Chapter

Origin of DNA Repair in the RNA World

Harris Bernstein and Carol Bernstein

Abstract

The early history of life on Earth likely included a stage in which life existed as self-replicating protocells with single-stranded RNA (ssRNA) genomes. In this RNA world, genome damage from a variety of sources (spontaneous hydrolysis, UV, etc.) would have been a problem for survival. Selection pressure for dealing with genome damage would have led to adaptive strategies for mitigating the damage. In today's world, RNA viruses with ssRNA genomes are common, and these viruses similarly need to cope with genome damage. Thus ssRNA viruses can serve as models for understanding the early evolution of genome repair. As the ssRNA protocells in the early RNA world evolved, the RNA genome likely gave rise, through a series of evolutionary stages, to the double-stranded DNA (dsDNA) genome. In ssRNA to dsDNA evolution, genome repair processes also likely evolved to accommodate this transition. Some of the basic features of ssRNA genome repair appear to have been retained in descendants with dsDNA genomes. In particular, a type of strand-switching recombination occurs when ssRNA replication is blocked by a damage in the template strand. Elements of this process appear to have a central role in recombinational repair processes during meiosis and mitosis of descendant dsDNA organisms.

Keywords: RNA world, RNA virus, recombination repair, copy-choice, synthesis-dependent strand annealing (SDSA), DNA repair, archaea, genome damage, strand-switching, self-replication, single-stranded RNA

1. Introduction

Protocellular organisms may have come into existence 2.5 to 3.5 billion years ago [1, 2]. Woese [3] proposed that the genomes of the early protocellular forms of life were individual strands of RNA rather than DNA, and that these RNA strands were present as separate genome segments, rather than being linked together end-to-end as is generally the case for genes in DNA. The idea that, during an early period in the evolution of life, genetic information was stored and transmitted solely by RNA molecules has come to be known as the "RNA world hypothesis." This hypothesis is currently being tested by many investigators. Of particular significance, Horning and Joyce [4] have demonstrated that the replication of genetic information and its conversion into functional molecules can be accomplished with RNA in the complete absence of protein. RNA molecules with catalytic activity are called ribozymes. An RNA ribozyme developed by Horning and Joyce can act as an RNA polymerase to replicate RNA [4].

Persistence and replication of even the simplest forms of RNA life must have depended on preserving the information content of the RNA genome from damage (a form of informational noise). Damage to the RNA genome likely occurred in a variety of ways including spontaneous hydrolysis, exposure to UV light and exposure to reactive chemicals. Natural selection would have acted to promote the evolution of RNA sequences that allowed solutions to this problem of informational noise. While free living organisms with ssRNA genomes are unknown in today's world, viruses with ssRNA genomes are currently common. The present day ssRNA viruses also need to cope with informational noise in the form of damage to their RNA genome. Therefore, such ssRNA viruses can serve as models for understanding the adaptive solutions that early ssRNA protocells may have developed for coping with genome damage. Numerous ssRNA viruses have been shown to be capable of exchanging sequence information between individual genomes within an infected cell [5]. This information exchange, or genetic recombination, can occur by reassortment of genome segments or during genome replication by a process of strand-switching to form a progeny genome with information from two parental genomes. The process of strand-switching is often referred to as "copy-choice" recombination. The term "copy-choice" embodies the idea of template-switching during genome replication, although the term was introduced before the DNA/RNA nature of genetic information was understood. Lederberg [6] and Bernstein [7] were among the first to explicitly propose copy-choice mechanisms of recombination. The two recombination processes, segment reassortment and copy-choice, allow the formation of an undamaged progeny genome even when one or both parental genomes contain damage. In the sense that both segment reassortment and copy-choice restore genetic sequence information that is damaged in the parental genomes, these are informational repair processes. Although information is restored in progeny, the parental genomes may retain their physical damage. Thus when "repair" is discussed at the level of ssRNA organisms it is the genetic information content of damaged parental genomes that is restored or "repaired" during formation of the progeny genome.

The role of RNA segment reassortment in genome repair is discussed by Bernstein et al. [8] and the role of copy choice recombination in an RNA genome repair is discussed by Hu and Temin [9].

As the early protocells with RNA genomes evolved they likely went through a series of adaptive transitions that eventually led to the double-stranded DNA (dsDNA) genome. The archaea are a group of prokaryotes with a dsDNA genome that likely evolved prior to the emergence of eukaryotes. These organisms are capable of a process, genetic transformation, during which cells exchange DNA to repair DNA double-strand breaks via homologous recombination [10]. In eukaryotes, during meiosis and mitosis, most recombination events occur by a repair process termed "synthesis-dependent strand annealing" (SDSA) [11] that is basically a form of copy-choice recombination (see Section 6.1.). In addition, single-strand damages that block the movement of the DNA polymerase during replication can be repaired by a mechanism that includes copy-choice recombination [12, 13]. Thus strand-switching copy-choice mechanisms that likely emerged in early ssRNA protocells appear to have evolved into fundamental processes for maintaining the information content of dsDNA genomes.

While the capability for recombinational repair is retained as a major mechanism for dealing with DNA damages, organisms with a dsDNA genome, including humans, have also evolved other repair processes that take advantage of the duplex nature of the DNA genome [14]. For such organisms, damages in one strand can be repaired by removal of the damaged section and its replacement by copying information from the other strand, as occurs in the well-studied processes of mismatch repair, nucleotide

excision repair and base excision repair [14]. Other processes for dealing with DNA damages in organisms with DNA genomes include direct reversal of UV photolesions and alkylated bases, repair of DNA crosslinks by Fanconi anemia proteins, and a mechanism for tolerating damages termed translesion synthesis [15].

The aim of this review is to outline how genome repair processes emerged in the earliest evolved protocells that likely had RNA genomes, and how these processes further evolved in the transition from the RNA world to the DNA world.

2. Genome repair in the RNA world

Since the actual sequence of evolutionary adaptive events in the RNA world that gave rise to genome repair occurred in organisms that are probably long extinct, and it is unlikely that events at the nucleic acid level are preserved in the fossil record, the sequence of evolutionary events proposed here is necessarily speculative. However, the proposed evolutionary sequence is based on the established activities of extant RNA viruses. These activities are reviewed in sections 2.1, 2.2, 3 and 4. Thus it is assumed, as discussed by Bernstein et al. (pgs. 342-345) [8], that the adaptations that extant RNA viruses use to repair genome damage can illuminate how early life in the RNA world also coped with genome damage.

In early protocellular organisms the genome is thought to have consisted of ssRNAs (genes) that formed folded structures with catalyic activity (ribozymes) [16]. If two or more such ssRNAs were present in a protocell they presumably functioned interdependently to promote the viability and reproduction of the protocell. A key ribozyme in early protocellular organisms would likely have been a polymerase that could catalyze RNA replication [4]. A persistent problem for early protocellular organisms would probably have been damage to their ssRNA genomes. The damaging stresses on protocellular organisms likely would have included hydrolytic reactions, exposure to UV light and interaction with reactive chemicals in the environment. For example, Sagan [17] analyzed the flux of solar UV light that penetrated the earth's primitive reducing atmosphere. His analysis indicated that unprotected microorganisms of the type existing today would receive a mean lethal dose at 2600 angstroms within 0.3 seconds and that this vulnerability could have posed a major problem during the early evolution of life. A protocell that has only one copy of each ssRNA (a haploid protocell) would be very vulnerable to damage, since damage to even one base in a ssRNA sequence might be lethal to the protocell by either blocking replication of the ssRNA or interfering with an essential ssRNA ribozyme function [8].

One possible adaptation for dealing with genome damage would be to maintain two or more copies of each ssRNA gene in each protocell, yielding a diploid or polyploid state. Genome redundancy would allow replacement of a damaged gene by an additional replication of an undamaged homologous gene. However, for a simple protocellular organism, the proportion of available resource budgeted to the maintenance of two or more genomes would have been a large portion of its total resource budget. When resources are limited, the protocell's reproductive rate would likely be inversely related to ploidy number. The fitness of the protocell would be diminished by the costs of genome redundancy. Coping with damage to the ssRNA genome while minimizing the costs of genome redundancy would likely have been a fundamental problem in the early evolution of cellular life [8].

