HUH site-specific recombinases for targeted modification of the human genome

Coral González-Prieto[1], Leticia Agúndez[2], Ralph Michael Linden[2*], and Matxalen Llosa[1*]

[1] Departamento de Biología Molecular (Universidad de Cantabria) and IBBTEC (UC, CSIC, SODERCAN), Santander, Spain.

[2] King's College London School of Medicine, and UCL Gene Therapy Consortium, University College London, London, UK

[*] Corresponding authors: llosam@unican.es, michael.linden@kcl.ac.uk

Abstract

Site-specific recombinases (SSR) have been crucial in the development of mammalian transgenesis. For gene therapy purposes, this approach remains challenging, as e.g. SSR delivery is largely unsolved and SSR DNA substrates must pre-exist in target cells. In this review, we discuss the potential of HUH recombinases to overcome some of the limitations of conventional SSR. Members of the HUH protein family cleave single-stranded DNA, but can mediate site-specific integration with the aid of the host replication machinery. AAV Rep remains the only known example to support site-specific integration in human cells, and AAV is an excellent gene delivery vector which can be targeted to specific cells and organelles. Bacterial protein TrwC catalyzes integration into human sequences and can be delivered to human cells covalently linked to DNA, offering attractive new features for targeted genome modification.

KEYWORDS: Genomic engineering / site-specific recombination / gene therapy / HUH family / AAV Rep / R388 TrwC

Highlights:

- HUH SSR offer attractive features for targeted genome modification
- HUH SSR can integrate ssDNA with the aid of host replication machinery
- AAV-Rep and R388-TrwC can be delivered in vivo
- Rep-mediated integration into AAVS1 does not have adverse effects

Glossary

Site-specific recombination (SSrec): also known as conservative site-specific recombination. Recombination process in which DNA exchange takes place between defined DNA sequences possessing only a limited degree of sequence homology, by a mechanism that conserves the phosphodiester bond energy. Depending on the initial arrangement of the two DNA partners, it can result in integration, excision, inversion, resolution, or translocation.

Site-specific recombinases (SSR): enzymes that catalyze cut-and-strand transfer reactions on specific DNA sequences, producing rearrangments of DNA segments. A distinctive feature of these recombinases is the formation of a covalent protein-DNA intermediate during the recombination process.

Tyr-SSR and **Ser-SSR:** two different families of SSR named after the nucleophilic amino acid residue that they use to attack the DNA and which becomes covalently linked to it during strand exchange. Although leading to the same practical outcomes, the two families are unrelated to each other, having different protein structures and reaction mechanisms (shown in Fig. 1).

HUH protein: defined by the presence of two conserved motifs: a His-hydrophobic-His (HUH) motif required for metal ion binding, and a motif containing one or two catalytic tyrosines for nucleophilic attack of the DNA. HUH family members are strand-transferases acting at target sites on a single DNA strand, and are involved preferentially in biological processes involving ssDNA intermediates, such as rolling-circle replication and transposition, or bacterial conjugation.

Rolling-circle replication (RCR): mechanism used for the replication of some circular molecules, such as plasmids and certain viruses. A Rep protein, belonging to the HUH family, cleaves the target oriV and remains covalently bound to the 5´end, providing a free 3'-OH end onto which nucleotides are added. This mechanism allows fast production of single-stranded replication products. RCR-transposition and processing of DNA during bacterial conjugation are related processes, also based on HUH proteins which catalyze the initial cleavage and final religation steps.

Adeno-associated virus (AAV): AAV is a small non-pathogenic human parvovirus whose life cycle consists of both a productive replicative phase and latent infection. It is known to need a helper virus for the productive life cycle. AAV-Rep proteins, belonging to the HUH family, are essential for initiation of replication of the viral genome and for site-specific integration of the virus into a single target site present in the human genome.

Bacterial conjugation: mechanism of horizontal DNA transfer from a donor to a recipient bacterium. The process involves the generation of a single-stranded DNA which is leaded into the recipient cell as a nucleoprotein complex by the conjugative relaxase. This HUH protein catalyzes cleavage and strand-transfer reactions at its target site to initiate and end ssDNA transfer. In addition, several relaxases can act as recombinases and integrases on double-stranded DNA substrates.

Genome modification of human cells through site-specific integration of foreign DNA

Gene therapy aims to treat disease through the alteration of genome content in target tissues and, in most cases, requires stable expression of exogenous DNA. Long-term expression is either achieved through extra-chromosomal persistence or by integration of the therapeutic DNA into the human genome, in particular, in proliferating cells. However, random integration in such scenarios, has been demonstrated to carry the risk of insertional mutagenesis, potentially leading to tumor growth [1]. An alternative approach, gene targeting via homologous recombination, has recently witnessed promising advances thanks to the design of synthetic nucleases with a high degree of target specificity [2]. However, their design is complex and questions regarding off-target activities have yet to be addressed [3]. Furthermore, methods to reliably predict these events are still missing ([4]; [5]).

The use of site-specific recombinases (SSR), which directly integrate foreign DNA into a specific site in the genome, could help overcome some of these hurdles. There are, however, some potentially limiting prerequisites (unidirectionality of the reaction, existence of a natural target site) as well as additional inherent problems (SSR toxicity, DNA rearrangements). These are discussed here and we will attempt to introduce possible solutions, thereby establishing this class of proteins as a viable addition to our tools to modify the human genome.

Site-specific recombination: 2+1=2+2

SSR catalyze the recombination between specific target DNA sequences. SSR mediate this process in a reaction involving a covalent intermediate with the target DNA. The final product is the recombinant molecule, which, depending on the orientation of the target sequences, can lead to integration, deletion, inversion, resolution, or translocation of the DNA between the crossover sites [6].

