG1* that will remove the 8-oxo-dG (Figure 3A) [40]. The remaining BER proteins, including *Pol β* , participate in repair of multiple oxidatively damaged bases.

Murine embryonic fibroblasts with deletions of *OGG1* or *MYH* and matched wild type cells were infected with an HIV based retroviral vector. The genomic DNA of infected cells was purified, integration sites were subcloned, sequenced, and mapped to the mouse genome [37]. The sequence preference of integration sites from wild type cells was similar to that previously observed during

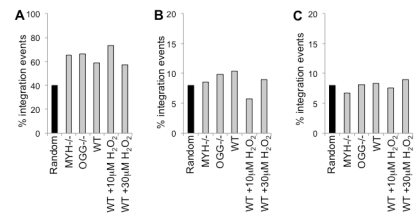


Figure 5. HIV integration near genomic elements. HIV integration sites were mapped to genomic elements in *OGG1* null cells, *MYH* null cells, untreated wild type cells, and wild type cells treated with 10 μM or 30 μM H_2O_2 . (A) HIV has a preference for integration to transcription sites. (B) HIV shows no preference for integration to promoters. (C) There is no preference for HIV integration within 5 kb of CpG islands.

OGG1 null cells integration sites showed a loss of any statistically significant preference for G at position -1 or C at position -2 (Figure 4). In *MYH* null cell lines there is a loss of base preference for C at positions -2 and 7 . Although C is not recognized by either repair protein, *MYH* or *OGG1* would recognize 8-oxo-dG on the opposite strand (Figure 3A). Thus the deletion of 8-oxo-dG specific glycosylase genes *MYH* and *OGG1* specifically alters the HIV integration site preference at G:C base pairs near the points of joining. These changes are reminiscent of the HIV PIC integration sequence preference *in vitro* which also loses preference for G:C base pairs at positions -2 , -1 , and 7 [25].

Wild type cells were also treated with the oxidative damage-inducing agent H_2O_2 immediately before HIV infection. Treatment of wild type cells with 10 μM H_2O_2 has no effect on viability, but treatment with 25 μM H_2O_2 reduces viability to less than 40% (Figure 3B). Thus 10 μM H_2O_2 is less than and 25 μM H_2O_2 is greater than the 50% lethal dose in this cell line. The sensitivity of these wild type cells to H_2O_2 is not affected by HIV infection

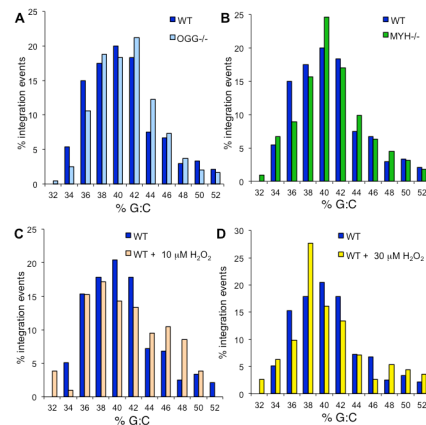


Figure 6. G:C content surrounding HIV integration sites. The percentage of HIV integration sites compared to the percentage of G:C base pairs within a 5 kb window is shown for wild type cells compared to (A) *OGG1* null cells, (B) *MYH* null cells, (C) wild type cells treated with 10 μM H_2O_2 , and (D) wild type cells treated with 30 μM H_2O_2 . The frequency of G:C base pairs in the mouse genome is 0.41. There is no significant difference between HIV integration sites in untreated wild type cells and *OGG1* null ($p=0.96$), *MYH* null ($p=0.99$), 10 μM H_2O_2 treated ($p=0.67$), or 30 μM H_2O_2 treated cells ($p=0.88$).

(Figure 3B). The major base damage induced by H_2O_2 is on guanines [41]. The integration site sequence preference for 10 μM or 30 μM H_2O_2 was similar to untreated wild type cells (Figure 4). However, the sequence preference observed in 30 μM H_2O_2 treated cells was expanded to the -3 and 8 positions. While T and A continue to be favored at positions -3 and 8, respectively, H_2O_2 treatment led to a statistically significant disfavor of C at position -3 and G at position 8 (Figure 4). Increased oxidative damage at the time of infection appears to augment the HIV integration site sequence preference.