When the costs of maintaining genome redundancy verses the costs of genome damage were balanced against each other in a cost–benefit analysis, it was found that under a wide range of conditions the selected strategy would be for each

protocell to be haploid, but to periodically fuse with another haploid protocell to form a transient diploid [18]. This strategy allows the haploid state to be retained to maximize reproductive rate, while the periodic fusions would allow otherwise lethally damaged protocells to be mutually reactivated. Reactivation can occur if at least one undamaged copy of each ssRNA gene is present in the transient diploid and this leads to production of a viable progeny protocell. In order for two (rather than just one) viable progeny protocells to be produced, an extra replication of the gene(s) homologous to damaged gene(s) would have to occur before division of the fused diploid protocell. The process of recovering from potentially lethal damage in one ssRNA genome by reassorting information with another homologous ssRNA genome can be regarded as a primitive form of genome repair [8, 18]. This proposed cycle for coping with genome damage, although hypothetical, is based on the way that ssRNA viruses with segmented RNA genomes deal with genome damage as discussed below in Section 2.1.

The events that contributed to the evolution of genomic repair in ssRNA protocells can also be viewed as an early stage in the evolution of sexual reproduction since these events include the coming together of two genomes from separate parents to generate progeny genomes containing shared genetic information [18].

2.1 Recombination in influenza virus and hantavirus

Influenza virus (Family *Bunyavirales*) is an example of a virus with a segmented ssRNA genome (**Figure 1**). Influenza virus has a genome comprised of eight physically separated ssRNA segments [19]. These eight segments of single-stranded RNA code for seven virion structural proteins and three non-structural proteins. During infection of a host cell by two viruses, recombinant progeny can be formed as the result of exchange of segments of the virus ssRNA, a process termed reassortment [19].

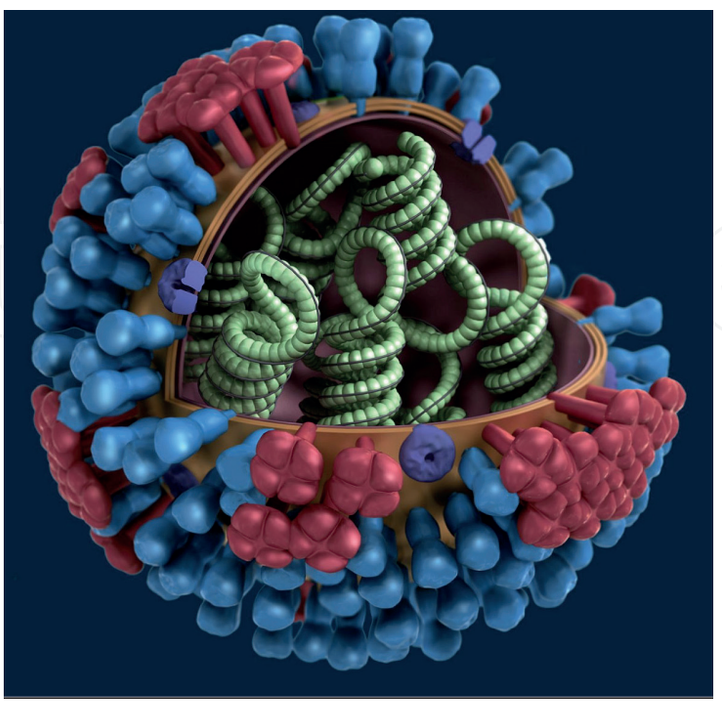


Figure 1.
Influenza virus. An enveloped virus with an outer lipid membrane and glycoprotein "spikes." Influenza A or B viruses have eight genome segments inside the virion. https://pixnio.com/science/microscopy-images/influenza/3-dimensional-model-of-influenza-virus In the public domain.

Upon infection, influenza virus induces a host response involving increased production of reactive oxygen species, and this can damage the virus genome [20]. Consider two individual viruses each with a lethal damage in its genome. If either of these viruses infects a host cell the infection aborts and no progeny viruses are produced. However, if these two damaged viruses infect the same host cell, the multiple infection may lead to reactivation (production of viable progeny). This phenomenon is known as "multiplicity reactivation" and is thought to reflect acts of recombination that allow an undamaged genome to be reconstituted from damaged ones [21]. Multiplicity reactivation has been demonstrated in influenza virus infections after induction of RNA damage by UV-irradiation [22] and ionizing radiation [23]. In these studies, recombination by reassortment of genome segments likely played a role in the observed multiplicity reactivation.

Hantaviruses (Order *Bunyavirales*; Family *Hantaviridae*), another group of segmented ssRNA viruses, are also able to undergo reassortment [24, 25]. Reovirus (Family *Reoviridae*), a segmented double-stranded RNA virus, can also undergo multiplicity reactivation after its genome is damaged by exposure to UV light [26]. Substantial evidence in model virus systems indicates that multiplicity reactivation is a recombinational repair process for overcoming a variety of types of genome damage (reviewed in [27, 28]). If, under natural conditions, virus survival is ordinarily vulnerable to oxidative or other damage, then multiplicity reactivation likely acts as an adaptive genomic repair process.

Recombination by reassortment is a simple way of restoring an undamaged genome from multiple lethally damaged genomes and thus is a primitive form of genomic repair. Lehman [29] has reviewed evidence supporting the view that recombination is an evolutionary development as ancient as the origins of life.

In addition to the role of recombination in genome repair, recombination also has a role in viral evolution by generating new genetic combinations that can be tested by natural selection. An infrequent new genetic combination may be selectively advantageous. However, RNA is very vulnerable to damage. Because of the reactivity of the oxygen and nitrogen atoms of the nucleobases [30], RNA molecules are especially susceptible to certain types of chemical damage from sources such as reactive oxygen species, UV light, and alkylating agents; and the oxygen atoms of the ribose and the phosphodiester backbone are also vulnerable to chemical damage [30]. In early protocells, repair of RNA genome damage likely provided a considerable and immediate selective advantage while new recombinant genetic combinations may have been adaptively beneficial only infrequently.

2.2 Intragenic recombination in segmented ssRNA influenza virus and hantavirus

In influenza virus infections, genome segment reassortment is not the only mechanism of recombination. Intragenic homologous recombination can also occur between a pair of homologous viral genes [31]. Homologous recombination occurs by template-switching (copy-choice) during viral genome replication [32].

In addition to influenza viruses, ssRNA hantaviruses are also capable of recombination by both segment reassortment and by homologous recombination [33].

In the evolution of repair processes in the RNA world, template-switching (copy-choice) recombination was likely an important advance since it allows two damaged homologous genes to generate an undamaged homolog. However, at present there is insufficient evidence available to determine whether copy-choice recombination emerged before or after the emergence of genome segment reassortment as a mechanism of genome repair [31].

3. Repair of RNA genomes by copy-choice recombination

3.1 Copy-choice recombination

Figure 2 indicates how an accurate undamaged progeny single-stranded genome can be generated from a damaged parental genome by strand-switching (copy-choice) recombination. As shown in this **Figure 2**, (1) during synthesis of a progeny strand by a replicative polymerase, a damage in the (green) template strand (strand being copied) blocks polymerase progression. (2) If another (orange) homologous template is available, the polymerase may switch templates, thereby bypassing the damage. (3) The newly synthesized strand may then release from the second template strand. (4) The newly synthesized strand can return and pair with the original template. (5) The polymerase may then complete the replication using the original template. (6) These steps can generate a new recombinant genome without damage [9, 34].

