

Localization of ASV Integrase-DNA Contacts by Site-Directed Crosslinking and their Structural Analysis

Elena Peletskaya^{1,9}, Mark Andrade^{2,9}, Alla Gustchina^{3,9}, George Merkel², Jerry Alexandratos³, Dongwen Zhou³, Ravi S. Bojja², Tadashi Satoh³, Mikhail Potapov⁴, Alex Kogon¹, Viktor Potapov⁴, Alexander Wlodawer^{3*}, Anna Marie Skalka²

1 Biolinx LLC, Alburgh, Vermont, United States of America, **2** Fox Chase Cancer Center, Institute for Cancer Research, Philadelphia, Pennsylvania, United States of America, **3** Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, Maryland, United States of America, **4** Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

Abstract

Background: We applied crosslinking techniques as a first step in preparation of stable avian sarcoma virus (ASV) integrase (IN)-DNA complexes for crystallographic investigations. These results were then compared with the crystal structures of the prototype foamy virus (PFV) intasome and with published data for other retroviral IN proteins.

Methodology/Results: Photoaffinity crosslinking and site-directed chemical crosslinking were used to localize the sites of contacts with DNA substrates on the surface of ASV IN. Sulfhydryl groups of cysteines engineered into ASV IN and amino-modified nucleotides in DNA substrates were used for attachment of photocrosslinkers. Analysis of photocrosslinking data revealed several specific DNA-protein contacts. To confirm contact sites, thiol-modified nucleotides were introduced into oligo-DNA substrates at suggested points of contact and chemically crosslinked to the cysteines via formation of disulfide bridges. Cysteines incorporated in positions 124 and 146 in the ASV IN core domain were shown to interact directly with host and viral portions of the Y-mer DNA substrate, respectively. Crosslinking of an R244C ASV IN derivative identified contacts at positions 11 and 12 on both strands of viral DNA. The most efficient disulfide crosslinking was observed for complexes of the ASV IN E157C and D64C derivatives with linear viral DNA substrate carrying a thiol-modified scissile phosphate.

Conclusion: Analysis of our crosslinking results as well as published results of retroviral IN protein from other laboratories shows good agreement with the structure of PFV IN and derived ASV, HIV, and MuLV models for the core domain, but only partial agreement for the N- and C-terminal domains. These differences might be explained by structural variations and evolutionary selection for residues at alternate positions to perform analogous functions, and by methodological differences: i.e., a static picture of a particular assembly from crystallography vs. a variety of interactions that might occur during formation of functional IN complexes in solution.

Citation: Peletskaya E, Andrade M, Gustchina A, Merkel G, Alexandratos J, et al. (2011) Localization of ASV Integrase-DNA Contacts by Site-Directed Crosslinking and their Structural Analysis. PLoS ONE 6(12): e27751. doi:10.1371/journal.pone.0027751

Editor: Ashok Aiyar, Louisiana State University Health Sciences Center, United States of America

Received: June 23, 2011; **Accepted:** October 24, 2011; **Published:** December 1, 2011

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This work was supported in part by the National Institutes of Health (NIH) grants AI064098 (EP), AI40385, CA71515, CA06927 (AS), by the Civilian Research and Development Foundation grant RUXO-2713-MO-06 (VP and AK), by an appropriation from the Commonwealth of Pennsylvania, and by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding received for this study.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: wlodawer@nih.gov

⁹ These authors contributed equally to this work.

Introduction

Retroviruses utilize the viral enzyme integrase (IN) for inserting DNA copies of their genomic RNA into host DNA. As this step is necessary for replication of pathogenic retroviruses such as HIV, integrase inhibitors are being developed as an important class of AIDS drugs [1–5]. Detailed structural data concerning IN-substrate interactions can contribute greatly to such efforts. Recent success in determining the structure of complexes of prototype foamy virus (PFV) IN with both the viral and target DNAs [6] has provided the foundation for a valuable HIV IN model [7]; however, experimental data for DNA complexes of HIV IN or other integrases from more closely related viruses are still lacking

[2]. On the other hand, a large number of structures have been published for the three domains of IN [N-terminal (NTD), core (CCD), and C-terminal (CTD)] from HIV, ASV and other retroviruses, either singly or in pairs [2]. Retroviral integrase is known to be a conformationally dynamic protein and current evidence indicates that it is capable of adopting a defined and/or active conformation only upon binding to its DNA substrate(s) and metal cofactors [8,9]. Several models of IN-DNA complexes have been developed, originally based on other transposase-DNA structures [10,11], and more recently on the structures of PFV IN [7,12], but the conformational variability of integrase, particularly within the inter-domain linkers, exacerbated by the significant differences in their lengths in various viruses, makes

such modeling efforts challenging. As detailed structural data on IN-DNA interactions are required to elucidate the molecular mechanism of catalysis and to facilitate drug development efforts, further studies of such complexes remain an imperative. Here we report the use of photoaffinity and chemical crosslinking methods to obtain insight into the interactions of avian sarcoma virus (ASV) IN with its DNA substrates.

Photoaffinity crosslinking and chemical crosslinking are essentially methods of measuring distances between points of interest in macromolecular complexes. By use of reagents with differing linker lengths it is possible to estimate the shortest distance between two sites on a protein or a protein complex. In photoaffinity crosslinking, a heterobifunctional reagent carries one functional group for chemical attachment to a specific target residue in a protein or nucleic acid molecule, and one photoactivatable group that can be triggered by mild UV irradiation into high reactivity, forming a covalent bond with the closest neighbor in a pre-formed complex.

Chemical crosslinking between DNA and target protein involves engineering of sulfhydryl groups into specific positions in the DNA, with the aim of forming disulfide bridges with the cysteine residues in the protein. The positioning of modified nucleotides to enable such chemical crosslinking relies on detailed knowledge of the most likely structure of the complexes. Crosslinking between two thiol groups through formation of the S-S bond serves as confirmation of their close proximity in the complex. However, if successfully prepared, chemically crosslinked protein-DNA complexes not only provide additional validation of the putative contact sites, but such complexes can be further purified in amounts sufficient for other structural studies.

Both of these approaches have been applied previously in studies of IN-DNA complexes, primarily with the HIV-1 IN protein. The feasibility of studying IN-DNA interactions using photoaffinity crosslinking was established in previous investigations in which DNA was modified with halogenated nucleoside-based photocrosslinking agents (I-dU, I-dC) [13,14] or azidophenacyl group attached to phosphorothioate-modified DNA oligonucleotides [15,16]. These studies have revealed several important features of HIV-1 IN-DNA binding. Determinants for recognition of viral DNA ends and for joining targets have been mapped to the CCD and CTD of HIV-1 IN. Most of previous studies were focused on HIV-1 IN and they were performed with crosslinking reagents attached to DNA [13–16]. The interaction sites were determined by mass-spectrometry and amino acid analysis after proteolytic digestion of the HIV-1 IN [14,15]. Because these detection methods require relatively large quantities of crosslinked material and their accuracy depends on protein composition, only crosslinks to major peptides can be detected and, in most cases, without amino acid localization.

In contrast, our experimental approach was designed to attach photoactivatable reagents at specified positions within IN for crosslinking to DNA substrates, as well as to utilize the more soluble ASV IN. Application of Cei 1 endonuclease then allowed for single nucleotide localization of the crosslinks. In one set of experiments described in this report, cysteine residues, either normally present or substituted at various positions in IN, have been used as attachment sites for carbene- and nitrene-generating photoreagents [17,18], whereas DNA was not modified beyond incorporation of radioactive markers. In the second set of experiments, a shorter, amino group-targeted carbene-generating photoreagent was attached to the positions on DNA identified in the first set, and modified DNA was crosslinked to wild type IN, in order to narrow down the most probable points of contact. Finally, in the third set of experiments sulfhydryl groups were engineered

into the identified most probable contact positions on DNA, with the aim of forming disulfide bridges with the cysteine residues in the protein. Formation of such bridges under mild conditions at high yields served as the most accurate confirmation of the discovered contacts. These results provide new information about the preferred sites of interaction within the ASV IN-DNA complex. This information is compared with published data on retroviral IN-DNA contacts obtained by the use of the same or other techniques, and the combined set has been compared with IN-DNA interactions observed in crystal structures of PFV IN-DNA complexes.

Results and Discussion

Photocrosslinking and chemical crosslinking techniques have been used in this study to map IN-DNA contacts with various substrates. Because photoactivatable reagents are fairly bulky, their introduction at or near the assumed sites of protein-DNA contact imposes a limit on distance resolution by this approach. Usually, multiple crosslinks are detected, dependent on spatial restrictions at a particular protein/DNA interface and the flexibility of the linker, on activated photocrosslinker preferences for certain chemistries of target groups, on general movements of the components of biomolecular complex, etc. To achieve higher resolution of localization of contact sites we employed three-step crosslinking. We first identified the nucleotides that were cross-linked by a long-linker photoactivatable reagent placed at selected positions in the ASV IN protein. In the second step, a short-linker photoreagent was placed at the most promising positions identified on DNA and crosslinked to IN protein for more accurate contact localization. Finally, the localization results of these two steps were refined by near-zero-length chemical crosslinking between unique cysteines on IN and unique SH-modified nucleotides on DNA substrates to confirm the positions of IN-DNA contacts.

Design of DNA substrates

In order to study different stages of the integration process, viral linear and Y-mer DNA substrates were employed to mimic the intermediate steps of processing viral DNA and joining the viral DNA substrate to host DNA. Specifically, blunt-end, unpaired end, and processed linear DNA substrates represented unprocessed, frayed, and cleaved U3 LTR viral end DNA, respectively (**Figure 1**). Y-mer substrates represent an integration intermediate in which one strand of a viral DNA end is joined to the host DNA (also known as a half-site strand-transfer intermediate) (**Figure 1**).

For the different crosslinking experiments, several modified DNA substrates were used: a) unmodified DNA, when a photoactivatable moiety was engineered into IN molecule; b) DNA with selected thymidines replaced by anchor 5-aminouridine residues for further attachment of amino-specific photocrosslinking reagent to crosslink to the IN molecule; c) DNA with selected adenosines and guanidines replaced by their corresponding 7-thio-derivatives in the mixed disulfide activated form (see **Methods S1**) for chemical crosslinking with target cysteine on the IN molecule.

In the discussion below, the nucleotide positions in both strands of the viral end substrate are numbered from the blunt end that contains the conservative CA dinucleotide preceding the scissile phosphate. This numbering is maintained in the viral end portion of the integration intermediate Y-mer substrate, so that the processed strand nucleotide that is the closest to the junction of the integration site is assigned #3. The first nucleotide position in the viral 5' overhang of the non-cleaved strand remains #1 (**Figure 1**, orange strand). For the host (target) portion of the Y-mer substrate

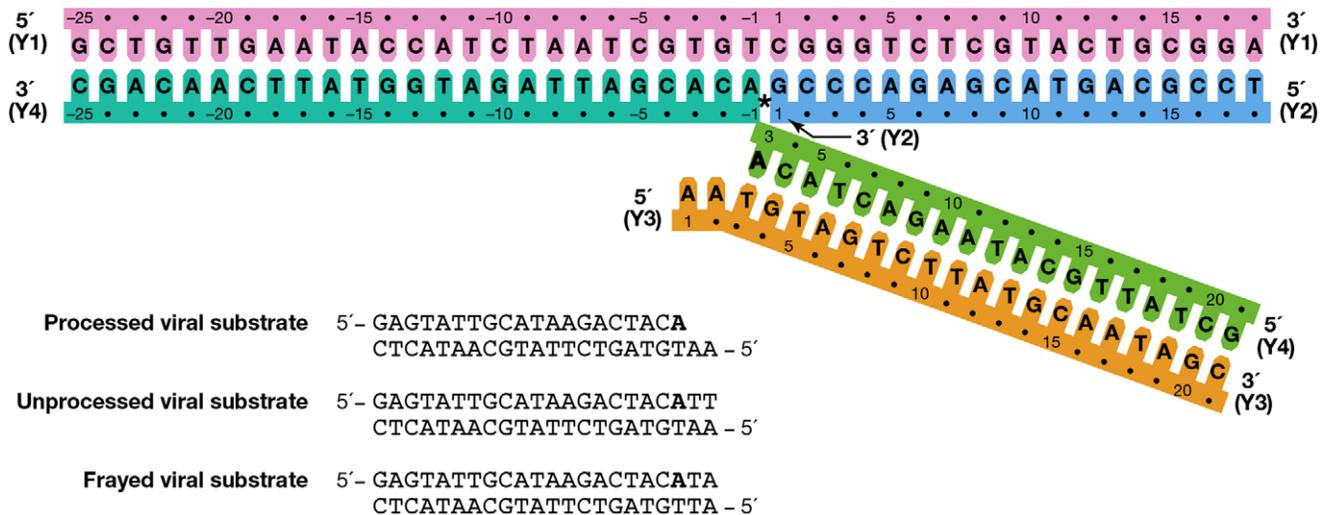


Figure 1. DNA substrates used for crosslinking to ASV IN. The indicated numbering is as used throughout the manuscript. The Y-mer strand Y4 has non-wild type bases at positions 20 and 21 in order to increase the stability of this arm of the substrate. DNA strands in the Y-mer substrate are colored for identification as in **Figures 3–6**. The host portion of strand Y4 is colored in lighter shade of blue to distinguish the host sequence from the viral sequence after covalent joining to viral DNA. In linear and Ymer DNAs, the conserved adenine preceding the scissile phosphate is shown in bold.

doi:10.1371/journal.pone.0027751.g001

the nucleotide numbering in both strands starts from the junction of the integration site (See **Figure 1**, pink and blue strands).

Design of Cys derivatives of ASV IN

Several IN derivatives with cysteine residues positioned at the putative points of contact with DNA substrates (**Figures 2, 3, 4, 5, 6**) were created by site-directed mutagenesis (**Table 1**). These cysteines were employed as “anchor” amino acids for attachment of the thiol-specific photoactivatable reagents. A single cysteine residue in the wild-type core domain of ASV IN (Cys125) was retained in some of the proteins, or replaced by serine in others. Positions 64, 124, 146, 157, and 244 were selected for substitution with cysteine, as follows:

The active site residues Asp64 and Glu157 were obvious choices for substitution with Cys due to their functional close contact with the DNA substrate. The other putative contact positions in the ASV IN-DNA complex were predicted based on crosslinking data [13,15,16,19], mutagenesis studies [20–22], and structure-based multiple sequence alignments involving analysis of superimposed 3D structures of individual and two-domain constructs of IN proteins (PDB codes 1EX4 [23], 1C6V [24], 1C0M [25], and 1K6Y [26], Mu transposase (PDB code 1BCO [10]), and the Tn5 transposase DNA complex (PDB code 1F3I [11]).

The choice to substitute Cys for Ile146, located in the flexible loop near the active site of ASV IN, was based on the results from a number of previous studies. Chemical crosslinking and photocrosslinking experiments [13,14,27–29] have suggested that the flexible loop near the IN active site is likely to make contact with the 5'-end of the non-processed viral DNA. Results of Esposito and Craigie [14] in which 5-iododeoxyuracil was placed at the 5'-end of HIV-1 non-processed DNA demonstrated a high efficiency (~10%) of photocrosslinking to residues Tyr143 and Gln148 in the flexible loop of HIV-1 IN. Q148C was also reported to chemically crosslink to thiol-modified 5'-end of viral DNA [27]. Johnson *et al.* [28] reported the formation of S-S bond between Y143C and position 2 next to 5'-end of the non-processed viral DNA. Similar experiments with murine leukemia virus (MuLV) IN [29] implicated Cys209 as another possible point of contact for the

cognate 5'-end. When aligned using the program CLUSTALW, the positions corresponding to MuLV IN residue 209 in HIV-1 and ASV IN are Ile141 and Ile146, respectively (**Figure 2**). These residues are located within the flexible loop region (amino acids 141–149), adjacent to the active site in the core domain of IN. Consequently, to establish covalent links to the end of the DNA substrate near the IN active site, we replaced Ile146 with cysteine (**Table 1**).

Although retroviral DNA can be inserted by IN into almost any site in cellular DNA, limited target site preferences have been described both *in vitro* and *in vivo*. Katzman and co-workers screened HIV-1 infected patient derived integrase sequences for amino acid changes in the catalytic core of HIV-1 IN and identified Ser119 as contributing to target site preferences [30], as assayed by integrase joining *in vitro*. These researchers were able to extend their findings to the integrases of a non-primate lentivirus Visna and the more distantly related alpharetrovirus, ASV [31]. Selection of target DNA sites is therefore likely to be a general property of the analogous residue in most retroviral integrases. Indeed, the corresponding residue in PFV is intimately involved in target DNA binding [6]. Non-conservative amino acid substitutions at this position in all three integrases exhibited a phenotype in which the processing activity was unaffected but the joining activity was significantly compromised, and it was hypothesized that this amino acid may be a critical component of the cellular DNA binding site on integrase proteins. To test this idea, we placed a photocrosslinker-anchoring cysteine residue at the analogous position in ASV IN, Ser124 (**Figure 2**).

