Integration of provirus

Bushman, FD., D'Eustachio, P., Gopinathrao, G., Jupe, S.

European Bioinformatics Institute, New York University Langone Medical Center, Ontario Institute for Cancer Research, Oregon Health and Science University.

The contents of this document may be freely copied and distributed in any media, provided the authors, plus the institutions, are credited, as stated under the terms of Creative Commons Attribution 4.0 International (CC BY 4.0) License. For more information see our license.

12/07/2022
Introduction

Reactome is open-source, open access, manually curated and peer-reviewed pathway database. Pathway annotations are authored by expert biologists, in collaboration with Reactome editorial staff and cross-referenced to many bioinformatics databases. A system of evidence tracking ensures that all assertions are backed up by the primary literature. Reactome is used by clinicians, geneticists, genomics researchers, and molecular biologists to interpret the results of high-throughput experimental studies, by bioinformatics seeking to develop novel algorithms for mining knowledge from genomic studies, and by systems biologists building predictive models of normal and disease variant pathways.

The development of Reactome is supported by grants from the US National Institutes of Health (P41 HG003751), University of Toronto (CFREF Medicine by Design), European Union (EU STRP, EMI-CD), and the European Molecular Biology Laboratory (EBI Industry program).

Literature references


Reactome database release: 81

This document contains 4 pathways and 6 reactions (see Table of Contents)
Integration of provirus

**Stable identifier:** R-HSA-162592

**Compartments:** nucleoplasm

**Diseases:** Human immunodeficiency virus infectious disease

For retroviral DNA to direct production of progeny virions it must become covalently integrated into the host cell chromosome (reviewed in Coffin et al. 1997; Hansen et al. 1998). Analyses of mutants have identified the viral integrase coding region (part of the retroviral pol gene) as essential for the integration process (Donehower 1988; Donehower and Varmus 1984; Panganiban and Temin 1984; Quinn and Grandgenett 1988; Schwartzberg et al. 1984). Also essential are regions at the ends of the viral long terminal repeats (LTRs) that serve as recognition sites for integrase protein (Colicelli and Goff 1985, 1988; Panganiban and Temin 1983).

The viral genomic RNA is reverse transcribed to form a linear double-stranded DNA molecule, the precursor to the integrated provirus (Brown et al. 1987, 1989; Fujiwara and Mizuuchi 1988). The provirus is colinear with unintegrated linear viral DNA (Dhar et al. 1980; Hughes et al. 1978) but differs from the reverse transcription product in that it is missing two bases from each end (Hughes et al. 1981). Flanking the integrated HIV provirus are direct repeats of the cellular DNA that are 5 base pairs in length (Vincent et al. 1990). This duplication of cellular sequences flanking the viral DNA is generated as a consequence of the integration mechanism (Coffin et al., 1997).

Linear viral DNA is found in a complex with proteins in the cytoplasm of infected cells. These complexes (termed "preintegration complexes", PICs) can be isolated and have been shown to mediate integration of viral DNA into target DNA in vitro (Bowerman et al. 1989; Brown et al. 1987; Ellison et al. 1990; Farnet and Haseltine 1990, 1991).