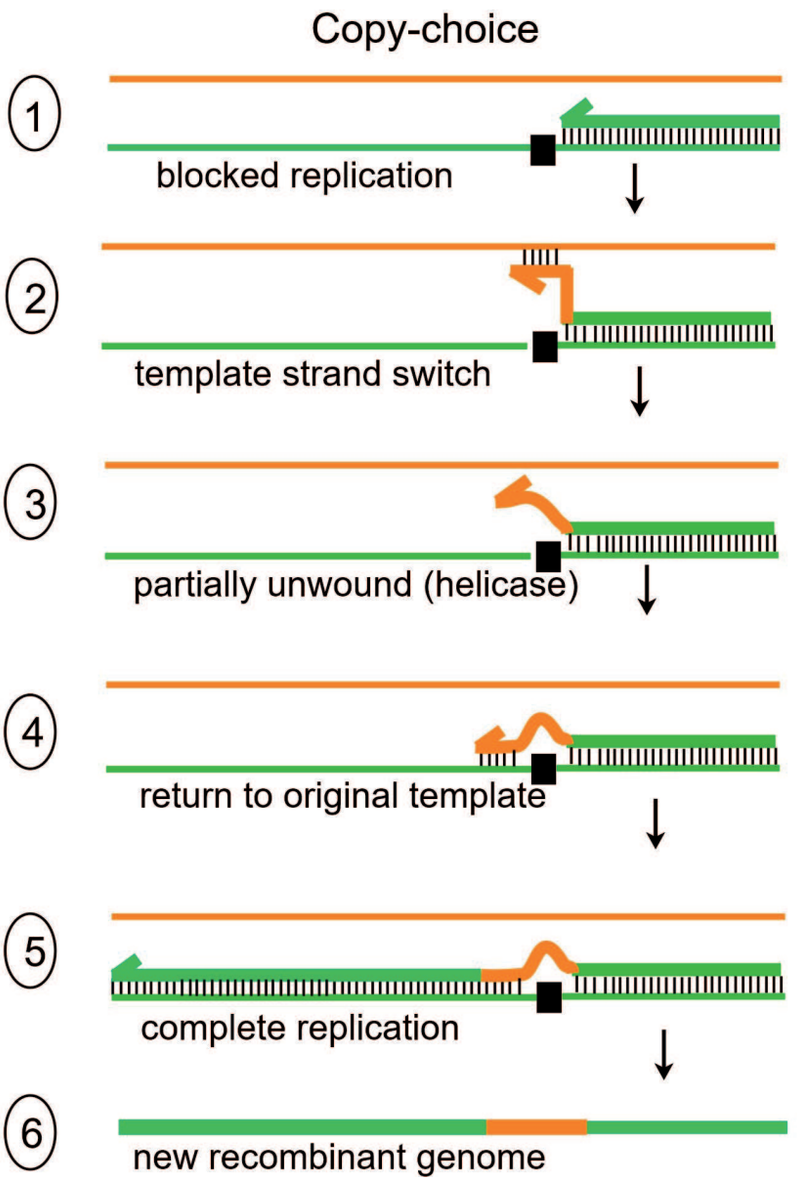


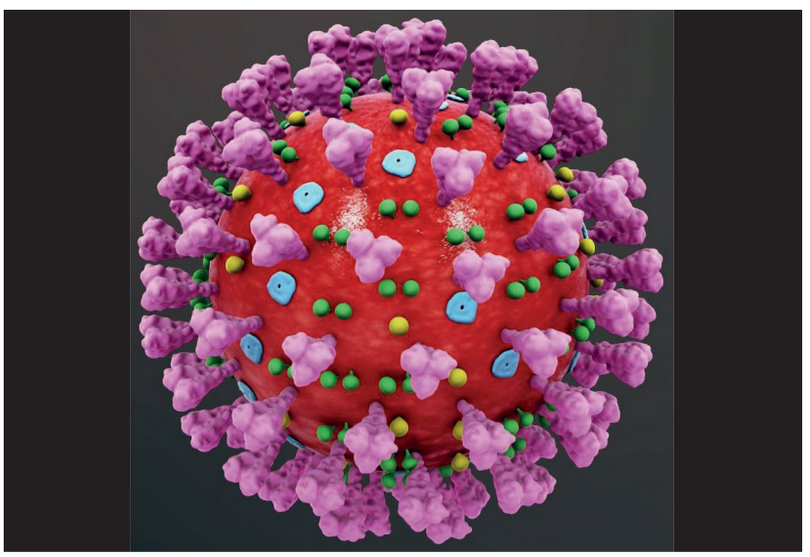
Figure 2.
Copy-choice recombination.

3.2 Poliovirus and coronavirus

Poliovirus (Family *Picornaviridae*; Genus *Enterovirus*) is a positive ssRNA ((+) ssRNA) virus that can undergo genetic recombination when there are at least two ssRNA viral genomes in the same host cell. RNA recombination is considered to be a major driving force in determining the course of poliovirus evolution [35]. RNA-dependent RNA polymerase (RdRp), an enzyme encoded in the viral genome, catalyzes genome replication. Kirkegaard and Baltimore [34] presented results strongly supporting a copy-choice mechanism for RNA recombination for poliovirus. By this mechanism the RdRp switches between (+)ssRNA templates during synthesis of the progeny negative strand (−)ssRNA (**Figure 2**). Recombination in RNA viruses is considered to be an adaptive mechanism for maintaining genome integrity [36].

To regenerate the next generation of (+)ssRNA strands, the (−)ssRNA strands are also copied and this may also be accompanied *infrequently* by strand switching [34].

When cells are infected by two or more viruses containing genome damage the viruses may undergo multiplicity reactivation. Polioviruses are able to undergo



Outer conformation of the coronavirus, above

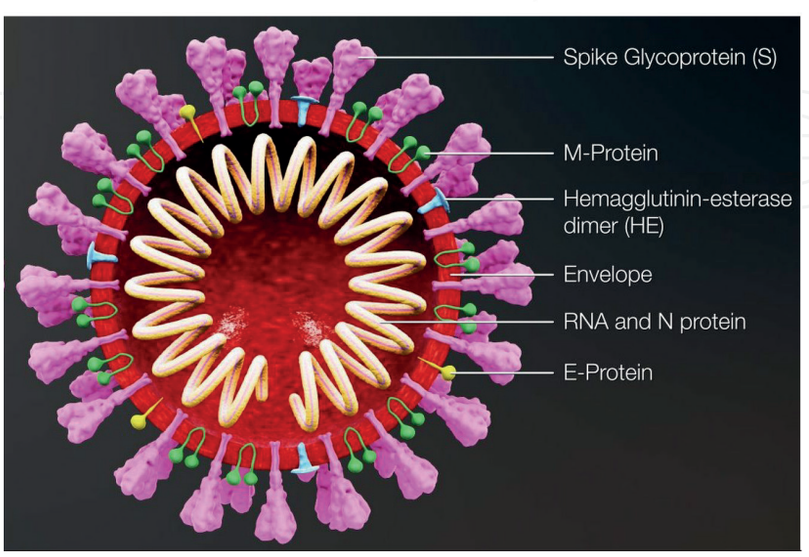


Figure 3.
Coronavirus. Modified from https://commons.wikimedia.org/wiki/File:3D_medical_animation_coronavirus_structure_vie.png with license https://www.scientificanimations.com/CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0)

multiplicity reactivation [37]. That is, when polioviruses were irradiated with UV light and then allowed to infect host cells at a multiplicity of two or greater, viable progeny are produced at UV doses that inactivate the virus in single infections. As noted above, multiplicity reactivation occurs in various different virus systems, and has been shown to be a form of recombinational repair [27, 28].

Coronaviruses (Family *Coronaviridae*) (see **Figure 3**) are (+)ssRNA enveloped viruses. The genome size of coronaviruses ranges from about 26 to 32 kilobases, one of the largest among RNA viruses. They have characteristic club-shaped spikes that project from their surface, which in electron micrographs create an image reminiscent of the solar corona, from which their name derives.

RNA recombination appears to be a major driving force in the evolution of (+) ssRNA coronaviruses. Recombination contributes to genetic variability within a coronavirus species, the capability of a coronavirus species to jump from one host to another and, infrequently, the emergence of a novel coronavirus [38]. The mechanism of recombination in coronaviruses likely involves template-switching during genome replication [38]. Also, the (+)ssRNA plant carmoviruses and tombusviruses frequently undergo recombination by RdRp template-switching (copy-choice) [39]. A key step in the evolution of repair in the RNA world appears to have been the emergence of template-switching (copy-choice) recombination as a major mechanism for dealing with genome damage.

4. Reverse transcription of the RNA genome to DNA in HIV

Human immunodeficiency virus (HIV (Family *Retroviridae*) (**Figure 4**) is a positive single-stranded RNA ((+)ssRNA) virus. Each HIV virus particle encapsidates two (+)ssRNA genomes.

