ABSTRACT

Illegitimate recombination (IR) is the process by which two DNA molecules not sharing homology to each other are joined. In *Kluyveromyces lactis*, integration of heterologous DNA occurred very frequently therefore constituting an excellent model organism to study IR. IR was completely dependent on the nonhomologous end-joining (NHEJ) pathway for DNA double strand break (DSB) repair and we detected no other pathways capable of mediating IR. NHEJ was very versatile, capable of repairing both blunt and non-complementary ends efficiently. Mapping the locations of genomic IR-events revealed target site preferences, in which intergenic regions (IGRs) and ribosomal DNA were overrepresented six-fold compared to open reading frames (ORFs). The IGR-events occurred predominantly within transcriptional regulatory regions. In a rad52 mutant strain IR still preferentially occurred at IGRs, indicating that DSBs in ORFs were not primarily repaired by homologous recombination (HR). Introduction of ectopic DSBs resulted in the efficient targeting of IR to these sites, strongly suggesting that IR occurred at spontaneous mitotic DSBs. The targeting efficiency was equal when ectopic breaks were introduced in an ORF or an IGR. We propose that spontaneous DSBs arise more frequently in transcriptional regulatory regions and in rDNA and such DSBs can be mapped by analyzing IR target sites.

INTRODUCTION

Living cells are challenged with a large number of DNA lesions in every cell cycle. Among these lesions, the DNA double strand break (DSB) is the most severe, leading to cell death if not repaired. Moreover, DSBs repaired in an incorrect manner can also be detrimental to the organism. Examples are the formation of oncogenic fusion genes or the juxtaposition of oncogenes to strong enhancers, both of which are a primary cause of malignancy. Studying how DSBs arise and how they are repaired is thus of vital interest. Cells have evolved two major pathways for DSB repair, the homologous recombination (HR) and the nonhomologous end-joining (NHEJ) pathways. The HR pathway (1,2) depends on the presence of a homologous chromosome or sister chromatid that is used as a template for homology-driven and mostly error-free repair. HR in fact refers to several different processes, including gene conversion and alternative pathways called break-induced replication (BIR) and single strand annealing (SSA). A key-protein in HR is Rad52, which forms a heptameric ring in *in vitro* (3), that interacts with single stranded DNA and facilitates the strand invasion reaction (4,5). Rad52 is thought to act early during HR and strains lacking Rad52 are completely deficient for all types of HR in *Saccharomyces cerevisiae*. Another important protein is Rad51 (similar to bacterial RecA), which facilitates the strand exchange reaction between the damaged substrate and the undamaged template molecule (6).

In contrast, the NHEJ pathway requires little or no homology and simply fuses two free DNA-ends, often generating small deletions and insertions. Among the proteins required for NHEJ are the Ku70/Ku80 heterodimer (7), which has high affinity for DNA ends. The heterodimer binds to the DSB protecting the free ends from extensive degradation and holds the DNA ends in an appropriate configuration for subsequent end-processing and ligation (8). Binding of Ku70/Ku80 is required for the recruitment of a second complex to the DSB (9). This complex consists of a DNA ligase, Dnl4 (or Lig4) (10–12) and Lif1 (Ligase four interacting factor 1) (13), the latter being the yeast functional homologue of mammalian Xrc4. Lif1 is required for the stability and full activity of Dnl4 (14). A third complex consisting of Mre11, Rad50 and Xrs2 (MRX), is also required for efficient NHEJ in *S.cerevisiae* (15,16), but not in mammals or fission yeast (17–20). Mre11 has an endo/exo nuclease activity, but this
activity appears to be dispensable for NHEJ in vivo (21). The MRX-complex promotes the end-joining reaction in vitro (22), but the exact role of this complex in NHEJ remains unclear. In S. cerevisiae, cell-type regulates NHEJ, haploid MATα or MATa strains perform NHEJ efficiently, but diploid MATα/MATa strains perform NHEJ inefficiently (23,24). In diploids, the a1/a2 repressor inhibits the transcription of the NEJ1 gene. Nej1 interacts with Lif1 and constitutes an essential component of the NHEJ pathway (25–28). Diploid cells can rely on the HR pathway to repair DSBs even in the G1 phase of the cell cycle, since these cells contain a homologous chromosome. Because HR is more accurate than NHEJ this cell-type regulation may have evolved to promote genome stability in S. cerevisiae.