Finally, Gao *et al.* [19] postulated that the C-terminal domain of HIV-1 IN makes most efficient contact with position +7 on the “non-cleaved” strand of viral DNA. The preferred site of interaction was identified as Glu246 of C-terminal domain of HIV-1 IN. Consequently, the corresponding ASV IN residue, Arg244, was replaced by a cysteine (**Figure 2** and **Table 1**).

Photocrosslinking of modified IN to DNA substrates

Modification of IN derivatives by photocrosslinkers. In the initial experiments, photocrosslinking to DNA substrates was

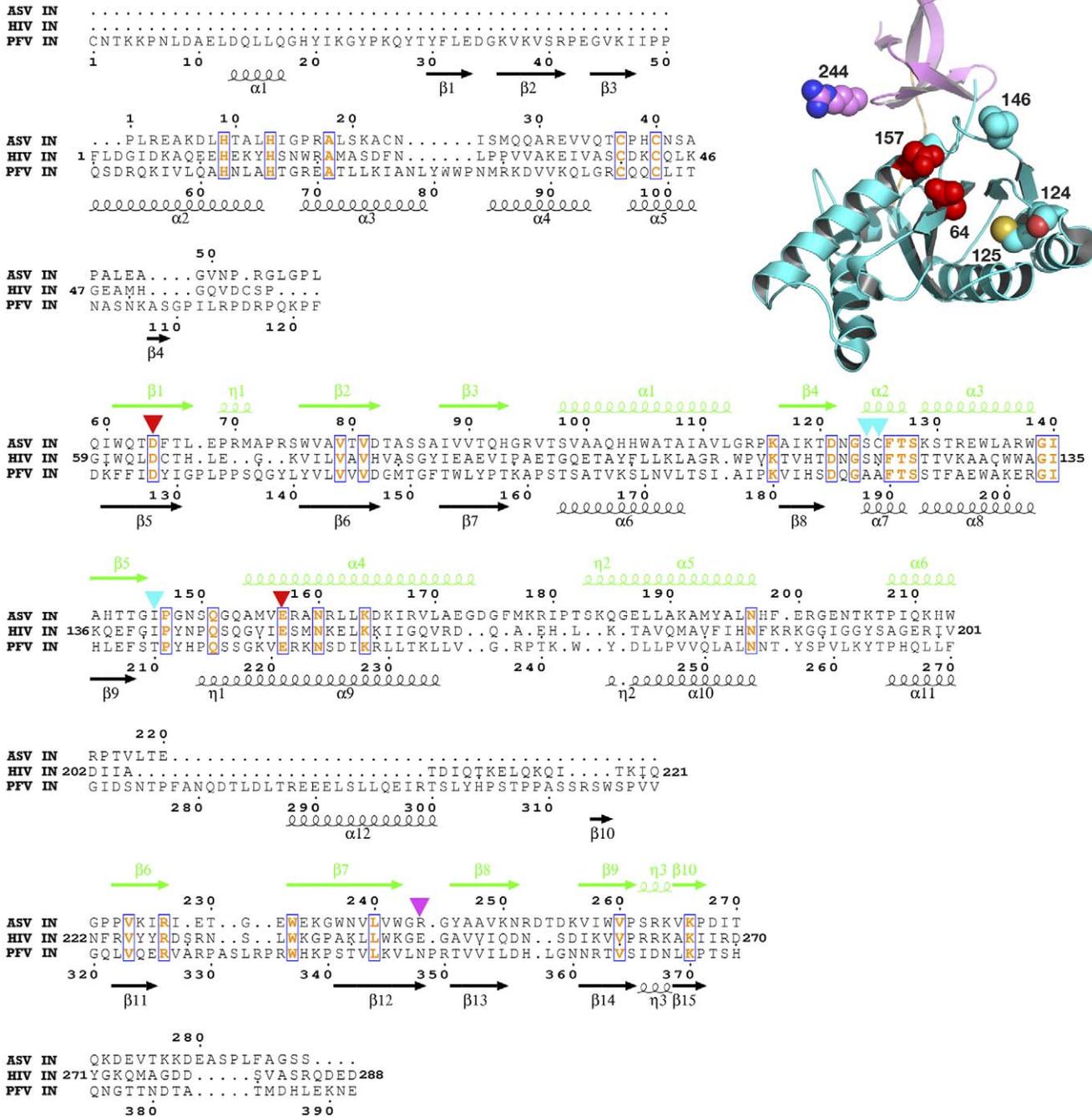


Figure 2. Structure-based sequence alignment of full-length ASV, HIV-1, and PFV IN proteins. ASV IN numbering is shown above the sequences and the structural elements are marked in green; PFV IN numbering and structural elements (black) are shown below. Numbering for HIV-1 IN is shown at the beginning and end of the lines only. The conserved amino acids, including the catalytic ASV IN residues Asp121 and Glu157, are red and boxed. Triangles mark residues that were changed to cysteines in ASV IN: red for the amino acids in the active site, cyan for other residues in the CCD, and magenta for the amino acids in the CTD. The structure of the ASV IN CCD and CTD (PDB code 1COM) with the location of the introduced cysteines is shown in the upper right corner, with the colors corresponding to the scheme described above.
doi:10.1371/journal.pone.0027751.g002

performed using wild type ASV IN and the cysteine substituted derivatives described above and listed in **Table 1**. These proteins were modified at one or two cysteine positions by coupling with photoactivatable thiol-specific compounds, either N-bromoacetyl-N'-{2,3-dihydroxy-3-[3-(3-(trifluoromethyl)diazirin-3-

yl)phenyl]propionyl}-ethylenediamine (BATDHP) [17,18], or azidophenacylthiopyridine (AFTP) [32] (see **Methods S1**).

Because the N-terminal domain (NTD) of ASV IN contains three cysteine residues that are either involved in Zn²⁺ coordination (Cys38, Cys40) or structurally important (Cys23),

| Integrase amino acids | | | | 3-D Structural models [6,7] | | | Photocrosslinking data | | | Chemical crosslinking data | | | Mutagenesis, MS, protease | Correlation with analogous residues in the PFV X-ray structure | PFV nt close | |
|-----------------------|-------|------|-------|-----------------------------|-----------------|-------|------------------------|------|----------------|----------------------------|-------|-----|---------------------------|--|---|--|
| PFV | ASV | MuLV | HIV-1 | PFV, ray | X- HIV-1, model | HIV-1 | ASV | MuLV | HIV-1 | ASV | HIV-1 | ASV | HIV-1 | | | |
| NTD | | | | | | | | | | | | | | | | |
| (L64) | (A11) | | K14 | | | | | | | | | | | #[48] | no contact in HIV-1 model | |
| (V90) | (R31) | | K34 | | | | | | | | | | | #[49] | nts 9-11 of non-cleaved viral strand are close in HIV model | |
| (51-62) | (1-9) | | 1-12 | | | | | | C2, C2- G3[15] | | | | | | no contact | |
| (51-62) | (1-9) | | 1-12 | | | | | | G1-G2[15] | | | | | | no contact | |

Nucleotides from different strands of DNA substrates are labeled by color:

Viral DNA non-cleaved strand Viral DNA cleaved strand Target DNA cleaved strand Target DNA non-cleaved strand

Figure 3. A comparison of all the IN-DNA contact points that have been determined experimentally or modeled for the NTD IN. The amino acid residues are presented in black, unless a particular residue comes from specified IN monomer in PFV intasome structure as in Refs. [6,7]. Specific residues shown to interact with DNA that are either in good correlation with the PFV structural results or do not contradict them are bolded. # -this amino acid is in contact with DNA, but the nucleotide is not determined. (G377) - The amino acid residues in parentheses indicate structural analogs to the ones implicated in DNA binding by experimental data. Nucleotides from different strands of DNA substrates are labeled by colors corresponding to the scheme used in Figure 1 and noted above. "G5[15]" - In this example and throughout Figures 3, 4, 5, 6 the nucleotide numbers correspond to the numbering scheme shown in Figure 1. The numbers in the curly brackets are as in the structure of PFV IN and the model of HIV-1 IN [6,7]. If listed, the letter designating a nucleotide comes from the original data. All reported contacts are references to original publications with numbers in brackets; our data are marked with asterisks (e.g. A3*). doi:10.1371/journal.pone.0027751.g003

| Integrase amino acids | | | | 3-D Structural models [6,7] | | | Photocrosslinking data | | | Chemical crosslinking data | | | Mutagenesis, MS, protease | Correlation with analogous residues in the PFV X-ray structure | PFV nt close | | |
|-----------------------|------------|------|-----------|-----------------------------|-----------------|-------|------------------------|--------|-------|----------------------------|-------|-----|---------------------------|--|---|---|---------|
| PFV | ASV | MuLV | HIV-1 | PFV, ray | X- HIV-1, model | HIV-1 | ASV | MuLV | HIV-1 | ASV | HIV-1 | ASV | HIV-1 | | | | |
| CCD | | | | | | | | | | | | | | | | | |
| R114 (D121) | (N51) D64 | | V54 (D64) | T5 | G4 | | | | | | | | | #[21] | do not contradict | | |
| (Y141) | (W76) | | V72 | | | | | | | | | | | #[21] | good correlation, contact via Mg ²⁺ , 4A Y141 (PFV) faces minor groove of tDNA, 4.1A, OK | 2(5) | |
| (K180) | (K116) | | K111 | | | | | | | | | | | #[49] | K180 (PFV) to bb of n2 - 9.6Å (do not contradict) | 2 | |
| (A188) | S124 | | (S119) | 3(-3) | | | 3, 8* | | | | | | | 3[31] | good correlation | | |
| (A188) | S124 | | (S119) | 3,4 (6,7) | | | | | | | | | | | good correlation, since Cel1 cuts both strands | | |
| (S193) | (K129) | | T124 | 4-5 (7-8) | | | | | | | | | | #[21] | 3.2 Å to phosphate; good correlation | | |
| (T194) | (S130) | | T125 | | | | | | | | | | | #[21] | 5.2 Å to phosphate; (does not contradict) | 4-5 (7-8) | |
| (H205) | (A141) | | K136 | | | | | | | | | | | #[49] | 4-5 Å to nucleotide 1-2 | 1-2 | |
| E207 | (T143) | | E138 | T2, T3 | C2, T3 | | | | | | | | | #[49] | do not contradict | | |
| T210 | I146 | C209 | I141 | T3 | T3 | | A3* | C1[29] | | | | A3* | C1[29] | | good correlation, ASV-different conformation | | |
| T210 | I146 | C209 | I141 | | | | | | | | | | | | good correlation, ASV-different conformation | | |
| (Y212) | (G148) | | Y143 | 1-3(-1--3) | | | A1[14] | | | | | | | | OH- of Y212 (PFV) - 2.9Å to bb n2-n3, NH - 2.8 Å to target bb n1-n2; | | |
| (S217) | (Q153) | | Q148 | | | | A1[14] | | | | | | | | no target DNA [14,27,28], substrate was blunt | | |
| E221 | E157 | | E152 | A3, A4 (17,16) | | | | | | | | | | A3* | v1-11[51] good correlation, contact via Mg ²⁺ , 3.8Å | | |
| R222 | (R158) | | S153 | | G5 (15) | | | | | | | | | | v1-11[51] do not contradict | | |
| R222 | (R158) | | S153 | T5, C6 | G4, C5 | | | | | | | | | | do not contradict | | |
| (N224) | (N160) | | N155 | A3, A4 (17,16) | | | | | | | | | | | 2.8Å to bb on rotamer of N224 (PFV), do not contradict | | |
| (S225) | (R161) | | K156 | | G5 (15) | | | | | | | | | | v1-11[51], v1-11[51] | 4.1Å tC-C, N224(PFV), do not contradict | A4 (16) |
| (D226) | (L162) | | E157 | | | | | | | | | | | | do not contradict | A7 | |
| K228 | (K164) | | K159 | A3(17) | A3 (17) | | A3[13] | | | | | | | | do not contradict | | |
| K228 | (K164) | | K159 | -1,-2 (0,1) | | | | | | | | | | | 3.3Å to bb, good agreement | | |
| R229 | (D165) | | K160 | T8 | G8 | | | | | | | | | A1[27] | do not contradict [21,48,49], but no correlation for x-linking | | |
| (230, 231) | (166, 167) | | 161, 162 | | | | | | | | | | | | v1-11[51], #[21] no contact | | |
| (K233) | (V169) | | Q164 | | | | | | | | | | | | 4.5 Å from K233(PFV) to bb of T8, no correlation | T8 | |
| (241,242) | (180,181) | | 171, 172 | | | | | | | | | | | | #[21] no contact | | |
| (N255) | (H198) | | K185 | A7, T8 | | | | | | | | | | | loop range do not contradict | A7, T8 | |
| (T256) | (F199) | | K186 | A7, T8 | | | | | | | | | | | loop range do not contradict | A7, T8 | |
| (Y257) | (E200) | | K188 | A7, T8 | | | | | | | | | | | S258 (PFV) hb 2.6Å to bb7 | A7 | |
| (V260) | (N204) | | I191 | | | | | | | | | | | A1, A7 [19] | 4.6Å to bb n7 in PFV, good correlation | -A7 | |
| K262 | (T205) | | G193 | T11(9) | | | | | | | | | | | do not contradict | | |
| (266,270) | (209,213) | | 197, 201 | | | | | | | | | | | | #[21,48] no contact | | |
| (301-317) | | | 207-219 | T7(13), C9(11) | T3, T6 | | | | | | | | | | #[48,49] linker region, OK correlation | | |

Figure 4. A comparison of all the IN-DNA contact points that have been determined experimentally or modeled for the CCD IN. For details, see legend to Figure 3. doi:10.1371/journal.pone.0027751.g004

| Integrase amino acids | | | | 3-D Structural models [6,7] | | | Photocrosslinking data | | | Chemical crosslinking data | | Mutagenesis, MS, protease | Correlation with analogous residues in the PFV X-ray structure | PFV nt close |
|-----------------------|-----------|------|---------|-----------------------------|----------|--------------|------------------------|-----|------|----------------------------|-----|---------------------------|---|---------------------------------|
| PFV | ASV | MuLV | HIV-1 | PFV, ray | X- model | HIV-1, model | HIV-1 | ASV | MuLV | HIV-1 | ASV | HIV-1 | | |
| CCD | | | | | | | | | | | | | | |
| (105-134) | (46-69) | | 49-69 | | | | C4-G5 [15] | | | | | | E69 (HIV-1), good correlation | |
| (105-134) | (46-69) | | 49-69 | | | | C4 [15] | | | | | | Linker region, H51(HIV-1), 10A, good correlation due to tether length | |
| (105-134) | (46-69) | | 49-69 | | | | A1 [15] | | | | | | P52 8A, I60 9A, D55(HIV-1), good correlation due to tether length | |
| (105-134) | (46-69) | | 49-69 | | | | G4-C5 [15] | | | | | | P52 8A, I60 9A, Q62(HIV-1), good correlation due to tether length | |
| (105-134) | (46-69) | | 49-69 | | | | C-2 - G-3 [15] | | | | | | H67-E69(HIV-1), good correlation | |
| (105-134) | (46-69) | | 49-69 | | | | G1-G2 [15] | | | | | | C65, Q62(HIV-1), good correlation | |
| (107-128) | (48-64) | | 51-64 | | | | C2[14] | | | | | | linker region, OK correlation | |
| (114-117) | (51-53) | | 54-57 | | | | | | | | | #[21] | do not contradict | |
| (197-199) | (133-135) | | 128-130 | | | | | | | | | #[21] | no contact | |
| (209-221) | (145-157) | | 139-152 | | | | C4[15] | | | | | | good correlation | |
| (209-221) | (145-157) | | 139-152 | | | | A6-A7 [15] | | | | | | no contact | |
| (209-221) | (145-157) | | 139-152 | | | | G1-G2 [15] | | | | | | no contact | |
| (209-221) | (145-157) | | 139-152 | | | | G4-C5 [15] | | | | | | S217(PFV), Q153(ASV), Q148(HIV-1), OK correlation | |
| (209-221) | (145-157) | | 139-152 | | | | C2-G3 [15] | | | | | | P211 and Y212(PFV), Q143 (ASV) | |
| (227-267) | (163-210) | | 158-198 | | | | A3[13] | | | | | | good correlation | |
| (232-234) | (168-170) | | 163-165 | | | | | | | | | #[21] | no contact | |
| | | | CCD | | | | A3[14] | | | | | | 3,4 [14] | good correlation, many contacts |
| | | | CCD | | | | | | | | | | 11,12,13 [14] | 207-209 (HIV-1) close to 12,13 |
| | | | CCD | | | | 1,2[14] | | | | | | 2 [14] | 207-209 (HIV-1) close to 12,13 |

Figure 5. A comparison of all the IN-DNA contact points that have been determined experimentally or modeled for the CCD IN (continued). For details, see legend to Figure 3. doi:10.1371/journal.pone.0027751.g005

reaction conditions were found that favored modification of the newly introduced solvent accessible cysteines and left those in the NTD unmodified. These conditions were found empirically by varying pH, temperature, and time of the reaction. The presence of the photocrosslinking moiety at selected positions and its absence in the NTD was confirmed by MALDI-TOF mass-

spectroscopy of tryptic peptides obtained from IN-DNA adducts excised from gels (data not shown).