During infection of a host cell, genome replication is catalyzed by reverse transcriptase, an RNA-dependent DNA polymerase [40]. During reverse transcription, recombination between the two genomes can occur [9]. The reverse transcriptase can switch between the two parental RNA genomes by copy-choice recombination [40], and such events may occur throughout the genome. Thus the two infecting genomes from each virus can cooperate to form a complementary negative single-strand DNA copy that has recombined information from the two parental RNA genomes. Recombination is necessary for efficient HIV replication and the maintenance of genome integrity [40]. During each replication cycle, from 5 to 14 recombination events may occur per genome [41]. The recombination events are "clustered" so that one recombination event is correlated with another that is close by. This clustering is apparently caused by correlated template-switches, known as high negative interference, during minus-strand DNA synthesis [42]. That is, once a switch is made from template *a* to template *b*, then another switch is made very soon (not at some random time) back to template *a*. Template-switching in HIV is considered to be a repair mechanism for salvaging damaged genomes that is essential for maintaining genome integrity [9, 40].

After the first single strand DNA copy is synthesized, another round of replication generates a duplex DNA molecule which can integrate into the host DNA genome to form a provirus [9].

4.1 HIV recombination can sometimes produce genetic variation

Recombination of the viral genomes can introduce genetic variation among progeny HIV that contributes to the evolution of resistance when humans are treated with anti-retroviral therapy [43]. Viral genome recombination may also

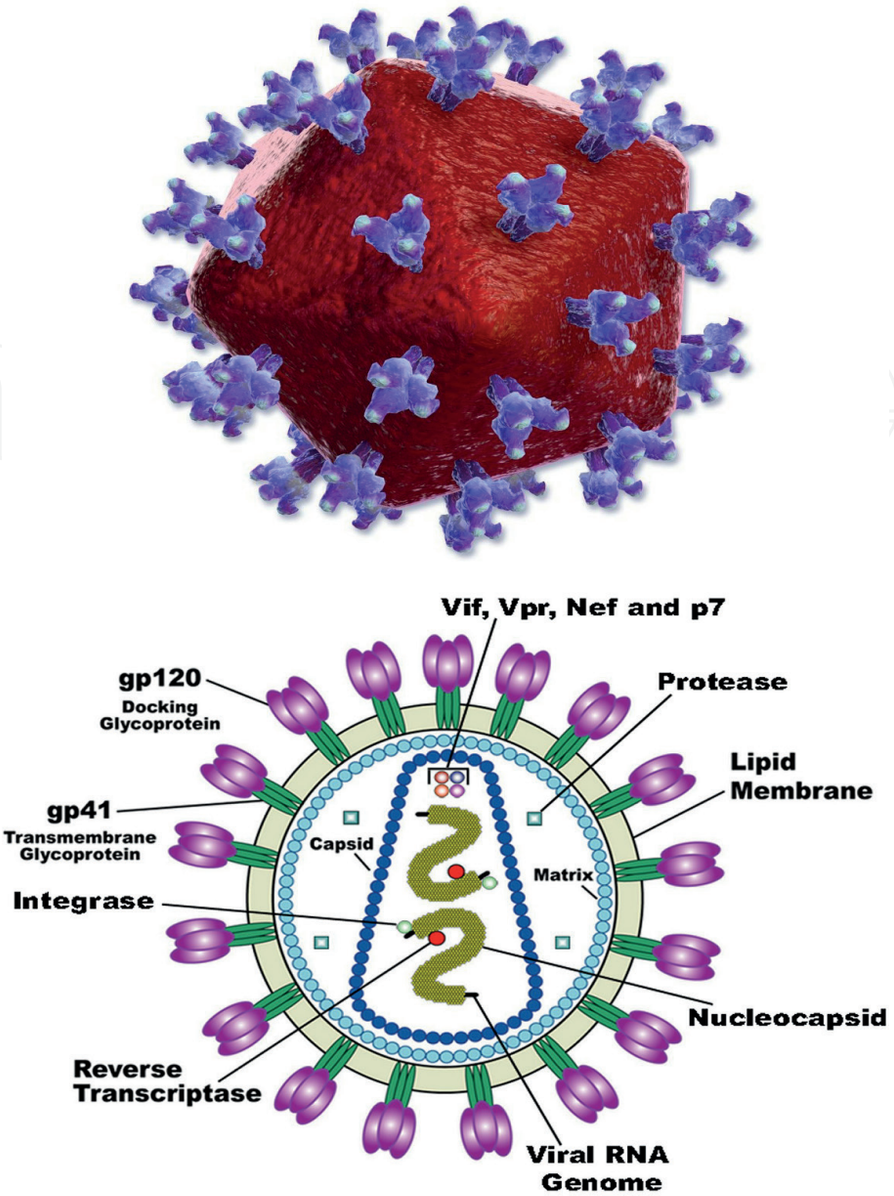


Figure 4.
Human Immunodeficiency virus (HIV). Top image indicates outer conformation of the virion. Lower image shows the two RNA genomes present within the virion, the reverse transcriptase and other components of the virion. Top image: https://commons.wikimedia.org/wiki/File:HIV.png BruceBlaus/CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0) Bottom image: https://commons.wikimedia.org/wiki/File:HI-Virion-en.png US National Institute of Health (redrawn by en:User:Carl Henderson) / Public domain.

play a role in overcoming the immune defenses of the human host. The sequence of events necessary to produce genetic variation by recombination that is adaptively beneficial to HIV are considered next.

For an adaptive benefit of genetic variation to be realized, the two RNA genomes contained in an individual infecting virus particle would have to be derived from separate progenitor viruses of differing genetic constitution. In general, only viruses that have packaged two genetically different RNA genomes can produce a recombinant genome with a genotype distinctly different from that of its parents [44] . For this to occur multiple events are required [44]. These events are: (1) A human host cell would need to be infected by two viruses of genetically different lineages, and the genomes of these two different viruses would have to produce progeny genomes. (2) Two different progeny RNA genomes produced from such an infection would have to be co-packaged into the same progeny virus particle. (3) When this progeny virus infects a new host cell, template-switching would have to occur during reverse

transcription to generate a recombinant DNA copy. (4) The recombinant DNA would then need to integrate into the DNA genome of the infected cell. (5) The recombinant provirus would next have to be able to produce replication-competent virus progeny for the impact of the recombination to be observed.

How often cells in HIV patients are infected by more than one HIV (double-infection) is not known, and it is unknown how often mixed packaging occurs under natural conditions [44, 45]. As discussed above, from 5 to 14 strand-switching recombination events occur in each infection cycle. These events, in most cases, occur between genomes with the same genetic constitution. Thus it is apparent that although recombination can, under some circumstances, produce variation that is adaptive, the great majority of recombination events do not produce significant adaptive variation.

4.2 Recombination as a repair process

Infection by HIV results in chronic ongoing inflammation associated with reactive oxygen species production [46]. Thus a strategy for dealing with oxidative damages to the HIV genome would be adaptively beneficial. Each HIV particle contains two homologous templates, rather than one. Temin [9] considered it likely that recombination is an adaptation for repair of damaged RNA genomes. Also, template-switching by the reverse transcriptase was suggested by Bonhoeffer et al. [47] to be a repair process for dealing with breaks in the ssRNA genome. Copy-choice recombination by the reverse transcriptase could produce a DNA copy of the genome that is free of damage even if both parental ssRNA copies in each virus are damaged. This benefit of recombination can be realized at each infection cycle even if, as is usually the case, the two genomes do not differ, or are closely similar genetically, and little if any new genetic variation will be produced [9, 45]. If recombination in HIV infections is primarily an adaptation for genome repair, the generation of recombinational variation would be an occasional natural consequence, but not the principle driving force, for the evolution of template-switching [47].

4.3 HIV as a model for the transition from ssRNA to dsDNA genomes

Early organisms may have evolved through a stage, like HIV, where their genome in the form of ssRNA was replicated to form a hybrid RNA: DNA duplex which upon further replication formed dsDNA. A laboratory evolved RNA polymerase ribozyme that synthesizes RNA has also been shown to act as a reverse transcriptase to synthesize DNA [48]. A ribozyme like this may have evolved in nature and been instrumental in the transition from the RNA to the DNA world. It could have arisen as a secondary function of an RdRp.

While oxidative stress appears to be a principle damaging stress for the HIV genome, the damaging stresses on organisms that were undergoing the early evolutionary transition from RNA to DNA genomes would likely have been different. The genome damages in the transition from RNA to DNA genomes could have arisen, as described above, from hydrolytic reactions, UV light or environmental reactive chemicals, but undoubtedly there would have been some kinds of significant damages. Thus during the transition from the RNA world to the DNA world there was very probably a continuous need to cope with genome damage. The copy-choice mechanisms that had a repair function in the RNA world may have continued to operate as repair functions during the transition to the dsDNA world. The selective pressure of genome damage on genome repair as the genetic material transitioned from RNA to DNA is discussed further in Bernstein et al. (pgs. 342-345) [8].