Integration sites were also analyzed for their proximity to genomic level elements. HIV shows a preference for integration in transcription units, but not near promoters. Deletion of the known HIV integrase co-factor LEDGF dramatically changes the integration preference for these elements [5]. The integration sites from all of these cell lines or wild type cells treated with H_2O_2 showed a preference for integration to transcription units (Figure 5A, $p < 0.05$ for all cell lines), but no preference for integration near promoters or CpG islands (Figure 5B and C, $p > 0.2$ for all cell lines at both promoters and CpG islands). Deletion of LEDGF significantly alters the integration profile of G:C content surrounding the integration sites [5]. The profiles of G:C content within 5 kb around the integration sites are similar for all the BER cell lines (Figure 6). Thus the chromatin markers of integration sites are not affected by deletion of *OGG1* or *MYH* or treatment with H_2O_2 .

Discussion

DNA repair proteins throughout the oxidative BER pathway were identified in an siRNA screen for factors affecting HIV infection [27]. Other members of the BER pathway that mediate repair of methylation, alkylation, or deamination base damage had no effect on HIV infection efficiency. Oxidative damage of DNA is always present in cells and BER proteins are expressed throughout the cell cycle [42,43]. Presence of oxidative damage and oxidative DNA damage repair proteins appear to correlate with successful HIV infection.

One obvious mechanism for the role of a DNA repair pathway during HIV infection could be the repair of the gapped integration intermediate. However, several observations argue against this hypothesis. First, deletion of BER proteins does not affect infection efficiency of a gamma retrovirus [28]. It seems probable that the DNA repair pathway responsible for repair of integration intermediates would broadly repair this structure for all retroviruses. Second, there is no obvious role for glycosylases during gap repair. Finally, the polymerase activity of Pol β has no effect on HIV infection efficiency (Figure 2). Instead the 5'dRP lyase activity of Pol β appears to be important. The Pol β 5'dRP lyase active site is not likely able to accommodate the 5' dinucleotide flap present in the HIV integration intermediate [44,45]. A previous study of this repair event shows that wild type recombinant Pol β was not able to digest the 5' dinucleotide flap of a model gapped integration intermediate *in vitro* [2]. Thus the role of the Pol β 5'dRP lyase activity is also not likely part of the integration gap repair event. While oxidative BER proteins may participate in repair of retroviral gapped integration intermediates, they are not absolutely required; there are likely multiple redundant proteins that participate in this repair [2]. Instead it appears that oxidative BER proteins identified in an siRNA screen may directly affect integration targeting at the sequence level.

The host protein LEDGF has been shown to directly bind the HIV integration complex and direct integration to sites within chromatin [46–48]. However, LEDGF has no effect on the

sequence preference at the site of integration [5,49,50]. Deletion or reduced expression of the *LEDGF* gene leads to significantly decreased HIV infection [4,5]. However, cytoplasmic or nuclear pre-integration complexes from *LEDGF* null cells show integration efficiency to a naked DNA target equal to PICs from wild type cells [5]. PICs from BER null cells show reduced integration efficiency to naked DNA targets compared to PICs from wild type cells, which can be rescued by addition of recombinant Pol β protein [28]. LEDGF appears to mediate HIV integration by tethering the PIC to chromatin, but not naked DNA [18]; BER proteins appear able to affect HIV integration targeting in naked DNA. Thus LEDGF affects HIV integration targeting on a chromatin scale, while BER proteins may affect integration targeting on a nucleotide scale.

One model for BER protein effects on targeting HIV integration is for direct protein binding with the integration complex. This model seems unlikely due to known binding of LEDGF to integrase and steric interference [46–48,51–54]. Alternatively, the oxidative BER proteins may affect the DNA or chromatin in a way that enhances HIV integration efficiency. Oxidative damage of genomic DNA is constant and BER proteins are constitutively expressed [42,43]. Glycosylases that do not recognize oxidative damage were shown to have no effect on HIV infection efficiency [27]. A major difference between oxidative glycosylases and all other glycosylases is their DNA product. All glycosylases remove a base leaving an abasic site. Oxidative glycosylases have AP (apurinic or apyridimic) lyase activity which nicks the sugar phosphate backbone at the abasic site. MYH is the only oxidative damage glycosylase that does not have AP lyase activity; however, MYH must always act in concert with OGG1 which does have AP lyase activity (Figure 3A). Hence, all oxidative base excision repair will result in an abasic site with a nick. HIV infection efficiency is reduced in the absence of glycosylases that have AP lyase activity or in the absence of Pol β 5'dRP lyase activity ([27] and this work).

