Reprod. Nutr. Dev. 41 (2001) 465–485

Review

Theoretical mechanisms in targeted and random integration of transgene DNA

Kevin SMITH*

School of Science and Engineering, University of Abertay Dundee, Kydd Building, Bell Street, Dundee DD1 1HG, UK

(Received 4 September 2001; accepted 10 December 2001)

Abstract — The genetic manipulation of mammalian cells and animals would be greatly expedited if gene targeting could be reliably achieved in the widest possible range of host cell types. This paper considers empirical evidence and theoretical considerations associated with transgene integration, and concludes that utilisation of gene targeting in non-selective systems awaits further progress in modelling homologous recombination.

transgenesis / gene targeting / gene therapy / integration / homologous recombination / concatenation

1. INTRODUCTION

In animal genetic manipulation (transgenesis) and also in gene therapy, the random nature of transgene integration poses several potential problems: (i) the transgene may be expressed poorly or inappropriately; (ii) an essential gene may be disrupted, or an oncogene activated; (iii) the outcome will be inconsistent between identically treated cells; and (iv) transgenesis is largely restricted to adding new functions rather than eliminating or altering endogenous gene expression.

Gene targeting is the key to surmounting these problems. To be of practical use, the process of gene targeting would comprise the ability to efficiently direct the transgene to a specific, precisely defined genomic site such that a particular sequence could be inserted or substituted at that target locus. This paper examines gene targeting at the mechanistic and phenomenological levels.

* Correspondence and reprints
E-mail: k.smith@abertay.ac.uk

1.1. Cell types

Gene targeting is an achievable goal in mammalian cells. However, progress has been limited by a lack of targeting efficiency. Studies on mammalian cells in vitro demonstrate that the vast majority of interactions between transgene and endogenous DNA result in random rather than targeted integrations. The reported ratio of random to targeted integration varies enormously, from around 1:4 to more than 1 000 000:1. In most cases, the ratio is between 1 000:1 and 10 000:1 [4, 8, 13, 35, 43, 46]. It is noteworthy that the low efficiency of targeted integration in mammalian cells is in marked contrast to that which occurs when DNA is transfected into lower eukaryotes such as yeast: under appropriate empirical conditions, targeting is the norm and random integration the exception for such organisms.

Due to the low efficiency of targeting, it is necessary to select for targeted outcomes against a background of random outcomes. Such selection is possible with cultured cells. Embryonic stem (ES) cells may be subjected to selection such that targeted cells survive at the expense of cells containing randomly integrated trangenes [55]. Because ES cells are totipotent, surviving (targeted) cells when transferred into early embryos give rise to founder animals able to produce transgenic offspring. The production of successfully targeted mice by the ES cell route is now fairly routine. However, this approach is at present limited until such time as ES cells are developed for non-murine species.

Targeted outcomes cannot be selected for in zygotes. This does not, however, mean that targeted events are impossible in zygotes. Very few studies have looked at gene targeting in zygotes, probably due to the expense of gene transfer and analysis. A landmark study by Brinster et al. [6] involved the analysis of 506 transgenic founder mice. These animals were produced by microinjecting zygotes from mice containing a deletion in the major histocompatibility (MHC) class II *E*α gene. The transgene construct was based on sequences from this gene and included the sequences absent in the host mice. A single mouse was found to have undergone targeted correction of the *E*α gene deletion. This study shows that gene targeting is possible in zygotes. Although it is not possible to determine an accurate frequency of gene targeting in zygotes from this work, it appears that the rate of targeting (1 in 506 animals), although quite high compared with cultured cells (see above), is too low to permit the efficient use of gene targeting in zygotes.

Nuclear transfer (NT) is the latest method for introducing targeted changes into the germline. NT involves replacing the oocyte's genome with that from another cell. The genetic material from the donor cell is "reprogrammed" into totipotency by the recipient oocyte, such that the "reconstructed" egg is able to develop into a viable animal [7]. NT per se is effective in a very broad range of animal types including cattle, goats, mice and sheep [9]. Transgenes can be introduced to donor cells in vitro, permitting the production of genetically modified animals by NT [33]. Because selection can be applied to cultured donor cells, NT can be used to produce gene-targeted transgenic animals. Although in its infancy, the use of NT for gene targeting certainly works, as demonstrated by the recent generation of the first gene targeted sheep [26]. The method's ability to work with many (possibly all) animal types indicates that NT holds great promise as a tool for gene targeting.

Gene targeting has great potential in gene therapy, because it offers the ability to precisely repair mutant genes to restore their normal functioning. Also, in contrast to gene therapy approaches involving randomly integrating transgenes, gene targeting is capable of correcting dominant, gain-of-function mutations. In situ gene targeting in humans is a distant prospect, because selection cannot be used in vivo. However, ex vivo

targeting approaches are under development. Hatada et al. [17] used gene targeting to correct a defective hypoxanthine phosphoribosyltransferase (HPRT) gene in hematopoietic progenitor cells. The approach was similar to gene targeting with ES cells or NT, in that selection was used to enrich for targeted outcomes. If similar successes can be obtained with pluripotent clonogenic cells such as hematopoietic stem cells (HSCs), it may be possible to return such targeted cells to the body of the patient such that repopulation by the corrected cells yields a therapeutic or curative outcome for hematological and other disorders.

1.2. Homologous recombination

Gene targeting depends upon homologous recombination (HR). HR refers to any process in which two similar DNA sequences interact and exchange genetic information. Cells have the inherent ability of performing HR (the most obvious natural occurrence of HR is in meiotic recombination; some DNA repair mechanisms undoubtedly use similar processes). Gene targeting seeks to harness HR such that transgenes can be induced to undergo recombination with their homologous endogenous counterpart sequences.

Molecular biology has not yet elucidated the details of HR. It remains conceivable that the efficiency of HR in mammals could be artificially enhanced in some way, such that selection would not be necessary. For example, if an appropriate mammalian recombinase enzyme (or enzyme complex) was to be discovered, the relevant gene for the recombinase might be co-introduced with the transgene molecules in order to boost the rate of HR. The expression of the recombinase would have to be tightly controlled, however, since excessive production of known recombinase enzymes is frequently associated with recombination abnormalities and cytotoxic effects. Nevertheless, if it were to become possible to use transfected recombinases to enhance the efficiency of HR, the utility of selection-free gene targeting would be quite significant. Possible benefits might include circumvention of the problem of non-availability of non-murine ES cells (by allowing gene targeting to be used with zygotes), and the use of gene targeting as part of in vivo gene therapies. However, although several candidate genes for enhancing HR have been described, the use of such genes in gene targeting is at an early experimental stage [46].

Novel alternative ways to improve targeting efficiencies have also been proposed. For example, triple helix-forming oligonucleotides (TFOs) are able to stimulate HR in mammalian cells [45]: it is possible that TFOs could be used as agents to augment targeting. Similarly, endonuclease molecules may be able to deliver double-stranded breaks to target DNA, leading to improved rates of HR [37]. However, such approaches are at an early experimental stage, and the possibilities presented by these approaches are beyond the scope of the present discussion.

2. MODELS OF HR

HR has proved to be highly recalcitrant to molecular/biochemical analysis. Indeed, the ability to perform an entire recombination reaction in the test tube remains an unachieved goal. Thus an understanding of HR depends primarily on genetic data. Most of this data has come from model systems based on lower eukaryotes such as moulds and yeasts. From such data several plausible hypotheses or models may be constructed. However, determining the most valid model(s) by distinguishing between specific mechanistic details will require advances at the level of precise molecular analysis. The following sections (2.1–2.6) review the major models of HR.

2.1. Simple "crossing-over"

In conceptual terms, the simplest mechanism of HR would involve the following events: (1) pairing of duplex DNA on the basis of homology; (2) double-strand cleavage at a homologous position; (3) duplex DNA strands "crossing-over" each other such that the broken ends become juxtaposed; and (4) ligation (Fig. 1). This "crossing-over" model is able to account for the "swapping" of entire chromatid arms or segments (swapped where two crossovers occur within a single chromatid) in meiotic recombination.