5. Recombination in archaea acts in DNA repair

In the previous sections it was proposed that genome repair processes emerged in the RNA world and that, after going through several evolutionary stages, such repair processes were present in organisms with DNA genomes. The archaea are single-celled microorganisms whose genome is DNA. These organisms are regarded as descendants of a form of life that arose subsequent to organisms with RNA genomes but prior to eukaryotes [49].

The evolution of the eukaryotic cell appears to trace back to the establishment of a symbiotic relationship between a host anaerobic archaeal cell and an internalized bacterium capable of aerobic metabolism [50]. The eukaryotic cell emerged at least 1.5 billion years ago [51]. Eukaryotic genes of archaeal origin appear to have a more central role in basic cellular functions than genes of eubacterial origin [49]. Thus the manner in which present day archaea deal with genome damage may throw light on how genome repair processes that arose in the RNA world became adapted for repair in both the archaeal and the eukaryote DNA world.

Recent findings show that cells of archaeal species, particularly *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius,* under stressful environmental conditions that cause DNA damage, aggregate and transfer DNA from one cell to another through direct contact [52, 53]. Exposure of *S. solfataricus* to UV irradiation strongly induces type IV pili formation which facilitates cellular aggregation [54, 55]. This induced cellular aggregation mediates intercellular chromosome marker exchange with high frequency. UV irradiated cultures were found to have recombination rates exceeding those of uninduced cultures by up to three orders of magnitude. The UV-inducible DNA transfer process and subsequent homologus recombination are considered to represent a repair mechanism for maintaining chromosome integrity [54, 56, 57]. Also in *S. solfataricus*, exposure to bleomycin or mitomycin C, agents that cause double-strand breaks and other damages, induces cellular aggregation [54]. In *S. acidoclaldarius,* genes that facilitate DNA transfer are upregulated by DNA damaging UV irradiation [52]. DNA damage can be lethal to a cell unless repaired. DNA transfer between neighboring archaeal cells appears to be an adaptation for aiding survival of nearby (and likely genetically related) damaged cells by facilitating recombinational repair.

The repair capabilities of archaea suggest that ancestral organisms arising early in the DNA world underwent processes that allowed DNA damage in one cell to be repaired by transfer of DNA sequence information from a neighboring cell in order to facilitate recombinational repair.

6. Eukaryotes

Eukaryotes are capable of several different types of DNA repair process:

a. The DNA damage may be enzymatically directly reversed. There are three known direct reversal mechananisms (Yi C) [58]: (1) Photolyase catalyzed direct reversal of UV light-induced photolesions; (2) O^6 alkylguanine-DNA alkytransferase catalyzed direct reversal of a set of of O^6 alkylated DNA damages; and (3) direct reversal of N-alkylated base adducts by AlkB family dioxygenases. Direct reversal mechanisms are specific for a small subset of DNA damages and thus have limited applicability.

b. Single-strand damages may be excised and the proper information restored by copying the other undamaged strand. This can occur by any one of several

well-studied processes. These include mismatch repair (MMR), nucleotide excision repair (NER) and base excision repair (BER) [14]. These processes appear to have arisen in the archaea [59], but are most well understood in the eukaryotes. This option was not available to organisms with ssRNA genomes because the double-stranded state exists only transiently during replication. In any case the enzymes that carry out such repair processes in organisms with DNA genomes are not known to be encoded in the ssRNA virus genomes. Thus this type of mechanism was not likely present during the early evolutionary stages in ssRNA genome containing organisms.

c. Double-strand damages in double-stranded DNA, such as double-strand breaks, can be repaired without the presence of an homologous template by such processes as non-homologous end joining (NHEJ) and microhomology mediated end joining (MMEJ). These processes depend on the duplex nature of DNA but not on strict homology. NHEJ can be accurate if the ends of the DNA in double-strand breaks do not need processing. However, if the ends need processing before rejoining then mutations are very likely to be introduced [60]. MMEJ is inaccurate and is always associated with a DNA deletion [61]. Thus these processes are inaccurate and generate mutations and are not applicable to ssRNA genomes.

d. Homologous recombinational repair is possible when two templates are present and adjacent. Such repair may occur for various types of DNA damage. For double-strand breaks in mitosis, homologous recombinational repair, either by the less common breakage and exchange mechanism or by the more frequently used SDSA (copy-choice) mechanism [11], are the only accurate forms of repair available. Template switching can occur during mitosis when two sister chromatids are present and adjacent after DNA synthesis and before cell division.

During meiosis homologous chromosomes originating from different parents align intimately with each other. This is followed by transfer of sequence information between homologs, homologous recombination. The main mechanism is SDSA (copy-choice recombination), a central characteristic of meiosis (see Section 6.1). Less frequent homologous recombination by breakage and exchange of chromosomes also occurs during meiosis.

Copy-choice recombination is also an important general mechanism for dealing with DNA damages that block the movement of the DNA polymerase during DNA replication (see Section 6.2).

6.1 Meiotic and mitotic recombination

The results of numerous studies in a wide range of eukaryotes indicate that during meiosis a variety of DNA damages are repaired by recombinational repair (reviewed in [62]). In somatic cells, mitotic recombination also facilitates DNA repair. Molecular models of recombination have been revised over the years as relevant evidence accumulated. Our current understanding of recombination reflects the work of several groups of investigators that have provided evidence that SDSA is a major mechanism of recombination [11, 63–65]. Furthermore, SDSA is a type of copy-choice mechanism since it involves switching from one template to another during strand synthesis and the return to the original template after a short distance (compare **Figure 5** to **Figure 2**).

Figure 5 illustrates the series of steps that occur by the meiotic SDSA process in the repair of a double-strand break (DSB) in one chromosome using information