An alternative model for the role of oxidative BER proteins during HIV integration is based on the nature of the oxidative BER DNA intermediates. In this model BER single base gap DNA intermediates, rather than the BER proteins, are the true mediators of enhanced HIV integration. The single base gap of short patch BER increases the local flexibility of DNA more than a nick or an abasic site [55–57]. HIV integration is enhanced by an obtuse angle of bent DNA and inhibited in an acute angle [58]. Thus HIV strand joining may be favored on a helical strand with local oxidative base repair (such as positions -1 and 6) and disfavored on the opposite strand (such as positions -3, -2, 7 and 8). The base preference of HIV integrase is subtle, requiring analysis of many integration sites. The observed differences in sequence preference are also subtle due in part to the apparent inherent promiscuity of HIV integrase as well as the redundancy of BER pathway proteins. For example, OGG1 only recognizes oxidatively damaged guanine, but other glycosylases also recognize damaged guanine.

Additional studies are required to elucidate the roles of DNA lyase activity during HIV integration. Only glycosylases with AP lyase activity enhance HIV infection efficiency. While it is impossible to isolate the lyase activity of bifunctional glycosylases, it is possible to differentiate the DNA synthesis and 5'dRP lyase activities of Pol β . Whether the AP lyase activity of glycosylases and the 5'dRP lyase activity of Pol β play the same function during integration also requires further investigation. However, this data suggest the intriguing possibility that DNA lyase activity may be a novel enzymatic target for anti-retroviral therapies.

Conclusions

HIV integration has a preference for both nucleotide sequence and chromatin features. The host protein LEDGF targets HIV to integrate near genomic elements but has no effect on the consensus sequence preference of integration sites. The oxidative BER pathway may influence the sequence preference at HIV integration sites. Additionally, host DNA lyase activity is a potential target for novel HIV inhibitors.

References

- Coffin JM, Hughes SH, Varmus HE (1997) Retroviruses. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Yoder KE, Bushman FD (2000) Repair of gaps in retroviral DNA integration intermediates. *J Virol* 74: 11191–11200.
- Cherepanov P, Maertens G, Proost P, Devreese B, Van Becumen J, et al. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* 278: 372–381.
- Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, et al. (2006) An essential role for LEDGF/p75 in HIV integration. *Science* 314: 461–464.
- Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, et al. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev* 21: 1767–1778.
- Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* 11: 1287–1289.
- Vandekerckhove L, Christ F, Van Maele B, De Rijck J, Gijssbers R, et al. (2006) Transient and stable knockdown of the integrase cofactor LEDGF/p75 reveals its role in the replication cycle of human immunodeficiency virus. *J Virol* 80: 1886–1896.
- Turlure F, Maertens G, Rahman S, Cherepanov P, Engelman A (2006) A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. *Nucleic Acids Res* 34: 1653–1665.
- Shun MC, Botbol Y, Li X, Di Nunzio F, Daigle JE, et al. (2008) Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. *J Virol* 82: 11555–11567.
- Gijssbers R, Vets S, De Rijck J, Owcieja KE, Ronen K, et al. (2011) Role of the PWWP domain of lens epithelium-derived growth factor (LEDGF/p75) cofactor in lentiviral integration targeting. *J Biol Chem* 286: 41812–41825.
- Llano M, Vanegas M, Hutchins N, Thompson D, Delgado S, et al. (2006) Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. *J Mol Biol* 360: 760–773.
- Vanegas M, Llano M, Delgado S, Thompson D, Peretz M, et al. (2005) Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. *J Cell Sci* 118: 1733–1743.
- Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, et al. (2005) The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J Biol Chem* 280: 17841–17847.
- Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, et al. (2005) Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *J Biol Chem* 280: 25517–25523.
- Maertens G, Cherepanov P, Pluyms W, Busschots K, De Clercq E, et al. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem* 278: 33528–33539.
- Hendrix J, Gijssbers R, De Rijck J, Voet A, Hotta J, et al. (2011) The transcriptional co-activator LEDGF/p75 displays a dynamic scan-and-lock mechanism for chromatin tethering. *Nucleic Acids Res* 39: 1310–1325.
- McKee CJ, Kessl JJ, Shkriabai N, Dar MJ, Engelman A, et al. (2008) Dynamic modulation of HIV-1 integrase structure and function by cellular lens epithelium-derived growth factor (LEDGF) protein. *J Biol Chem* 283: 31802–31812.
- Botbol Y, Raghavendra NK, Rahman S, Engelman A, Lavigne M (2008) Chromatinized templates reveal the requirement for the LEDGF/p75 PWWP domain during HIV-1 integration in vitro. *Nucleic Acids Res* 36: 1237–1246.
- McNeely M, Hendrix J, Busschots K, Boons E, Deleersnijder A, et al. (2011) In vitro DNA tethering of HIV-1 integrase by the transcriptional coactivator LEDGF/p75. *J Mol Biol* 410: 811–830.
- Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, et al. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* 2: E234.
- Barr SD, Leipzig J, Shinn P, Ecker JR, Bushman FD (2005) Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J Virol* 79: 12035–12044.
- Crise B, Li Y, Yuan C, Morcock DR, Whitby D, et al. (2005) Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. *J Virol* 79: 12199–12204.