In contrast to homologous recombination, site-specific recombination (SSrec) has different biological roles (see **Table 1**). Examples in prokaryotes, which illustrate the potential of SSRec, include the integration-excision cycles of bacteriophages and other mobile genetic elements, the resolution of plasmid multimers, or the control of gene expression through the repositioning of control elements. Among those are the inversion-mediated alternate expression of flagellins, or the assembly of active genes by irreversible deletion of a DNA segment leading to the expression of a reconstructed open reading frame (ORF) in non-vegetative cells, such as cyanobacterial heterocysts or *Bacillus* mother cells during sporulation [6].

SSR usually refer to the canonical conservative recombinases that act on two target sites and catalyze their recombination by performing four nicks and two strand-exchange reactions. They are divided into two main families, Tyr-SSR (such as Cre recombinase from phage P1) and Ser-SSR (such as Int from the *Streptomyces* phage ΦC31), which differ in their structure, catalytic residues and mode of action, yet perform essentially identical reactions with the same outcome, i.e. recombination (**Figure 1**). For a detailed description of the recombination mechanisms of Tyr- and Ser- SSR, the reader is referred to a recent review [7].

The integrases responsible for gene cassette insertion/excision in integrons, also termed Int, belong to the family of Tyr-SSR, yet they perform a single strand exchange reaction. These SSR recognizes a folded single-stranded intermediate, thus forming an atypical Holliday junction (Fig. 1). It has recently been demonstrated that recombination is resolved by the host replication machinery [8]. Thus, a reaction starting with a double-stranded (ds) target DNA and a single-stranded (ss) donor DNA (2+1) results in a recombination product identical to those of SSR acting on two dsDNA molecules (2+2). This apparently surprising "2+1=2+2" equation is not so novel: old experiments have shown that Tn7 could switch from conservative to replicative transposition with a single mutation abolishing the second nicking reaction [9]. This observation suggests a remarkable plasticity in the ability to recruit host functions.

Within this scenario, a new unexpected family of SSR has been found among members of the HUH family of enzymes, defined by the presence of a series of motifs involved in their catalytic activity [10]. These proteins are sequence-specific single-strand endonucleases which perform strand-transfer reactions on ssDNA, required in processes such as bacterial conjugation, and rolling-circle replication or transposition. In addition, some of them have been shown to catalyze site-specific recombination and integration reactions on dsDNA substrates, and thus can also be considered as SSR. These include conjugative relaxase TrwC of plasmid R388 [11]; rolling circle replicase Rep from adeno-associated virus, AAV [12]; and transposase TnpA(REP) from *E. coli* K12 [13]. To date, all proposed models involving ss-based SSR include a replication step, carried out by the host machinery, in order to complete the reaction ([14]; [11]; [15]) (Fig. 1).

SSR are characteristically sequence-specific, very efficient, and often initiate recombination without cofactors or host-cell components. These features make them potentially useful tools for genome engineering [16]. Outstanding examples are the Tyr-SSR Cre from phage P1 or the Ser-SSR Int from phage ΦC31. Interestingly, most widely used SSR come from bacterial phages and have been shown to be functional in the eukaryotic environment. For instance, ΦC31 Int has been shown to catalyze site-specific integration in plant, *Drosophila*, murine and human cells [17]. This striking promiscuity hints at a considerable evolutionary conservation and possibly an as yet underappreciated function of these mechanisms in higher eukaryotes.

Prerequisites: unidirectionality and target site

SSR can perform reversible or irreversible reactions. An example of an SSR which catalyzes recombination in both senses with equal efficiency is Cre from phage P1, which has been widely used for the construction of transgenic animals, including the generation of conditional phenotypes [18]. However, the reversibility of the recombination reaction can be a limitation if the goal is to obtain stable integration of a foreign DNA into the host genome, since the recombinase can catalyze the excision of this DNA at any moment. Some strategies, such as transient expression of the recombinase, can partially overcome this limitation.

In contrast, other SSR perform a unidirectional reaction. Most SSR determine the directionality of the reaction by recognizing specific target sites, which, upon recombination, generate new sites that can no longer serve as substrates. This is the case of phage integrases such as Int-ΦC31, which convert the phage attP and host attB sites into two new hybrid sites attL and

attR, upon integration. Additional factors are required to catalyze recombination on these newly created sites [19], thus allowing a control of the directionality of the reaction.

One of the main limiting factors for the use of SSR genome modification is the potential absence of a naturally occurring target within the human genome. To date, most proof-of concept studies use cells that have been engineered to contain respective SSR target sequences; this approach, however, is not applicable for gene therapy purposes.

The integrase of phage ΦC31 catalyzes unidirectional phage integration. This integrase also recognizes target sequences in many eukaryotic genomes, including the human genome [20]. A large number of pseudo-attP sites have been characterized (**Table 2**) into which this SSR can integrate any incoming DNA containing an attB motif. However, not all sites are used with the same efficiency: the ψA site in chromosome 8 was reported to be used preferentially [20] and a recent report finds an additional hotspot at 19q13.31 [21].

The use of Int-ΦC31 for integration of exogenous DNA in mammals originated more than a decade ago. The system is very efficient, rendering a high percentage of viable transformed cells. It has been used for genetic correction in mice and also in cultured human cells, including stem cells. Notably, in the last few years this system has allowed phenotypic correction of hemophilia A and B in mice through the expression of human clotting factors ([22]; [23]). It has also been used for targeted integration in human muscle and cardiac progenitor cells ([24]; [25]), stem cell lines [26], as well as in approaches to generate mouse iPS cells ([27];[28]). It was not until recently that the additional protein required for phage excision was determined [19], which will allow further optimization of this tool for mammalian cells [29].

Other Ser-SSR phage integrases have been characterized and used successfully in a mammalian environment. R4 and A118 integrases can integrate an incoming plasmid into endogenous pseudo *att* sites in the human genome, although aberrant chromosomal events were found to be associated with R4, and for A118, four out of fifteen integration events at pseudo *attB* sites showed imperfect junctions ([30]; [31]).