The development of in vitro assays with purified integrase has allowed its enzymatic functions to be elu-
The provirus is formed by two reactions catalyzed by the viral integrase: terminal cleavage and strand transfer. Studies with purified integrase have shown that it is sufficient for both 3' end cleavage (Bushman and Craigie 1991; Craigie et al. 1990; Katzman et al. 1989; Sherman and Fyfe 1990) and joining of the viral DNA to the cellular chromosome or naked target DNA (Bushman et al. 1990; Craigie et al. 1990; Katz et al. 1990). HIV integrase catalyze the removal of two bases from the 3' end of each viral DNA strand, leaving recessed 3’ hydroxyl groups (Brown et al. 1989; Fujiwara and Mizuuchi 1988; Roth et al. 1989; Sherman and Fyfe 1990). This terminal cleavage reaction is required for proper integration. It may allow the virus to create a standard end from viral DNA termini that can be heterogeneous due to the terminal transferase activity of reverse transcriptase (Miller et al. 1997; Patel and Preston 1994). In addition, the terminal cleavage step is coupled to the formation of a stable integrase-DNA complex (Ellison and Brown 1994; Vink et al. 1994). Following terminal cleavage, a recessed hydroxyl is exposed that immediately follows a CA dinucleotide. More internal LTR sites are also important for integration (Balakrishnan and Jonsson 1997; Bushman and Craigie 1990; Leavitt et al. 1992). After end processing, integrase catalyzes the covalent attachment of hydroxyl groups at the viral DNA termini to protruding 5' phosphoryl ends of the host cell DNA (Brown et al. 1987; Brown et al. 1989; Fujiwara and Mizuuchi 1988). The DNA cleavage and joining reactions involved in integration are shown in the figure below. Both the viral DNA 3' end cleavage and strand transfer reactions are mediated by single-step transesterification chemistry as shown by stereochemical analysis of reaction products (Engelman et al. 1991). Biochemical analysis of purified integrase revealed that it requires a divalent metal - either Mg2+ or Mn2+ - to carry out reactions with model substrates, that probably mediate the reaction chemistry (Bushman and Craigie 1991; Craigie et al. 1990; Katzman et al. 1989; Sherman and Fyfe 1990; Gao et al. 2004).

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-05-19</td>
<td>Authored, Edited</td>
<td>Gopinathrao, G., D'Eustachio, P.</td>
</tr>
<tr>
<td>2006-10-31</td>
<td>Reviewed</td>
<td>Bushman, FD.</td>
</tr>
</tbody>
</table>
The DNA in the Pre-initiation complex is considerably compacted relative to its length when fully extended, probably due to binding of proteins in addition to the viral integrase. These proteins are not fully clarified, due to the difficulty of biochemical analysis of small amounts of material, but candidates include the viral NC and MA proteins, and the cellular HMGA, BAF, and PSIP1/LEDGF/p75 proteins. Purified integrase is capable of carrying out the terminal cleavage and initial strand transfer reactions.

Followed by: Formation of Pre-Integration Complex (PIC)

Literature references


Concomitant with the completion of reverse transcription, the pre-integration complex is formed by shedding of some viral proteins from the viral core, and binding of cellular proteins, thereby yielding complexes capable of integration. The terminal cleavage reaction takes place in the cytoplasm, where two nucleotides are removed from each viral DNA 3' end. This serves to remove heterogeneous extra bases from the viral DNA ends occasionally added by reverse transcription, thereby yielding a homogeneous substrate for downstream steps, and also serves to stabilize the PIC. The DNA in PICs is considerably compacted relative to its length when fully extended, probably due to binding of proteins in addition to the viral integrase. These proteins are not fully clarified, due to the difficulty of biochemical analysis of small amounts of material, but candidates include the viral NC and MA proteins, and the cellular HMGA, BAF, and PSIP1/LEDGF/p75 proteins. Purified integrase is capable of carrying out the terminal cleavage and initial strand transfer reactions.

**Preceded by:** Formation of RTC with integration competent viral DNA:BANF1:HMGA1:PSIP1

**Followed by:** Integrase binds viral DNA ends

**Literature references**


<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-05-19</td>
<td>Authored, Edited</td>
<td>Gopinathrao, G., D'Eustachio, P.</td>
</tr>
<tr>
<td>2006-10-31</td>
<td>Reviewed</td>
<td>Bushman, FD.</td>
</tr>
</tbody>
</table>
Integrase binds viral DNA ends

Location: Integration of provirus

Stable identifier: R-HSA-164514

Type: transition

Compartments: cytosol

Diseases: Human immunodeficiency virus infectious disease

Upon completion of reverse transcription, the viral integrase protein (IN) becomes bound to the ends of the viral DNA. This is inferred by the fact that this is the site of integrase action, and several biochemical studies have documented integrase interactions with the terminal DNA.