Despite the high degree of molecular conservation in NHEJ mechanisms between yeast and mammalian cells, there are some striking differences regarding the in vivo function. In mammals, NHEJ is the major pathway for DSB repair. In S. cerevisiae, the HR pathway for DSB repair is very active and the NHEJ pathway has largely a back-up function at least under laboratory conditions. Consequently, S. cerevisiae mutant strains deficient for NHEJ do not have any obvious growth disadvantages. In contrast, in mice some NHEJ proteins (Ligase IV and Xrcc4) are required for embryonic development (29,30). In addition, NHEJ proteins (Ligase IV and Xrcc4) are required for obvious growth disadvantages. In contrast, in mice some S. cerevisiae genes are not more sensitive to DNA damaging agents than a wild-type strain, whereas mammalian cells lacking NHEJ components are highly sensitive to DNA damage (31).

Since both the HR and NHEJ pathways use the same substrate, a DSB, a competition between them can be envisaged. Thus, inactivation of the NHEJ pathway would lead to an increase of homologous integration events upon introduction of a gene targeting cassette. This hypothesis was tested in mouse ES-cells, but NHEJ mutant cells did not have a more favorable gene targeting frequency (32). In S. cerevisiae, the reciprocal experiment showed that the NHEJ efficiency did not increase in the absence of HR-components (23,33). A moderate competition between NHEJ and HR has been observed, however, when genomic DSBs were introduced in vivo using inducible endonucleases (20,32,34,35).

NHEJ, the direct rejoining of DNA DSBs, is closely associated with illegitimate recombination (IR) and chromosomal rearrangements. Hence, IR between two genomic loci results in deletions, duplications, insertions or translocations. In S. cerevisiae, the frequency of IR is reduced in rad50, mre11, xrs2 and hfd1 (yku70) mutant strains (36,37), indicating that IR takes place through the NHEJ pathway. Target sites of IR correlate with consensus sites for Topoisomerase I (38) and are dependent on the TOP1 gene. In addition, overexpression of Top1 leads to increased IR, demonstrating that the nicking activity of Top1 promotes IR (38).

In this study, we investigated IR and NHEJ in Kluyveromyces lactis (milk yeast). Similarly to mammalian cells, milk yeast NHEJ was capable of repairing blunt and noncohesive ends efficiently. Integration of a nonhomologous DNA molecule into the genome was 1000-fold more frequent than in S. cerevisiae and ectopically introduced DSBs were hotspots for integration. We present evidence suggesting a model in which IR takes place at spontaneous mitotic DSB and that these breaks occurred more frequently within promoter regions and rDNA.

**MATERIALS AND METHODS**

**Plasmids**

For the gene disruption of K. lactis NEJ1, two PCR-fragments corresponding to the nucleotides −166 to +250, (the ATG start codon corresponding to nucleotides +1 to +3) and +474 to +895 were digested with SacI–BamHI and BamHI–XhoI, respectively. Both PCR-fragments were combined with a SacI–XhoI digested plBlueScript SK(+) in a three-factor cloning, generating plasmid pAKP68. Targeting constructs for two-step gene disruption of the RAD52, RAD51 and YKU80 genes were all generated using a similar approach. For RAD52, two PCR fragments corresponding to nucleotides −329 to +112 (SacI–BamHI) and +835 to +1353 (BamHI–XhoI) were combined with SacI–Sall digested pRS406 (39) generating vector p580. For RAD51, two PCR fragments corresponding to nucleotides −536 to +65 (XhoI–BamHI) and +435 to +1085 (BamHI–XbaI) were combined with XhoI–XbaI digested pRS406 generating vector p182. For YKU80, two PCR fragments corresponding to nucleotides +5 to +177 (XhoI–BamHI) and +1267 to +1866 (BamHI–XbaI) were ligated with XhoI–XbaI digested pRS406 generating plasmid pAKP120. Into the resulting BamHI site of pAKP68, p580, p182 and pAKP120, a 2.9 kb BglII-fragment containing the LEU2 gene from vector pCXJ20 (40) was cloned, generating plasmid pAKP69 (pBlueScript SK(+)-nej1::LEU2), p582 (pRS406-rad52::LEU2), p183 (pRS406-rad51::LEU2) and pAKP122 (pRS406-yku80::LEU2), respectively.