Apart from the activity changes due to Cys substitutions, the introduction of photocrosslinkers did not result in significant changes in protein function, as measured by comparison of the enzymatic activities of the modified and non-modified IN proteins,

| Integrase amino acids | | | | 3-D Structural models [6,7] | | | Photocrosslinking data | | | Chemical crosslinking data | | Mutagenesis, MS, protease | Correlation with analogous residues in the PFV X-ray structure | PFV nt close |
|-----------------------|---------|------|---------|-----------------------------|----------|--------------|------------------------|-----|----------|----------------------------|------|---------------------------|--|--------------|
| PFV | ASV | MuLV | HIV-1 | PFV, ray | X- model | HIV-1, model | HIV-1 | ASV | MuLV | HIV-1 | ASV | HIV-1 | | |
| CTD | | | | | | | | | | | | | | |
| (S332) | (E229) | | S230 | | | | | | | A1, A7 [19] | | | S332 (PFV) close to junction host n(-2) | |
| R336 | (E232) | | L234 | C4[16] | | | | | | | | #[55] | do not contradict | |
| (V346) | (W242) | | K244 | | | | | | | | | #[48] | no contact | |
| N348 | R244 | | E246 | T2 | T3 | | | | T11-A12* | A7[19] | A12* | #[49] | contact not at the same position | |
| N348 | R244 | | E246 | | | | | | A10-T12* | | A10* | | contact not at the same position | |
| (N366) | (S262) | | R262 | G4 | | | | | | | | #[21,48,55] | do not contradict | |
| (D367) | (R263) | | R263 | G4 | | | | | | | | #[21] | viral DNA 1, do not contradict | |
| (N368) | (K264) | | K264 | | | | | | | 6-7[50] | | #[48] | no contact | |
| (K370) | (K266) | | K266 | | | G10 | | | | 6-7[50] | | | viral DNA 2, no contact in PFV | |
| S373 | (I269) | | R269 | G10 | G10 | | | | | | | #[48,49] | viral DNA 2, do not contradict | |
| (G377) | (D273) | | K273 | | | | | | | | | #[49] | 8-9A | |
| 307-348 | 221-244 | | 213-246 | | | | 6-7[15] | | | | | | no contact; R329 (PFV) -n1 cleaved, (2.8A) and n4 non-cleaved | |
| 307-348 | 221-244 | | 213-246 | | | | 1-2[15] | | | | | | no contact | |
| 350-374 | 245-270 | | 247-270 | | | | 77[14] | | | | | | not close to base | |
| 360-374 | 245-270 | | 247-270 | | | | 6-7[15] | | | | | | 263, 266 (HIV-1) may be OK | |
| 375-392 | 271-289 | | 271-288 | | | | 6-7[15] | | | | | | no contact | |
| 375-392 | 271-289 | | 271-288 | | | | 4[15] | | | | | | residues 375-392 are not resolved in PFV structure | |
| 375-392 | 271-289 | | 271-288 | | | | 1[15] | | | | | | residues 375-392 are not resolved in PFV structure | |
| 375-392 | 271-289 | | 271-288 | | | | 2-3[15] | | | | | | residues 375-392 are not resolved in PFV structure | |
| 375-392 | 271-289 | | 271-288 | | | | -2 [15] | | | | | | residues 375-392 are not resolved in PFV structure | |
| | | | CTD | | | | T6[14] | | | | | | T6[14] | no contact |

Figure 6. A comparison of all the IN-DNA contact points that have been determined experimentally or modeled for the CTD IN. For details, see legend to Figure 3. doi:10.1371/journal.pone.0027751.g006

Table 1. Substitutions in the ASV IN derivatives and their enzymatic activities (crosslinkers placed at residues that are bold).

| Protein | % WT processing ¹ | % WT joining ² |
|---|------------------------------|---------------------------|
| Wild-type | 100 | 100 |
| F199K | 238 | 102 |
| C125S, I146C , F199K | 80 | 72 |
| C125S, F199K, R244C | <1 | <1 |
| F199K, R244C | <5–12 | <1 |
| C125S, I146C , F199K, R244C | 4 | 2 |
| I146C , F199K, R244C | <1 | <1 |
| D64C , F199K | 4 | 20 |
| E157C , F199K | 6 | 1 |
| C23S, C125S, E157C F199K | ND | <4 |
| C23S, C125S, E157C F199K, W259A | <4 | <4 |
| S124C , C125S, F199K | 38 | 16 |

The dominant Cys for crosslinking is **bolded**.

¹Processing of viral ends as assayed either by a fluorescence anisotropy release assay or a radioactively labeled processing assay in 10 mM Mg²⁺ for 30 min.

²Single end joining assessed by a fluorescent joining assay [33].

R244C substitution severely compromises activity in all IN derivatives that contain it.

doi:10.1371/journal.pone.0027751.t001

using a standard disintegration assay [33] (**Figure S1A** – schematics; **Figure S1B** – results).

The IN-DNA complexes prepared from the modified IN derivatives and oligonucleotide substrates were irradiated with long-wavelength UV-light, to activate the photocrosslinkers and produce covalent links between the IN protein and DNA. The products were then separated by gel electrophoresis, visualized, and quantified with a PhosphorImager (see Photocrosslinking in Materials and Methods).

Procedures for localization of photocrosslinks

To detect the preferred sites of IN photocrosslinking on the DNA substrates, a photoactivatable reagent with a *cis*-diol bond in the linker, BATDHP, was specifically cleaved by mild periodate treatment. This resulted in the transfer of the bulky aromatic part of the photocrosslinking reagent from IN to the DNA, thereby producing a cleavage site for the endonuclease Ccl1 (Transgenomics, Inc. [34–36]), which cuts double-stranded DNA at mispaired bases or sites with bulky nucleotide adducts essentially reporting their location (**Figure 7A**). Because Ccl1 cleaves both DNA strands at such sites, a separate set of experiments was performed in which each of the four strands that constituted the Y-mer (labeled 1–4 as in **Figure 1**) were separately labeled with ³²P at the 5'-end. The crosslinked strands were then identified after denaturing polyacrylamide gel electrophoresis by their reduced mobility due to the covalent attachment of IN (**Figure S2**). As Ccl1 occupies about a 10-bp stretch of the DNA substrate and requires substrates longer than 20 bp to process consistently, this approach was only used for detection of photocrosslinks on the Y-mer DNA and not on the shorter linear substrates. Because most of the crosslinks to the Y-mer DNA mapped to the viral portion, these results were combined with crosslink locations identified in the linear substrates and, together with the data published by others, were used to select positions on linear DNA substrates for placement of photocrosslinking reagents and chemical crosslinking moieties.

Proximities identified from crosslinking IN residues to the DNA substrates

WT IN (containing Cys125 as the only cysteine available for modification with photocrosslinking reagents) was used as a negative control in all our crosslinking experiments. As that position was not expected to be in contact with the DNA substrate, it is not surprising that no significant photocrosslinking was observed with WT IN (see **Fig. S2** and **Results S1**). Strand 4 on the Y-mer (see also **Figure 1**, green strand) was found to be the most likely target for crosslinking for modified IN derivatives with Cys residues at positions 146, 244, and 146 plus 244 (**Figures S2, S3** and **Results S1**). This strand of DNA is analogous to the newly joined viral DNA strand. Photocrosslinking from Cys124 resulted in covalent binding to the host portion of the Y-mer substrate, specifically 3 and 8 nucleotides away from the integration junction (**Figure 7C–E**). These IN-DNA contacts are in good agreement with the suggested role of ASV IN residue Ser124 in host site binding/selectivity [31].

Photocrosslinking from Cys146 resulted in covalent binding to the viral portion of strand Y4 (**Figure 7B–E**), primarily one nucleotide to the 5' side of the scissile phosphate (position 3). Analysis of phenol/chloroform-separated covalent complexes of IN-DNA also showed interactions at position 3 of this strand in a linear substrate (L3). Cell cleavage of photocrosslinked products obtained with the Cys244 derivative uncovered a range of sites predominantly around positions 9–12 in Y4; 7–10 and 12 in Y3 (**Figure 7C–E**). Such variability may be due to mobility of the CTD. The results of these and additional experiments with ASV IN derivatives are summarized in **Table 2**.

Photocrosslinking from specific nucleotides in linear DNA substrates to IN

In order to refine IN-DNA contact localization data for the CTD, we attached a photoactivatable reagent with a shorter linker to selected nucleotides on linear substrates for crosslinking to IN. Three different synthetic DNA substrates were designed with amino-modified nucleotides introduced in positions 8 and 11 of strand L3 and position 12 of strand L4 (see Materials and methods). The amino groups served as specific anchors for DNA modification with the NHS ester of the carbene-generating (N-hydroxysuccinimidyl-3-[3-(trifluoromethyl)diazirin-3-yl]benzoate). The resulting modified DNA oligonucleotides were labeled with ³²P and annealed to the corresponding complementary oligonucleotide to form 22 bp linear DNA substrates. The highest efficiency of crosslinking to WT IN was found for position 11 on strand L3 and position 12 on strand L4. Efficiency of crosslinking from position 8 of strand L3 was less than half of that for position 11 on strand L3 (**Figure S4**). These data show that these positions are in close contact with IN and together with the results from the previous experiment suggest a contact between nucleotides at positions 11-L3 and 12-L4 of the linear substrate and IN position 244 in the CTD.

Chemical crosslinking of modified DNA substrates to residues near the active center of ASV IN

Mixed disulfide chemical crosslinking has been used previously to locate points of contact between HIV-1 reverse transcriptase and DNA with better accuracy and to obtain preparative quantities of tethered RT-DNA complexes [37]. The information derived from our photocrosslinking experiments was used to incorporate mixed disulfide-activated thiol-containing nucleotide derivatives at specific positions of synthetic 22 bp DNA oligonucleotides, representing the U3 viral end (see Materials and

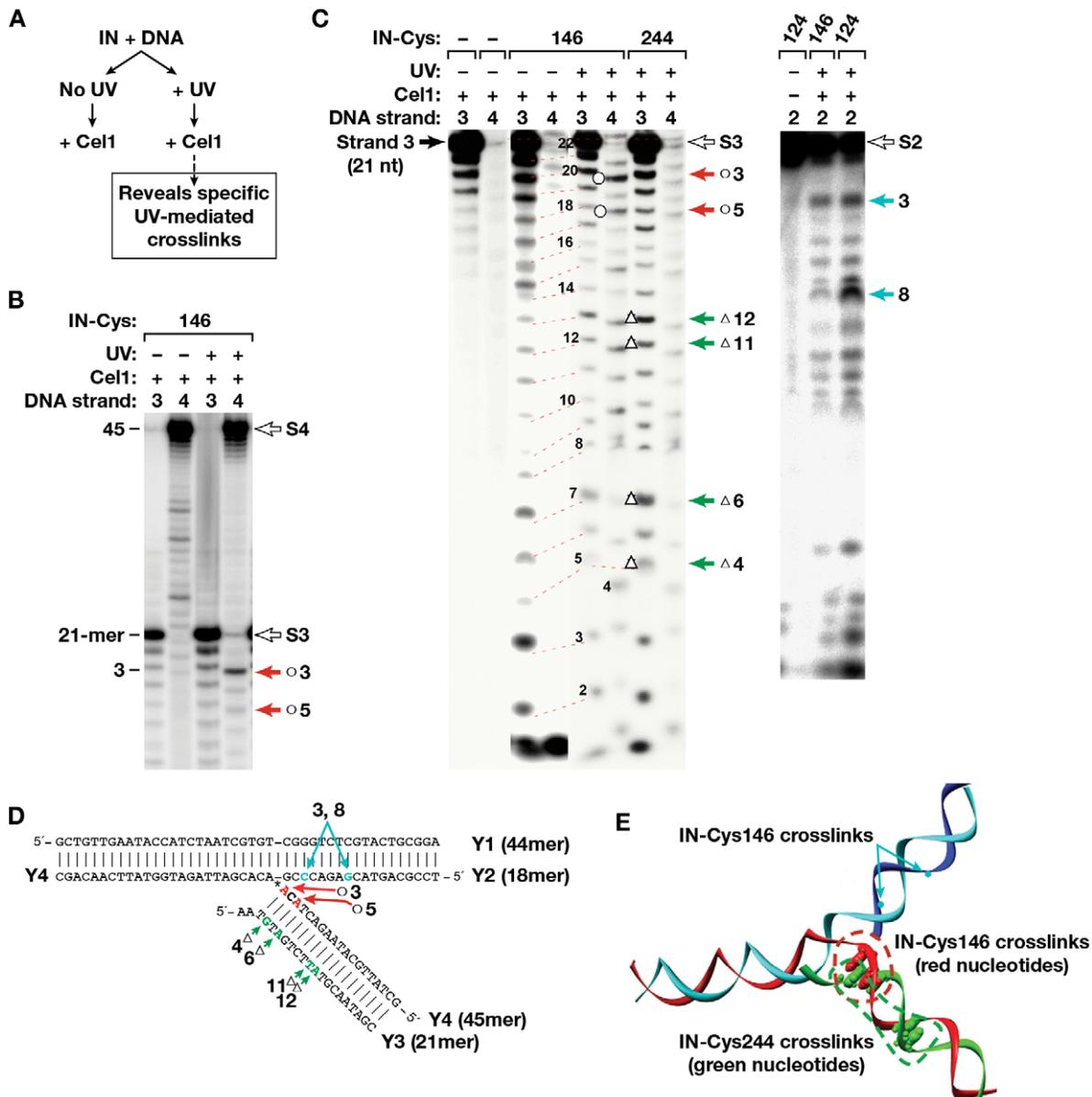


Figure 7. Cel 1-based localization of the crosslinking sites on the Y-mer DNA. (A) The method for detecting specific UV-mediated crosslinks in Ymer DNA is outlined. (B) Cel 1 cleavage of the photocrosslinked complex of IN I146C. Products from Y-mer DNAs labeled at the 5'-end of strand 3 or 4 (marked above each lane) are shown. The filled arrows point to prominent Cel 1 products, indicative of a bulky adduct at the conserved viral CA dinucleotide in strand 4. Open arrows mark the position of non-cleaved substrate strands. (C) Cel 1 cleavage of various photocrosslinked complexes of Cys-modified derivatives of IN with Y-mer DNA labeled at strands 2, 3, or 4. Numbers above the gels indicate which DNA strand in the Y-mer was labeled. Open arrows mark the position of non-cleaved substrate strands. Numbers to the left of the gel indicate length in nucleotides, and arrows to the right mark the positions of adducts of IN with DNA. In both B and C, products were separated by denaturing gel electrophoresis and then visualized with a PhosphorImager. (D) Y-mer DNA sequence with positions of preferred crosslinking detected by Cel1 indicated for each IN derivative by red (I146C), green (R244C) and teal (S124C) arrows; (E) 3-D model of the Y-mer DNA with positions of preferred crosslinking detected by Cel1 indicated for each IN derivative by green (C244), red (C146) and teal (C124) dots.

doi:10.1371/journal.pone.0027751.g007

Methods). In addition, the 5'-end on the non-cleaved strand of viral DNA was chosen for S-S chemical crosslinking because various lines of evidence have indicated that its binding increases the stability of IN-DNA complex [38–40]. Double-stranded Y-mer and linear DNA substrates prepared with these oligonucleotides were subjected to chemical crosslinking with each of the cysteine derivatives of ASV IN (Table 1). As previous results have indicated that the viral end binding is facilitated by the “breathing” (or fraying) that normally occurs preferentially at

DNA termini [41–44] most of the linear substrates for S-S crosslinking were prepared with “frayed” ends (Figure 1).