Adeno-associated virus (AAV) is the only known virus capable of targeted integration in human cells. AAV-Rep protein, an HUH protein, recognizes a unique target sequence within *AAVS1* which is located on human chromosome 19 (19q13.3-qter) (Table 2), and several studies have demonstrated that insertion at this site poses no apparent risks [15]. The AAV-Rep-mediated site-specific integration reaction has extensively been studied in tissue culture. The characteristic Tyr of HUH SSR orchestrates the nicking in the target locus, *AAVS1*. Subsequently, a DNA strand exchange between the viral and human chromosomal sequences is proposed to form a covalent junction. This junction formation is then followed by the resolution of the intermediate by the host cell replication machinery [15], as outlined in Fig. 1D. AAV-Rep mediated site-specific integration of foreign DNA has been achieved in mouse [15] and human embryonic stem cells (hESCs) [32], opening the possibility of using this system in replacement therapies in several human diseases.

A number of conjugative relaxases possess site-specific recombinase activity. Notably, the conjugative relaxase TrwC is also able to integrate the DNA to which it is covalently attached into its target sequence present in the recipient bacterial cell [33]. TrwC has two active Tyr

residues which are both involved in the integration reaction [34]. Several sequences resembling its natural target exist in the human genome, and it has been shown that TrwC can catalyze integration into two of these sites [34]. Interestingly, integration was stable, since the target sequences were not substrates for the excision reaction [34]. These features confer potential as a genomic modification tool; however, it remains to be demonstrated whether this activity can be transposed into human cells.

Inherent risks: off-target insertion and genotoxicity

A key issue for the successful use of SSR in human genomic modification protocols is their sequence specificity. SSR are generally very specific enzymes, however they can act on pseudo-target sites with sufficient similarity to their natural targets. This is particularly relevant when protein levels are high, under which conditions these enzymes can introduce nicks and ds-breaks, thus possibly initiating unintended recombination events. Therefore, a main concern continues to be the inherent risk of introducing undesired DNA rearrangements during integration. Overall, the genotoxicity of SSR has not been sufficiently studied to allow for reliable predictions about the safety of their use in humans.

The main drawback of using Int-ΦC31 for gene therapy, for example, is the presence of too many possible insertion sites (pseudo-att sites; Table 1). This problem may be partially overcome by the use of mutant integrases, obtained by directed evolution, that show increased specificity for the main ψA site [35]. Another hurdle is based on the observation that integration may lead to chromosomal rearrangements. This point is open to debate, since there are conflicting reports showing precise integration of Int-mediated transformed cells, while others demonstrate the presence of rearrangements upon Int-ΦC31 mediated integration in pseudo-attP sites [36]. It remains to be determined whether these differences are simply due to the cellular levels of the integrase.

AAV-Rep mediated integration was reported to be targeted to *AAVS1* in early works ([37]; [38]). Integration was analyzed in several latently infected human cell lines and 78% of these cells showed integration in *AAVS1*, highlighting the specificity of the system for the human target. Rep-mediated integration into pseudo-AAV sites has been reported [39]. However, as these studies were performed in HeLa cells, it remains to be shown whether these events were due to the well-documented propensity of AAV and AAV vectors to insert into pre-existing double-strand breaks, rather than due to a Rep-specific off-target event. Other reports have also challenged site-specificity, arguing that Rep-mediated integration was close to random [40]. However, a recent, unbiased analysis of integration sites, has provided further evidence for the specificity of Rep-mediated integration [41].

AAVS1 is located within gene *PPP1R12C,* in a very gene-dense region in chromosome 19, making it more likely that possible rearrangements may have deleterious effects for the cell upon integration. An exhaustive study addressing potential adverse effects resulting from *AAVS1* integration has recently been presented. *AAVS1*-targeted mouse embryonic stem cells showed that, despite of the resulting rearrangement, the cells maintained multi-lineage differentiation potential and contributed successfully to mouse development when injected into blastocysts [15]. A potential explanation for the lack of adverse effects in spite of significant associated rearrangements in this stringent model system has been provided by the

observation of a duplication resulting from the integration mechanism, thereby preserving functional expression from the disrupted allele. These data underscore the potential of this locus as a suitable safe harbor for therapeutic transgene insertion when Rep is used to mediate integration. Several recent reports have also used *AAVS1* as the target for designed synthetic nucleases [42, 43]. These studies further underscore the suitability of *AAVS1* for transgene expression. However, similarly stringent assays have yet to be employed to assess whether gene addition to this locus is inert in the absence of Rep-mediated target gene duplication.

As highlighted above, the cellular level of integrase is possibly a key factor in the fidelity of the integration mechanism and thus the associated genotoxicity. Expression of Int-ΦC31 induces a DNA damage response and chromosomal rearrangements in human cells [44], but Int is expressed for only a few hours in both mouse liver and human cultured cells [45], making it unlikely that the integrase would induce such effects. Overexpression of Rep regulates the expression of cellular and viral genes and may induce apoptosis [46], DNA damage and cell cycle arrest [47]. This challenge underlines the necessity to strictly control recombinase expression. In this context, protein delivery by fusing a TAT domain has been reported for Int-ΦC31 [48], but this approach requires protein purification, and renders lower recombination efficiency.

Delivery of SSR and DNA to the human cell

Any strategy for genome modification must include a way to deliver foreign DNA and the integration system to the target cells. DNA can be introduced into human cells by a variety of methods, including naked DNA, synthetic vectors, viral vectors, or bactofection [49]. SSR from any source can be introduced by any of these methods. **Figure 2** compares the entry pathways for various integrases.

Int- ΦC31 is routinely introduced by plasmid transfection of cultured cells, or by hydrodynamic tail-vein injection in mice. Introduction with adenovirus vectors allows for site-specific integration of large DNA fragments with low genotoxicity [50].