Preceded by: Formation of Pre-Integration Complex (PIC)

Followed by: Terminal (3' end) cleavage of viral DNA

Literature references


Editions

<table>
<thead>
<tr>
<th>Edition</th>
<th>Updated</th>
<th>Author/Editor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-05-19</td>
<td>Authored, Edited</td>
<td>Gopinathrao, G., D'Eustachio, P.</td>
</tr>
<tr>
<td>2006-10-31</td>
<td>Reviewed</td>
<td>Bushman, FD.</td>
</tr>
</tbody>
</table>
Terminal (3' end) cleavage of viral DNA

Location: Integration of provirus

Stable identifier: R-HSA-164522

Type: transition

Compartments: cytosol

Diseases: Human immunodeficiency virus infectious disease

Prior to integration, two nucleotides are removed from each 3' end of the linear viral DNA, thereby exposing recessed 3' hydroxyls. This reaction may serve to remove heterogenous extra bases from the viral DNA end, and to stabilize the IN-DNA complex. The chemistry of cleavage is a simple hydrolysis by single-step transesterification.

Preceded by: Integrase binds viral DNA ends

Followed by: Import of PIC to the Host Nucleus

Literature references


Editions

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-05-19</td>
<td>Authored, Edited</td>
<td>Gopinathrao, G., D'Eustachio, P.</td>
</tr>
<tr>
<td>2006-10-31</td>
<td>Reviewed</td>
<td>Bushman, FD.</td>
</tr>
</tbody>
</table>
HIV can infect non-dividing cells, implying that the PIC must be able to traverse the nuclear membrane. In contrast, simple retroviruses such as MLV can only infect cells once they have passed through mitosis, potentially because they require breakdown of the nucleus to access chromosomal integration sites. The mechanism of nuclear localization is controversial. A variety of proposals have been made for nuclear localization sequences (NLS) in the PIC, but most of those have now been shown to be dispensable for HIV integration. According to a new idea from Yamashita and Emerman, it may be that the PIC is imported into the nucleus by a default pathway, while MLV PICs are retained in the cytoplasm because capsid protein is stably associated with PICs.

**Preceded by:** Terminal (3' end) cleavage of viral DNA

**Followed by:** 1-LTR circle formation

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-05-19</td>
<td>Authored, Edited</td>
<td>Gopinathrao, G., D'Eustachio, P.</td>
</tr>
<tr>
<td>2006-10-31</td>
<td>Reviewed</td>
<td>Bushman, FD.</td>
</tr>
</tbody>
</table>
Integration of viral DNA into host genomic DNA

Location: Integration of provirus

Stable identifier: R-HSA-175567

Compartments: nucleoplasm

Diseases: Human immunodeficiency virus infectious disease

Following nuclear entry, the viral preintegration complex (PIC) must select a site for integration in a host cell chromosome, and then carry out the chemical steps of the reaction.

At the chromosomal level, HIV has been found to favor active transcription units for integration. Subsequent studies established that the cellular PSIP1/LEDGF/p75 protein is important in this reaction. PSIP1/LEDGF/p75 binds tightly to HIV integrase, and also to chromatin. Knocking down PSIP1/LEDGF/p75 in cells resulted in several perturbations of integration targeting in vivo, including reduced integration in transcription units. Thus PSIP1/LEDGF/p75 has been hypothesized to act as a tethering factor that dictates at least in part the placement of HIV integration sites.

The integration target DNA is also expected to be coated with nucleosomes. Tests of integration into mononucleosomes in vitro have shown that wrapping integration target DNA actually boosts integration activity. Kinked positions on the DNA gyre are particularly favored for integration.

Integration does not take place at a unique sequence in the integration target DNA (i.e. it is not like a restriction enzyme). However, favored and disfavored primary sequences can be detected when many integration sites are aligned. Synthesis and testing of favored HIV integration sites showed that they were favored for integration by PICs in vitro.