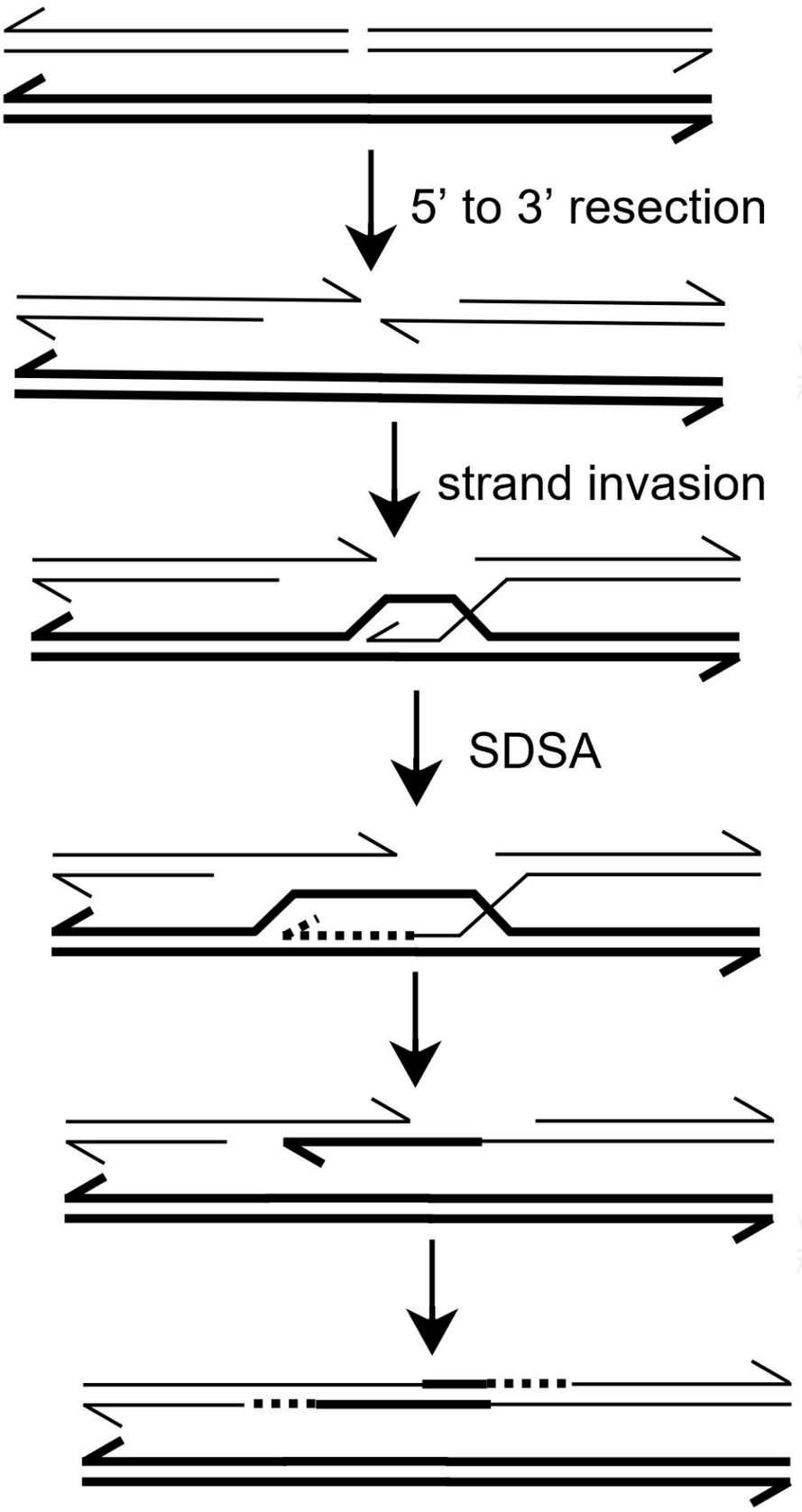


Figure 5.
Synthesis-dependent strand annealing (SDSA) in the repair of a double-strand break.

from an adjacent undamaged homologous chromosome. As shown in the figure, the steps include strand invasion by a broken strand to form a D-loop, the further extension of the strand by DNA synthesis, and then the reassociation of the transferred strand with its original pairing partner. These strand-switching and DNA synthesis events associated with repair of a damage are similar to the copy-choice

recombination described above for ssRNA viruses. Thus a central feature of eukaryotic recombination in meiosis and mitosis, strand-switching copy-choice recombinational repair, may have evolved from the simpler repair-related copy-choice events postulated above for ssRNA protocells based on the known processes in ssRNA viruses. Experimental evidence demonstrating that SDSA is a major recombination pathway in meiosis was presented by McMahill et al. [64].

The process of SDSA can accurately repair genome damage by copying the information lost in a damaged template strand from another intact homologous template strand without the need for physical breakage and exchange of DNA. Evidence bearing on the role of SDSA during meiotic recombination was reviewed by Bernstein et al. [66]. An alternative mechanism for recombinational repair termed the Double-Strand Break Repair (DSBR) model also explains some types of recombination events, but in contrast to SDSA recombination, the DSBR model does require physical breakage and exchange of DNA strands [67]. However,

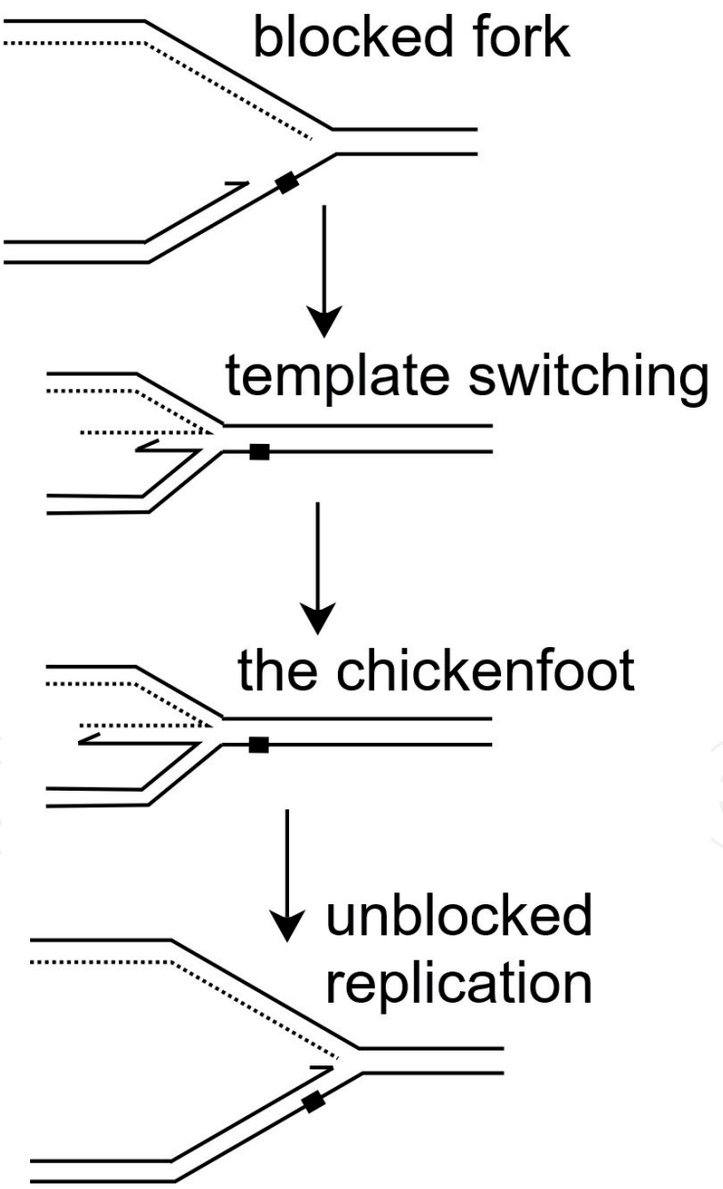


Figure 6.
Bypassing a DNA damage during replication. This mechanism involves reversal of the replication fork, where the newly replicated strands dissociate from their previous templates and anneal to form a cruciform intermediate, known as the "chicken foot" structure. Further replication of the previously blocked strand can then continue, leading to the bypass of the damaged site.

both the SDSA and DSBR models include a step in a which a DNA strand switches at a site of damage from one complementary partner strand to another and then continues synthesis with the new partner as template. Thus both models have elements of copy-choice recombination.

With respect to mitotic recombination in somatic cells, Andersen and Sekelsky [11] reviewed evidence that DSBR is a minor pathway for recombinational repair, and that the SDSA model appears to describe mitotic repair more accurately.

6.2 DNA replication

During DNA replication, a DNA damage in a template strand may be present and act as roadblock to the movement of the DNA polymerase as it extends synthesis of a new complementary strand. A blocked replication fork may be accurately bypassed by the mechanism illustrated in **Figure 6** [12, 13]. When movement of the replicative polymerase is blocked by a damage, the polymerase can switch template strands (mediated by a helicase) [12, 13] to form a structure referred to as a "chickenfoot" intermediate. As synthesis of the new strand proceeds along the alternate template it synthesizes the DNA region that is complementary to the damaged site in its original partner strand. The newly forming strand may then unwind and then re-associate with its original partner to continue synthesis along its original track. Polymerase-mediated strand-switching to deal with a damaged template during DNA synthesis appears to be an important general mechanism in eukaryotic cells [64]. This mechanism can be regarded as a type of copy-choice recombinational repair, and it too may have evolved from simpler copy-choice processes in ssRNA protocells.

7. Conclusions

Given the copy-choice genomic repair mechanism present in today's ssRNA viruses, it appears that copy choice as a repair process may have emerged as early as 3.5 to 2.5 billion years ago when RNA was apparently the only genetic material. It is possible that the capability for strand-switching was a property of the earliest ribozyme polymerases.

In early protocells, the ssRNA genomes may have been segmented, as some ssRNA viruses are in the present day. Two protocells with damaged segmented genomes could have been able to generate undamaged progeny after fusion and then reassortment of segments. Present day ssRNA segmented genome viruses can repair damage in their genomes through both copy choice and segment reassortment.

The early stages of the evolution of genome repair proposed here are based on known capabilities of extant RNA viruses. Currently it is not known if these RNA viruses are the actual evolutionary descendants of early RNA life forms, or if they arose later. It has only been assumed here that the problem of dealing with damage to an RNA genome arises in the two cases, and that the solutions to this problem would be similar.

The earliest ssRNAs that formed folded structures that acted as ribozymes can be designated plus (+) strands. Such a ribozyme strand could have had polymerase activity and acted as an RdRp. The progeny ssRNAs that it synthesizes would be complementary to the corresponding parental (+) strands, and can be designated minus (−) strands. During the synthesis of (−) strands template-switching may have occurred.