A key advantage of the AAV-Rep system is that the virus naturally infects human cells and the site-specific integration potential has been retained throughout evolution. In addition, AAV poses no known safety issues. The main hurdle of AAV as a targeting vector is the strict space limitation imposed by its capsid. In order to overcome this problem, attempts have been made to use dual infections [51] or to incorporate Rep into other viruses, such as adenovirus or herpesvirus [52, 53]. An additional benefit of AAV is that recombinant vectors (rAAV) have been widely used in preclinical as well as some clinical trials [54]. In addition, the availability of capsids from different AAV serotypes allows for targeting specific tissues, such as skeletal muscle, liver, central nervous system, retina and heart [54]. Furthermore, the addition of a mitochondrial targeting sequence to AAV capsids redirects the virus inside the organelle, achieving correction of Leber's hereditary optic neuropathy in a mouse model [55].

The conjugative relaxase TrwC also shows a potential for *in vivo* delivery into human cells. A recent report showed that this SSR can be delivered to specific human cells using Type IV Secretion Systems (T4SS) of pathogenic bacteria [56]. T4SS are encoded by many human pathogens, each targeting specific cell types, thereby introducting the potential for some tissue specificity for *in vivo* gene therapy [57]. A major advantage, however, might be that the SSR enters the cell in a covalent complex with the transgene and thus overcomes the need for recombinase expression in the cell and favors irreversible integration of the incoming DNA.

Many artificial delivery systems can reach the cytoplasm, yet integration takes place within the nucleus of the cell. Accordingly, AAV-Rep has a Nuclear Localization Signal (NLS) for nuclear targeting [58]. SSR from bacteria or phage are not expected to target the nucleus, but it appears feasible to engineer an approach that includes nuclear import. For example, while TrwC localizes to the cytoplasm, a mutant shows nuclear localization [59]. Similarly, the addition of an NLS to Int-ΦC31 increases integration [60]. In addition, a TAT-Int-NLS was shown to recombine more efficiently than TAT-Int in mammalian cells [48]. In contrast nuclear localization provides little or no benefit to ϕC31 integrase for liver directed gene therapy, even in the absence of cell division [61], suggesting that Int nuclear entry is not the limiting factor for integration. A small fraction of the integrase can enter the nucleus bound to the DNA, as occurs in the case of retroviral integrases [62].

Future directions

The main conclusions and outstanding questions on the use of SSR for human genomic modification are outlined in Box 1. Taken together, no simple solution has been put forward to address the challenge that any system ideally would i) efficiently deliver DNA to the target tissue *in vivo*, ii) allow there for efficient integration, iii) express only transiently the required exogenous recombinase, and iv) evade significant immune detection, and thus ensure the survival of the modified target cell. SSR are a potential tool for *in vivo* and *ex vivo* genome modification. While tailor-made nucleases show great potential, we propose that the inherent and unique characteristics of SSR might provide distinct benefits that warrant further investigation.

ss-dependent SSR from the HUH family such as AAV-Rep have shown to be as efficient in integration as conservative SSR, and proof of principle together with evidence for safety and utility of this approach have been provided in mouse and human embryonic stem cells [15, 32]. Rep-mediated modification of human iPS cells is ongoing in our laboratory. The underlying strategy is that iPS cells can be obtained from patients and subsequently the genetic defect can conceivably be corrected by AAV-mediated site-specific integration of the un-mutated gene where appropriate, resulting in a suitable cell population for differentiation and subsequent transplantation.

Bacterial conjugative relaxases may represent promising new tools due to their site-specific integrase activity and the presence of potential target sites within the human genome. In addition, new substrate specificities can be engineered [63], thus broadening the possibility to find the adequate insertion site. In vivo delivery through bacterial T4SS as a covalent protein-

DNA complex constitutes a unique feature conferring added value. However, to date, proof of concept for bacterial SSR-mediated site-specific integration into the human genome has yet to be provided.

HUH recombinases may represent a family of moonlighting proteins evolutionarily selected to perform site-specific integration, in addition to their role in viral replication, bacterial conjugation, or transposition; this molecular strategy has been preserved from bacteria, to plants, to mammalian viruses. The intricacies of this approach include efficient cooperation with host enzymes (thus only one exogenous protein is required) and, in the case of AAV Rep, a mechanism that includes partial gene duplication through which functional expression from both target alleles is retained. Among this new family of recombinases is the possibility to overcome such problems as target infidelity and thereby genotoxicity. The main limiting factor for the use of SSR is the presence of a target sequence in the human genome. However, these enzymes are highly prevalent in bacteria and viruses and, as TrwC demonstrates, it is likely that suitable candidates with human target sequences can be identified.

Acknowledgements

Work in our labs is supported by grant BIO2010-11623E from the Spanish Ministry of Science and Innovation to ML and UK Medical Research Council grant 1001764 to RML. CGP was a recipient of a predoctoral fellowship from the University of Cantabria (Spain).

Box 1. Conclusions and outstanding questions

Conclusions

- Site-specific recombinases are valuable tools for human genomic modification
- SSR belonging to the HUH protein family catalyze integration of ssDNA with the aid of the host replication machinery
- AAV-Rep catalyzes integration into *AAVS1* with no known additional effect on the recipient genome, partly due to reconstruction of the target gene by partial duplication upon integration
- Several conjugative relaxases have been shown to act as SSR, providing new sources of potential integration sites. TrwC from plasmid R388 can integrate DNA into two sequences from the human genome which resemble its natural target
- TrwC can be delivered as a protein-DNA complex into specific human cell types through bacterial Type IV secretion systems. These machines are present in bacteria targeting different tissues
- AAV is an excellent vector for delivery of the transgene and Rep. It can be targeted to different cellular types and even to mitochondria
- The introduction of the SSR protein in place of the gene may be the best way to avoid genotoxicity

Outstanding questions

- Will it be possible to target different human cellular types through the Type IV secretion systems of different human pathogens?
- Can TrwC integrate foreign DNA into its specific targets in the human genome? If so, what will be the effect of integration into these sites?
- Is it possible to obtain Rep-mediated modification of human iPS cells, allowing correction of diseases such as SCID-X1?
- What will be a suitable system to transiently deliver AAV Rep to target cells in order to mediate site-specific integration?