2.2. The Holliday model

In the "crossing-over" model, each recombined allele should segregate with equal frequency, since genetic information is neither created nor destroyed in a reciprocal "crossing-over" event. Genetic crosses, however, do not always show this 1:1 segregation outcome: at low frequency, "aberrant" segregation is observed, where the ratio is skewed [18, 19]. Tetrad analysis in moulds such as neurospora illustrate this well. For example, in the cross between loci AB × ab, "crossing-over" should give equal numbers of the progeny Ab and aB, represented in the eight-spored ascus as a ratio of 4:4. When aberrant segregation occurs, outcomes such as 6:2 or 5:3 are observed [34]. Aberrant segregation indicates that nonreciprocal transfer of genetic information may occur in HR, a process known as gene conversion. The "crossing-over" model cannot account for gene conversion.

In 1964, Robin Holliday proposed a model that is able to explain the existence of gene conversion [18]. The Holliday model involves the following events: (1) homology pairing, as per the "crossing-over" model; (2) single-strand cleavage ("nicking") of both duplexes at a homologous site; (3) "strand invasion" of free ends between duplexes, such that the crossed strands unite the duplexes in a structure called a Holliday junction; (4) "branch migration" of the Holliday junction, moving the crossover point away from its original position; and (5) "resolution" of the Holliday junction by single strand nicking (followed by ligation of free ends) (Fig. 2).

The special features of the Holliday model are: (a) the Holliday junction; and (b) branch migration. The existence of the former structure permits the latter process to occur. Holliday junctions are envisaged as symmetrical structures in which all four (single) strands of DNA are equivalent, with the junction being in a state of rapid equilibrium in vivo (Fig. 3). In branch migration, the bases on each side of the junction exchange places. Since breakage of base pairing is balanced by formation of new base pairing, the exchange process is thermodynamically neutral: thus it may be envisaged that the crossover point is able to easily and quickly migrate along the paired duplexes. As branch migration occurs, in its wake heteroduplex DNA will be formed.

Assuming that the homologous alleles undergoing HR are nonidentical, the heteroduplex DNA formed by branch migration will contain mismatched bases. Such mismatches may be repaired. Any such repair would involve bases being replaced using the opposite strand as a donor of sequence information. The repair of mismatched bases in heteroduplex DNA would explain the occurrence of aberrant segregation of the sort described above. Thus, the Holliday model is able to account for gene conversion.

Implications for transgenesis can be drawn from the central features of the Holliday model, in that gene conversion may occur between transgene and endogenous DNA. Indeed, depending upon the "plane" of resolution of the Holliday junction, the targeted transgene may not actually become integrated into the target genome; however the target sequences may have undergone gene conversion-mediated alteration. If this

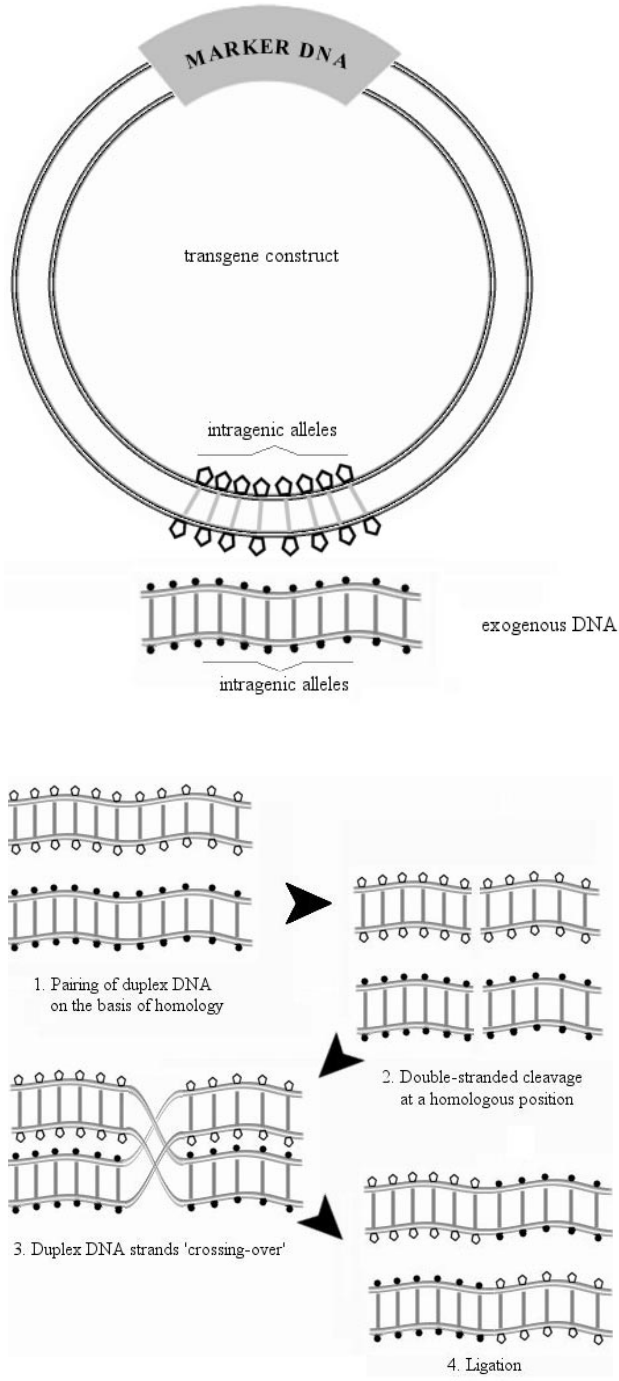

Figure 1. Simple Crossing-over Model.

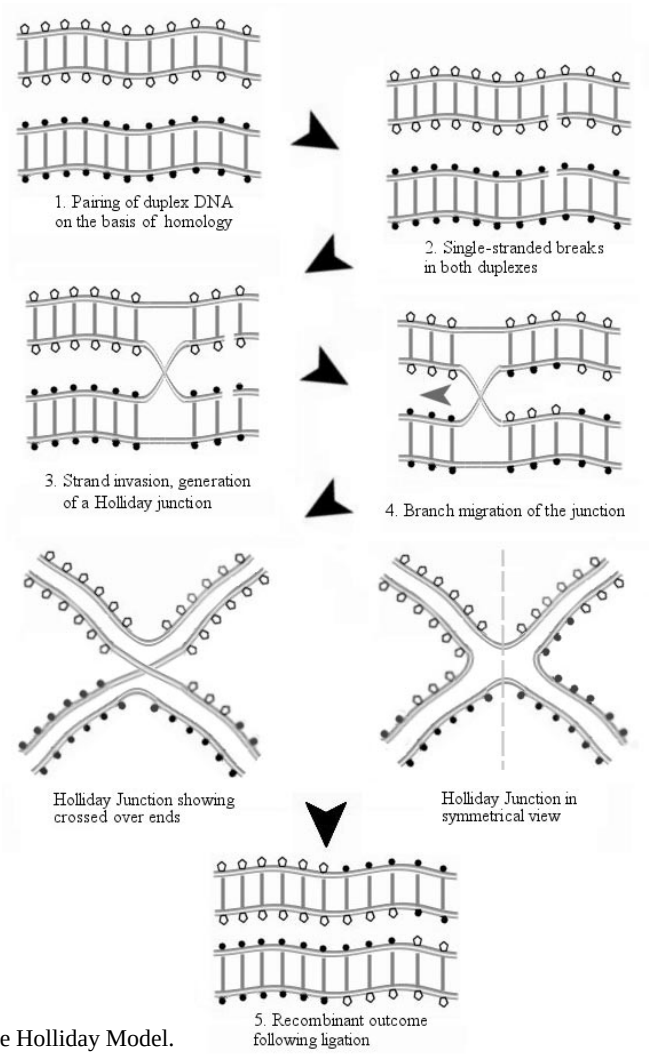

Figure 2. The Holliday Model.

process could be controlled (or its desired outcomes selected for), it would provide the means to directly introduce subtle changes into endogenous genes. To harness gene conversion in this way would be of major importance for gene therapy.

2.3. The Meselson-Radding model

Analysis of gene conversion genetic data indicates that relative segregation ratios often vary significantly from predicted values, the latter values being based on the assumption that the direction of mismatch repair is random. The apparent breakdown of randomness is most simply explained by postulating the existence of asymmetric heteroduplex DNA at the HR initiation region.