After a target DNA is bound, the integration reactions take place via a single-step transesterification.
Integration of both ends of the viral DNA, followed by melting of the target DNA segments between the points of joining, yields single stranded gaps at each host-virus DNA junction, and a two base overhang derived from the viral DNA. The manner by which this intermediate is subsequently repaired to yield the fully integrated provirus is unclear. For many parasitic DNA replication reactions, the parasite carries out reaction steps only up to a point that the host cannot easily reverse, forcing the host to complete the job (Bushman 2001; Craig et al. 2002). For retroviral integration, it is reasonable to infer that host DNA repair enzymes complete provirus formation. DNA gap repair enzymes are known to be involved in a variety of DNA repair pathways, so their recruitment to gaps at host-virus DNA junctions is readily envisioned. Consistent with this, known gap repair enzymes have been shown to act on model host-virus DNA junctions in vitro (Yoder and Bushman, 2000).

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author-Editor</th>
<th>Reviewer</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-05-19</td>
<td>Authored, Edited</td>
<td>Gopinathrao, G., D'Eustachio, P.</td>
<td></td>
</tr>
<tr>
<td>2006-10-31</td>
<td>Reviewed</td>
<td>Bushman, FD.</td>
<td></td>
</tr>
</tbody>
</table>
The 1-LTR circle can be formed by either of two pathways. The first involves a failure to complete reverse transcription; the second, annotated here, follows the completion of reverse transcription and is mediated by cellular enzymes. In this pathway, the action of host cell homologous recombination enzymes on the long terminal repeat (LTR) termini of the viral DNA results in formation of a single LTR. This reaction probably takes place after partial or complete disassembly of the PIC to expose the viral DNA. Repair of this intermediate as in the late stages of homologous recombination pathways results in formation of the 1-LTR circle. Mutations in the Mre11/Rad50/NBS pathway influence the formation of 1-LTR circles.

**Preceded by:** Import of PIC to the Host Nucleus

**Literature references**


2-LTR circle formation

Location: Integration of provirus

Stable identifier: R-HSA-164843

Compartments: nucleoplasm

Diseases: Human immunodeficiency virus infectious disease

The formation of 2-LTR circles requires the action of the cellular non-homologous DNA end-joining pathway. Specifically the cellular Ku, XRCC4 and ligase IV proteins are needed. Evidence for this is provided by the observation that cells mutant in these functions do not support detectable formation of 2-LTR circles, though integration and formation of 1-LTR circles are mostly normal. The reaction takes place in the nucleus, and formation of 2-LTR circles has been used as a surrogate assay for nuclear transport. It has also been suggested that the NHEJ system affects the toxicity of retroviral infection.

Literature references


Editions

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-05-19</td>
<td>Authored, Edited</td>
<td>Gopinathrao, G., D'Eustachio, P.</td>
</tr>
<tr>
<td>2006-10-31</td>
<td>Reviewed</td>
<td>Bushman, FD.</td>
</tr>
</tbody>
</table>

https://reactome.org
**Autointegration results in viral DNA circles**

**Location:** Integration of provirus

**Stable identifier:** R-HSA-177539

**Compartments:** nucleoplasm

**Diseases:** Human immunodeficiency virus infectious disease

In this pathway, the viral integration machinery uses a site within the viral DNA as an integration target. This results in a covalent rearrangement of the viral DNA. The resulting DNA forms are not substrates for integration.

It has been suggested that the cellular BAF protein binds to viral DNA and diminishes autointegration by coating and condensing the viral DNA, thereby making it a less efficient integration target.

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-05-19</td>
<td>Authored, Edited</td>
<td>Gopinathrao, G., D'Eustachio, P.</td>
</tr>
<tr>
<td>2006-10-31</td>
<td>Reviewed</td>
<td>Bushman, FD.</td>
</tr>
</tbody>
</table>

[https://reactome.org](https://reactome.org)
Table of Contents

Introduction 1

Integration of provirus 2
  ▶ Formation of RTC with integration competent viral DNA:BANF1:HMGA1:PSIP1 4
  ▶ Formation of Pre-Integration Complex (PIC) 5
  ▶ Integrase binds viral DNA ends 7
  ▶ Terminal (3' end) cleavage of viral DNA 8
  ▶ Import of PIC to the Host Nucleus 9
  ▶ Integration of viral DNA into host genomic DNA 10
  ▶ 1-LTR circle formation 12
  ▶ 2-LTR circle formation 13
  ▶ Autointegration results in viral DNA circles 14

Table of Contents 15