REFERENCES

1 McCormack, M.P. and Rabbitts, T.H. (2004) Activation of the T-cell oncogene LMO2 after gene therapy for X-linked severe combined immunodeficiency. *The New England journal of medicine* 350, 913-922

2 Mussolino, C. and Cathomen, T. (2012) TALE nucleases: tailored genome engineering made easy. *Current opinion in biotechnology* 23, 644-650

3 Handel, E.M. and Cathomen, T. (2011) Zinc-finger nuclease based genome surgery: it's all about specificity. *Current gene therapy* 11, 28-37

4 Gabriel, R.*, et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature biotechnology* 29, 816-823

5 Pattanayak, V.*, et al.* (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature methods* 8, 765-770

6 Craig, N.L.*, et al.* (2002) *Mobile DNA II*. ASM Press

7 Hirano, N.*, et al.* (2011) Site-specific recombinases as tools for heterologous gene integration. *Applied microbiology and biotechnology* 92, 227-239

8 Loot, C.*, et al.* (2012) Replicative resolution of integron cassette insertion. *Nucleic acids research* 40, 8361-8370

9 May, E.W. and Craig, N.L. (1996) Switching from cut-and-paste to replicative Tn7 transposition. *Science* 272, 401-404

10 Ilyina, T.V. and Koonin, E.V. (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic acids research* 20, 3279-3285

11 Cesar, C.E.*, et al.* (2006) A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. *Molecular microbiology* 62, 984-996

12 Smith, R.H. and Kotin, R.M. (2000) An adeno-associated virus (AAV) initiator protein, Rep78, catalyzes the cleavage and ligation of single-stranded AAV ori DNA. *Journal of virology* 74, 3122-3129

13 Ton-Hoang, B.*, et al.* (2012) Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences. *Nucleic acids research* 40, 3596-3609

14 Mendiola, M.V.*, et al.* (1994) Differential roles of the transposon termini in IS91 transposition. *Proceedings of the National Academy of Sciences of the United States of America* 91, 1922-1926

15 Henckaerts, E.*, et al.* (2009) Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proceedings of the National Academy of Sciences of the United States of America* 106, 7571-7576

16 Turan, S. and Bode, J. (2011) Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 25, 4088-4107

17 Geisinger, J.M. and Calos, M.P. (2013) Site-Specific Recombination Using PhiC31 Integrase. In *Site-directed Insertion of Transgenes* (Renault, S. and Duchateau, P., eds), pp. 211-239, Springer

18 Sorrell, D.A. and Kolb, A.F. (2005) Targeted modification of mammalian genomes. *Biotechnology advances* 23, 431-469

19 Khaleel, T.*, et al.* (2011) A phage protein that binds phiC31 integrase to switch its directionality. *Molecular microbiology* 80, 1450-1463

20 Thyagarajan, B.*, et al.* (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Molecular and cellular biology* 21, 3926-3934

21 Chalberg, T.W.*, et al.* (2006) Integration specificity of phage phiC31 integrase in the human genome. *Journal of molecular biology* 357, 28-48

22 Keravala, A., *et al.* (2011) Long-term phenotypic correction in factor IX knockout mice by using PhiC31 integrase-mediated gene therapy. *Gene therapy* 18, 842-848
23 Chavez, C.L., *et al.* (2012) Long-term expression of human coagulation factor VIII in a tolerant mouse model using the phiC31 integrase system. *Human gene therapy* 23, 390-398
24 Quenneville, S.P., *et al.* (2007) Dystrophin expression in host muscle following transplantation of muscle precursor cells modified with the phiC31 integrase. *Gene therapy* 14, 514-522
25 Lan, F., *et al.* (2012) Safe genetic modification of cardiac stem cells using a site-specific integration technique. *Circulation* 126, S20-28
26 Thyagarajan, B., *et al.* (2008) Creation of engineered human embryonic stem cell lines using phiC31 integrase. *Stem cells* 26, 119-126
27 Ye, L., *et al.* (2010) Generation of induced pluripotent stem cells using site-specific integration with phage integrase. *Proceedings of the National Academy of Sciences of the United States of America* 107, 19467-19472
28 Karow, M., *et al.* (2011) Site-specific recombinase strategy to create induced pluripotent stem cells efficiently with plasmid DNA. *Stem cells* 29, 1696-1704
29 Farruggio, A.P., *et al.* (2012) Efficient reversal of phiC31 integrase recombination in mammalian cells. *Biotechnology journal* 7, 1332-1336
30 Keravala, A., *et al.* (2006) A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. *Molecular genetics and genomics : MGG* 276, 135-146
31 Olivares, E.C., *et al.* (2001) Phage R4 integrase mediates site-specific integration in human cells. *Gene* 278, 167-176
32 Yang, L., *et al.* (2008) Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature* 453, 524-528
33 Draper, O., *et al.* (2005) Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. *Proceedings of the National Academy of Sciences of the United States of America* 102, 16385-16390
34 Agundez, L., *et al.* (2012) Site-specific integration of foreign DNA into minimal bacterial and human target sequences mediated by a conjugative relaxase. *PloS one* 7, e31047
35 Sclimenti, C.R., *et al.* (2001) Directed evolution of a recombinase for improved genomic integration at a native human sequence. *Nucleic acids research* 29, 5044-5051
36 Ehrhardt, A., *et al.* (2006) Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. *Human gene therapy* 17, 1077-1094
37 Kotin, R.M., *et al.* (1990) Site-specific integration by adeno-associated virus. *Proceedings of the National Academy of Sciences of the United States of America* 87, 2211-2215
38 Kotin, R.M., *et al.* (1992) Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *The EMBO journal* 11, 5071-5078
39 Huser, D., *et al.* (2010) Integration preferences of wildtype AAV-2 for consensus rep-binding sites at numerous loci in the human genome. *PLoS pathogens* 6, e1000985
40 Drew, H.R., *et al.* (2007) Increased complexity of wild-type adeno-associated virus-chromosomal junctions as determined by analysis of unselected cellular genomes. *The Journal of general virology* 88, 1722-1732
41 Ward, P. and Walsh, C.E. (2012) Targeted integration of a rAAV vector into the AAVS1 region. *Virology* 433, 356-366
42 Hockemeyer, D., *et al.* (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature biotechnology* 27, 851-857
43 Lombardo, A., *et al.* (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nature methods* 8, 861-869
44 Liu, J., *et al.* (2009) PhiC31 integrase induces a DNA damage response and chromosomal rearrangements in human adult fibroblasts. *BMC biotechnology* 9, 31