Because IN-DNA binding efficiency differs from one IN derivative to another, the crosslinking data can be interpreted only by comparing the crosslinking yields with substrates modified at different nucleotide positions (also see Results S1). All analytical experiments were carried out in physiological buffers, at low IN concentrations, with the IN:DNA ratio reflecting theoretical stoichiometry (2:1 for linear DNA and 4:1 for Y-mer

Table 2. Preferred position for photocrosslinking of modified IN Cys derivatives to a Y-mer DNA substrate.

| ASV IN derivative and DNA type | Cys146 | | Cys146,244 | | Cys244 | | Cys124 | |
|--|-------------|------------------|-----------------|----------|------------------|------------------|--------|---------------|
| | Y4 | Y3 | Y4 | Y3 | Y4 | Y3 | Y1 | Y2 |
| Untreated crosslinking reaction | 5, 3 | 12 | 5, 3 ,1t | 3 | 10,11 ,1t | 6, 12 ,16 | | 3t ,8t |
| NaIO₄- treated crosslinking reaction | 3 | 3 ,11,12, | | 3 | 3–5, 1t | 8–10 , 18 | | 3t ,8t |
| Phenol fraction treated with NaIO₄ | 9, 3 | 3 , 7, 8 | 11 | | 9–12 | 1–4, 7 | 4t* | 3t ,8t |

Bold letters stand for the most prominent contacts revealed by Cel 1 endonuclease analysis, *numbers followed by "t" represent positions in the target DNA portion of the Y-mer substrate.
doi:10.1371/journal.pone.0027751.t002

DNA). The results of these experiments are summarized in **Table 3**. Representative data are provided in **Figure S5**. Following are our observations with respect to each of the relevant Cys-substituted IN derivatives:

Cys125. Only low amounts of crosslinking were seen with linear and Y-mer substrates containing a thiol modification near the 5' nucleotide on strand 3 and the opposite nucleotides on strand 4 (**Figure 1**, orange and green strands, **Figure S5**, **Table 3**). These data were consistent with our photocrosslinking results suggesting no direct contact between Cys125 and viral DNA. While small, the Cys125 contribution to IN-DNA crosslinking was still taken into account in all other chemical crosslinking experiments where Cys125 remained intact.

Cys146. The most prominent contact with Cys146 (residing in the flexible loop near the active site in core domain) was observed at the 3'-end nucleotide of the strand L4 (**Table 3**). Significant crosslinking was also detected at positions 1 and 2 of strands L3 and Y3 (**Figure S5** and **Table 3**). These data are in good agreement with our photocrosslinking results and with previously reported involvement of the flexible loop with the viral end of DNA close to scissile phosphate [13,14,27–29].

Cys244. The C-terminal domain Cys244 was found to crosslink with the viral end of DNA at positions 10 of strand 4 or position 12 of strand 3 in both linear and Y-mer oligonucleotides, in agreement with our photocrosslinking data (**Figure S5**, **Table 3**). These contact positions differ from the chemical crosslinking results [19] that placed the homologous amino acid residue 246 of HIV-1 IN in contact with position 7 of the non-cleaved strand of viral DNA. This discrepancy could be attributed to the significant differences in the lengths of the linker regions between the CCD and CTD in HIV-1 IN (18 aa) and ASV

IN (7 aa), compared to that in PFV IN (48 aa), possibly resulting in different relative positioning of their CTDs in an intasome.

Chemical crosslinking of modified DNA substrates to catalytic residues in ASV IN

In order to find the best approach for producing stable IN-DNA complexes for structural studies, we compared the crosslinking efficiencies of several full-length ASV IN derivatives carrying Cys substitutions in the active site, including the metal cofactor binding residues Asp64 or Glu157, and the Cys already present at position 125. The same substitutions were introduced into the core domain which was then expressed separately. In some constructs the Cys125 was substituted with serine, and a W259A substitution was included. The W259A replacement has been shown to block formation of ASV IN dimers [45].

The 22-mer dsDNA substrates used in these experiments were designed to represent the processed (recessed) U3 portion of the viral genome and contained modified adenosine in the 3' position of the processed strand (**Figure 1**). One modified adenosine contained 3-mercaptopropanol phosphodiester at 3' position of the 3'-terminal deoxyribose; in a second substrate the same deoxyribose was substituted by N-mercaptopropanol derivative of morpholine. To find the optimal position for the thiol group in the nucleotide, the structure of TN5 transposase complexed with Tn5 transposon end DNA (PDB code 1MUH) was used as a reference. Superposition of the active sites of TN5 and core domain of ASV IN (PDB code 1VSH) allowed modeling of the 3'-end nucleotide in the active site of ASV IN (**Figure 8**). Both modified oligonucleotides were designed to present their thiols for direct interaction with a Cys residue introduced in the active site of the ASV IN at the positions of the catalytic residues, Asp64 and Glu157.

Table 3. A comparative summary of IN-DNA S-S crosslinking with mixed disulfide-modified substrates.

| Substrate | L4-3 | L4-10 | L4-12 | L3-2 | Y3-2 | L3-12 | Y3-12 | L4-3 processed |
|---------------------|-----------|-------|-----------|-----------|-----------|-------|-------------|----------------|
| Cys position | | | | | | | | |
| 146 | ++ | | | | + | +/- | | |
| 146, 244 | +/- | +/- | ++ | ++ | ++ | | | |
| 244 | +/- | +/- | | | | +d* | ++d* | |
| 125 | + | | | + | + | | | |
| 125, 157 | ++ | | +/- | | + | + | | ++ |

The efficiency of crosslinking is shown by a number of (+) signs. The most efficient contact sites are shown in bold (++) . The DNA substrates used are shown in the top row; d* indicates preference for dimer formation. Linear(L) and Y-mer(Y) dsDNA substrates with thiol groups introduced in the bases of certain strand (e.g. L4 carries a modification in strand 4 (processed strand of viral end) at certain position (e.g. L3-12 is modified at 12th nucleotide from 5'-end of the non-processed strand, or L4-3 – at 3rd position (A) of processed strand that is next to scissile phosphate.

doi:10.1371/journal.pone.0027751.t003

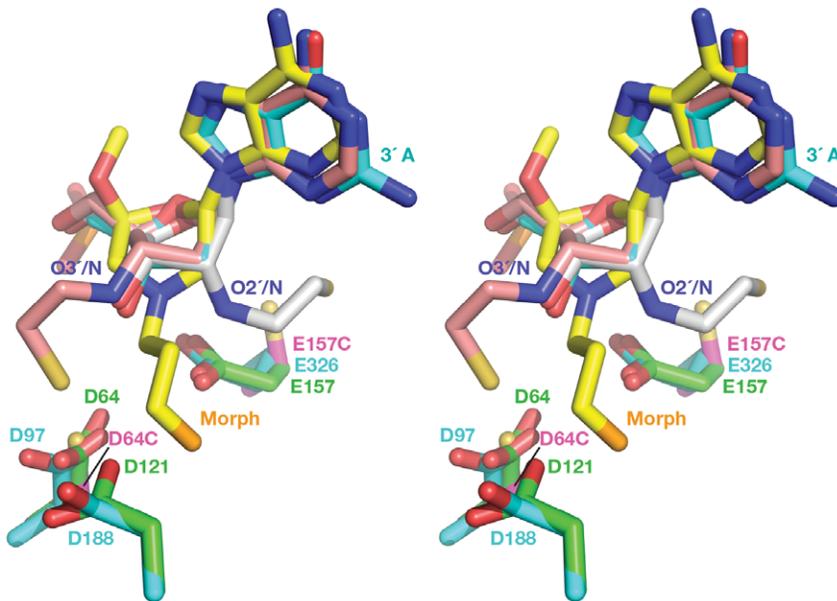


Figure 8. Design of modified 3'-end adenine. A nucleotide at the 3'-end of DNA in TN5 transposase structure is shown in blue, morpholino adenosine analog in yellow, nucleotides with modifications on the C3' and on C2' of ribose, are shown in pink and gray respectively. Catalytic residues are shown in blue for transposase and in green for ASV IN. Substituted catalytic residues D64C and E157C are shown in magenta. doi:10.1371/journal.pone.0027751.g008

The most efficient crosslinking to E157C and D64C was observed in the presence of 10 mM $MgCl_2$, indicating that, in contrast to other IN-DNA contact sites, crosslinking to these derivatives required the presence of Mg^{2+} (data not shown). The fact that the E157C IN construct is capable of binding a metal cation suggests that the ion binds in site I (Asp64–Asp121), as seen in previous structures of IN with a disordered region encompassing the Glu157 residue [46,47]. We also showed in previous experiments with a D64N derivative that Asp121 alone can bind a single Zn^{2+} cation in site I [47]. It is therefore quite plausible that the D64C derivative could likewise coordinate Mg^{2+} with Asp121 in site I alone (or perhaps with Glu157 in site II) in the presence of additional contacts with DNA. Such interactions, in turn, could stabilize the DNA-IN complex at the active site. A crystal structure

of an ASV IN-DNA complex is required to confirm this hypothesis.

All active site substituted derivatives were subjected to pH-dependent and DTNB-mediated protocols to promote formation of S-S bonds with the DNA substrates, and the results are summarized in **Figure 9**. For experiments performed with the full length E157C IN, the highest yields were observed with the 3'-attached 3-mercaptopropanol phosphodiester modified substrates (P-SS), similar for both pH and DTNB activation. The C23S/C125S/E157C/F199K IN derivative (**Figure 9A**) produced higher yields of crosslinking than the single E157C IN derivative (not shown) with both modified DNA substrates, regardless of the activation method (pH-dependent or DTNB). Crosslinking of the C23S/C125S/E157C/F199K/W259A IN derivative with both

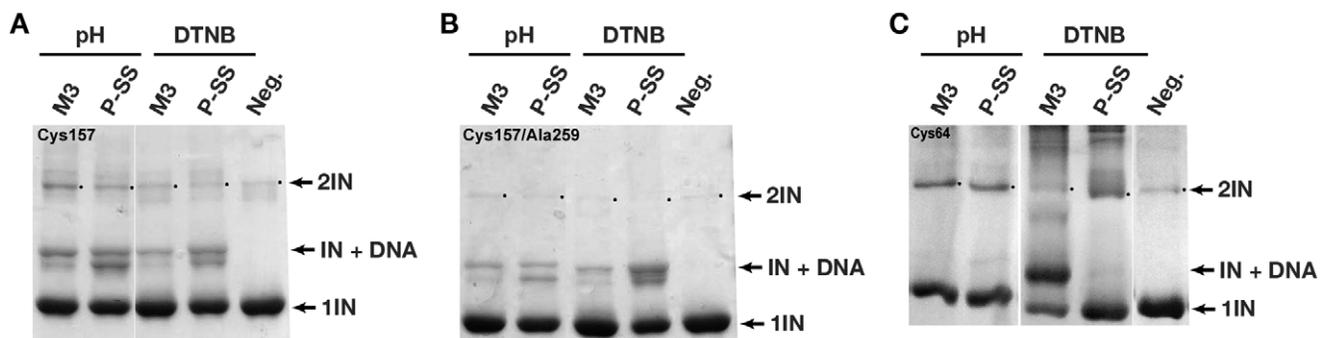


Figure 9. S-S crosslinking of the ASV IN active site derivatives to modified linear dsDNA substrates. Sample Coomassie-stained polyacrylamide gels with IN-DNA crosslinks. pH indicates pH-induced crosslinking; DTNB, DTNB-mediated crosslinking; lanes labeled P-SS correspond to crosslinking to DNA carrying thiol modification on the 3' phosphate, lanes labeled M3 correspond to crosslinking to DNA carrying thiol modification on the morpholino adenosine analog. The lane marked Neg represents samples with no DNA. 2IN, IN+DNA and 1IN designate dimeric IN bands, adduct bands and monomeric IN bands, respectively. Molecular weight marker lanes are not shown as IN-DNA the monomer (lower) and dimer (upper) bands provide internal calibration. The figure shows the results only for full length INs. Panel A shows crosslinking to the C23S/C125S/E157C/F199K ASV IN derivative (labeled Cys157 after the key mutation); panel B to C23S/C125S/E157C/F199K/W259A (labeled Cys157/Ala259), and panel C to D64C/F199K (labeled Cys64). doi:10.1371/journal.pone.0027751.g009

modified DNA substrates using the pH activation method produced slightly lower yields than crosslinking of the C23S/C125S/E157C/F199K IN derivative (Compare **Figure 9A and 9B** pH-P-SS lanes), and no adduct band was observed above the position of dimeric IN in **Figure 9B**. Protein migrating at the 2IN position and weak bands above this on SDS PAGE represent covalently linked IN dimers and IN dimers linked to DNA, respectively. Because the W259A substitution has been shown to impair dimer formation [45], this result was anticipated. However, even if the majority of IN was dimeric in complex with DNA (i.e. W259 containing protein in **Figure 9A**), the predominant adduct band is expected to migrate in an SDS gel as a monomer+DNA adduct, as crosslinks between IN proteins are unlikely with this experimental design.

After the structure of the PFV intasome became available we verified that the position of the 3' nucleotide in the active site of TN5 transposase is similar to its counterpart in PFV IN. Although the orientation of the 3'-end nucleotide is slightly different in PFV IN, the presence of the flexible linkers carrying thiol groups is likely to have allowed successful crosslinking of both modified nucleotides to ASV IN D64C and E157C derivatives. The requirement for metal ions for the efficient crosslinking of Cys derivatives to substrates containing thiol at the 3'-end of the processed strand indicates that binding to the viral DNA substrate is preserved upon replacement of one of the catalytic residues of ASV IN with Cys.

Rationalization of the crosslinking data in the context of currently available structural information

Photocrosslinking and chemical crosslinking data available to date, combined with results presented in this study, were compared with the interactions observed in the recently solved structures of the PFV intasome. In order to identify corresponding residues, a structure-based sequence alignment of ASV IN, HIV-1 IN, and PFV IN was created by superimposing the coordinates of the individual domains of the ASV and HIV-1 INs on the structure of full-length PFV IN complexed with the viral and target DNAs (**Figure 2**). A summary of our analyses is presented in **Figures 3, 4, 5, 6**. Comparison of the data from different sources was complicated by the fact that different ways of numbering of the nucleotides in the DNA substrates have been used by various investigators. For example, in several studies numbering of the cleaved strand starts with the first adenine on the 3'-end, resulting in the assigning of the numbers “-1” and “-2” to the two extra nucleotides on the 5'-end of the non-cleaved strand, (i.e. Gao *et al.* [19]). In the structures of PFV IN complexed with DNA, numbering from the 5'-end was introduced for the cleaved strand of viral DNA, placing the 3'-end adenine under number 17. Because the length of the oligonucleotides used in different studies varies, numbering from the 5'-end introduces additional confusion, as the number designations for the structurally equivalent nucleotides in the cleaved strands of different length would be different. We, as well as some others, elected to number the non-cleaved viral DNA strand from the first nucleotide at the 5'-end. The first nucleotide on the 3'-end of the cleaved strand of processed substrate (closest to the junction in Y-mer or X-mer integration intermediate substrate) is assigned #3 (**Figure 1**, green strand). For the target DNA, numbering of both strands starts from the junction of the integration site (**Figure 1**, pink and blue strands). In order to compare our crosslinking results with IN-DNA contact data from other laboratories, we have translated all nucleotide numbering of the strands that vary in substrate DNAs into this format. However, as a reference, we have included in curly brackets the original numbers from Maertens

et al. [6] and Krishnan *et al.* [7] for the nucleotides shown to interact with PFV IN.

To identify the functionally equivalent residues in ASV, HIV, and PFV INs, the structures of individual NTD, CCD, and CTD domains were superimposed upon the structure of the complex of PFV IN with DNA (PDB code 3OS0). Some chemical and photocrosslinking data identify the individual points of contact between the proteins and DNA. If a method does not allow one to specify a single contact point in both protein and DNA, then these data are not sufficient to establish the exact correlation with results from crystallography, even when they do not contradict them. Such data can be categorized only as either “do not contradict,” if IN and DNA are in proximity to each other in the PFV IN structure, or “no contact,” if IN is remote from substrate DNA in the PFV intasome. Specific residues shown to interact with DNA that are either in good correlation with the PFV structural results or do not contradict them are bolded in **Figures 3, 4, 5, 6**. The tabulated results show that the correlation between the PFV crystal structures and experimental data from crosslinking, mutagenesis, protease mapping, and mass spectrometry for ASV, MuLV, and HIV-1 IN proteins is highest for the CCD (**Figures 4,5** with color coding as in **Figure 1**). The crosslinking results that pinpoint individual IN-DNA contacts in the NTD and CTD of HIV-1 and ASV IN proteins show low correlation with the interactions observed in the structure of PFV IN complexed with DNA.

Interactions between DNA and the NTD. Very limited crosslinking data are available for the NTD. Both subdomains of the NTD of PFV (NED and the conserved subdomain) interact with viral DNA in the PFV intasome crystal. There is no target DNA in the proximity of the NTD of PFV, and no viral DNA in the proximity to the region analogous to the HIV-1 NTD peptide (amino acids 1–12) reported to interact with viral DNA by Heuer *et al.* [15,16]. Two HIV-1 IN amino acids in the NTD, Lys14 and Lys34, were implicated as having contacts with DNA by mass spectrometry [48] and proteolysis mapping [49], respectively, but only Lys34 (Val90 in PFV) appears to be relatively close (~9 Å) to nucleotides 9–11 of the non-cleaved viral strand in the PFV IN structure, whereas Lys14 lies more than 20 Å away from the DNA. It should be noted that only two of the four PFV NTDs are visible in the intasome crystal complexes, and alternate NTD positions for these unseen NTDs could account for proximity data for various solution experiments.