When the ssRNA genome evolved to a dsDNA form, elements of the earlier copy-choice recombinational repair processes appear to have been retained. In addition, the informational redundancy inherent in double-stranded DNA allowed the emergence of novel excision repair pathways (MMR, BER and NER) that could use the information in one strand to repair damage in the other strand. Other mechanisms (e.g. NHEJ and MMEJ) also emerged to deal with double-strand damages when an homolgous genome was not available. As eukaryotes evolved from unicellularity to multicellularity, and within an organism the germline became segregated from the somatic cell line, copy-choice recombinational repair was retained in the germline as a central feature of meiosis. Recombinational repair was also retained during mitosis, and as a general process for overcoming damage roadblocks to DNA replication.

Author details

Harris Bernstein and Carol Bernstein*
Department of Cellular and Molecular Medicine, College of Medicine, University of Arizona, Tucson, Arizona, USA

*Address all correspondence to: bernstein324@yahoo.com

References

[1] Schopf, JW. The evolution of the earliest cells. Sci. Am. 1978;239(3):110-139. DOI:10.1038/scientificamerican0978-110

[2] Sklarew DS, Nagy B. 2,5-Dimethylfuran from approximately 2.7 billion year-old Rupemba-Belingwe stromatolite, Rhodesia: Potential evidence for remnants of carbohydrates. Proc Natl Acad Sci USA. 1979;76(1):10-14. DOI:10.1073/pnas.76.1.10

[3] Woese CR. The primary lines of descent and the universal ancestor. In: Bendall DS editor. Evolution from molecules to men. Cambridge, UK: Cambridge University Press. 1983. p. 209-233. ISBN 978-0-521-28933-7. (https://archive.org/details/evolutionfrommol0000unse).

[4] Horning DP, Joyce GF. Amplification of RNA by an RNA polymerase ribozyme. Proc Natl Acad Sci USA. 2016;113(35):9786-9791. DOI:10.1073/pnas.1610103113

[5] Agol VI, Gmyl AP. Emergency services of viral RNAs: Repair and remodeling. Microbiol Mol Biol Rev. 2018;82(2):e00067-17. DOI:10.1128/MMBR.00067-17

[6] Lederberg J. Recombination mechanisms in bacteria. J Cell Physiol. 1955;45 (suppl. 2):75-107. DOI: 10.1002/jcp.1030450506. PMID:13242626

[7] Bernstein H. On the mechanism of intragenic recombination. I. Phage T4. J Theor Biol. 1962;3:335-353.

[8] Bernstein H, Hopf FA, Michod RE. The molecular basis of the evolution of sex. *Adv Genet*. 1987;24:323-370. doi:10.1016/s0065-2660(08)60012-7

[9] Hu WS, Temin HM. Retroviral recombination and reverse transcription. Science. 1990;250(4985):1227-1233. DOI:10.1126/science.1700865. PMID:1700865

[10] van Wolferen M, Shajahan A, Heinrich K, Brenzinger, S, Black IM, Wagner A, Briegel A, Azadi P, Albers SV. Species-specific recognition of Sulfolobales mediated by UV-inducible pili and S-layer glycosylation patterns. mBio. 2020;11(2):e03014-19. DOI:10.1128/mBio.03014-19

[11] Andersen SL, Sekelsky J. Meiotic versus mitotic recombination: two different routes for double-strand break repair: the different functions of meiotic versus mitotic DSB repair are reflected in different pathway usage and different outcomes. *Bioessays. 2010;32(12):1058-1066. DOI:* 10.1002/bies.201000087. *PMID:20967781*

[12] Ralf C, Hickson ID, Wu L. The Bloom's syndrome helicase can promote the regression of a model replication fork. J Biol Chem. 2006;281(32):22839-22846. DOI:10.1074/jbc.M604268200

[13] Gildenberg MS, Washington MT. Conformational flexibility of fork-remodeling helicase Rad5 shown by full-ensemble hybrid methods. PLoS One. 2019;14(10):e0223875. Published 2019 Oct 18. DOI:10.1371/journal.pone.0223875

[14] Fleck O, Nielsen O. DNA repair. J Cell Sci. 2004;117(Pt 4):515-517. DOI:10.1242/jcs.00952

[15] Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen. 2017;58(5):235-263. DOI:10.1002/em.22087

[16] Tjhung KF, Shokhirev MN, Horning DP, Joyce GF. An RNA polymerase ribozyme that synthesizes its own ancestor. Proc Natl Acad Sci U S

A. 2020;117(6):2906-2913. DOI:10.1073/pnas.1914282117

[17] Sagan C. Ultraviolet selection pressure on the earliest organisms. J Theor Biol. 1973;39(1):195-200. DOI:10.1016/0022-5193(73)90216-6

[18] Bernstein H, Byerly HC, Hopf FA, Michod RE. Origin of sex. J Theor Biol. 1984;110(3):323-351. DOI: 10.1016/s0022-5193(84)80178-2. PMID:6209512

[19] Lamb RA, Choppin PW. The gene structure and replication of influenza virus. Annu Rev Biochem. 1983;52:467-506. DOI:10.1146/annurev.bi.52.070183.002343. PMID:6351727

[20] Peterhans E. Oxidants and antioxidants in viral diseases: disease mechanisms and metabolic regulation. J Nutr. 1997;127(5 Suppl):962S-965S. DOI:10.1093/jn/127.5.962S PMID:9164274

[21] Michod RE, Bernstein H, Nedelcu AM. Adaptive value of sex in microbial pathogens. Infect Genet Evol. 2008;8(3):267-285. doi:10.1016/j.meegid.2008.01.002

[22] Barry RD. The multiplication of influenza virus. II. Multiplicity reactivation of ultraviolet irradiated virus. Virology. 1961;14:398-405. DOI:10.1016/0042-6822(61)90330-0. PMID:13687359

[23] Gilker JC, Pavilanis V, Ghys R. Multiplicity reactivation in gamma irradiated influenza viruses. Nature. 1967;214(5094):1235-1237. DOI:1038/2141235a0. PMID:6066111

[24] Lee SH, Kim WK, No JS, Kim JA, Kim JI, Gu SH, Kim HC, Klein TA, Park MS, Song JW. Dynamic circulation and genetic exchange of a shrew-borne hantavirus, Imjin virus, in the Republic of Korea. Sci Rep. 2017;7: 44369. DOI:10.1038/srep44369. PMC:5353647 PMID:28295052

[25] Lv Q, Zhang H, Tian L, Zhang R, Zhang Z, Li J, Tong Y, Fan H, Carr MJ, Shi W. Novel sub-lineages, recombinants and reassortants of severe fever with thrombocytopenia syndrome virus. Ticks Tick Borne Dis. 2017;8(3):385-390. DOI:10.1016/j.ttbdis.2016.12.015. PMID:28117273

[26] McClain ME, Spendlove RS. Multiplicity reactivation of reovirus particles after exposure to ultraviolet light. J Bacteriol. 1966;92(5):1422-1429.

[27] Bernstein C. Deoxyribonucleic acid repair in bacteriophage. Microbiol. Rev. 1981;45(1), 72-98.

[28] Bernstein C, Bernstein H. DNA repair in bacteriophage. In: Nickoloff JA. Hoekstra MF, editors. DNA Damage and Repair. Advances from Phage to Humans 3. Humana Presss, Totowa , NJ, 2001. p. 1-19 (ISBN 978-0896038035).