45 Chavez, C.L.*, et al.* (2010) Kinetics and longevity of PhiC31 integrase in mouse liver and cultured cells. *Human gene therapy* 21, 1287-1297
46 Schmidt, M.*, et al.* (2000) Adeno-associated virus type 2 Rep78 induces apoptosis through caspase activation independently of p53. *Journal of virology* 74, 9441-9450
47 Berthet, C.*, et al.* (2005) How adeno-associated virus Rep78 protein arrests cells completely in S phase. *Proceedings of the National Academy of Sciences of the United States of America* 102, 13634-13639
48 Zhang, M.X.*, et al.* (2009) TAT-phiC31 integrase mediates DNA recombination in mammalian cells. *Journal of biotechnology* 142, 107-113
49 Seow, Y. and Wood, M.J. (2009) Biological gene delivery vehicles: beyond viral vectors. *Molecular therapy : the journal of the American Society of Gene Therapy* 17, 767-777
50 Robert, M.A.*, et al.* (2012) Efficacy and site-specificity of adenoviral vector integration mediated by the phage PhiC31 integrase. *Human gene therapy methods*
51 Ghosh, A.*, et al.* (2008) A hybrid vector system expands adeno-associated viral vector packaging capacity in a transgene-independent manner. *Molecular therapy : the journal of the American Society of Gene Therapy* 16, 124-130
52 Saydam, O.*, et al.* (2012) Construction and packaging of herpes simplex virus/adeno-associated virus (HSV/AAV) Hybrid amplicon vectors. *Cold Spring Harbor protocols* 2012, 352-356
53 Sitaraman, V.*, et al.* (2011) Computationally designed adeno-associated virus (AAV) Rep 78 is efficiently maintained within an adenovirus vector. *Proceedings of the National Academy of Sciences of the United States of America* 108, 14294-14299
54 Asokan, A.*, et al.* (2012) The AAV vector toolkit: poised at the clinical crossroads. *Molecular therapy : the journal of the American Society of Gene Therapy* 20, 699-708
55 Yu, H.*, et al.* (2012) Gene delivery to mitochondria by targeting modified adenoassociated virus suppresses Leber's hereditary optic neuropathy in a mouse model. *Proceedings of the National Academy of Sciences of the United States of America* 109, E1238-1247
56 Fernandez-Gonzalez, E.*, et al.* (2011) Transfer of R388 derivatives by a pathogenesis-associated type IV secretion system into both bacteria and human cells. *Journal of bacteriology* 193, 6257-6265
57 Llosa, M.*, et al.* (2012) New perspectives into bacterial DNA transfer to human cells. *Trends in microbiology* 20, 355-359
58 Cassell, G.D. and Weitzman, M.D. (2004) Characterization of a nuclear localization signal in the C-terminus of the adeno-associated virus Rep68/78 proteins. *Virology* 327, 206-214
59 Agundez, L.*, et al.* (2011) Nuclear targeting of a bacterial integrase that mediates site-specific recombination between bacterial and human target sequences. *Applied and environmental microbiology* 77, 201-210
60 Andreas, S.*, et al.* (2002) Enhanced efficiency through nuclear localization signal fusion on phage PhiC31-integrase: activity comparison with Cre and FLPe recombinase in mammalian cells. *Nucleic acids research* 30, 2299-2306
61 Woodard, L.E.*, et al.* (2010) Effect of nuclear localization and hydrodynamic delivery-induced cell division on phiC31 integrase activity. *Gene therapy* 17, 217-226
62 Whittaker, G.R. and Helenius, A. (1998) Nuclear import and export of viruses and virus genomes. *Virology* 246, 1-23
63 Gonzalez-Perez, B.*, et al.* (2009) Changing the recognition site of a conjugative relaxase by rational design. *Biotechnology journal* 4, 554-557
64 Henckaerts, E. and Linden, R.M. (2010) Adeno-associated virus: a key to the human genome? *Future virology* 5, 555-574
65 Ding, W.*, et al.* (2005) Intracellular trafficking of adeno-associated viral vectors. *Gene therapy* 12, 873-880

66 Duan, D.*, et al.* (1998) Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. *Journal of virology* 72, 8568-8577
67 Daya, S.*, et al.* (2009) Adeno-associated virus site-specific integration is mediated by proteins of the nonhomologous end-joining pathway. *Journal of virology* 83, 11655-11664
68 Thorpe, H.M. and Smith, M.C. (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proceedings of the National Academy of Sciences of the United States of America* 95, 5505-5510
69 Sakuma, T.*, et al.* (2012) Lentiviral vectors: basic to translational. *The Biochemical journal* 443, 603-618
70 Bouyac-Bertoia, M.*, et al.* (2001) HIV-1 infection requires a functional integrase NLS. *Molecular cell* 7, 1025-1035
71 Engelman, A. and Cherepanov, P. (2012) The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature reviews. Microbiology* 10, 279-290
72 Abremski, K. and Hoess, R. (1984) Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *The Journal of biological chemistry* 259, 1509-1514
73 Thyagarajan, B.*, et al.* (2000) Mammalian genomes contain active recombinase recognition sites. *Gene* 244, 47-54
74 Groth, A.C.*, et al.* (2000) A phage integrase directs efficient site-specific integration in human cells. *Proceedings of the National Academy of Sciences of the United States of America* 97, 5995-6000
75 Feng, D.*, et al.* (2006) A 16bp Rep binding element is sufficient for mediating Rep-dependent integration into AAVS1. *Journal of molecular biology* 358, 38-45