Introduction of the I-SceI-endonuclease recognition site within the VMRI promoter region and within the VMRI coding sequence was accomplished as follows. Two PCR fragments corresponding to nucleotides +1455 to −189, and to nucleotides −187 to −974 with respect to VMRI start codon were digested with XhoI–HindIII and HindIII–NotI, respectively and ligated into XhoI–NotI digested pRS406, generating plasmid pAKP175. Two oligonucleotides (5′ pmol/μl) (5′-agctagctaggtgatacagaagggtaatac-3′) and (5′-agctagctaggtgatacagggtagctgct-3′) were annealed in 2 mM Tris–HCl (pH 7.4), 0.2 mM MgCl2 and 5 mM NaCl and cloned into the HindIII site of pAKP175, generating plasmid pAKP179. A PCR fragment corresponding to nucleotide +2190 to +5097 with respect to the VMRI start codon was digested with SacI–XhoI and ligated into SacI–XhoI digested pRS406, creating plasmid pPMB19. Two oligonucleotides, (5′-ggccggctagctgatacagaagggtaatac-3′) and (5′-ggccggctagctgatacagggtagctgct-3′), were annealed and cloned into the VMRI endogenous KasI site at position +3871, creating plasmid pPMB20. To express the I-SceI endonuclease in K. lactis, an XhoI–SacI fragment containing pGAL1-I-SCEI from plasmid pTW406 (kindly provided by T. E. Wilson) was ligated into Sall digested pCJX18 (40), generating plasmid pAKP164. To measure IR at a specific DSB the promoter regions of two S. cerevisiae genes ADH1 (−1 to −479) and TEF2 (−1 to −1000) were PCR-amplified and digested with XbaI–BamHI and XhoI–BamHI, respectively. Both PCR-fragments were combined with Spel–XhoI digested pRS405 in a three-factor cloning,
generating plasmid pPMB18. The oligonucleotide sequences used in this study are available upon request.

### Yeast strains

The strains used in this study are listed in Table 1. All null alleles generated were confirmed by genomic DNA-blot analysis or locus specific PCR. Crossing the strain SAY119 with CK213-4c generated a diploid MATα/MATα strain (SAY681). Strain SAY572 (nej1::LEU2) was generated by a one-step gene disruption procedure (41) by transforming CK213-4c with plasmid pPMB18. The oligonucleotide sequences used in this study are available upon request.

Strain SAY684 and PMY34 were generated in a two-step gene replacement procedure transforming CK213-4c with either Clal-linearized pAKP179 or AgeI-linearized pPMB20, respectively. Crossing strain SAY684 either with SAY509 or SAY687 generated the mutant segregants SAY685 (nej1::LEU2) and SAY686 (rad52::LEU2) both containing the I-SceI site in the VMRI promoter. Strain SAY687 (MATα rad52::LEU2) is a segregant from the cross of strains SAY119 and SAY507. Strain PMY2 (MATα/MATα rad52::LEU2/rad52::LEU2 mutant strain was generated by crossing SAY684 and PMY34.