Acknowledgments

We thank Randi Mackler, Miguel Lopez, and Molly Osterhage for critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: RWS RB RF KEY. Performed the experiments: GRB RP XW RB KEY. Analyzed the data: GRB RP JH RWS RB RF KEY. Contributed reagents/materials/analysis tools: XW JH RWS RB. Contributed to the writing of the manuscript: GRB RP RWS RB RF KEY.

- Trobridge GD, Miller DG, Jacobs MA, Allen JM, Kiem HP, et al. (2006) Foamy virus vector integration sites in normal human cells. *Proc Natl Acad Sci U S A* 103: 1498–1503.
- Holman AG, Coffin JM (2005) Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc Natl Acad Sci U S A* 102: 6103–6107.
- Wu X, Li Y, Crise B, Burgess SM, Munroe DJ (2005) Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J Virol* 79: 5211–5214.
- Wang H, Jurado KA, Wu X, Shun MC, Li X, et al. (2012) HRP2 determines the efficiency and specificity of HIV-1 integration in LEDGF/p75 knockout cells but does not contribute to the antiviral activity of a potent LEDGF/p75-binding site integrase inhibitor. *Nucleic Acids Res* 40: 11518–11530.
- Espeseth AS, Fishel R, Hazuda D, Huang Q, Xu M, et al. (2011) siRNA screening of a targeted library of DNA repair factors in HIV infection reveals a role for base excision repair in HIV integration. *PLoS One* 6: e17612.
- Yoder KE, Espeseth A, Wang XH, Fang Q, Russo MT, et al. (2011) The base excision repair pathway is required for efficient lentivirus integration. *PLoS One* 6: e17862.
- Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, et al. (2006) DNA Repair and Mutagenesis. Washington, D.C.: ASM Press.
- Hegde ML, Hazra TK, Mitra S (2008) Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res* 18: 27–47.
- Svilar D, Goellner EM, Almeida KH, Sobol RW (2011) Base excision repair and lesion-dependent subpathways for repair of oxidative DNA damage. *Antioxid Redox Signal* 14: 2491–2507.
- Gao Y, Katyal S, Lee Y, Zhao J, Rehg JE, et al. (2011) DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair. *Nature* 471: 240–244.
- Sobol RW, Prasad R, Evenski A, Baker A, Yang XP, et al. (2000) The lyase activity of the DNA repair protein beta-polymerase protects from DNA-damage-induced cytotoxicity. *Nature* 405: 807–810.
- Russo MT, De Luca G, Casorelli I, Degan P, Molatore S, et al. (2009) Role of MUTHY and MSH2 in the control of oxidative DNA damage, genetic instability, and tumorigenesis. *Cancer Res* 69: 4372–4379.
- Naldini L, Blomer U, Galloway P, Ory D, Mulligan R, et al. (1996) In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* 272: 263–267.
- Follenzi A, Ailles LE, Bakovic S, Geuna M, Naldini L (2000) Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nat Genet* 25: 217–222.
- Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110: 521–529.
- Ciuffi A, Barr SD (2011) Identification of HIV integration sites in infected host genomic DNA. *Methods* 53: 39–46.
- Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12: 656–664.
- Sobol RW (2012) For MutY, it's all about the OG. *Chem Biol* 19: 313–314.
- Dizdaroğlu M, Nackerdien Z, Chao BC, Gajewski E, Rao G (1991) Chemical nature of in vivo DNA base damage in hydrogen peroxide-treated mammalian cells. *Arch Biochem Biophys* 285: 388–390.
- Boldogh I, Milligan D, Lee MS, Bassett H, Lloyd RS, et al. (2001) hMYH cell cycle-dependent expression, subcellular localization and association with replication foci: evidence suggesting replication-coupled repair of adenine: 8-oxoguanine mispairs. *Nucleic Acids Res* 29: 2802–2809.
- Zmudzka BZ, Fornace A, Collins J, Wilson SH (1988) Characterization of DNA polymerase beta mRNA: cell-cycle and growth response in cultured human cells. *Nucleic Acids Res* 16: 9587–9596.
- Prasad R, Batra VK, Yang XP, Krahn JM, Pedersen LC, et al. (2005) Structural insight into the DNA polymerase beta deoxyribose phosphate lyase mechanism. *DNA Repair (Amst)* 4: 1347–1357.
- Sawaya MR, Pelletier H, Kumar A, Wilson SH, Kraut J (1994) Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism. *Science* 264: 1930–1935.