Figures

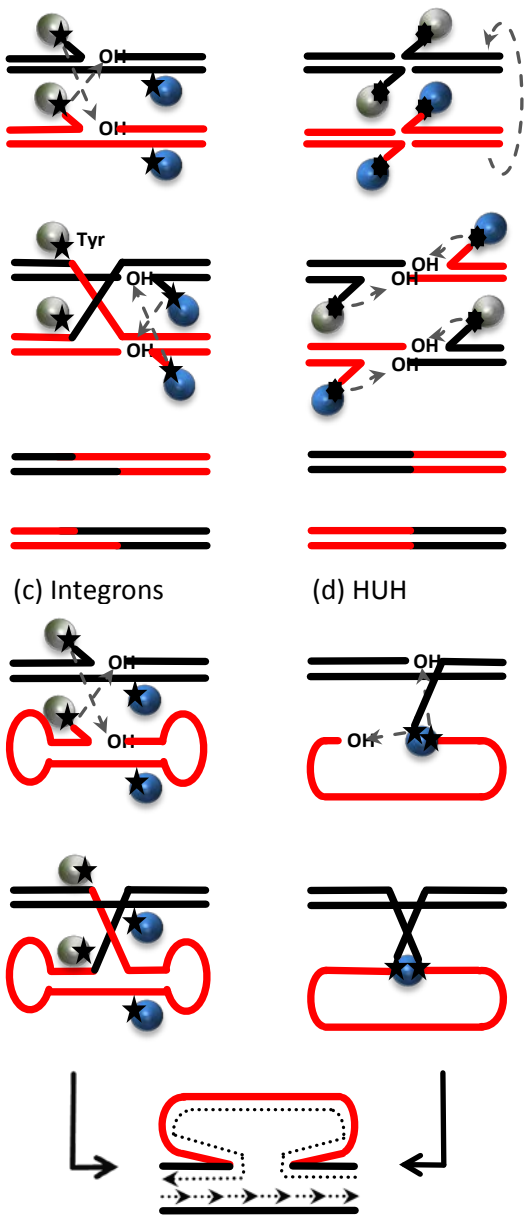
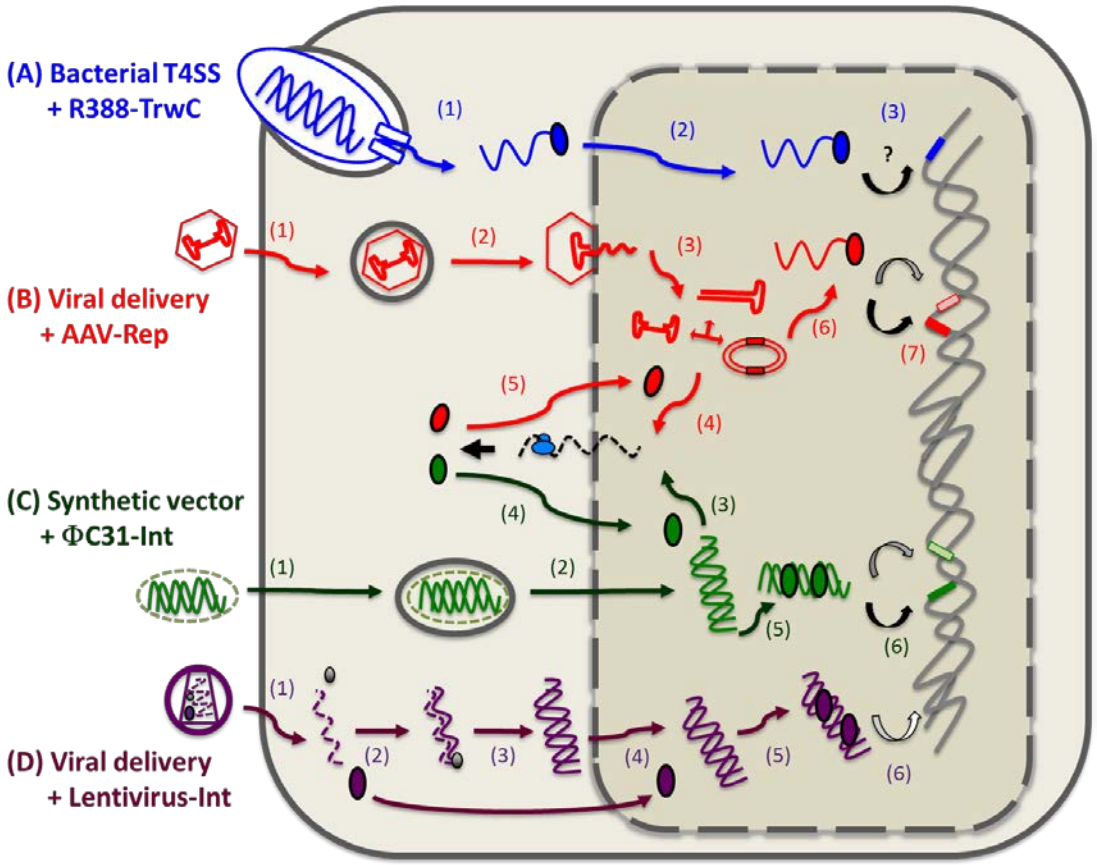


Figure 1. Comparison of recombination reactions catalyzed by SSR belonging to different families. A) Tyr-SSR mediate conservative recombination by sequential action of two pairs of monomers, each catalyzing a strand-transfer reaction. The first crossover leads to the formation of a holiday junction (HJ). B) Ser- SSR mediate conservative recombination by concerted cleavage of four monomers, followed by a conformational switch of two of them, and both strand exchanges. C) In the case of integrons, Int is an atypical Tyr-SSR which recognizes a folded single-stranded substrate and catalyzes its integration through a single crossover which leads to the formation of an "atypical holiday junction", aHJ [8]. This intermediate is resolved by replication. D) Proteins of the HUH family of single-strand transferases such as R388-TrwC have been shown to catalyze also the integration of single-stranded substrates through two transesterification steps which can be catalyzed by the two catalytic Tyr residues of the protein. The intermediate would be resolved by replication as in C).