In 1975, Matthew Meselson and Charles Radding proposed a model of HR that incorporates asymmetric heteroduplex DNA [27]. The Meselson-Radding model involves the

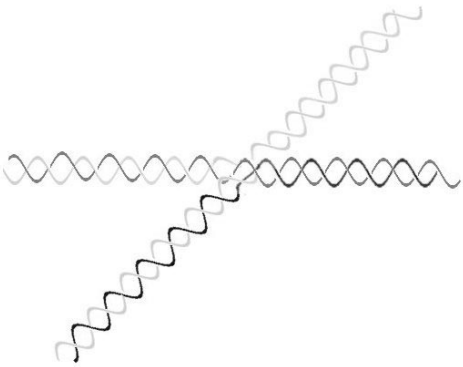

Figure 3. Holliday junction envisioned as a symmetrical structure.

following events: (1) homology pairing, as per the previous models; (2) nicking of (one strand of) one duplex; (3) strand invasion of a free end from the nicked strand into the intact duplex; (4) "strand exchange" whereby the invading strand undergoes progressive homology pairing with one strand of the invaded duplex (the other, displaced strand is assumed to play no further part in the recombination event); (5) Holliday junction formation, as per the Holliday model; (6) branch migration; and (7) resolution (Fig. 4).

The special features of the Meselson-Radding model are: (a) the single nick (2, above); and (b) strand exchange (4, above). Essentially, this creates a single-stranded gap in one of the recombining duplexes: the loss of sequence information in this region generates the previously mentioned asymmetric heteroduplex DNA. Mismatch repair, of the sort envisaged for the Holliday model, cannot occur within the gapped region. Instead, "gap repair" could occur, with sequence information being donated from the single-strand within the gapped region (Fig. 5). Gap repair is thus unidirectional, and therefore may account for the breakdown of randomness observed in gene conversion genetic data. The Meselson-Radding model fits with the observation that most gene conversion genetic data is indicative of random, bidirectional repair, because mismatch repair is assumed to occur following branch migration: thus the model envisions both asymmetric and symmetric heteroduplex DNA formation. The Meselson-Radding model is able to account for virtually all data from meiotic recombination. Additionally, the model simplifies the concept of initiation of recombination, in that only a single nick is required. This contrasts with the Holliday model, where the nuclear machinery must somehow be able to precisely localise homologous sequences on the recombining duplexes prior to nicking. Conceptually, single-stranded DNA seems to be an essential prerequisite for any sequence-level homology recognition, otherwise it is very difficult to envisage the transient base-to-base interactions that presumably must be an integral part of the search for homology. It is noteworthy that subsequent to the elucidation of the Meselson-Radding model, the *E. coli* recombinase enzyme RecA was discovered: this enzyme promotes strand exchange between single-stranded DNA molecules and homologous DNA duplex molecules. More recently, eukaryotic enzymes with abilities similar to RecA have been described. Of particular interest is Rad51p, a homologue of RecA: this enzymes also binds to single-stranded DNA and catalyses strand exchange [28, 39].

Implications for transgenesis can be drawn from the special features of the Meselson-Radding model. Firstly the concept of a single nick is encouraging in that it suggests that incoming transgene molecules may be able to search through the endogenous duplex DNA for homologous sequences, without requiring any preliminary homology recognition/alignment (for meiotic recombination, it is conceptually possible to rely to an extent on the gross-level homology alignment of homologous chromosomes). Secondly, the single nick concept suggests that it may be possible to enhance targeting frequencies by generating a single nick in a transgene prior to its introduction to the cell. Finally, the concept

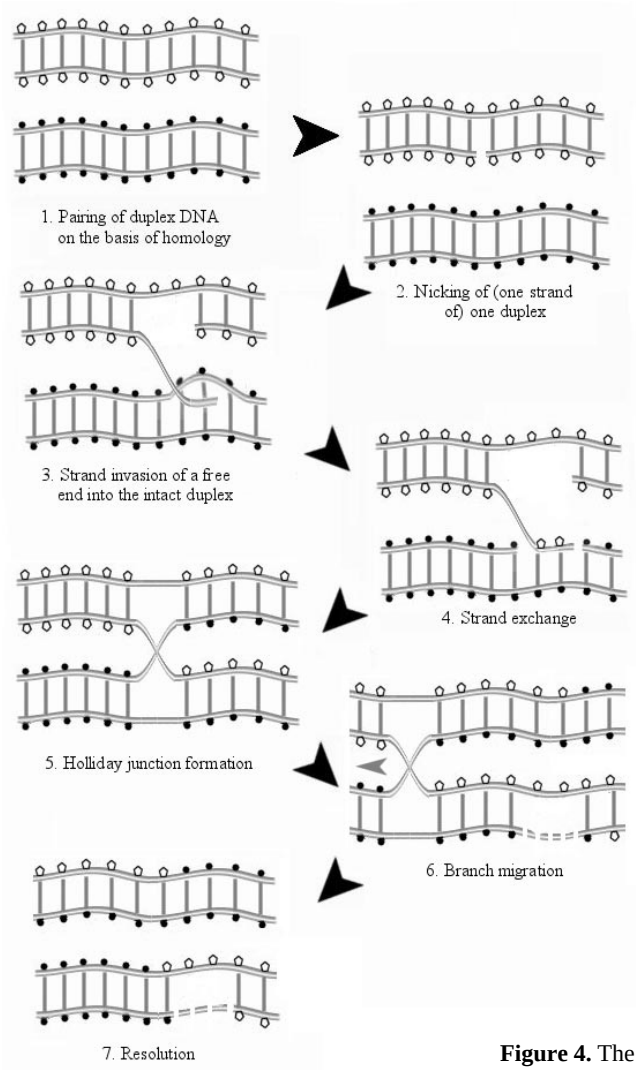

Figure 4. The Meselson-Radding Model.

of strand transfer/gap repair suggests that it may be possible to ensure that the flow of gene conversion information is in the desired direction (at least for sequences close to the single nick), i.e. from transgene to endogenous gene.

2.4. The double-strand-break repair model

Yeast cells are amenable to transfection by plasmids. Transfected plasmids undergo gene targeting if they contain sequences homologous to yeast endogenous sequences. The Meselson-Radding model is unable to explain certain findings from yeast transfection experiments [29, 40]. Firstly, introduction of a double-strand break in the region of homology on the plasmid leads to products of HR that are identical to those from unbroken (i.e. closed circular) plasmids; however, the "frequency" of targeting is markedly increased. Secondly, introduction to the plasmid of a gap (in the region

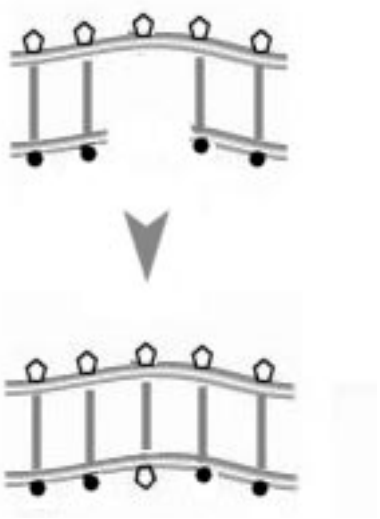

Figure 5. Gap Repair.

of homology) yields the same HR products as does the break, however the HR frequency is markedly elevated.

In 1983, Jack Szostak, Terry Orr-Weaver, Rodney Rothstein and Frank Stahl proposed a model of HR that accounts for the above observations [40]. The Szostak, Orr-Weaver, Rothstein and Stahl, or double-strand-break repair (DSBR) model of HR involves the following events: (1) homology pairing; (2) double-strand breakage of one duplex; (3) generation of a gap by exonuclease action on the double-stranded break; (4) strand invasion of a free end from the gapped strand into the intact duplex; (5) strand exchange, accompanied by displacement of a "D loop" of single-stranded DNA from the uncut duplex; (6) invading strand acting as primer for DNA synthesis using nondisplaced strand as template, leading to D loop enlargement; (7) D loop "invasion" into gap site on opposite duplex, forming asymmetric duplex DNA and generating two Holliday junctions; (8) branch migration (of both Holliday junctions); and (9) resolution (by cutting/ligation at both Holliday junctions) (Fig. 6).

The special features of the DSBR model are: (a) the double-strand break/gap (2 and 3, above); and (b) the generation of two Holliday junctions (7, above). As with the Meselson–Radding model, both gap repair and mismatch repair are envisaged, leading to gene conversion.

Implications for transgenesis can be drawn from the special features of the DSBR model, in addition to the implications arising from the Meselson-Radding model. Firstly, the double-strand break concept suggests that it may be possible to enhance targeting frequencies by generating a break in a transgene prior to its introduction to the cell. Secondly, the concept of gap generation/repair suggests that information flow at the region of the gap may go in the "wrong" direction, i.e. from endogenous gene to transgene, at least in cases where the transgene construct has been introduced as a linear molecule. In practice this need not always be a problem, because the transgene can be designed such that although its recombining (homologous) element(s) may undergo such conversion, the net result is a targeted integration of the desired (nonhomologous) sequences.