Interactions between DNA and the CTD. For the CTD, none of the individual contacts revealed in ASV and HIV-1 IN proteins by crosslinking or other methods can be correlated with those observed in the crystal structures of PFV IN-DNA complexes (**Figure 6**). Our crosslinking results with ASV IN show contacts of Arg244 to both strands of viral DNA at positions 10–12. However, in the PFV intasome structure [6], the equivalent residue, Asn348, is separated from the corresponding positions on DNA by the linker regions that connect the CCD with the NTD and CTD (**Figure 10A**). We note that while not seen in the PFV intasome structure, CTD interactions with the trans viral DNA remain a possibility and could be accomplished with minor movement of the domain. Results of Gao *et al.* [19], indicate that residues Ser230 and Glu246 of HIV-1IN interact with bases 1 and 7 of the non-cleaved strand of viral DNA, respectively. Crosslinking experiments based on the electron microscopy model obtained by Michel *et al.* [50] provided evidence for contact between Lys266 in HIV-1 IN and nucleotides 6–7 in the non-cleaved strand of viral DNA. These results are not in agreement with the HIV-1 model of Krishnan *et al.* [7], which was derived from the PFV crystal structure. Contact between the CTD of HIV-1 IN and the base of thymidine 6 of the non-cleaved strand

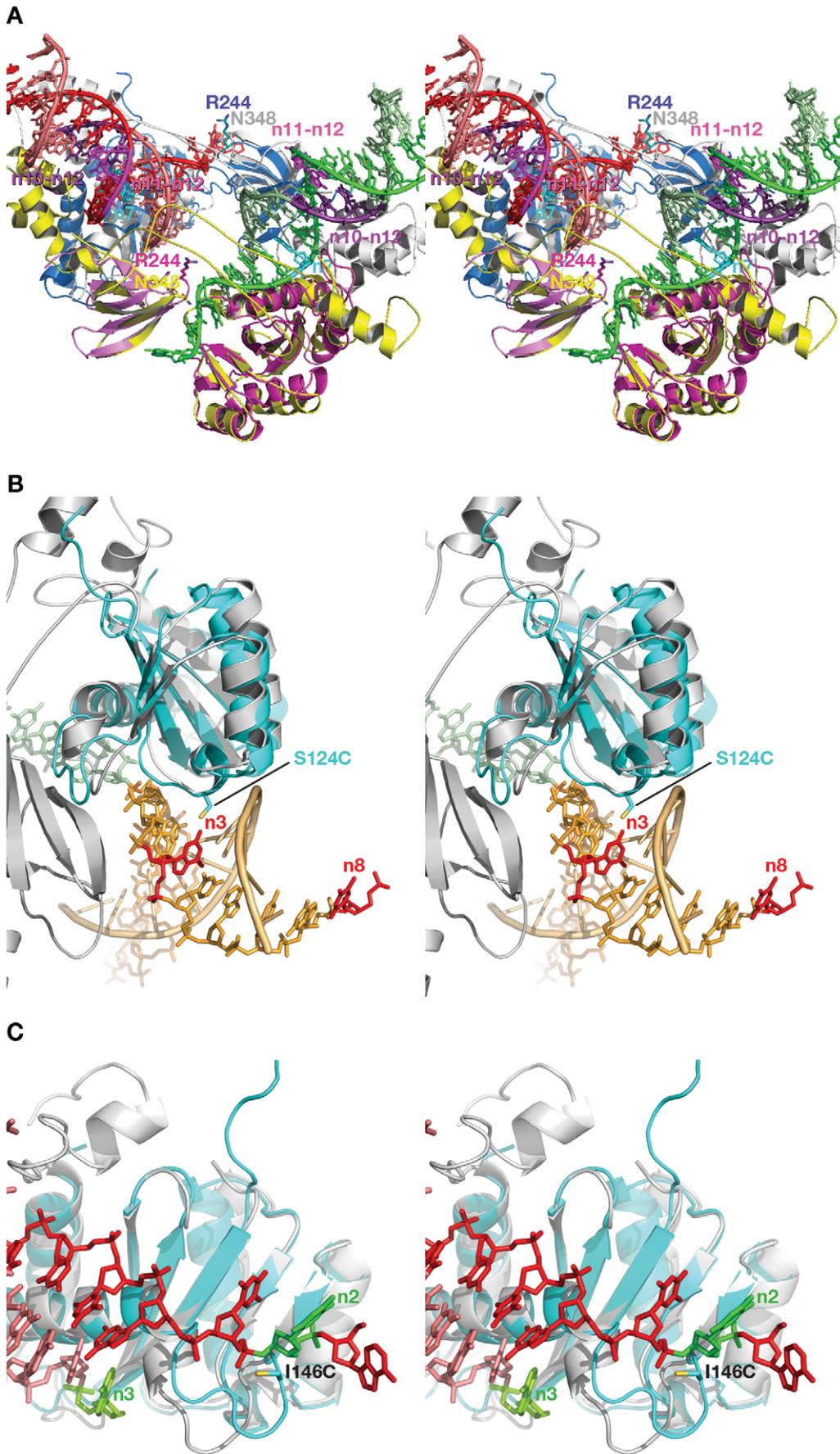


Figure 10. Structural interpretation of crosslinking data for ASV IN. (A) Superposition of a section of the DNA complex of PFV IN (cartoon) containing the two viral DNAs (sticks) with the individual CCD and CTD domains of ASV IN. Individual protomers of the PFV IN tetramer are colored gray and yellow, with Asn348 shown as sticks in corresponding colors. CCDs of ASV IN are magenta and blue and one CTD of ASV is shown in magenta, with Arg244 shown in sticks in corresponding colors. The linker region of PFV is shown in yellow. Nucleotides 10–12 of the non-cleaved viral DNA strand are shown in dark violet, nucleotides 11 and 12 of the cleaved viral DNA strand are shown in magenta, nucleotide 7 of the non-cleaved viral DNA strand is shown in cyan. (B) Superposition of the CCDs of ASV IN (blue) onto PFV IN (gray). The viral DNA after the integration step is shown in green and the host target DNA in orange. Nucleotides 3 and 8 on the host DNA are red and the substituted residue S124C is shown as sticks. (C) An analogous superposition, with viral DNA in red (non-cleaved strand) and pink (processed strand). A flexible loop contains the substituted residue I146C. Nucleotide 2 on the non-cleaved strand and the first nucleotide on the 3'-end of the cleaved strand are shown in green. doi:10.1371/journal.pone.0027751.g010

of viral DNA as reported by Esposito *et al.* [14], faces the same problems as the contact of nucleotide 7 with residue 246 [19], as the linker regions separate the protein and DNA in the PFV intasome.

Residues Leu234, Arg262, Arg263 and Arg269 in the CTD of HIV-1 IN, which have been shown to interact with DNA by modeling and/or experimental studies, were also implicated as interacting with DNA by several mass-spectrometry and mutagenesis studies (Figure 6). Residues that are structurally equivalent to Arg262 in HIV-1 IN are Ile366 and Ser262 in PFV and ASV INs, respectively. Because of the different sizes of the side chains of these residues in the three INs, analogous contacts cannot be made with PFV IN, and for ASV IN this seems problematic. Similarly, the capability of Leu234 and Arg263 in HIV-1 IN to contact DNA appears to correlate with the presence of arginine at structurally equivalent positions in ASV and MuLV IN proteins. However, HIV-1 IN Arg269 and PFV IN Ser373 both interact with DNA. The segment containing Ser373 is located at the very end of the visible CTD of PFV IN, and the flexibility of this part of the protein may facilitate interaction with DNA.

Heuer *et al.* [15] showed that the azidophenacyl photocrosslinker, attached to unique phosphorothioate located between nucleotides 6 and 7 of the cleaved strand of viral DNA, could be crosslinked to the peptide comprising residues 247–270 of HIV-1 IN. While some residues from the corresponding range in PFV IN are within reach of equivalent nucleotides 6 and 7 in the crystal structure, the specific residues in HIV-1 IN that are involved in these interactions are unknown.

Interactions between DNA and the CCD. Much more information regarding the sites of contact with DNA is available for the CCDs of various INs. Out of twenty-seven individual residues and 7 peptide ranges identified in 50 experimental data points that were analyzed and presented in Figures 4,5 as making contact between the CCD and DNA, thirty-seven IN-DNA contacts corresponded to residues analogous with those observed to interact with DNA in the crystal structure of the PFV intasome.

Our photocrosslinking data indicate that S124C of ASV IN makes contact with the third nucleotide of the cleaved strand of target DNA, and a minor contact with nucleotide 8 of the same strand (Figure 10B). In the crystal structure of the PFV intasome the analogous residue makes contacts with nucleotides 3 on the cleaved and 6 on the non-cleaved strands of the target DNA (shown in red in Figure 10B, they correspond to nucleotides 3 and 3, respectively, in the numbering system used here). The nucleotide corresponding to nucleotide 8 on host DNA complexed to ASV IN (minor contact) is not visible in the structure of the PFV intasome due to the mobility of the ends of the host DNA in the absence of contacts with the protein. This crosslink might be attributed to the flexibility of the photocrosslinking tether combined with mobility of the ends of host DNA (see Materials and Methods).

Photo- and chemical crosslinking data for I146C of ASV IN identified nucleotide 3 of the cleaved strand of viral DNA as the point of contact. Contact between I146C and nucleotide 2 of the non-cleaved strand of viral DNA was also detected by chemical crosslinking (Tables 2, 3). In MuLV, the structural equivalent of this residue is Cys209. Photo- and chemical-crosslinking experiments on MuLV by Vera *et al.* [29] confirmed the involvement of this residue in the interactions with the viral end of DNA in the active site area. Cys209 in MuLV IN is reported to make contact with nucleotide 1 on the non-cleaved strand of viral DNA (Figure 4). The corresponding residue in PFV IN, Thr210, also contacts the base of nucleotide 3, as in ASV IN, but in the non-cleaved DNA strand. All chemical crosslinks involving the ASV I146C derivative are maintained with the bases of the corresponding nucleotides. The contacts between Thr210 and DNA in PFV IN are localized in the minor groove between two strands; therefore the data from ASV IN correlate reasonably well with the PFV structure (Figure 10C). Residue 146 in ASV IN and the corresponding residues in HIV-1 and PFV INs are located within the active site flexible loop, which has been shown to adopt multiple conformations in different IN structures with various inhibitor, substrate, and pH/buffer conditions. The tip of this loop can move up to 7 Å under conditions that do not alter the overall three-dimensional structure of the CCD. In the PFV intasome, this loop is inserted between the ends of the complementary strands of viral DNA (Figure 10C). Therefore, if a similar position is assumed by the ASV loop when complexed with viral DNA, I146C would be able to interact with nucleotides on both strands.

Photo- and chemical crosslinking data for CCD-DNA contacts have been reported by several other groups. The contact for the HIV-1 residue Lys159 reported by Jenkins *et al.* [13] is with A3 nucleotide at the 3'-end of the processed strand. This amino acid is equivalent to Lys228 in PFV IN, and it interacts with the phosphate backbone between the nucleotides 3 and 4. The crosslinking observed in [13] between HIV-1 Lys159 and N7 of the base of A3 requires some adjustment of the orientation of A3 base, as seen in the PFV intasome structure.

The results of S-S crosslinking [28] of both blunt and processed DNA substrates and the results of photocrosslinking [14] of blunt DNA substrates to HIV-1 IN Q148 implicate two neighboring nucleotides of the non-cleaved strand of viral DNA, #2 and #1, respectively, for interaction. Although these nucleotides are found in the crystal structure of PFV IN in the vicinity of S217 (analogous to Q148 in HIV-1 IN), their bases, modified for crosslinking experiments, point away from the side chain of S217. As suggested by Krishnan *et al.* [7], such discrepancies can be attributed to the experimental setup (blunt *vs.* processed substrates) or to conformational mobility of the crosslinker.

Several amino acid residues of HIV-1 IN were reported by Alian *et al.* [27] to be involved in crosslinking, but these results do not match the IN-DNA contacts found in the PFV intasome structure for the corresponding pairs. There was a very low correlation of crosslinking data for HIV-1 residues 143, 160, and 164 using processed DNA with the model of HIV-1, which is

derived from the PFV intasome structure [7]. Nucleotide A1 of the non-cleaved strand of viral DNA identified by crosslinking to interact with all HIV-1 IN three residues (**Figure 4**) cannot reach the corresponding residues in the crystal structure of PFV IN. Only one contact was detected between HIV-1 Y143 and nucleotide A1 when a blunt ended substrate was used [27]. The same contact was identified by Esposito *et al.* in photocrosslinking experiments with blunt DNA substrates [14]. Alian *et al.* [27] suggested that the loop comprising HIV-1 residues 160–164 comes in close proximity to the 5'-end of the non-cleaved strand of viral DNA only during strand transfer. This hypothesis is inconsistent with the HIV-1 IN model [7]; as Lys 160 lies within contact range of G8 and quite far from the integration center, HIV-1 Y143 is not listed as a possible contact with viral end DNA by Krishnan *et al.* [7] but is positioned in close proximity to processed target DNA nucleotides (#1, 2, 3, -1, -2, and -3) closest to the integration site. It should be noted that, under some conditions, DTNB activation can produce nonspecific crosslinks [19].

Gao *et al.* detected contacts between HIV-1 I191C and two nucleotides, 1 and 7 of non-processed viral DNA by S-S crosslinking [19]. In the PFV intasome structure [6], the amide of V260 (I191 in HIV-1 IN) is located 4.5 Å away from the phosphate of nucleotide #7 of the non-cleaved strand of viral DNA, which is reasonable if the length of the thiopropyl linker is taken into account.

While the photocrosslinking experiments in which interactions between specific modified nucleotides and HIV-1 IN [15] in most cases do not provide exact localization of the contact sites on the IN protein, comparison of the relative positions of identified peptides (49–69, 51–64, 139–152, and 158–198) and DNA show good correlation for 11 out of 13 reported crosslinking contacts when compared to the PFV intasome structure [7], the ASV IN two-domain structure (PDB code1C0M) [7] superimposed on the corresponding domains of the PFV intasome, and the model of the HIV-1 intasome [7]. Some of these peptides have been targeted from multiple locations on DNA. For example, HIV-1 peptide 49–69 comes into close proximity to the viral processed DNA (phosphate between C4 and G5, G5 base [15]), non-processed viral DNA (A1 base, and the phosphate between G4 and C5 [15]), and non-cleaved strand of target DNA (backbone of G1, G2, and C(-2), G(-3) [15]). The latter contacts are located on the opposite sides of the same strand of target DNA from the integration site (a similar spatial relationship is illustrated in **Figure 1** for ASV IN substrate) and are made with residues from two IN monomers in the model of HIV-1 IN [7]. Introduction of the photoactivatable nucleotide analogs I-dU and I-dC into positions 3 of the cleavable strand and 1 and 2 of non-cleavable strand of blunt viral DNA substrates resulted in the crosslinks with CCD, although the exact positions in the protein were not elucidated [15]. Nucleotides in these positions are also found to be in close proximity to the active site of the CCD in the PFV intasome [7].

Mutagenesis experiments carried out by Chen *et al.* [21] on HIV-1 IN provided a list of residues (V54, V72, T124, T125, S153, K156, E157, K160, G193, 54–57) likely to be important for DNA binding and substrate specificity. Circular dichroism, fluorescence, and NMR experiments involving a synthetic analog of α 4 helix of HIV-1 CCD and U5 LTR end [51] revealed that the HIV-1 IN residues E152, S153, N155, K156, and K159 were likely to make contact with DNA. Protease mapping with HIV-1 IN [49] assigned a similar role to the residues K111, K136, K159, E138, K185, K186, and K188, and mass spectrometry footprinting experiments [48] indicated that K159 and K160 are involved DNA interactions. The corresponding residues in the PFV IN-DNA complex structure are within range to establish contacts with target or viral DNAs.

However, the PFV equivalents of some residues in HIV-1 IN implicated in DNA binding in these experiments (e.g. 161, 162, 171, 172, 197 and 201 and peptides 128–130, 163–165), are not in a suitable range to contact DNA in the PFV intasome. Several positions in the fragment comprising residues 207–219, shown to interact with DNA by protease mapping [49] and mass spectrometry [48], belong to the linker region between the CCD and CTD. This region differs in length in HIV-1, ASV, and PFV INs and exhibits little sequence homology. The HIV-1 IN model built by Krishnan *et al.* [7] allows for the residues from this fragment to maintain contacts with non-cleaved strand of viral DNA (**Figures 4,5**), correlating with the mapping data listed above.