### Growth media and standard methods

Media for growth of yeast and bacteria, protocols for RNA and DNA-blot analysis, and RNA preparation, yeast and bacterial transformations were carried out as described elsewhere (43–45). For detecting the NEJ1 transcript a [α-32P]dCTP-labeled PCR-fragment corresponding to the NEJ1 open reading frame (ORF) was used. The RNA-blot was analyzed using a phosphorimager. NHEJ efficiency was tested in a plasmid rescue assay (16). The plasmid pCXJ18 was cleaved either with HindIII, SmaI, PstI or HindIII and PstI, generating a linearized vector. Crossing strain SAY684 and PMY34 was performed with the same amount of super-coiled pPMB18 as template. Strain SAY573 (ka80::LEU2) was generated by a two-step gene disruption procedure (42) transforming CK213-4c with NheI-linearized pAKP172. A segregant from the cross of strain SAY119 with SAY687 generated a MATα/MATα strain (SAY509). Strain SAY509 was crossed to SAY573 and from this cross the double mutant nej1::LEU2 ka80::LEU2 segregant (SAY574) was recovered by screening for a non-parental dittype (NPD) for Leu". Strain SAY516 (rad51::LEU2) and strain SAY507 (rad52::LEU2) were generated in a two-step gene disruption procedure transforming CK213-4c with PmlI-linearized p183, and BglII-linearized p582, respectively. Crossing strain SAY516 with PMY34 and strain SAY507 with PMY34 were generated in a two-step gene replacement procedure transforming CK213-4c with PmlI-linearized p183, and BglII-linearized p582, respectively. Crossing strain SAY516 with PMY34 and strain SAY507 with PMY34 were generated in a two-step gene replacement procedure transforming CK213-4c with PmlI-linearized p183, and BglII-linearized p582, respectively. Crossing strain SAY516 with PMY34 and strain SAY507 with PMY34 were generated in a two-step gene replacement procedure transforming CK213-4c with PmlI-linearized p183, and BglII-linearized p582, respectively. Crossing strain SAY516 with PMY34 and strain SAY507 with PMY34 were generated in a two-step gene replacement procedure transforming CK213-4c with PmlI-linearized p183, and BglII-linearized p582, respectively.

### Gene targeting in K. lactis

A general method for performing accurate gene targeting was developed in K. lactis, similar to a previously described method for S. cerevisiae (46), but using the modifications.
described below. For each side of the target gene two 70mer oligonucleotides were designed sharing ~50 bp of homology to the target locus (LIG4, MRE11 and RAD50). The stretches of homology were chosen so that the resulting deletions were from the start codon to the stop codon. On the 3’ end of the oligonucleotides 20 bp of homology to the KANMX gene was included (5’ n50-ccagcgcgtggagccgccag-3’ and 5’ n50-ggtaggccgttgtatcg-3’). PCR amplifications using these oligonucleotides and pFA6a-KANMX (47) as template were performed. The resulting fragments were cloned into the pGEM-T Easy vector. An aliquot of 4–10 µg of the resulting plasmid was digested with EcoRI and Spel and the linear DNA fragments used to transform strain SAY572 (nei1::LEU2). Large colonies, indicating the integration of the KANMX gene into the genome, were picked after 48 to 72 h of incubation at 30°C. For confirmation of the correct genotypes, a genomic-specific oligonucleotide homologous to a region 300–500 bp upstream of the target gene and an oligonucleotide (5’-cggccccttagaatgccggag-3’) complementary to the KANMX gene were used in PCR amplifications, using genomic DNA as template.

**Plasmid rescue and inverse PCR**

To analyze the genomic target sites of IR of pFA6a-KANMX, the vector was digested with SacII and the enzyme was heat inactivated at 65°C for 20 min. The linearized pFA6a-KANMX was transformed either into wild-type cells CK213-4c or PMY2. Cells were plated on rich medium and incubated at 30°C for 24 h, replica-printed on rich medium containing 200 µg/ml G418 and incubated for 72 h. Approximately 250 colonies were re-streaked on fresh 200 µg/ml G418 plates. Chromosomal DNA was prepared and 1–5 µg of genomic DNA was digested with EcoRI for 3 h at 37°C in a total volume of 20 µl in plasmid rescue experiments. The samples were heated at 65°C for 20 min and then diluted to 100 µl. Following addition of T4 DNA ligase the samples were incubated at 16°C for 12 h. After adding 0.5 vol of 7.5 M NH4Ac and 3 vol of 2-propanol, the DNA was recovered by centrifugation. The DNA pellet was washed once in 70% ethanol, air-dried and resuspended in 10 µl of sterile water. For plasmid rescue, 1–5 µl of DNA was transformed into Escherichia coli DH5α by electroporation. Plasmid DNA was prepared from a single bacterial colony and the SacII junction was DNA sequenced (Macrogen-South Korea) using a custom primer (5’-ctaaagccgacatggagcc-3’). For sequencing of both ends of the target sites an identical procedure was used, but the restriction enzymes used for plasmid rescue were HpaI, XbaI or XhoI and an additional sequencing primer (5’-cggccgttagctgctggtc-3’) was included. The data obtained were compared with the Genolevures database (48) to assign a genomic coordinate to the insertions.