2.5. Chimeric oligonucleotides and HR

Chimeric oligonucleotides (COs) are small (ca. 50 bp), self-complementary DNA-RNA oligonucleotides with a double-hairpin configuration. These highly specialized transgene molecules have been developed as tools for gene targeting [51]. By virtue of their inherent structure, COs cannot produce large (several nucleotides) changes in targeted genes; however they are able to generate base substitutions at precisely defined genomic positions. An important feature of COs is that they have a higher structural integrity than conventional transgene molecules, and are thus less prone to destruction within the nucleus prior to targeting. Following initial reports of success

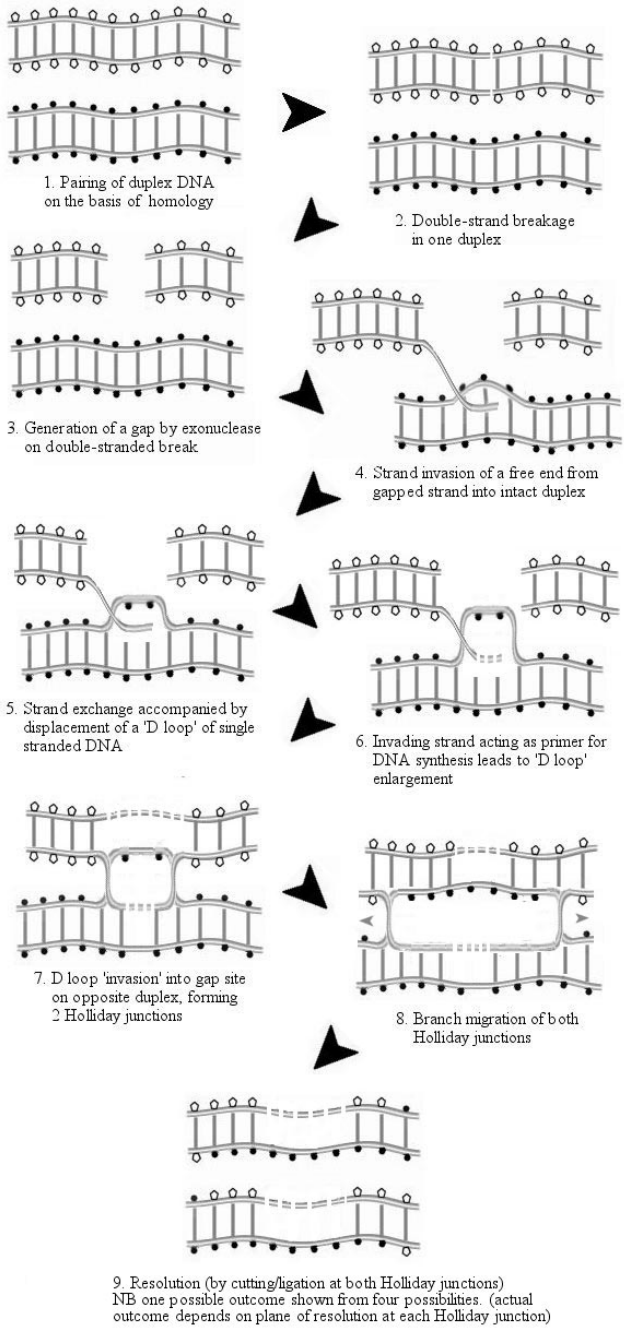

Figure 6. Double Strand Break Repair model.

in 1996 [10, 52], COs have been used to correct various point mutations in a range of mammalian cells. However, reported rates of gene repair have been highly variable (ranging from zero to ca. 40%), and occasional unexpected mutational effects have been found [11, 50]. These problems have engendered a certain amount of controversy concerning the use of COs in gene targeting, particularly in the context of gene therapy.

The mechanism of CO action has not yet been fully elucidated. However, gene correction experiments designed to explore this question indicate that targeting by COs involves several of the fundamental features found in the foregoing models of HR, such as homology pairing and strand exchange [14, 15]. Thus, the action of COs represents a special case of HR. Extending from the findings of Gamper et al, a provisional model for the HR involved in CO may be envisaged, as follows: (1) homology pairing, involving a stabilised displacement loop in which the entire CO resides; (2) strand invasion of the free end of the CO into one target strand, leading to Holliday junction formation; (3) branch migration; and (4) resolution (Fig. 7). If the CO includes a single base difference compared with the target site, the result of steps 1–7 above will be a mismatch. As with the previously described models of HR, such mismatches may be repaired. In 50% of cases in which the mismatch is repaired, the target sequence should be converted to that of the targeting CO.

2.6. Conclusion

The simple "crossing-over" and Holliday models are inadequate as explanations of HR. The Meselson-Radding model is able to account for data from genetic crosses equally as successfully as the DSBR model, and has the advantage of being conceptually the simpler of the two models. However, only the DSBR model is able to account for yeast transfection data. In terms of transgenesis, deciding between the two models is important, in that the implications differ between the two models. These implications include: (a) the form of DNA that may be used for the transgene; and (b) the direction of information flow between transgene and endogenous gene (and the associated transgene design implications).

At the moment, the fine molecular/biochemical details of HR are poorly understood, therefore it is not yet possible to definitively decide which model (Meselson-Radding or DSBR) is most accurate. Indeed, it has been suggested that "multiple" recombination pathways may exist [34]. It is possible that different classes of organisms have their own specific pathways; but it is also possible that each class of organism is able to carry out HR by the use of more than one pathway. In simple systems such as *E. coli*, many recombination-disabled mutants are known. However, not all mutants block all HR events: one mutation may block interplasmid recombination but leave conjugational recombination unaffected, for example. If such multiple pathways exist in higher eukaryotic cells, it may in principle be possible to influence the host cell (for example by altering the structure of the transgene molecule) such that one particular pathway is used. Such "pathway selection" could influence the outcome of a targeting event, for example by ensuring that information flow occurs in the desired direction.

It is conceptually possible that independent, different HR pathways may be operational during a single targeting event. This is probably most likely with replacement ("ends-out" or Ω-type) transgenes, in which the homology regions are present at the ends of the transgene, separated by heterologous DNA. In such cases, an independent HR event is expected at each end of the homologously aligned transgene: such events would not necessarily be identical in nature. Indeed, it is also possible that HR may occur at one transgene end and nonhomologous

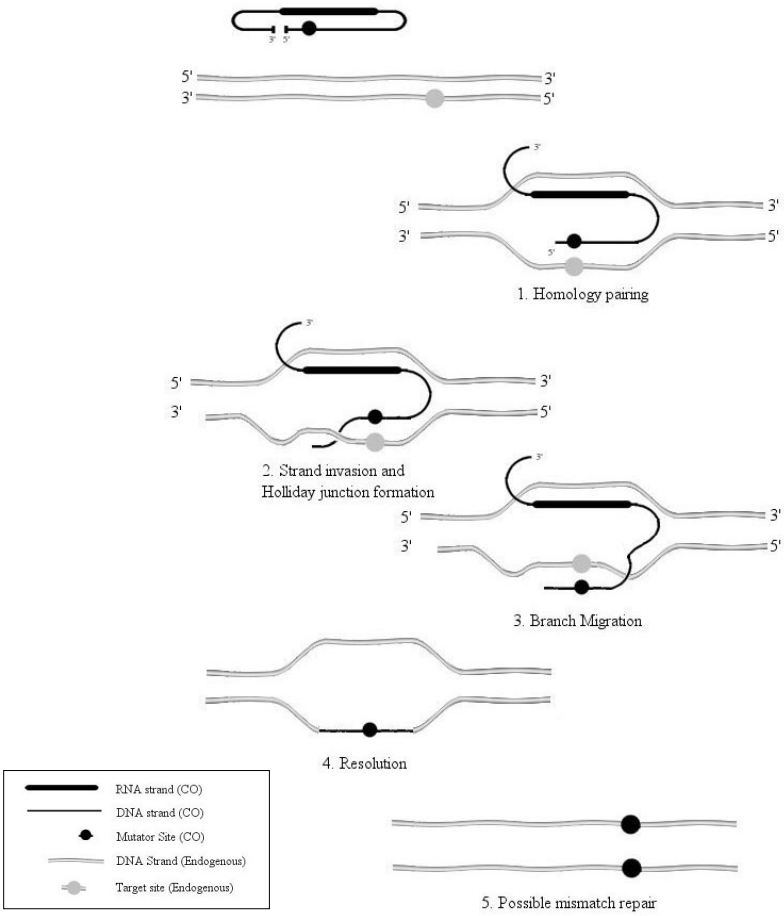

Figure 7. Proposed model for gene targeting by chimeric oligonucleotides.

integration at the other [1]. A further possibility, involving a variant form of DSBR has been suggested by Li et al. [24]. In this model, the transgene ends are envisaged invading the homology sites within the target duplex, to produce one Holliday junction at each site. Thus, the resulting structure (encompassing the entire transgene) would be equivalent to a single unresolved DSBR intermediate (Fig. 8).