574

575

576 **Figure 2. Different pathways for DNA delivery and gene targeting for human genomic**
577 **modification.** Representation of the different tools described in the text, which are under
578 investigation or already being used for gene therapy purposes. **(A) Bacterial T4SS + R388-**
579 **TrwC**: (1) The human pathogen *Bartonella henselae* transfers a plasmid which codes for the
580 gene of interest together with the R388 conjugative relaxase TrwC. TrwC (blue sphere)
581 covalently attached to the single stranded DNA (ssDNA) is secreted via the bacterial T4SS [56].
582 (2) Once in the cytoplasm TrwC-ssDNA complex has to reach the nucleus; TrwC can be targeted
583 to the nucleus [59]. (3) It is expected that the integrase catalyses a site-specific integration
584 reaction (curved arrow in black) of the attached DNA into either Hu5 or HuX sites (represented
585 by a blue square), both previously shown to be targets for TrwC integrase in *in vitro* assays
586 [34], but this activity has yet to be proven in vivo. **(B) Viral delivery + AAV-Rep**: (1) AAV virus
587 binds to the cell using different surface glycans as receptors and specific coreceptors for
588 efficient infection [64]. (2) AAV is internalized by endocytosis via clathrin-coated vesicles,
589 followed by escape from the vesicles [65]. It is likely that AAV injects its genome into the
590 nucleus. (3) The ssDNA genome is replicated to dsDNA, required for gene expression. During
591 replication intermediates such as circular double strand molecules are presumably assembled,

allowing episomal persistence to the viral genome [66]. (4) RNA synthesis (dashed black line) and subsequent translation (ribosome represented in light blue) is necessary to provide Rep integrase (red sphere). (5) An NLS targets Rep to the nucleus [58]. (6) The integrase targets its viral origin (presumably ssDNA [67]) in order to form protein-DNA complexes. (7) Site-specific integration (curved arrow in black) mediated by Rep occurs into *AAVS1* (red square, [15]). Other pseudo-sites (light red square and curved arrow in grey) have also been reported to act as targets with lower efficiency [39]. **(C) Synthetic vector + ϕC31-Int**: (1) DNA coding for the transgene and ϕC31 Int can be transfected directly to the target cell or introduced with synthetic vectors such as polymers or liposomes which interact with cellular receptors to achieve internalization through endocytosis. (2) After escaping from the vesicles, the vector is disassembled and the dsDNA (double line in green) reaches the nucleus. (3) This is followed by transcription and translation of the DNA. (4) ϕC31 Int (green sphere) is expressed in the cytoplasm and the protein has to find its way towards the nucleus. (5) Once inside, the integrase binds its target dsDNA and (6) catalyzes its site-specific integration [68] (curved arrow in black) into specific hotspots (green square), as well as into many pseudo-sites (curved arrow in grey and light green square) in the human genome [21]. **(D) Viral delivery + Lentivirus-Int**: (1) The lentivirus RNA genome (dashed purple line) is contained in the capsid (purple trapezoid) together with the reverse transcriptase (black spheres) and the viral integrase (purple sphere). Upon cell entry through binding to receptors and coreceptors, capsid proteins are uncoated, resulting in the release of the RNA genome together with the viral proteins into the cytoplasm [69]. (2) Reverse transcription takes place, giving rise to an RNA-DNA hybrid structure, which is subsequently converted to dsDNA (3) (double purple line). The dsDNA enters the nucleus, and the integrase leaded by an NLS [70] is imported to the nucleus (4). (5) The integrase binds the viral origins within the dsDNA and (6) catalyzes random integration (curved arrow in white) in the human genome [71].

Table 1. A comparison of the most relevant SSR discussed in this work

SSR	Family [1]	Source	Host	Biological role	Stable integration [2]
Cre	Tyr-SSR	P1 phage	*Escherichia coli*	resolution of phage genome multimers	NO
Int-ΦC31	Ser-SSR	ΦC31 phage	*Streptomyces lividans*	Integration and excision of phage genome	YES [3]
AAV-Rep	HUH	Adeno-associated virus	*Homo sapiens*	Replication and Integration of phage genome	YES [4]
TrwC	HUH	R388 Plasmid	*Escherichia coli*	Processing and Leading DNA during bacterial conjugation	YES [5]

[1] See text for details and Figure 1 for description of SSrec reaction mechanisms.

[2] Ability of the SSR to catalyze integration of foreign DNA on a target site present in a different genome, without catalyzing its excision subsequently.

[3] Int requires extra factors to catalyze excision.

[4] The rescue of AAV proviruses is thought to be mediated by the initiation of replication of integrated viral genomes. For this the cellular replication machinery and helper virus co-factors are required.

[5] TrwC-mediated reaction is reversible, but integration into the human targets could be irreversible (see text for details).

628 **Table 2. SSR target sites in their host genomes and in the human genome**

SSR	Natural target		Targets in human genome		Refs
			bona fide	***pseudo-sites***	
Cre	*loxP*	ATAACTTCGTATAGCATACATTATACGAAGTTAT	none	4 ψlox	[72], [73]
Int-ΦC31	*attP* *attB*	CCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGG GTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCG	none	101 ψattP	[74], [21]
AAV-Rep	RBE_{itr} RBE_{P5}	GAGCGAGCGAGCGCGC GCCCGAGTGAGCACGC	*AAVS1*	*AAVS2, AAVS3*	[75], [38], [39]
TrwC	*nic*	GGTGCGTATTGTCTATA	none	2	[11], [59]

629