The PMY2 transformants were analyzed by inverse PCR. Genomic DNA was prepared and digested with either HpaII or MspI and following intramolecular ligation, inverse PCR was performed with primer PMO5-F (5’-ctatgctagctgctgctgctc-3’) and PMO6-R (5’-tgacgcgctgctgctgctgctc-3’) using standard conditions. The SacII junction was sequenced and analyzed as described above.

**Statistical methods**

To identify any unusual clusters of insertions (or regions with higher insertion intensity), χ2-tests and scan statistics were used. Under the null hypothesis, insertions occurred independently, completely at random and were equally likely to occur anywhere on the chromosomes. χ2-tests were used to investigate if there were differences in insertion intensity between the six chromosomes and between other divisions of the genome (rDNA/not rDNA, ORF/ not ORF). To discover any unexpected clusters within a chromosome we looked at the maximum number of insertions on a fixed proportion of the chromosome length. This number is a scan statistic, and the probability to find k or more insertions in any such subinterval of the chromosome was calculated with an approximate formula (49). To calculate the probability for the observed smallest distance between any of the insertions on a chromosome an exact formula (50) was used. All protein coding gene sequences from K. lactis were downloaded from http://cbl.ucsf.edu/Genolevures/raw/seq/K_lactis.rc2.nt. The codon adaptation index (CAI) values were calculated with the web-based software ‘The CAI calculator version 2’ (http://www.evolvecode.net/codon/CalculateCAIs.php) using the ribosomal protein genes as the reference set. The ribosomal protein coding genes were determined based on the genome annotation.

**RESULTS**

**Identification of K. lactis NHEJ genes**

Based on sequence conservation, we identified the potential MRE11 (KLLA0C06930g), RAD50 (KLLA0C02915g), YUK80 (KLLA0B12672g), NEJ1 (KLLA0F20339g) and LIG4 (KLLA0D01089g) genes from K. lactis using the Genolevures database (51). The predicted proteins encoded by these genes shared homology with their S. cerevisiae orthologs; Yku80 was 29% identical and 55% similar, Mre11 (53/72), Rad50 (50/71), Lig4 (41/63) and Nej1 (18/42). To investigate the role of these genes in maintaining genome stability we generated yku80::LEU2, nei1::LEU2, rad50::KANMX, nei1::KANMX, and lig4::KANMX null mutant strains. We also generated strains compromised for HR carrying rad51::LEU2 and rad52::LEU2 null alleles (52,53). Strains carrying the above mentioned mutations were crossed to each other and following tetrad analysis different double mutant combinations were recovered.

**The molecular requirements for NHEJ were conserved in K. lactis**

We investigated the molecular requirements for NHEJ in K. lactis. In particular, we wanted to see if the genes identified by us were required for efficient NHEJ, like their S. cerevisiae orthologs. Furthermore, we wanted to investigate whether haploids and diploids performed NHEJ with different efficiency. For this purpose, we performed a plasmid rejoining assay (16) using a centromeric plasmid linearized in the multiple cloning cassette, a region of the plasmid not sharing homology with the milk yeast genome. To obtain stable transformants the plasmid must be re-circularized using the NHEJ pathway, and the number of transformants using the linear vector was normalized to parallel transformations using circular plasmid. The ratio of linearized/circular transformants is a measure of NHEJ efficiency (16). In contrast to S. cerevisiae, the NHEJ efficiency in a MATa haploid and a MATa/MATa diploid strain was similar (Figure 1A),
transformation efficiency of supercoiled plasmid pCXJ18 with 
K.lactis
To test if it is substantially less efficient at joining non-cohesive ends (55).
NHEJ is efficient in rejoining ends with cohesive overhangs, suggesting that none of the proteins required for NHEJ was regulated by cell-type. Strains lacking the genes encoding Nej1, Yku80, Lig4, Mre11 and Rad50 were severely deficient for NHEJ. The nej1 yku80 double mutant strain was not more deficient than either single mutant strain indicating that Nej1 and Yku80 acted in the same pathway. As expected, Rad51 and Rad52 were not required for efficient NHEJ.