Finally, the special case of COs fits well with the concept of multiple HR pathways. Although the mechanism of targeting used by COs has not yet been fully elucidated, it seems certain that key steps from the general models of HR are utilised by COs to yield gene correction outcomes.

3. TRANSGENE INTEGRATION IN MAMMALIAN CELLS

This section addresses the issue of what happens to transgene molecules upon introduction to the nucleus or pronucleus of the host cell. First, the events that may lead to

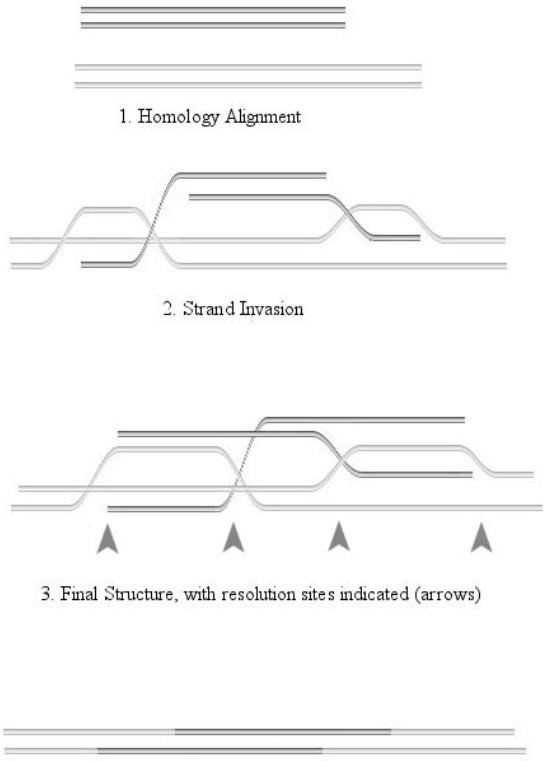

Figure 8. Lin et al. (2001) model for mammalian gene replacement.

non-targeted (i.e. random) integration are considered. Second, the phenomenology of targeted integration is reviewed.

3.1. Random integration

The arrangements of exogenous DNA integrated into the chromosomes of cultured mammalian somatic cells are very similar to the arrangements found in transgenic animals [5, 12, 16]. Furthermore, this is true regardless of whether the transgenic animals have been derived from one-cell embryos or ES cells. This suggests that the underlying molecular mechanisms of random integration are essentially the same in all cell types.

The main features of randomly integrated exogenous DNA are as follows:

• Integration occurs at a frequency of between 10–30% of cells in which transgene DNA is delivered to the nucleus;

• Integration occurs at one or, rarely, a few chromosomal sites per nucleus;

• Integrated DNA is usually present in the form of a multicopy array;

• The vast majority of arrays consist of head-to-tail associations.

The precise molecular mechanisms of random integration are not known. However, experimental data such as that outlined above has allowed the construction of models of random integration. A general model is presented below (Sects. 3.1.1 and 3.1.2).

3.1.1. Concatenation

The fact that transgenes are usually present as arrays (concatemers) indicates that

extrachromosomal events (concatenation) take place prior to chromosomal integration.

Random end-to-end joining (ligation) of transgene molecules should generate head-to-tail, head-to-head and tail-to-tail associations in a ratio of 2:1:1. However, as mentioned above (Sect. 3.1), the vast majority of arrays take the form of head-to-tail associations. Therefore, end-to-end joining cannot be an adequate explanation of concatenation. However, rare head-to-head/tail-to-tail associations do occur, so ligation would appear to be a possibility. The simplest explanation lies in the kinetics of free transgene molecules: it must be stochastically infrequent for any two transgene molecules to meet together in an end-to-end fashion, and stay together long enough for a molecule of DNA ligase to unite them [2].

If end-to-end joining cannot explain the majority of concatenation events, another form of interaction between transgene molecules must be operative. The most likely process is extrachromosomal HR between circular and linear molecules. This mechanism can be shown, in formal geometric terms, to generate exclusively head-to-tail concatemers [2].

A prerequisite for concatenation by extrachromosomal HR is the co-existence of both circular and linear transgene molecules. Experimental data shows that head-to-tail arrays result (with equal frequency) following the introduction of either circular or linear molecules [5, 12]. Therefore it is necessary to postulate the existence of two nuclear processes: (1) circularisation (by ligation of the free ends of a transgene molecule); and (2) linearisation (by random nuclease action).

Linearisation would generate circularly permuted molecules. HR could then occur between circularly permuted and circular molecules, or between circularly permuted and input linear molecules, or between individual (different) circularly permuted molecules: in all cases the effect would be the formation of a head-to-tail concatemer. Repeated rounds of HR would extend the array (i.e. increase the number of transgene copies therein) (Fig. 9).

Several cultured cell studies have demonstrated that linear DNA molecules are circularised by intracellular ligation, and a number of similar studies have indicated that circular DNA molecules are randomly cleaved [3]. Thus, circularly permuted molecules will be produced following introduction of exogenous DNA molecules.

Concatenation of circularly permuted molecules can most simply be explained by the following events: (1) homology pairing; (2) exposure of single-strand substrates at the end of each duplex; (3) formation of a duplex between exposed complementary strands; and (4) resolution (by repair of the duplex) (Fig. 10).

The fact that extrachromosomal HR occurs with high frequency amongst individual transgene molecules contrasts sharply with the low frequency of HR between transgenes and endogenous chromosomal sequences (see Sect. 1.1). There would appear to be two possible explanations, as follows: (a) the “free” nature of the interacting transgene molecules in some way enables HR to proceed very efficiently; and/or (b) the free (non-telomeric) ends of the interacting transgene molecules are very good substrates for recombinase enzyme activities. However, the molecular/biochemical details of HR remain to be elucidated (Sect. 2); thus it is not yet possible to give a precise explanation for the contrast that exists between extrachromosomal HR and the HR that underlies gene targeting.

3.1.2. Illegitimate recombination

Transgene or transgene array integration only occurs in a minority of surviving transfected cells, suggesting that (nontargeted) integration is the result of a rare intranuclear event. The simplest model would suppose that the rare intranuclear event is