Mutagenesis experiment by Esposito *et al.* [14] indicated that nucleotides 3, 4, 12, and 13 of the cleaved strand of viral DNA and nucleotide 2 of the non-cleaved strand participate in CCD-DNA interactions. The contacts of the nucleotides 2, 3, and 4 are in good agreement with the model of the HIV-1 intasome and structural data from PFV IN. Similarly, the loop comprising residues 207–209 of HIV-1 IN is in close proximity to nucleotides 12 and 13 of the cleaved strand. While the mutagenesis results [14] do not contradict the structural data, they do not locate the contact residues in the protein. In contrast, our S-S crosslinking data identify both counterparts in the ASV IN-DNA interactions. For example, results with the I146C derivative of ASV IN implicate this residue in interactions with nucleotide 3 of the cleaved strand and nucleotide 2 of the non-cleaved strand of viral DNA.

In conclusion, the high degree of correlation between the structural and biochemical data on IN-DNA contacts in the CCD indicate that the mode of binding DNA to this domain is highly conserved in PFV, HIV-1, and ASV INs. Differences in protein structure and composition may explain the lack of correspondence in details of DNA binding by the NTD and CTD of PFV in the crystal structure of the intasome, when compared with data obtained from analysis of crosslinking and other experiments performed with ASV and HIV-1 IN proteins. The presence of an additional domain at the N terminus of PFV IN (the NED) certainly sets it apart from the other two retroviral IN proteins. In addition, variations in length and sequence of the linker regions between the NTD and CCD, and the CCD and CTD, suggests that residues at different positions in these domains could have been selected to perform analogous functions during the course of evolution of these viruses. On the other hand, depending on the concentration, IN proteins can exist in a variety of multimers in solution (dimers, tetramers, and higher forms), each of which might interact with DNA in unique ways during the assembly of a functional intasome. Such interactions may be detected in biochemical experiments, but not represented in the intasome crystal. Furthermore, the same amino acid in individual subunits might make different contacts with DNA in one or more of these multimers. We note that the NTDs and CTDs of only two of the four component subunits are visible in the crystal of the PFV intasome, and it is unknown if or how these domains in the other two subunits might interact with DNA. Additional crystal structures, including those of other retroviral intasomes, could help to resolve some of these issues. However, until we understand more about the dynamic properties of IN, and the conformational changes that accompany intasome assembly, it will be important to keep all of these factors in mind when interpreting both structural and biochemical data.

Materials and Methods

Photocrosslinkers

Heterobifunctional photoactivatable thiol-specific reagents, a carbene-generating N-bromoacetyl-N'-{2,3-dihydroxy-3-[3-(3-

(trifluoromethyl)diazirin-3-yl]phenyl]propionyl} ethylenediamine (BATDHP) [17] and nitrene-generating azidophenylthiophthalimide (AFTP) from Sigma [32] were used. Photocrosslinking reagents were prepared as 10–20 mM stock solutions in DMSO and stored in the dark at -20°C for no longer than 30 days. These reagents were coupled to the SH- group of the engineered Cys-containing IN derivatives. Amino-reactive photocrosslinking reagent N-hydroxysuccinimidyl-3-[3-(trifluoromethyl)diazirin-3-yl]benzoate was used for modification of NH_2 -derivatized thymidines in DNA substrates to complement our data obtained with Cys-substituted modified proteins.

Thiol modification

Modification and crosslinking procedures with IN were performed as previously described [17,18]. The IN proteins were modified with the photocrosslinking reagents via a single Cys residue. 50 μL of 30 μM solutions of IN were treated with 5 mM DTT on ice for 30 min to reduce the SH group. DTT was then removed by gel filtration using Sephadex G50 Centriscip desalting columns (from Princeton Separations, Adelphia, NJ) in buffer 1 (40 mM HEPES pH 6.5, 500 mM NaCl, 5% glycerol). The reduced IN was allowed to react with 10-fold molar excess of a photoreagent in dark vials on ice for 12 hr after raising the pH of reaction mixtures from 6.5 to 7.8 by addition of 1 M Tris-HCl pH 8.0. The appropriate amounts were extrapolated for small volume reactions (20–100 μL) from test-titration of a 100 ml mixture without protein and DNA. Excess photocrosslinking reagent was removed by gel filtration with buffer 1. All subsequent manipulations were carried out under reduced light levels.

Mass spectrometry

Mass spectrometry, performed at the Fox Chase Cancer Center Shared Facility, was used to determine the number and position(s) of modifications to ensure that Cys residues in the Zn-coordinating motif of the NTD were not modified with crosslinking reagents. IN modified with either BATDHP or AFTP was digested with trypsin. Tryptic peptides containing modifications were identified by matrix-assisted laser desorption time-of-flight (MALDI-TOF) spectrometry.

DNA substrates

Amino-derivatized and non-modified DNA oligonucleotides synthesized using the phosphoroamidite method with subsequent PAGE purification were obtained from commercial sources (Oligos, Etc.). Oligonucleotides were tagged by 5' labeling with $\gamma^{32}\text{P}$ -ATP using T4 polynucleotide kinase (T4 PNK) obtained from Boehringer Mannheim.

The following sequences were prepared:

1 5' GCTGTTGAATACCATCTAATCGTGTCGGGTCT-CGTACTGCCGAA 3'
 2 5' TCCGCAGTACGAGACCCG 3'
 3 5' AATGTAGTCTTATGCAATAGC 3'
 3' 5' AATGTAGTCTTATGCAATACTC 3'
 3f 5' ATTGTAGTCTTATGCAATACTC 3'
 4 5' GCTATTGCATAAAGACTACAACAGATTAGATG-GTATTCAACAAGC 3'
 4' 5' GAGTATTGCATAAAGACTACATT 3'
 4f 5' GAGTATTGCATAAAGACTACAAT 3'
 NH2-3.8 5' AATGTAGTCTTATGCAATACTC 3'
 NH2-3.11 5' AATGTAGTCTTATGCAATACTC 3'
 NH2-4.12 5' GAGTATTGCATAAAGACTACATT 3'

where amino-modified nucleotides are shown underlined and in bold.

DNA strands 1–4 were mixed in equal concentrations and annealed to prepare the Y-mer substrate; strands 3' and 4' were mixed in equal concentrations and annealed to prepare the linear substrate. Oligonucleotides 3f and 4f were used for preparation of frayed-end substrates with the appropriate modified complementary strands. Amino-modified oligonucleotides were used to introduce the NHS-[3-(trifluoromethyl)diazirin-3-yl]benzoate photoreagent by a procedure similar to modification of IN, except the reducing step.

For chemical crosslinking, oligonucleotides with thiol-modified adenosines and guanidines were prepared similarly to the method of Erlandson et al. [52] (See **Methods S1** for details). Oligonucleotide SH 4.3-P carried a mercaptopropanol phosphate ester $-\text{O}_3\text{P}-\text{O}-(\text{CH}_2)_3-\text{SH}$ in place of scissile phosphate. In SH 4.3-M the 3' terminal desoxyribose was substituted with N-mercaptoethyl morpholine. Modified positions below are bolded and underlined; numbering is as in **Figure 1**. For description of synthetic pathways and structures, see **Methods S1**.

SH3.1,7 5' AATGTAGTCTTATGCAATACTC 3'
 SH3.1,12 5' AATGTAGTCTTATGCAATACTC 3'
 SH3.2 5' AATGTAGTCTTATGCAATACTC 3'
 SH3.12 5' AATGTAGTCTTATGCAATACTC 3'
 SH4.13 5' GAGTATTGCATAAAGACTACATT 3'
 SH4.11 5' GAGTATTGCATAAAGACTACATT 3'
 SH4.3 5' GAGTATTGCATAAAGACTACATT 3'
 SH4.3-P 5' GAGTATTGCATAAAGACTACA 3'
 SH4.3-M 5' GAGTATTGCATAAAGACTACA 3'

Photocrosslinking

10 μM IN and 0.05 μM DNA substrate (5'-labeled with $\gamma^{32}\text{P}$ ATP at one of the component oligonucleotides) were incubated in buffer 2 (40 mM HEPES pH 6.5, 150 mM NaCl, 5% glycerol) for 15 min at 0°C and then UV-irradiated with a hand-held lamp placed 1 cm away from the samples on ice for 15 min using a glass plate as additional filter (cutoff 315 nm). Non-reducing denaturing PAGE was used to separate crosslinked IN from the non-crosslinked protein, as well as to remove any DNA that was not crosslinked. The products were visualized and quantified with a PhosphorImager (Storm 860 from Molecular Dynamics, Inc., Sunnyvale, CA). The efficiency of crosslinking was calculated as the percent of radioactivity in the IN-DNA bands relative to the total amount of radioactivity in the lane. As an excess of IN protein was used, and both the DNA and IN were present at concentrations significantly higher than the IN-DNA binding constant, all DNA is assumed to be bound to the enzyme. The negative control samples were obtained by UV-irradiation of reaction mixtures with non-modified INs and by analyzing non-irradiated samples.

Localization of the preferred sites of crosslinking

Localization of the preferred sites of IN photocrosslinking to different DNA substrates under various conditions was performed using Cel 1 "Surveyor" endonuclease from Transgenomics, Inc. (Rockville, MD) [34–36,53]. Samples of the UV-crosslinked IN-DNA complexes were prepared and 2–3 μL aliquots were used to analyze the crosslinking efficiency by PAGE and PhosphorImager. The reaction mixtures were extracted twice by phenol-chloroform. Aqueous fractions that contained non-crosslinked DNA were pooled together, precipitated with ethanol and saved as a negative control for Cel1 analysis. Phenol/chloroform fractions containing covalent IN-DNA complexes were pooled together, divided in half, and one half was treated with 10 mM NaIO_4 (30 min. at room temperature), while the other half was left untreated. Both phenol-extracted samples were ethanol-precipitated and dissolved

in 20 microliters of water. We used non-treated ^{32}P -labeled DNA substrates, as well as crosslinking reaction mixtures and crosslinking reaction mixtures after treatment with 10 mM NaIO_4 as controls for each Cell assay. The Cell assays were performed according to manufacturer's instructions. The reactions were terminated by adding the formamide gel loading buffer and heating to 90°C . Reaction mixtures were separated by Urea-PAGE using 6% and 20% gels and analyzed with Phosphor-Imager.

Chemical crosslinking

Reducing denaturing PAGE was used to break the covalent linkage between DNA and IN in order to confirm that the DNA was in fact crosslinked by a disulfide linkage. Crosslinking reactions were performed by mixing $25\ \mu\text{M}$ IN with corresponding concentration of DNA to produce a desired ratio of IN-DNA in 40 mM HEPES pH 6.5, 150 mM NaCl, 1 mM EDTA and 5% glycerol. After 1 hr preincubation on ice, the pH of the reaction mixture was adjusted to 7.8 by addition of 1 M Tris-HCl pH 8.0 and then left on ice for 10–15 hr to allow crosslinking. PAGE gel analysis with Coomassie staining was used to separate and quantify the products of reactions by densitometry.

Crosslinking experiments with IN derivatives that contained Cys substitutions in catalytic residues D64 and E157 of ASV IN were performed essentially as described above, with minor modifications. The processed, recessed linear DNA substrate (**Figure 1**) was used with the “processed” strand comprising either SH4.3-M or SH4.3-P.

The oligonucleotides were mixed in equimolar quantities and annealed prior to reduction with 100 mM β -ME on ice for 12 hr. The excess of reducing agent was removed by gel filtration on Centriscap spin-columns in 150 mM NaCl. IN was concentrated to ~ 5 – $6\ \text{mg/ml}$ in Buffer 1. The concentrated protein was treated with β -mercaptoethanol on ice for 12 hr for reduction of the surface Cys. The excess of reducing agent was removed by gel filtration into Buffer 1. The DNA was then added to a protein solution for a final molar ratio of protein to DNA of 2:1 or 1:1. The complex was incubated on ice for 30 min in Buffer 1 before adjustment of NaCl concentration to 250–300 mM and pH to 7.5. The reactions were carried out with and without 10 mM MgCl_2 .

Alternatively, for the catalytic Cys derivatives, S-S bond formation was facilitated with DTNB as in [27]. 1 mM DTNB was added to each of the reduced oligonucleotide substrates and the mixture was incubated for 2 hr at room temperature. Excess DTNB was also removed by gel filtration in 20 mM Tris-HCl pH 8.0, 150 mM NaCl. IN proteins were prepared as above in Buffer 1. $50\ \mu\text{M}$ DNA was added to $50\ \mu\text{M}$ IN protein at either 2:1 or 1:1 molar ratio and incubated on ice for 10 min, diluted 1.5-fold with 20 mM Tris-HCl pH 8.0, 150 mM NaCl. Reactions were initiated by adding MnCl_2 to a final concentration of 10 mM. After overnight incubation on ice, the yields of crosslinking were determined by PAGE after overnight incubation on ice.

Supporting Information

Figure S1 Disintegration reactions of the modified IN proteins with Y-mer substrate. A) Schematic depicting integrase-catalyzed disintegration and joining reactions of the Y-mer substrate superimposed on a generic tetramer model for integrase (left). The liberated viral DNA (right) is the same DNA as seen in a pre-cleaved end, with the cleaved portion of strand 4 indicated by “4-”. B) Disintegration activity of WT and ASV IN derivative proteins before and after modification with photo-

crosslinking reagents [54]. Upper band on each gel represents the 44-mer product of reaction, lower bands correspond to the 19-mer substrate (5' ^{32}P -labeled Y-mer strand 2*). Due to increased exposure times of the four gels on the right, the contaminating 18-mer fragments of the substrate are also visible. Reaction times were 0–1200 s; increasing time is indicated by the wedge above gel lanes. “noXL” stands for no IN modification with photocrosslinker.

(TIF)

Figure S2 Strand specificity of photocrosslinking of modified Cys residues in substituted derivatives of ASV IN. A sample of gels accompanied by bar graphs with the photocrosslinking yields (%) presented. The position of modified Cys residues in the IN derivatives is noted above the bar graph. The crosslinker used is also listed above the graph (noXL stands for no IN modification with photocrosslinker). DNA strands in the Ymer substrate are labeled according to **Figure S1**, and the labeled strands are indicated below each graph. Brackets on the right show the bands corresponding to IN-DNA adducts and free DNA.

(TIF)

Figure S3 Comparison of the efficiency of photocrosslinking of APTP- and BATDHP-modified ASV IN derivatives to Y-mer (Y3,Y4) and linear (L3, L4) DNA substrates. Labeling is as in **Figure S2**.

(TIF)

Figure S4 Comparison of UV photocrosslinking yields of wild type ASV IN to dsDNA substrates modified with diazirine photocrosslinkers. Positions of the modified bases are bolded and the conserved A is underlined.

(TIF)

Figure S5 A sample of Coomassie-stained gels with S-S crosslinked IN-DNA complexes. The substrates are labeled above the lanes. Y stands for Y-mer DNA, L for linear; letters are followed by strand numbers, thiol modified positions are shown after dash. Bands corresponding to crosslinked complexes are marked by arrows. The negative controls are marked with a dash above the left-most lane of each gel. 1IN and 2IN with arrowheads designate monomeric and dimeric IN bands, respectively. The stemmed arrows point to IN-DNA adduct bands. Molecular weight marker lanes are not shown, since the monomer (lower) and dimer (upper) strong bands provide internal calibration.

(TIF)

Methods S1 A more detailed description of the materials and methods utilized in this work.

(PDF)

Results S1 A more detailed analysis of the enzymatic activity of IN and of the chemical crosslinking of modified DNA substrates to ASV IN.

(PDF)

Acknowledgments

We thank Steve Seeholzer and Yibai Chen (Fox Chase Mass Spectrometry and Proteomics Facility) for assistance in identifying positions of modification within integrase, and Jane Sayer and Donald Jerina (Laboratory of Bioorganic Chemistry, National Institute of Diabetes and Digestive and Kidney Diseases, NIH) for synthesis of several modified oligonucleotides. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of any sponsoring organization.

Author Contributions

Conceived and designed the experiments: EP AG AK AW AS. Performed the experiments: EP MA GM JA DZ TS AK RB. Analyzed the data: EP

MA AG AK. Contributed reagents/materials/analysis tools: MP VP. Wrote the paper: EP MA AG AK AW AS.