NHEJ was efficient with a wide variety of DNA ends and NEJ1 transcription was not regulated by cell-type

In S.cerevisiae, utilization of either NHEJ or HR is regulated by the nature of the DSB and by cell-type. In S.cerevisiae, NHEJ is efficient in rejoining ends with cohesive overhangs, but ~40-fold less efficient in rejoining blunt ends (13,54) and substantially less efficient at joining non-cohesive ends (55). To test if K.lactis had a similar specificity, we compared the transformation efficiency of supercoiled plasmid pCXJ18 with the same plasmid digested with various restriction enzymes (Figure 1B). No significant differences were detected for blunt ends (SmaI digested), S' (HindIII) or 3' (PstI) cohesive ends. Strikingly, the rejoining efficiency was high even for ends with opposite overhangs (HindIII PstI double digest), reduced only about 2-fold compared to the efficiency for cohesive ends. To analyze how the 3' protruding single strand (3' PSS) and 5' PSS were joined we rescued the plasmid from 13 independent transformants and sequenced the joint. All events recovered had deletions of the 3' PSS and several of the 5' PSS or both (Figure 1C). In all events except one that had a 17 bp deletion, at least one base from the 3' PSS was present in the repair product.

In S.cerevisiae, transcription of NEJ1 is regulated by the a1/a2 repressor, leading to an absence of Nej1 in MATa/ MATα diploids (25–28). Diploid cells are therefore inefficient at NHEJ. This regulation is likely to be conserved among closely related yeasts, since a multiple sequence alignment of the promoter sequences of the NEJ1 genes from S.cerevisiae,
Saccharomyces paradoxus, Saccharomyces mikita  e, Saccharomyces bayanus and Saccharomyces kudriavzevii revealed the presence of the a1/a2 operator (data not shown). We previously showed that K. lactis contains a regulator similar to the a1/a2 repressor (56), but an a1/a2 operator site could not be identified in the KINEJ1 promoter region. This could be due to either that KINEJ1 was not regulated by cell-type or that the a1/a2 operator site was different compared to S. cerevisiae and thus not found using our search criteria. To distinguish between these possibilities, we performed an RNA-blot analysis on total RNA from the appropriate K. lactis strains. We could detect the NEJ1 transcript in all cell-types investigated (Figure 1D), both in the presence (MATa sir2) and in the absence (MATa, MATa hmlΔp and MATa sir2 hmlΔp) of the a1/a2 repressor. Hence, NEJ1 transcription was not regulated by cell-type in K. lactis, consistent with the observation that both haploid and diploid cells performed NHEJ with similar efficiency (Figure 1A). We concluded that NHEJ was very versatile in K. lactis, capable of handling a wide variety of ends and insensitive to cell-type.

IR and NHEJ had identical genetic requirements

IR is defined as the joining of two DNA molecules not sharing extensive homology to each other. To test the genetic requirements for IR wild-type strain CK213-4C and mutant strains compromised for NHEJ (nej1, ku80, nej1 ku80, lig4, mre11 and rad50) or HR (rad51 and rad52) were transformed with linearized plasmid pRS406. Plasmid pRS406 has no homology to the K. lactis genome and is unable to replicate as an episome. Thus, Ura+ stable transformants only arise as a result of integrative IR in the genome. IR occurred with a frequency of ~1 x 10^3 transformants/μg of plasmid DNA, which can be compared to 1.5 x 10^4 in S. cerevisiae using a plasmid of similar size (57). Integrative transformation by IR thus occurred ~1000-fold more frequently in K. lactis than in S. cerevisiae. No transformants were obtained with NHEJ mutant strains, but mutations in the HR pathway did not affect the IR frequency (Table 2). In addition, all strains investigated had approximately the same transformation efficiencies using a circular centromeric plasmid. The results clearly showed that IR was completely dependent on the NHEJ pathway and independent of the HR pathway.