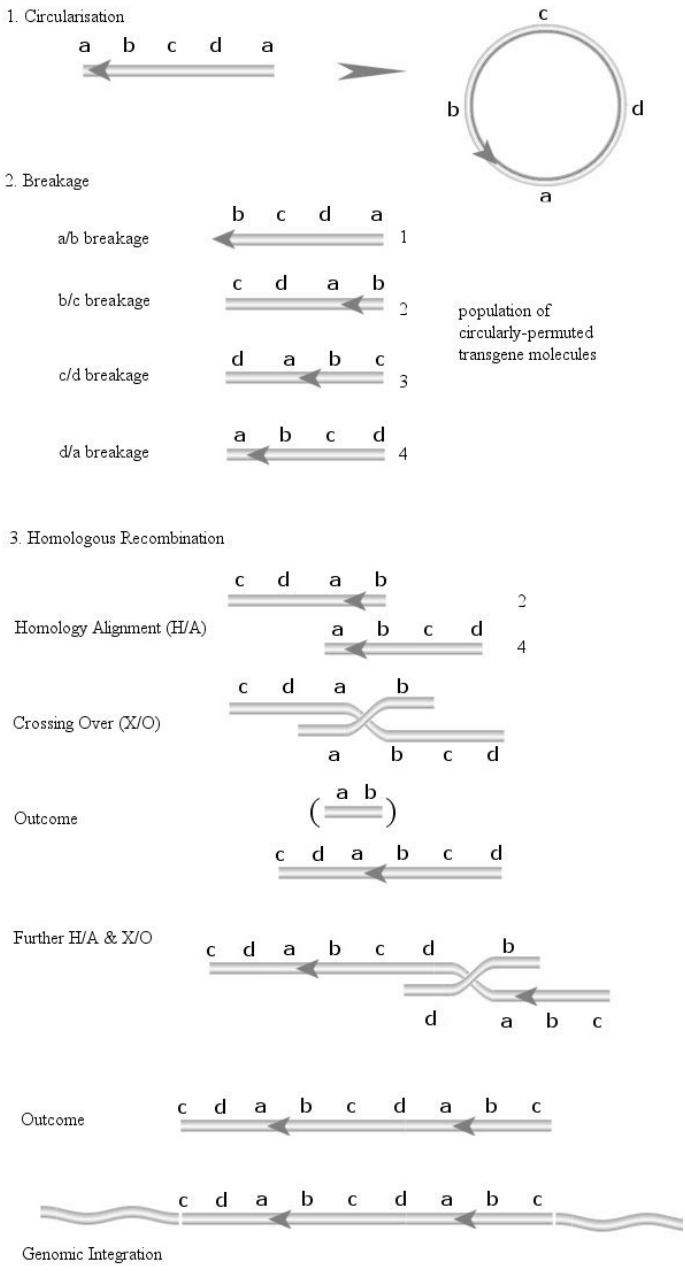

Figure 9. Model for concatenation.

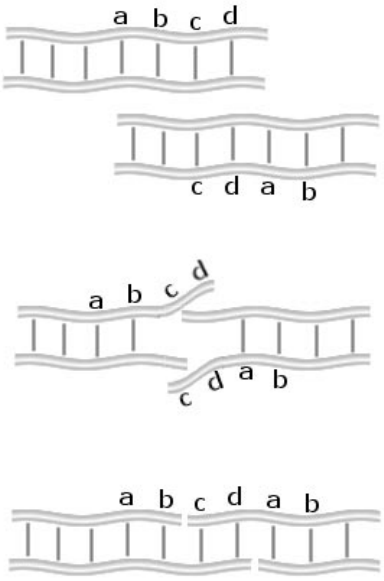

Figure 10. Model for nonconservative HR between circularly permuted (extrachromosomal) transgene molecules.

chromosomal (double-strand) breakage followed by end-joining between the transgene ends and the chromosomal broken ends. Certainly, the frequency of DNA integration is increased by irradiation of the transfected cells [30].

An alternative mode of chromosomal integration could be "illegitimate" recombination between very poorly matched sequences. Studies of the nucleotide sequences at exogenous-endogenous DNA junctions have shown illegitimate recombination in a number of cases, although the overall number of studies is small [2].

Interestingly, a diverse range of chromosomal sequence disturbances has been found in junctional studies. These include, in order of frequency: deletions, duplications, inversions and more complex rearrangements including the appearance of sequences from elsewhere in the genome or even of unknown origin. Bishop [2] has proposed an explanation of these observations that sees exposed single-stranded ends of transgene molecules initiating recombination by invading DNA duplexes.

It is now known that the majority of transgenic founder animals are mosaics. This suggests that integration occurs during DNA replication. Wilkie and Palmiter [48] have proposed that the free ends of the transgene initiate recombination by invading a replication "eye".

It may be that, in a situation analogous to that of deciding between alternative pathways of HR (see Sect. 2.5), more than one integration route is possible for randomly integrated transgenes. As with gene targeting, a full understanding of the mechanisms of random transgene integration awaits further study at the molecular/biochemical level. Such understanding is important from the perspective of gene targeting, because it may be that targeting frequencies could be enhanced by somehow blocking "background" (i.e. random) integration events.

3.2. Targeted integration

The utility of gene targeting as a means of gene therapy requires systematic study into the mechanisms of the process. Given the many variables involved in any gene targeting experiment (i.e. cell type, transfection method, transgene design, target site, etc.), it is unsurprising that progress towards a detailed understanding has been relatively slow.

Nevertheless, various studies have provided important insights into several aspects of mammalian gene targeting. Initial studies involved the use of artificially introduced selectable target sites in mammalian cell lines, such that rare targeting events could be recovered. Later studies have used targets of natural loci in mammalian cell lines, ES cells and mammalian zygotes. Various observations and inferences from such studies are considered in the following sections (3.2.1–3.2.8).

3.2.1. Transfection method

In methods such as co-precipitation, the exogenous DNA molecules must somehow migrate through the cytoplasm of the host cell in order to reach the nucleus. Of the DNA that survives this journey, a substantial proportion sustains some degree of endonucleolytic or exonucleolytic damage. In contrast, virtually no damage occurs to DNA delivered directly into the nucleus by microinjection [22, 47].

Gene targeting using a damaged transgene is unlikely to be desirable in gene therapy. Beyond this concern, it may be the case that damage sustained by incoming transgene molecules renders them less able to undergo HR. The reasons for this are not known, but the effect seems to exist. For example, separate studies were conducted by Lin et al. [25] and Thomas et al. [42] both involving gene targeting of artificially introduced defective genes in mouse fibroblasts. Lin et al. reported a ratio of random integration to gene targeting of 100 000:1 whereas Thomas et al. reported a ratio of 100:1. The major difference between the two sets of studies lies in the method of transfection, with Lin et al. using $CaPO_4$ co-precipitation and Thomas et al. using microinjection. Similarly, in a systematic study using the adenine phosphoribosyltransferase (APRT) locus in Chinese hamster ovary cells, Vasquez et al. [46] compared the targeting efficiencies associated with various transfection methods. In these experiments, mass-delivery methods (electroporation, co-precipitation, liposomes) yielded an average ratio of random integration to gene targeting of 200 000:1 (range 2 400:1 to 350 000:1) compared with a ratio of 1:15 for microinjection. In addition to the possibility that HR frequency is reduced due to DNA damage associated with the mass-transfection methods, it has been suggested that the larger numbers of transgene molecules delivered by the mass–transfection methods may overwhelm the HR machinery [46].

3.2.2. Transgene sequences

Gene targeting is dependent on HR, which is in turn dependent on shared homology between recombining DNA sequences. The question is, how much homology is required for optimal efficiency of gene targeting? There is at present no complete answer to this question, because systematic studies are lacking, and comparison between separate studies is very problematic due to the existence of several variables other than the extent of homology. Such variables include other transgene sequences, the physical state of the transgene, the cell type used, the target gene and the actual nature (rather than simply the extent) of homology. Nevertheless, several studies have provided a partial answer. Thomas and Capecchi [41], targeting the HPRT gene in ES cells, found that targeting efficiency appeared to be strongly dependent upon the degree of homology possessed by the transgene (and shared with the target locus). Specifically, an increase in homology from 4 kb to 9.1 kb correlated with 40-fold increase in the rate of targeting, as measured by the ratio of targeted:random integration. Shulman et al. [36], targeting an immunoglobulin gene in hybridoma cell lines, varied the extent of homology from 1.2 to 9.5 kb. Again the degree of homology correlated with targeting efficiency, with a 25-fold increase seen over the range from 2.5 to 9.5 kb. No increase in targeting efficiency was observed between 1.2 and 2.5 kb. From these studies it can be concluded that the frequency of gene targeting is roughly proportional to the extent of homology shared by the transgene and its target locus. However, it is notable that the effects of very large (> 9.5 kb) homologies are not known.

Besides homology length, base pair variation may affect the rate of targeting. This is evident from experiments comparing isogenic and nonisogenic transgenes. The homology region(s) in an isogenic transgene is derived from the same (syngenic) laboratory animal strain as the target

animal. Therefore, isogenic trangenes contain homology blocks that are genetically identical (or virtually identical) to the target homology regions. By contrast, a homology stretch in a nonisogenic transgene will typically be interrupted by a number of slight sequence divergences, such as base-pair mismatches and small deletions/insertions. In a series of experiments designed to compare isogenic and nonisogenic transgenes, Riele et al. [31] reported a 20-fold improvement in targeting efficiency when an isogenic transgene was used, yielding a remarkably favourable ratio of random to targeted integration (approximately 1:4). The target site was the retinoblastoma susceptibly gene (Rb) in an ES cell line derived from mouse strain 129. The isogenic and nonisogenic transgene constructs contained 17 kilobases of homology, derived respectively from (a) mouse strain 129 and (b) mouse strain BALB/c. Similar results were obtained from a systematic study by Van Deursen and Wieringa [44], in which the creatine kinase M gene (CKM) in ES cells was targeted with transgenes sharing 9 kb of homology with the target site. In these experiments, an increase in targeting efficiency of approximately 25-fold was observed when isogenic transgenes were used. Thus, the use of isogenic DNA in transgenes appears to hold promise for improving gene–targeting efficiencies. However, it has not been established whether the outcomes described above are applicable to other target genes in other cell types. Moreover, there exists a dearth of systematic knowledge concerning the nature, frequency and extent of heterologies that may affect targeting efficiencies. Nevertheless, it is reasonable to conclude that, all other factors being equal, transgenes employing perfect homology are likely to yield better targeting efficiencies in comparison with those using interrupted homology.