References

- Pommier Y, Pilon AA, Bajaj K, Mazumder A, Neamati N (1997) HIV-1 integrase as a target for antiviral drugs. *Antivir Chem Chemother* 8: 463–483.
- Jaskolski M, Alexandratos JN, Bujacz G, Wlodawer A (2009) Piecing together the structure of retroviral integrase, an important target in AIDS therapy. *FEBS J* 276: 2926–2946.
- Prada N, Markowitz M (2010) Novel integrase inhibitors for HIV. *Expert Opin Investig Drugs* 19: 1087–1098.
- Powderly WG (2010) Integrase inhibitors in the treatment of HIV-1 infection. *J Antimicrob Chemother* 65: 2485–2488.
- Al-Mawsawi LQ, Neamati N (2011) Allosteric inhibitor development targeting HIV-1 integrase. *ChemMedChem* 6: 228–241.
- Maertens GN, Hare S, Cherepanov P (2010) The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* 468: 326–329.
- Krishnan L, Li X, Naraharisetty HL, Hare S, Cherepanov P, Engelman A (2010) Structure-based modeling of the functional HIV-1 intasome and its inhibition. *Proc Natl Acad Sci USA* 107: 15910–15915.
- Asante-Appiah E, Skalka AM (1997) A metal-induced conformational change and activation of HIV-1 integrase. *J Biol Chem* 272: 16196–16205.
- Asante-Appiah E, Seeholzer SH, Skalka AM (1998) Structural determinants of metal-induced conformational changes in HIV-1 integrase. *J Biol Chem* 273: 35078–35087.
- Rice P, Mizuuchi K (1995) Structure of the bacteriophage Mu transposase core: A common structural motif for DNA transposition and retroviral integration. *Cell* 82: 209–220.
- Davies DR, Goryshin IY, Reznikoff WS, Rayment I (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* 289: 77–85.
- Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464: 232–236.
- Jenkins TM, Esposito D, Engelman A, Craigie R (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J* 16: 6849–6859.
- Esposito D, Craigie R (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J* 17: 5832–5843.
- Heuer TS, Brown PO (1997) Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry* 36: 10655–10665.
- Heuer TS, Brown PO (1998) Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry* 37: 6667–6678.
- Peletskaya EN, Boyer PL, Kogon AA, Clark P, Kroth H, et al. (2001) Cross-linking of the fingers subdomain of human immunodeficiency virus type 1 reverse transcriptase to template-primer. *J Virol* 75: 9435–9445.
- Peletskaya EN, Kogon AA, Tuske S, Arnold E, Hughes SH (2004) Nonnucleoside inhibitor binding affects the interactions of the fingers subdomain of human immunodeficiency virus type 1 reverse transcriptase with DNA. *J Virol* 78: 3387–3397.
- Gao K, Butler SL, Bushman F (2001) Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes. *EMBO J* 20: 3565–3576.
- Lutzke RA, Vink C, Plasterk RH (1994) Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res* 22: 4125–4131.
- Chen A, Weber IT, Harrison RW, Leis J (2006) Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat Ends. *J Biol Chem* 281: 4173–4182.
- Dolan J, Chen A, Weber IT, Harrison RW, Leis J (2009) Defining the DNA substrate binding sites on HIV-1 integrase. *J Mol Biol* 385: 568–579.
- Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, et al. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc Natl Acad Sci USA* 97: 8233–8238.
- Chen Z, Yan Y, Munshi S, Li Y, Zugay-Murphy J, et al. (2000) X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50–293)—an initial glance of the viral DNA binding platform. *J Mol Biol* 296: 521–533.
- Yang ZN, Mueser TC, Bushman FD, Hyde CC (2000) Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J Mol Biol* 296: 535–548.
- Wang JY, Ling H, Yang W, Craigie R (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J* 20: 7333–7343.
- Alian A, Griner SL, Chiang V, Tsiang M, Jones G, et al. (2009) Catalytically active complex of HIV-1 integrase with a viral DNA substrate binds anti-integrase drugs. *Proc Natl Acad Sci USA* 106: 8192–8197.
- Johnson AA, Santos W, Pais GC, Marchand C, Amin R, et al. (2006) Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J Biol Chem* 281: 461–467.
- Vera J, Valenzuela B, Roth MJ, Leon O (2008) Characterization of the long-terminal repeat single-strand tail-binding site of Moloney-MuLV integrase by crosslinking. *Biol Res* 41: 69–80.
- Harper AL, Skinner LM, Sudol M, Katzman M (2001) Use of patient-derived human immunodeficiency virus type 1 integrases to identify a protein residue that affects target site selection. *J Virol* 75: 7756–7762.
- Nowak MG, Sudol M, Lee NE, Kongsavage WM, Katzman M (2009) Identifying amino acid residues that contribute to the cellular-DNA binding site on retroviral integrase. *Virology* 389: 141–148.
- Moreland RB, Smith PK, Fujimoto EK, Dockter ME (1982) Synthesis and characterization of N-(4-azidophenylthio)phthalimide: A cleavable, photoactivable crosslinking reagent that reacts with sulfhydryl groups. *Anal Biochem* 121: 321–326.
- Merkel G, Andrade MD, Ramcharan J, Skalka AM (2009) Oligonucleotide-based assays for integrase activity. *Methods* 47: 243–248.
- Kulinski J, Besack D, Oleykowski CA, Godwin AK, Yeung AT (2000) CEL I enzymatic mutation detection assay. *BioTechniques* 29: 44–46.
- Oleykowski CA, Bronson Mullins CR, Chang DW, Yeung AT (1999) Incision at nucleotide insertions/deletions and base pair mismatches by the SP nuclease of spinach. *Biochemistry* 38: 2200–2205.
- Oleykowski CA, Bronson Mullins CR, Godwin AK, Yeung AT (1998) Mutation detection using a novel plant endonuclease. *Nucleic Acids Res* 26: 4597–4602.
- Huang H, Chopra R, Verdine GL, Harrison SC (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* 282: 1669–1675.
- Ellison V, Brown PO (1994) A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro. *Proc Natl Acad Sci USA* 91: 7316–7320.
- Mazumder A, Uchida H, Neamati N, Sunder S, Jaworska-Maslanka M, et al. (1997) Probing interactions between viral DNA and human immunodeficiency virus type 1 integrase using dinucleotides. *Mol Pharmacol* 51: 567–575.
- Miller MD, Farnet CM, Bushman FD (1997) Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* 71: 5382–5390.
- Katz RA, DiCandeloro P, Kukulj G, Skalka AM (2001) Role of DNA end distortion in catalysis by avian sarcoma virus integrase. *J Biol Chem* 276: 34213–34220.
- Scottoline BP, Chow S, Ellison V, Brown PO (1997) Disruption of the terminal base pairs of retroviral DNA during integration. *Genes Dev* 11: 371–382.
- Renisio JG, Cosquer S, Cherrak I, El Antri S, Mauffret O, Fermandjian S (2005) Pre-organized structure of viral DNA at the binding-processing site of HIV-1 integrase. *Nucleic Acids Res* 33: 1970–1981.
- Katz RA, Merkel G, Andrade MD, Roder H, Skalka AM (2011) Retroviral integrases promote fraying of viral DNA ends. *J Biol Chem*. In press.
- Bojja RS, Andrade MD, Weigand S, Merkel G, Yarchivskva O, et al. (2011) Architecture of a full-length retroviral integrase monomer and dimer, revealed by small angle X-ray scattering and chemical cross-linking. *J Biol Chem* 286: 17047–17059.
- Lubkowski J, Yang F, Alexandratos J, Merkel G, Katz RA, et al. (1998) Structural basis for inactivating mutations and pH-dependent activity of avian sarcoma virus integrase. *J Biol Chem* 273: 32685–32689.
- Lubkowski J, Dauter Z, Yang F, Alexandratos J, Merkel G, et al. (1999) Atomic resolution structures of the core domain of avian sarcoma virus integrase and its D64N mutant. *Biochemistry* 38: 13512–13522.
- Zhao Z, McKee CJ, Kessl JJ, Santos WL, Daigle JE, et al. (2008) Subunit-specific protein footprinting reveals significant structural rearrangements and a role for N-terminal Lys-14 of HIV-1 Integrase during viral DNA binding. *J Biol Chem* 283: 5632–5641.
- Dirac AM, Kjems J (2001) Mapping DNA-binding sites of HIV-1 integrase by protein footprinting. *Eur J Biochem* 268: 743–751.
- 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.
- Hobaika Z, Zargarian L, Boulard Y, Maroun RG, Mauffret O, et al. (2009) Specificity of LTR DNA recognition by a peptide mimicking the HIV-1 integrase α 4 helix. *Nucleic Acids Res* 37: 7691–7700.
- Erlanson DA, Chen L, Verdine GL (1993) Allosteric effects on enzyme action in disulfide-crosslinked DNA. *J Am Chem Soc* 115: 12583–12584.
- Yang B, Wen X, Kodali NS, Oleykowski CA, Miller CG, et al. (2000) Purification, cloning, and characterization of the CEL I nuclease. *Biochemistry* 39: 3533–3541.

54. Jenkins TM, Engelman A, Ghirlando R, Craigie R (1996) A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J Biol Chem* 271: 7712–7718.
55. Lutzke RA, Plasterk RH (1998) Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. *J Virol* 72: 4841–4848.

Supporting Information

Localization of ASV Integrase-DNA Contacts by Site-Directed Crosslinking and their Structural Analysis

**Elena Peletskaya^{1*}, Mark Andrade^{2*}, Alla Gustchina^{3*}, George Merkel², Jerry
Alexandratos³, Dongwen Zhou³, Ravi S. Bojja², Tadashi Satoh³, Mikhail Potapov⁴, Alex
Kogon¹, Viktor Potapov⁴, Alexander Wlodawer³, Anna Marie Skalka²**

*¹Biolinx LLC, Alburgh, VT, USA; ²Fox Chase Cancer Center, Institute for Cancer Research,
Philadelphia, PA, USA; ³Macromolecular Crystallography Laboratory, Center for Cancer
Research, National Cancer Institute, National Institutes of Health, Frederick, MD, USA;
⁴Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences,
Moscow, Russia.*

Running Title: Crosslinking ASV integrase and DNA

Address correspondence and proofs to: Alexander Wlodawer, MCL, NCI, Bldg. 536, Frederick,
MD 21702; Tel. 301-846-5036; E-mail: wlodawer@nih.gov

* These authors contributed equally

Supporting Materials and Methods

Photocrosslinking to DNA substrates

The ASV IN derivatives were modified by coupling with two photoactivatable thiol-specific compounds, a carbene-generating compound, N-bromoacetyl-N'-{2,3-dihydroxy-3-[3-(3-(trifluoromethyl)diazirin-3-yl)phenyl]propionyl}ethylenediamine (BATDHP) [1] or a nitrene-generating compound, azidophenacylthiopyridine (AFTP) [2]. Carbenes are among the most reactive moieties known. They are capable of insertion into any chemical bond present in a biomolecule, including aliphatic chains and aromatic rings. Having nanosecond half-lifetimes, carbenes rapidly form covalent bonds with neighboring atoms. Although electrophilic, carbenes are so highly reactive that in the absence of strong nucleophiles they attack even the C-H bonds (over 80% crosslinking to cyclohexane) [3,4]. The high level of reactivity of carbenes with buffer components usually precludes high yields of crosslinked products but their low specificity makes them ideal for identification and localization of members and parts of affinity complex. Nitrenes such as those generated from the azide reagent, AFTP, are less reactive and tend to undergo intramolecular rearrangements that lead to even less reactive products. They crosslink primarily to nucleophiles such as amino groups and, in a non-nucleophilic environment, can remain active for periods of up to several minutes. This makes them less reliable for detection of close interactions since selective crosslinking may preferentially occur to a relatively distant nucleophilic group that is only occasionally in the vicinity of the crosslinker. The efficiency of crosslinking with nitrenes is thus higher, but there is a possibility of bias towards interactions with nucleophilic groups that are not relevant to the complex under study [4].

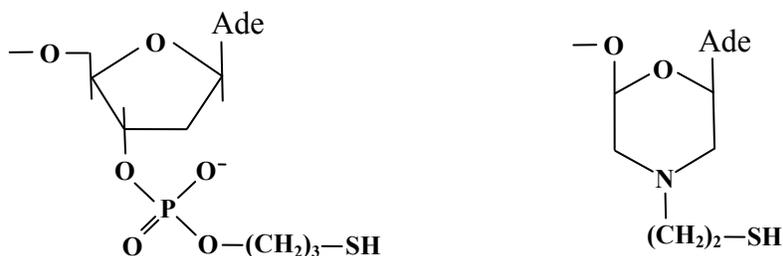
Oligonucleotide synthesis with 8-amino-3,4-dithiaoctyl tether at N^2 of dG

The oligonucleotides were prepared on commercial dC-cpg (40-50 $\mu\text{mol/g}$) using a standard synthesizer protocol for generation of the 21-base sequence at the 3'-end. The 5'-DMT protected phosphoramidite [5] derived from O^6 -(2-*p*-nitrophenylethyl)-2-fluoro-2'-deoxyinosine was coupled manually to the 5'-end of the oligonucleotide bound to the support [6]. A typical 2- μmole synthesis utilized 20 mg (21 μmol) of the phosphoramidite and 150 μL of 0.5 M 4,5-dicyanoimidazole in acetonitrile for 16 h at rt.; yield was estimated from the recovery of DMT cation after deprotection. End capping with acetic anhydride was omitted after the manual coupling step [6], and the support-bound oligonucleotide was directly oxidized (1 M *tert*-butyl hydroperoxide in dichloromethane, 30 s) and returned to the synthesizer for addition of the remaining 8 residues by the standard automated synthesis procedure. After removal of the 5'-DMT protecting group, the linker was coupled to the support-bound oligonucleotide using a

modification of the procedure of Erlanson et al. [5]. The support-bound oligonucleotide was treated with 45 mg (178 μmol) of 3,3'-dithiobis(propylamine) dihydrochloride [7], which was prepared as described for the 4-carbon homolog [8] 60 μL (600 μmol) triethylamine and 100 μL H_2O for 16 h at room temperature. In the course of preparing O^6 -(2-*p*-nitrophenylethyl)-2-fluoro-2'-deoxyinosine, we observed cleavage of the nitrophenylethyl protecting group in the presence of wet *tert*-butyl ammonium fluoride. After reaction with dithiobis(propylamine), addition of concentrated NH_4OH (1.5 mL) containing 20 μmol *tert*-butyl ammonium fluoride to the beads and solution, followed by heating at 60 $^\circ\text{C}$ for 3 days, resulted in complete deblocking of the oligonucleotide; this procedure avoided the DBU/formamide cleavage step and accompanying formylation [5] of the free amino group of the tether. After filtration, the oligonucleotide solution was dialyzed against 0.1 M triethylammonium acetate buffer (pH 6.0) overnight to remove excess amines. The oligonucleotide was purified by HPLC on a Hamilton PRP-1 column (7 μm , 10 x 250 mm) eluted at 3 mL/min with a linear gradient of acetonitrile in 0.1 M $(\text{NH}_4)_2\text{HCO}_3$ buffer (pH 7.5) that increased the acetonitrile concentration from 0 to 17.5% over 20 min; t_r 16.3 min.

Synthesis of the 3' modified viral end substrates for crosslinking to the active site IN derivatives.

Oligonucleotides 5'-GAGTATTGCATAAGACTAC-A*-3' where A* represents the following structures:

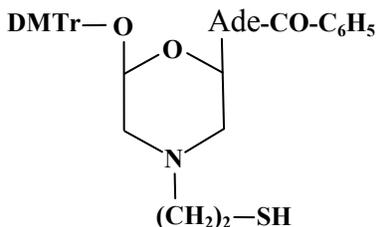


were synthesized following a standard protocol using modified solid support of the following structure: CPG-700-LCAA-NH-CO-(CH_2)₂-COO-(CH_2)₂-SS-(CH_2)₃-O-DMTr

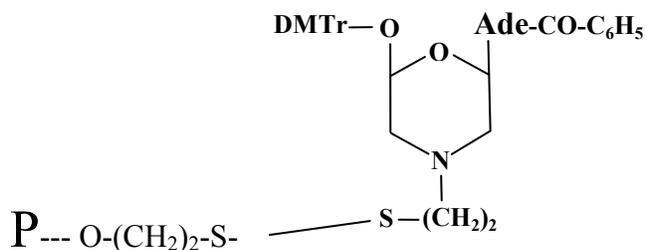
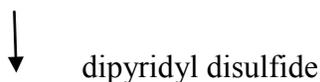
For deblocking a standard mixture of ammonia and methylamine also contained 50mM DTT for disulfide cleavage.

Schematics of nucleotide synthesis:

3'-end N-mercaptoethyl morpholine-adenosine was synthesized according to Stirchak et al. [9]. To complete the morpholine ring we used cystamine acetate, $(\text{NH}_2-(\text{CH}_2)_2-\text{SS}-(\text{CH}_2)_2-\text{NH}_2)\text{Ac}_2$, for amino component. The disulfide was cleaved by DTT to yield corresponding mercaptane:



that was then attached to modified solid support. The support modification consisted of mercaptoethanol treatment to achieve corresponding polymer-mercaptane, followed by activation with dipyridyl disulfide.



Isolation of the synthesized oligonucleotide was performed as described above. Both oligonucleotides were purified by RP-HPLC and their structures confirmed by mass-spectrometry (MALDI-TOF).

Supporting Results

Modification of Cys residues does not compromise enzymatic activity of ASV IN

Modification of wild type (WT) ASV IN (Cys125) with cysteine-specific photo-crosslinker caused only minor decrease in the rate of the disintegration reaction. Similar results have been obtained for both I146C and R244C proteins (see Table 1). Modification of wild type (WT) ASV IN (Cys125) with cysteine-specific photo-crosslinker caused only minor decrease in the rate of the disintegration reaction. These proteins retained 20% and 2% of the WT activity prior to modification, respectively, but suffered no additional loss of activity after the modification; disintegration activity of the protein with a modified I146C is even slightly higher than before modification (Fig. S1). Similar results have been obtained for processing and joining activities for both I146C and R244C containing proteins (see Table 1).