[29] Lehman N. A case for the extreme antiquity of recombination. J Mol Evol. 2003;56(6):770-777. DOI:10.1007/s00239-003-2454-1. PMID:12911039

[30] Yan LL, Zaher HS. How do cells cope with RNA damage and its consequences?. J Biol Chem. 2019;294(41):15158-15171. DOI:10.1074/jbc.REV119.006513

[31] He CQ, Xie ZX, Han GZ, Dong JB, Wang D, Liu JB, Ma LY, Tang XF, Liu XP, Pang YS, Li GR. Homologous recombination as an evolutionary force in the avian influenza A virus. Mol Biol Evol. 2009;26(1):177-187. DOI:10.1093/molbev/msn238. PMID: 18931384

[32] De A, Sarkar T, Nandy A. Bioinformatics studies of Influenza A hemagglutinin sequence data indicate recombination-like events leading to segment exchanges. BMC Res Notes. 2016;9:222. DOI:10.1186/s13104-016-2017-3 PMC:4832483. PMID:27083561

[33] He CQ, Ding NZ. Discovery of severe fever with thrombocytopenia syndrome bunyavirus strains originating from intragenic recombination. J Virol. 2012;86(22):12426-12430. DOI: 10.1128/JVI.01317-12. PMID:22933273

[34] Kirkegaard K, Baltimore D. The mechanism of RNA recombination in poliovirus. Cell. 1986;47 (3):433-443. DOI:10.1016/0092-8674(86)90600-8. PMC:7133339. PMID:3021340

[35] Muslin C, Mac Kain A, Bessaud M, Blondel B, Delpeyroux F. Recombination in enteroviruses, a multi-step modular evolutionary process. Viruses. 2019;11(9):859. DOI:10.3390/v11090859 PMC:6784155 PMID:31540135

[36] Barr JN, Fearns R. How RNA viruses maintain their genome integrity. J Gen Virol. 2010;91 (Pt 6): 1373-1387. DOI:10.1099/vir.0.020818-0. PMID:20335491

[37] Drake JW. Interference and multiplicity reactivation in polioviruses. Virology. 1958;6 (1): 244-264. DOI:10.1016/0042-6822(58)90073-4. PMID:13581529

[38] Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y, Gao GF. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol. 2016;24(6):490-502. DOI:10.1016/j.tim.2016.03.003. PMID:27012512

[39] Cheng CP, Nagy PD. Mechanism of RNA recombination in carmo- and tombusviruses: evidence for template switching by the RNA-dependent RNA polymerase in vitro. J Virol. 2003;77(22):12033-12047. DOI:10.1128/jvi.77.22.12033-12047.2003 PMC:254248. PMID:14581540

[40] Rawson JMO, Nikolaitchik OA, Keele BF, Pathak VK, Hu WS. Recombination is required for efficient HIV-1 replication and the maintenance of viral genome integrity. Nucleic Acids Res. 2018;46(20):10535-10545. DOI:10.1093/nar/gky910 PMID:30307534

[41] Cromer D, Grimm AJ, Schlub TE, Mak J, Davenport MP. Estimating the in-vivo HIV template switching and recombination rate. AIDS. 2016;30(2):185-192. DOI:10.1097/QAD.0000000000000936. PMID:26691546

[42] Anderson JA, Teufel RJ 2nd, Yin PD, Hu WS. Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. J Virol. 1998;72(2):1186-1194. PMID:9445017

[43] Nora T, Charpentier C, Tenaillon O, Hoede C, Clavel F, Hance AJ. Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. J Virol. 2007;81(14):7620-7628. DOI:10.1128/JVI.00083-07

[44] Hu WS, Hughes SH. HIV-1 Reverse transcription. Cold Spring Harb Perspect Med. 2012;2(10): a006882. DOI:10.1101/cshperspect.a006882. PMID:23028129 PMCID:PMC3475395

[45] Chen J, Powell D, Hu WS. High frequency of genetic recombination is a common feature of primate lentivirus replication. J Virol. 2006;80(19):9651-9658. DOI:10.1128/JVI.00936-06

[46] Israël N, Gougerot-Pocidalo MA. Oxidative stress in human immunodeficiency virus infection. Cell Mol Life Sci. 1997;53(11-12):864-870. DOI:10.1007/s000180050106

[47] Bonhoeffer S, Chappey C, Parkin NT, Whitcomb JM, Petropoulos CJ. Evidence for positive

epistasis in HIV-1 [published correction appears in Science. 2008 May 9;320(5877):746]. Science. 2004;306(5701):1547-1550. DOI:10.1126/science.1101786 PMID:15567861

[48] Samanta B, Joyce GF. A reverse transcriptase ribozyme. Elife. 2017;6:e31153. Published 2017 Sep 26. DOI:10.7554/eLife.31153

[49] Cotton JA, McInerney JO. Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. Proc Natl Acad Sci U S A. 2010;107(40):17252-17255. doi:10.1073/pnas.1000265107.

[50] Speijer D. Birth of the eukaryotes by a set of reactive innovations: New insights force us to relinquish gradual models. Bioessays. 2015;37(12):1268-1276. DOI:10.1002/bies.201500107

[51] Javaux EJ, Knoll AH, Walter MR. Morphological and ecological complexity in early eukaryote ecosystems. Nature. 2001;412(6842):66-69. DOI:10.1038/35083562

[52] van Wolferen M, Wagner A, van der Does C, Albers SV. The archaeal Ced system imports DNA. Proc Natl Acad Sci USA. 2016;113(9):2496-2501. DOI:10.1073/pnas.1513740113

[53] Bernstein H, Bernstein C. Sexual communication in archaea, the precursor to eukaryotic meiosis. In: Witzany G, editor. Biocommunication of Archaea, Springer International Publishing 2017. p. 103-117. DOI:10.1007/978-3-319-65536-9 7

[54] Fröls S, Ajon M, Wagner M, Teichmann D, Zolghadr B, Foles M, Boekema EJ, Driessen AJ, Schleper C, Albers SV. UV-inducible cellular aggregation of the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by pili formation. Mol Microbiol. 2008;70(4):938-952. DOI:10.1111/j.1365-2958.2008.06459.x

[55] Allers T. Swapping genes to survive - a new role for archaeal type IV pili. Mol Microbiol. 2011;82(4):789-791. DOI:10.1111/j.1365-2958.2011.07860.x

[56] Fröls S, White MF, Schleper C. Reactions to UV damage in the model archaeon Sulfolobus solfataricus. Biochem Soc Trans. 2009;37(Pt 1):36-41. DOI:10.1042/BST0370036

[57] Ajon M, Fröls S, van Wolferen M, Stoecker K, Teichmann D, Driessen AJ, Grogan DW, Albers SV Schleper C. UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. Mol Microbiol. 2011;82(4):807-817. DOI:10.1111/j.1365-2958.2011.07861.x

[58] Yi C, He C. DNA repair by reversal of DNA damage [published correction appears in Cold Spring Harb Perspect Biol. 2014 Apr;6(4):a023440]. Cold Spring Harb Perspect Biol. 2013;5(1):a012575. DOI:10.1101/cshperspect.a012575

[59] White MF, Allers T. DNA repair in the archaea-an emerging picture. FEMS Microbiol Rev. 2018;42(4):514-526. DOI:10.1093/femsre/fuy020

[60] Bétermier M, Bertrand P, Lopez BS. Is non-homologous end-joining really an inherently error-prone process?. PLoS Genet. 2014;10(1):e1004086. DOI:10.1371/journal.pgen.1004086

[61] Seol JH, Shim EY, Lee SE. Microhomology-mediated end joining: Good, bad and ugly. Mutat Res. 2018;809:81-87. DOI:10.1016/j.mrfmmm.2017.07.002

[62] Bernstein H, Bernstein C. Evolutionary origin and adaptive function of meiosis. In: Bernstein C, Bernstein H, editors. Meiosis. InTech

Publ, Croatia (open access); 2013 p. 41-75. http://dx.doi.org/10.5772/56557

[63] Allers T, Lichten M. Differential timing and control of noncrossover and crossover recombination during meiosis. Cell. 2001;106(1):47-57. DOI:10.1016/s0092-8674(01)00416-0. PMID:11461701

[64] McMahill MS, Sham CW, Bishop DK. Synthesis-dependent strand annealing in meiosis. PLoS Biol. 2007;5(11):e299. DOI:10.1371/journal.pbio.0050299

[65] Miura T, Yamana Y, Usui T, Ogawa HI, Yamamoto MT, Kusano K. Homologous recombination via synthesis-dependent strand annealing in yeast requires the Irc20 and Srs2 DNA helicases. Genetics. 2012;191(1):65-78. DOI:10.1534/genetics.112.139105 PMC:3338270 PMID:22367032

[66] Bernstein H, Bernstein C, Michod RE. Meiosis as an evolutionary adaptation for DNA repair. In: Inna Kruman, editor. DNA Repair. Intech Publ, Chapter 19: 2011; p. 357-382 (ISBN 978-953-307-697-3).

[67] Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair model for recombination. Cell. 1983;33(1):25-35. DOI:10.1016/0092-8674(83)90331-8