3.2.3. Physical state of the transgene

In keeping with yeast data (see Sect. 2.4), all studies agree that linearization of the transgene (in the region of homology) greatly enhances targeting efficiency (see [21] for example). This finding is supportive of the DSBR model (Sect. 2.4) as an explanation of the mechanism of gene targeting (but see also Sect. 2.6). Beyond linearization, stripping the transgene ends to expose around 200 nucleotides of single-stranded DNA appears to further enhance targeting [46]. Although the underlying mechanism is not understood, the finding that single-stranded transgene tails enhance targeting fits well with the notion that HR involves single-stranded DNA ends invading target duplex DNA [38].

3.2.4. Transgene copy number

A targeting transgene molecule presumably has to "search" through the host genome until it "finds" its target sequence. Therefore, it might be expected that targeting efficiency would be enhanced by increasing the number of transgene molecules introduced to the host cell. An increasing cytotoxic effect is observed where increasingly large quantities of DNA are microinjected into the nucleus. However, the targeting efficiency can still be obtained, by calculating the proportion of surviving cells that have been successfully targeted. However, no study has demonstrated a correlation between transgene copy number and targeting frequency. This area has not been extensively researched, but at least two studies have positively determined that there appears to be no relationship whatsoever between the number of copies introduced and the efficiency of [32, 41]. The inference must be that the initial search for homology does not seem to be the rate-limiting step for targeting. This conclusion is also supported by experiments involving amplification of the target site (see Sect. 3.2.8): an increased target copy number does not appear to enhance the frequency of targeting. Indeed, if Vasquez et al. [46] are correct in postulating that too many transgene molecules may overwhelm the

HR machinery (see Sect. 3.2.1 above), it may turn out to be the case that an inverse relationship exists between transgene copy number and targeting efficiency.

3.2.5. Position of target site

The position of the target site within the genome does not strongly influence the frequency of HR. For example, twelve independent recipient cell lines were produced by Thomas et al. [41], each line containing a defective neomycin-resistance gene integrated at a different chromosomal position. Introduction of targeting DNA constructs gave similar gene targeting frequencies in all twelve lines.

3.2.6. Recombination hotspots

Targeting the endogenous β_2-microglobulin gene in ES cells, Zijlstra et al. [54] achieved a very high frequency of gene targeting (a ratio of about 1: 25 targeting to random integration). Other investigators for the β_2-microglobulin target gene [20] and for the Hox 3.1 target gene [23] have reported similarly high targeting frequencies. Such studies support the existence of recombination "hotspots". However, the sequences involved in such hotspots remain to be elucidated.

3.2.7. Target gene activity

There is no evidence that the level of expression of the target gene correlates with the frequency of gene targeting (see [20] for example).

3.2.8. Target copy number

As noted in Sect. 3.2.4, experimental amplification of the target site does not appear to enhance the frequency of targeting. For example, Zheng and Wilson [53] used two mammalian cell lines, one of which contained 2 target gene copies. The second line contained around 800 target gene copies, located in three clusters on different chromosomes. Gene targeting rates were the same in both cell lines. Similarly, Thomas et al. [41] used three cell lines containing integrated target plasmid sequences present as one copy, four dispersed copies or five tandem copies. Again, the rates of gene targeting were similar in all three lines. Such results infer, as suggested previously (Sect. 3.2.4), that the initial search for homology does not appear to be the rate–limiting step for targeting. Interestingly, recent (target amplification) experiments in yeast have suggested that the frequency of gene targeting does depend on the number of target copies. Indeed, Wilson et al. [49] report a linear relationship between target site copy number and the rate of targeting in yeast. The reason for this difference between yeast cells and mammalian cells remains to be established.

4. CONCLUDING REMARKS

Although major progress in model building has been made in recent decades, extensive biochemical analysis of the molecular mechanism(s) of HR will be required if gene targeting is to be better understood. On a different level, the phenomenology of gene targeting also requires extensive systematic analysis.

Absolute frequencies of gene targeting in mammalian cells remain low, and the ratio of targeted to random integration is still heavily weighted in favour of the latter. Until the frequencies are improved, the potential use of gene targeting in non-selective systems will be limited. Such improvement is likely to depend upon a more detailed understanding of gene targeting, which is in turn dependent upon systematic analysis of the sort described above.

ACKNOWLEDGEMENTS

Artwork by 1st Class Media: http://www.1stclass.uk.com.

REFERENCES

[1] Berinstein N., Pennell N., Ottaway C.A., Shulman M.J., Gene Replacement with One-Sided Homologous Recombination, Mol. Cell. Biol. 12 (1992) 360–367.

[2] Bishop J.O., Chromosomal insertion of foreign DNA, Reprod. Nutr. Dev. 36 (1996) 607–618.

[3] Bishop J.O., Smith P., Mechanism of Chromosomal Integration of Microinjected DNA, Mol. Biol. Med. 6 (1989) 283–298.

[4] Bollag R.J., Waldman A.S., Liskay R.M., Homologous Recombination in Mammalian-Cells, Annu. Rev. Genet. 23 (1989) 199–225.

[5] Brinster R.L., Chen H.Y., Trumbauer M.E., Yagle M.K., Palmiter R.D., Factors Affecting the Efficiency of Introducing Foreign DNA into Mice by Microinjecting Eggs, Proc. Natl. Acad. Sci. USA 82 (1985) 4438–4442.

[6] Brinster R.L., Braun R.E., Lo D., Avarbock M.R., Oram F., Palmiter R.D., Targeted Correction of a Major Histocompatibility Class–Ii E-Alpha-Gene by DNA Microinjected into Mouse Eggs, Proc. Natl. Acad. Sci. USA 86 (1989) 7087–7091.

[7] Campbell K.H.S., McWhir J., Ritchie W.A., Wilmut I., Sheep cloned by nuclear transfer from a cultured cell line, Nature (1996) 64–65.

[8] Capecchi M.R., Altering the Genome by Homologous Recombination, Science 244 (1989) 1288–1292.

[9] Clark A.J., Burl S., Denning C., Dickinson P., Gene targeting in livestock: a preview, Transgenic Res. 9 (2000) 263–275.

[10] Cole-Strauss A., Yoon K., Xiang Y., Byrne B.C., Rice M.C., Gryn J., Holloman W.K., Kmiec E.B., Correction of the Mutation Responsible for Sickle Cell Anemia by an RNA-DNA Oligonucleotide, Science 273 (1996) 1386–1388.

[11] Cole-Strauss A., Gamper H., Holloman W.K., Munoz M., Cheng N., Kmiec E.B., Targeted gene repair directed by the chimeric RNA/DNA oligonucleotide in a mammalian cell-free extract, Nucleic Acids Res. 27 (1999) 1323–1330.

[12] Folger K.R., Wong E.A., Wahl G., Capecchi M.R., Patterns of Integration of DNA Micro-Injected into Cultured Mammalian Cells – Evidence for Homologous Recombination between Injected Plasmid DNA-Molecules, Mol. Cell. Biol. 2 (1982) 1372–1387.

[13] Frohman M.A., Martin G.R., Cut, Paste, and Save – New Approaches to Altering Specific Genes in Mice, Cell 56 (1989) 145–147.

[14] Gamper H.B., Cole-Strauss A., Metz R., Parekh H., Kumar R., Kmiec E.B., A plausible Mechanism for Gene Correction by Chimeric Oligonucleotides, Biochemistry-US 39 (2000) 5808–5816.

[15] Gamper H.B., Parekh H., Rice M.C., Bruner M., Youkey H., Kmiec E.B., The DNA strand of chimeric RNA/DNA oligonucleotides can direct gene repair/conversion activity in mammalian and plant cell-free extracts, Nucleic Acids Res. 28 (2000) 4332–4339.