Photocrosslinking studies were performed with APTP and BATDHP-modified Cys- and WT ASV IN proteins, complexed with Y-mer DNA substrates, representing integration intermediate, that were separately radioactively labeled at the 5' end of each of the four component strands (labeled 1-4 in Fig. S2). Because an excess of IN was used and both the DNA and IN were present at concentrations significantly higher than the IN-DNA binding constant, all the DNA was assumed to be bound to the enzyme. UV-irradiated reaction mixtures with non-modified IN and non-irradiated reaction mixtures of modified samples served as negative controls. As expected, 340-370 nm UV does not cause any significant crosslinking of DNA to IN, here demonstrated by the results with non-modified IN proteins (Fig. S2, lanes noXL).

As shown in Figure S2, the IN derivative that carries Cys at position 146 in the active site loop gave the highest yields of crosslinking, almost twice as high as the other two proteins tested. Significant differences in crosslinking among the three proteins tested confirm that the Zn-binding Cys residues in the NTD, present in all cases, were not modified and/or did not contribute significantly to the IN-DNA crosslinking. WT IN crosslinking (Cys at position 125) was largely non-specific as demonstrated by the low crosslinking yields for both modified and non-modified proteins. With all the IN proteins analyzed, oligonucleotide 4 (see Fig. S2) was preferred for crosslinking, although the derivative with Cys in the CTD (R244C) has significantly less pronounced preference for strand 4. This strand of the bound DNA is analogous to the newly joined viral and host DNA strand, catalyzed by IN. The diazirine

photocrosslinker BATDHP shows higher crosslinking efficiency than the azide-based APTP, possibly because it is 8 Å longer. Since guanidines are the best (if not the only) nucleophilic target for nitrenes to crosslink to, lower azide efficiency can be interpreted either as absence of G bases in the section of DNA within reach of APTP, or that the overall distance to DNA was too long for APTP to crosslink but still reachable for BATDHP.

We also compared photocrosslinking efficiencies of Y-mer and linear viral DNA substrates (see Fig. S3). For the IN derivatives with the following modified Cys residues: I146C, I146C/R244C, and R244C, strand 4 of the Y-mer substrate was found to be the most likely target for crosslinking. Two strands of the short linear dsDNA substrate representing viral end (L3 and L4) showed no clear preference for interaction with modified INs. Overall crosslinking efficiency to Cys at position 244 was low. This may be explained by reduced DNA-binding capacity of the IN derivative with an R244C substitution.

Chemical crosslinking of modified DNA substrates to ASV IN

The chemical crosslinking experiments were designed so that the reaction (S-S crosslinking) involved a particular chemical group on both biomolecules (IN and DNA). The thiol groups have only a minimal reactivity to other chemical moieties including buffer components and water. As there are no other targets in the reaction mixture, this type of chemical crosslinking is efficient. Unlike photoactivatable reagents the chemical crosslinkers are reactive during the whole time of an experiment. This means crosslinking can occur whenever the modified elements are close enough for interaction, which does not necessarily require that these elements be appropriately complexed. However, in the publication in which an SH- group on DNA was crosslinked to the thumb of HIV-1 RT, crosslinking was specific and depended on proper alignment of the SH-groups on the protein and the nucleic acid [10]. We used similar methods to obtain IN-DNA complexes and analyze localization of the contact sites.

Supporting Figures

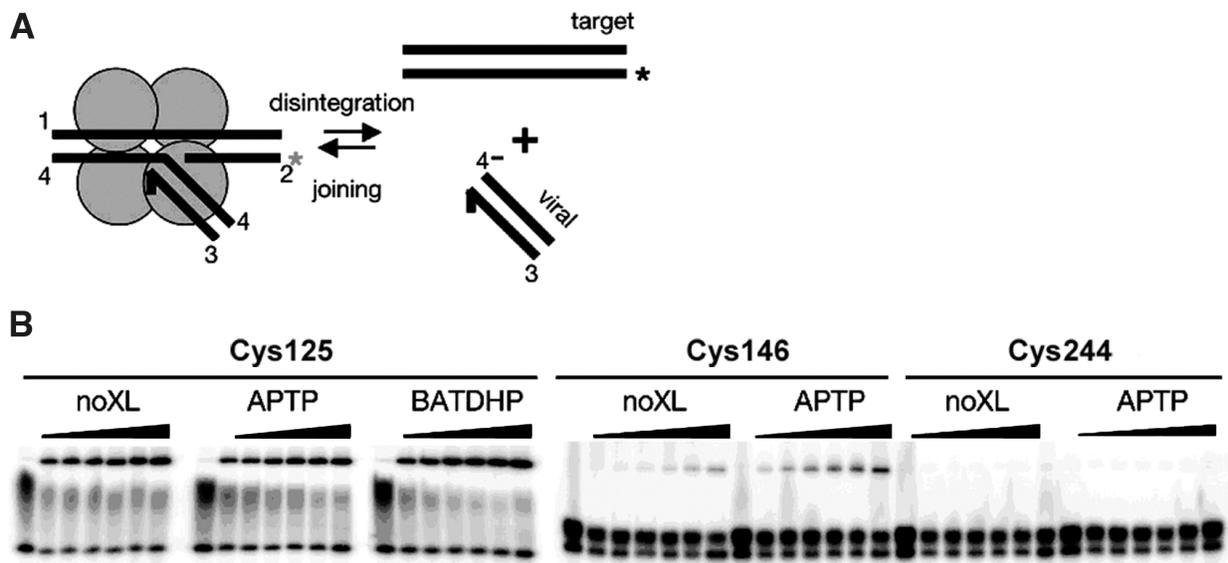


Figure S1. Disintegration reactions of the modified IN proteins with Y-mer substrate.

A) Schematic depicting integrase-catalyzed disintegration and joining reactions of the Y-mer substrate superimposed on a generic tetramer model for integrase (left). The liberated viral DNA (right) is the same DNA as seen in a pre-cleaved end, with the cleaved portion of strand 4 indicated by “4-“.

B) Disintegration activity of WT and ASV IN derivative proteins before and after modification with photocrosslinking reagents [11]. The position of modified Cys residues in the IN derivatives is noted above the bar graph. Upper band on each gel represents the 44-mer product of reaction, lower bands correspond to the 19-mer substrate (5' 32 P-labeled Y-mer strand 2*). Due to increased exposure times of the four gels on the right, the contaminating 18-mer fragments of the substrate are also visible. Reaction times were 0-1200 s; increasing time is indicated by the wedge above gel lanes. “noXL” stands for no IN modification with photocrosslinker.

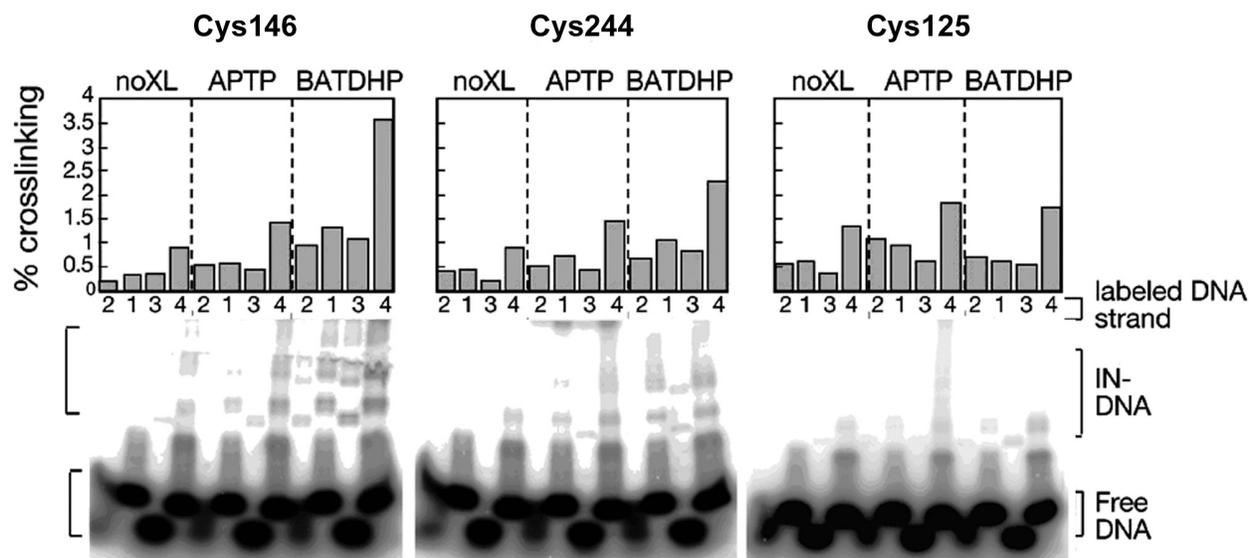


Figure S2. Strand specificity of photocrosslinking of modified Cys residues in substituted derivatives of ASV IN. A sample of gels accompanied by bar graphs with the photocrosslinking yields (%) presented. The position of modified Cys residues in the IN derivatives is noted above the bar graph. The crosslinker used is also listed above the graph (noXL stands for no IN modification with photocrosslinker). DNA strands in the Ymer substrate are labeled according to Fig. S1, and the labeled strands are indicated below each graph. Brackets on the right show the bands corresponding to IN-DNA adducts and free DNA.

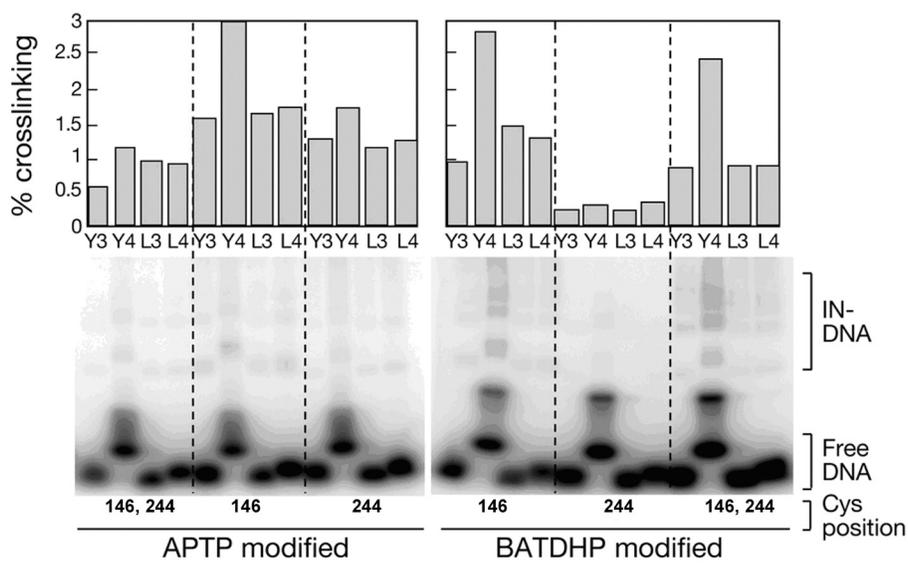


Figure S3. Comparison of the efficiency of photocrosslinking of APTP- and BATDHP-modified ASV IN derivatives to Y-mer (Y3, Y4) and linear (L3, L4) DNA substrates. Labeling is as in Fig. S2.

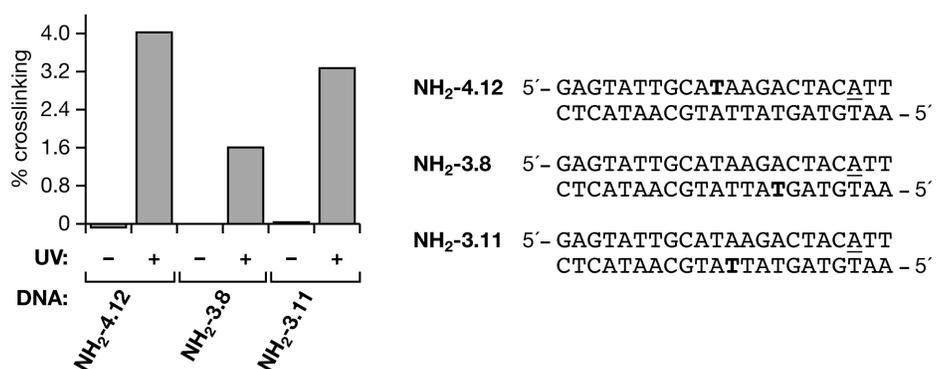


Figure S4. Comparison of photocrosslinking yields of unmodified wild type ASV IN to dsDNA substrates modified with diazirine photocrosslinkers. Positions of the modified residues are bolded and the conserved A is underlined.

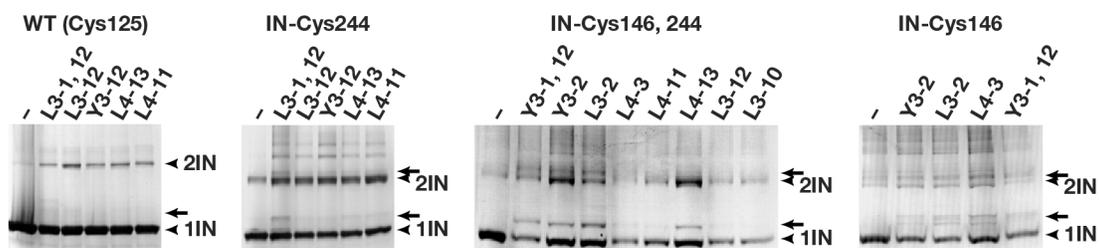


Figure S5. A sample of Coomassie-stained gels with S-S crosslinked IN-DNA complexes. The IN substrates are labeled above the lanes as in Fig. S1. Y stands for the Y-mer DNA, L for linear; letters are followed by strand numbers; thiol modified positions are shown after dash. Bands corresponding to crosslinked complexes are marked by arrows. The negative controls are marked with a dash above the left-most lane of each gel. 1IN and 2IN with arrowheads designate monomeric and dimeric IN bands, respectively. The stemmed arrows point to IN-DNA adduct bands. Molecular weight marker lanes are not shown, as the monomer (lower) and dimer (upper) strong bands provide internal calibration.

Reference List

1. Peletskaya EN, Boyer PL, Kogon AA, Clark P, Kroth H, Sayer JM, Jerina DM, Hughes SH (2001) Cross-linking of the fingers subdomain of human immunodeficiency virus type 1 reverse transcriptase to template-primer. *J Virol* 75: 9435-9445.
2. Moreland RB, Smith PK, Fujimoto EK, Dockter ME (1982) Synthesis and characterization of N-(4-azidophenylthio)phthalimide: A cleavable, photoactivable crosslinking reagent that reacts with sulfhydryl groups. *Anal Biochem* 121: 321-326.
3. Weber PJ, Beck-Sickinger AG (1997) Comparison of the photochemical behavior of four different photoactivatable probes. *J Pept Res* 49: 375-383.
4. Tate JJ, Persinger J, Bartholomew B (1998) Survey of four different photoreactive moieties for DNA photoaffinity labeling of yeast RNA polymerase III transcription complexes. *Nucleic Acids Res* 26: 1421-1426.
5. Erlanson DA, Chen L, Verdine GL (1993) Allosteric effects on enzyme action in disulfide-crosslinked DNA. *J Am Chem Soc* 115: 12583-12584.
6. Custer L, Zajc B, Sayer JM, Cullinane C, Phillips DR, Cheh AM, Jerina DM, Bohr VA, Mazur SJ (1999) Stereospecific differences in repair by human cell extracts of synthesized oligonucleotides containing trans-opened 7,8,9, 10-tetrahydrobenzo[a]pyrene 7,8-diol 9,10-epoxide N2-dG adduct stereoisomers located within the human K-ras codon 12 sequence. *Biochemistry* 38: 569-581.
7. Evans BJ, Doi JT, Musker WK (1990) Kinetics of the aqueous periodate oxidation of aliphatic disulfides and thioethers. *J Org Chem* 55: 2580-2586.
8. Dirscherl W, Weingarten FW (1951) Synthese von Homologen des Cystamins. *Justus Liebigs Ann Chem* 574: 131-139.
9. Stirchak EP, Summerton JE, Weller DD (1989) Uncharged stereoregular nucleic acid analogs: 2. Morpholino nucleoside oligomers with carbamate internucleoside linkages. *Nucleic Acids Res* 17: 6129-6141.
10. Huang H, Chopra R, Verdine GL, Harrison SC (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* 282: 1669-1675.
11. Bao KK, Skalka AM, & Wong I (2002) Presteady-state analysis of avian sarcoma virus integrase. II. Reverse-polarity substrates identify preferential processing of the U3-U5 pair. *J Biol Chem* 277: 12099-12108.