Target sites of IR

Since the NHEJ pathway repairs DSBs and IR in K. lactis completely depended on NHEJ, we reasoned that the mechanism of IR in K. lactis involved DSB repair events during which the plasmid becomes captured in-between two DNA ends. Consistent with this model, K. lactis strains transformed with circular pFA6α-KANMX resulted in very low transformation efficiency (0–2 transformants/μg of plasmid DNA), indicating that free DNA ends were necessary for efficient integration. If this model was correct, then mapping the locations of the IR-events should reveal the loci where mitotic DSBs repaired by NHEJ occur. The linearized pFA6α-KANMX was transformed into a wild-type strain. To explore the target sites, we carried out plasmid rescue from several independent integrants and sequenced the flanking genomic regions (see Materials and Methods). For 38 out of 164 successful plasmid rescue events, we could not determine an exact chromosomal location. Twenty-three were insertions in the rDNA loci on chromosome IV, three were in pKD1, an endogenous episomal plasmid, and four insertions were in repetitive DNA other than rDNA (subtelomeric repeats or transposons). For eight of the plasmids we did not find homology to K. lactis chromosomal DNA, but two of these showed significant homology to K. lactis mitochondrial DNA. By crossing the strains containing these insertions, followed by tetrad analysis, we determined that the insertions segregated as nuclear single copy genes. Our interpretation of these events was that fragmented mitochondrial DNA was captured between the ends of the plasmid and the genomic target locus during IR, thus leading to transfer of mitochondrial DNA into the nucleus. Regarding the remaining six events, we speculate that salmon sperm DNA introduced during the transformation procedure was inserted at the junctions, as observed previously in Schizosaccharomyces pombe (58). For 126 insertions we could determine an exact location in the genome using the newly released K. lactis genome sequence, and these insertions are schematically shown in Figure 2 (exact locations are available on request).

The target sites were largely randomly distributed with similar insertion intensities on all chromosomes (Table 3). The distances between the insertions ranged from 180 bp to 343 kb, with a mean distance of 84.4 kb and a median distance of 62 kb. A statistical calculation showed that the smallest distance of 180 bp was not unexpected ($P = 0.25$) assuming uniformly distributed insertions. Only 31% of the insertions (39) were in ORFs, an unexpectedly low number given that 71.6% of the K. lactis genome is occupied by protein encoding genes (48). The difference in insertion intensity (Insertions/Mbp) between ORFs and intergenic regions (IGRs) was highly significant ($P < 0.001$). In addition, the number of insertions observed in the rDNA locus on chromosome IV was higher than would be expected by chance ($\chi^2$-test, $P = 0.0014$). We noted that the insertion intensity in IGRs and rDNA was similar (Table 4), suggesting that they may arise from a similar mechanism. In summary, IR-target sites were overrepresented in IGRs and in the rDNA loci, but did not have a preference for a particular chromosome.

### Table 2. IR was completely dependent on the NHEJ pathway

<table>
<thead>
<tr>
<th>Strains</th>
<th>Number of Ura+ colonies/μg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pRS406</td>
</tr>
<tr>
<td>Wild-type</td>
<td>1.2 x 10^3</td>
</tr>
<tr>
<td>nej1</td>
<td>0</td>
</tr>
<tr>
<td>yku80</td>
<td>0</td>
</tr>
<tr>
<td>nej1 yku80</td>
<td>0</td>
</tr>
<tr>
<td>lig4</td>
<td>0</td>
</tr>
<tr>
<td>mre11</td>
<td>0</td>
</tr>
<tr>
<td>rad50</td>
<td>0</td>
</tr>
<tr>
<td>rad51</td>
<td>1.1 x 10^3</td>
</tr>
<tr>
<td>rad52</td>
<td>1.1 x 10^3</td>
</tr>
<tr>
<td>nej1 rad51</td>
<td>0</td>
</tr>
<tr>
<td>nej1 rad52</td>
<td>0</td>
</tr>
</tbody>
</table>