[16] Gordon J.W., Ruddle F.H., DNA-Mediated Genetic Transformation of Mouse Embryos and Bone Marrow – a Review, Gene 33 (1985) 121–136.

[17] Hatada S., Nikkuni K., Bentley S.A., Kirby S., Smithies O., Gene correction in hematopoietic progenitor cells by homologous recombination, Proc. Natl. Acad. Sci. USA 97 (2000) 13807–13811.

[18] Holliday R., A mechanism for gene conversion in fungi, Genet. Res. (1964) 282–304.

[19] Holliday R., Genetic recombination in fungi, in: Peacock J.W., Brook R.D. (Eds.), Replication and recombination of genetic material, Canberra, Australian Academy of Science, 1968, pp. 157–174.

[20] Koller B.H., Smithies O., Inactivating the Beta-2-Microglobulin Locus in Mouse Embryonic Stem-Cells by Homologous Recombination, Proc. Natl. Acad. Sci. USA 86 (1989) 8932–8935.

[21] Kucherlapati R.S., Eves E.M., Song K.Y., Morse B.S., Smithies O., Homologous Recombination between Plasmids in Mammalian Cells can be Enhanced by Treatment of Input DNA, Proc. Natl. Acad. Sci. USA 81 (1984) 3153–3157.

[22] Lebkowski J.S., Dubridge R.B., Antell E.A., Greisen K.S., Calos M.P., Transfected DNA is Mutated in Monkey, Mouse, and Human Cells, Mol. Cell. Biol. 4 (1984) 1951–1960.

[23] Lemouellic H., Lallemand Y., Brulet P., Targeted Replacement of the Homeobox Gene Hox-3.1 by the Escherichia-Coli LacZ in Mouse Chimeric Embryos, Proc. Natl. Acad. Sci. USA 87 (1990) 4712–4716.

[24] Li J., Read L.R., Baker M.D., The Mechanism of Mammalian Gene Replacement Is Consistent with the Formation of Long Regions of Heteroduplex DNA Associated with Two Crossing-Over Events, Mol. Cell. Biol. 21 (2001) 501–510.

[25] Lin F.L., Sperle K., Sternberg N., Recombination in Mouse L-Cells between DNA Introduced into Cells and Homologous Chromosomal Sequences, Proc. Natl. Acad. Sci. USA 82 (1985) 1391–1395.

[26] McCreath K.J., Howcroft J., Campbel K.H.S., Colman A., Schnieke A.E., Kind A.J., Production of gene-targeted sheep by nuclear transfer from cultured somatic cells, Nature 405 (2000) 1066–1069.

[27] Meselson M., Radding C.M., A general model for genetic recombination, Proc. Natl. Acad. Sci. USA. 73 (1975) 358–361.

[28] Ogawa T., Yu X., Shinohara A., Egelman E.H., Similarity of the Yeast Rad51 Filament to the Bacterial Reca Filament, Science 259 (1993) 1896–1899.

[29] Orrweaver T.L., Szostak J.W., Fungal Recombination, Microbiol. Rev. 49 (1985) 33–58.

[30] Perez C.F., Botchan M.R., Tobias C.A., DNA-Mediated Gene Transfer Efficiency Is Enhanced by Ionizing and Ultraviolet-Irradiation of Rodent Cells In vitro, Radiat. Res. 104 (1985) 200–213.

[31] Riele H.T., Maandag E.R., Berns A., Highly Efficient Gene Targeting in Embryonic Stem-Cells through Homologous Recombination with Isogenic DNA Constructs, Proc. Natl. Acad. Sci. USA 89 (1992) 5128–5132.

[32] Rommerskirch W., Graeber I., Grassmann M., Grassmann A., Homologous Recombination of SV40 DNA in Cos7 Cells Occurs with High–Frequency in a Gene Dose Independent Fashion, Nucleic Acids Res. 16 (1988) 941–952.

[33] Schnieke A.E., Kind A.J., Ritchie W.A., Mycock K., Scott A.R., Ritchie M., Wilmut I., Colman A., Campbell K.H.S., Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts, Science 278 (1997) 2130–2133.

[34] Sedivy J.M., Joyner A.L., Gene Targeting, Oxford University Press USA, Publ., New York, 1993.

[35] Sedivy J.M., Dutriaux A., Gene targeting and somatic cell genetics – a rebirth or a coming of age? Trends Genet. 15 (1999) 88–90.

[36] Shulman M.J., Nissen L., Collins C., Homologous Recombination in Hybridoma Cells – Dependence on Time and Fragment Length, Mol. Cell. Biol. 10 (1990) 4466–4472.

[37] Smih F., Rouet P., Romanienko P.J., Jasin M., Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells, Nucleic Acids Res. 23 (1995) 5012–5019.

[38] Sun H., Treco D., Szostak J.W., Extensive 3'-Overhanging, Single-Stranded-DNA Associated with the Meiosis-Specific Double-Strand Breaks at the Arg4 Recombination Initiation Site, Cell 64 (1991) 1155–1161.

[39] Sung P., Catalysis of ATP-Dependent Homologous DNA Pairing and Strand Exchange by Yeast Rad51 Protein, Science 265 (1994) 1241–1243.

[40] Szostak J.W., Orrweaver T.L., Rothstein R.J., The Double-Strand-Break Repair Model for Recombination, Cell 33 (1983) 25–35.

[41] Thomas K.R., Capecchi M.R., Site-Directed Mutagenesis by Gene Targeting in Mouse Embryo-Derived Stem-Cells, Cell 51 (1987) 503–512.

[42] Thomas K.R., Folger K.R., Capecchi M.R., High-Frequency Targeting of Genes to Specific Sites in the Mammalian Genome, Cell 44 (1986) 419–428.

[43] Thomas K.R., Deng C.X., Capecchi M.R., High-Fidelity Gene Targeting in Embryonic Stem-Cells by Using Sequence Replacement Vectors, Mol. Cell. Biol. 12 (1992) 2919–2923.

[44] Van Deursen J., Wieringa B., Targeting of the Creatine Kinase-M Gene in Embryonic Stem-Cells Using Isogenic and Nonisogenic Vectors, Nucleic Acids Res. 20 (1992) 3815–3820.

[45] Vasquez K.M.,Wilson J.H., Triplex-directed modification of genes and gene, activity, Trends Biochem. Sci. 23 (1998) 4–9.

[46] Vasquez K.M., Marburger K., Intody Z., Wilson J.H., Manipulating the mammalian genome by homologous recombination, Proc. Natl. Acad. Sci. USA 98 (2001) 8403–8410.

[47] Wake C.T., Gudewicz T., Porter T., White A., Wilson J.H., How Damaged Is the Biologically-Active Subpopulation of Transfected DNA, Mol. Cell. Biol. 4 (1984) 387–398.

[48] Wilkie T.M., Palmiter R.D., Analysis of the Integrant in Myk-103 Transgenic Mice in Which Males Fail to Transmit the Integrant, Mol. Cell. Biol. 7 (1987) 1646–1655.

[49] Wilson J.H., Leung W.Y., Bosco G., De D., The frequency of gene targeting in yeast depends on the number of target copies, Proc. Natl Acad. Sci. USA 91 (1994) 177.

[50] Xiang Y., Cole-Strauss A., Yoon K., Gryn J., Kmiec E.B., Targeted gene conversion in a mammalian CD34(+)-enriched cell population using a chimeric RNA/DNA oligonucleotide, J. Molec. Med. 75 (1997) 829–835.

[51] Ye S., Cole-Strauss A., Frank B., Kmiec E.B., Targeted gene correction: a new strategy for molecular medicine, Mol. Med. Today 4 (1998) 431–437.

[52] Yoon K., Cole-Strauss A., Kmiec E.B., Targeted gene correction of episomal DNA in mammalian cells mediated by a chimeric RNA/DNA oligonucleotide, Proc. Natl. Acad. Sci. USA 93 (1996) 2071–2076.

[53] Zheng H., Wilson J.H., Gene Targeting in Normal and Amplified Cell-Lines, Nature 344 (1990) 170–173.

[54] Zijlstra M., Li E., Sajjadi F., Subramani S., Jaenisch R., Germ-Line Transmission of a Disrupted Beta-2-Microglobulin Gene Produced by Homologous Recombination in Embryonic Stem-Cells, Nature 342 (1989) 435–438.

[55] Zimmer A., Manipulating the Genome by Homologous Recombination in Embryonic Stem-Cells, Annu. Rev. Neurosci. 15 (1992) 115–137.