46. Cherepanov P, Sun ZY, Rahman S, Maertens G, Wagner G, et al. (2005) Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat Struct Mol Biol* 12: 526–532.
47. Cherepanov P, Devroe E, Silver PA, Engelman A (2004) Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J Biol Chem* 279: 48883–48892.
48. Busschots K, Voet A, De Macyer M, Rain JC, Emiliani S, et al. (2007) Identification of the LEDGF/p75 binding site in HIV-1 integrase. *J Mol Biol* 365: 1480–1492.
49. Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, et al. (2007) Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS One* 2: e1340.
50. Schrijvers R, Vets S, De Rijck J, Malani N, Bushman FD, et al. (2012) HRP-2 determines HIV-1 integration site selection in LEDGF/p75 depleted cells. *Retrovirology* 9: 84.
51. Maillot B, Levy N, Eiler S, Crucifix C, Granger F, et al. (2013) Structural and functional role of IN1 and LEDGF in the HIV-1 preintegration complex. *PLoS One* 8: e60734.
52. Michel F, Crucifix C, Granger F, Eiler S, Mouscadet JF, et al. (2009) Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. *EMBO J* 28: 980–991.
53. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc Natl Acad Sci U S A* 102: 17308–17313.
54. Hare S, Di Nunzio F, Labeja A, Wang J, Engelman A, et al. (2009) Structural basis for functional tetramerization of lentiviral integrase. *PLoS Pathog* 5: e1000515.
55. Guo H, Tullius TD (2003) Gapped DNA is anisotropically bent. *Proc Natl Acad Sci U S A* 100: 3743–3747.
56. Mills JB, Cooper JP, Hagerman PJ (1994) Electrophoretic evidence that single-stranded regions of one or more nucleotides dramatically increase the flexibility of DNA. *Biochemistry* 33: 1797–1803.
57. Roll C, Ketterle C, Faibis V, Fazakerley GV, Boulard Y (1998) Conformations of nicked and gapped DNA structures by NMR and molecular dynamic simulations in water. *Biochemistry* 37: 4059–4070.
58. Bor YC, Bushman FD, Orgel LE (1995) In vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. *Proc Natl Acad Sci U S A* 92: 10334–10338.