One μg of plasmid pRS406 was linearized with SacI and transformed into the indicated strains. The number of Ura+ colonies was determined by growth on SC-Ura plates representing IR events. Parallel transformations of supercoiled pCXJ18 (1 μg) into the strains revealed no significant differences in transformation efficiency.
IR was associated with insertions and deletions, but not with microhomology and topoisomerase I cleavage sites

Next we investigated if sequence alterations could be found close to the sites of integration. We performed plasmid rescues of 31 independent insertions using restriction enzymes that did not cleave pFA6a-KANMX, thus rescuing genomic DNA from both sides of the insertions. The flanking sequences were then compared to the undisrupted wild-type sequence in the Genolevures database. Several different mutagenic events occurred close to the breakpoints and 10 representative examples are shown in Figure 3. Small deletions or insertions (less than 5 bp) or a combination of both were the most common sequence alterations representing 15 out of 31 samples. In four cases, we observed a target site duplication of 2 bp, an event resulting in a 2 bp insertion. In five cases the alterations were large deletions or insertions (more than 100 bp). Seven of the samples revealed no sequence alteration on either side of the insertion.

In *S. cerevisiae*, the most common IR events involve base pairing between the target sequence and the terminal few base pairs of the transforming DNA (57). This microhomology mediated IR requires at least two contiguous bases of homology between the target and the plasmid end. In addition, IR in *S. cerevisiae* often occurs close to consensus cleavage sites for topoisomerase I (Top1). Top1 has a degenerate recognition motif (C/G T/A T) shared by all eukaryotes investigated (38) and thus probably also shared by *K. lactis*. For one of the integration events (3%) we found microhomology between both ends of the integrating plasmid and the target site. In two cases (6%), we found microhomology between one of the ends of the integrating plasmid and the target site. In 28 cases (90%), no microhomology between either ends was found, suggesting that base pairing between the integrating plasmid and the genomic target site was of little relevance during *K. lactis* IR. Furthermore, only five (16%) of the integration events analyzed had an adjacent Top1 site, a number close to what would be expected on a random basis. Thus, IR was highly mutagenic, but did not correlate with microhomology or Top1 sites.

IR-target sites were linked with transcriptional regulatory regions

To further explore the intergenic insertions we investigated the transcriptional orientation of the ORFs flanking the insertion points. There were three possibilities, (type I) IGRs representing the 5’ ends of the flanking genes, (type II) IGRs representing one 5’ end and one 3’ end of the flanking genes and (type III) IGRs representing the 3’ end of the flanking genes. To determine the genome wide size of each type of IGR, we analyzed the total length of type I, II and III on six randomly selected 100 kb segments of the genome (one on each chromosome). Out of 298 analyzed IGRs, type I represented 38.5%, type II 45.4% and type III 16.0%
The high frequency of IR-events in the promoter regions prompted us to examine whether they correlated with transcript levels. We therefore determined the CAI of the genes targeted by IR. The CAI is a single value measurement that summarizes the codon usage of a gene relative to the codon usage of a reference set of genes. This value ranges from 0 to 1. A high CAI value suggests that the gene of interest has similar codon usage compared to the reference genes.

Previously, this value was suggested as a useful predictor of gene expression level when highly expressed genes were used as reference genes (59). We computed the CAI values of all protein coding genes in the K. lactis genome (N = 5330) using ribosomal protein genes, assumed to be highly expressed, as the reference set. The CAI values were plotted against the GC content at the third codon position (GC3) (Figure 4B). Indeed, most of the ribosomal protein coding genes exhibit high CAI values and cluster on top of the plot, consistent with the assumption that they are highly expressed. Moreover, a lack of correlation between the genome CAI values and GC3 suggested that CAI values were not dependent upon the GC composition due to mutational bias (60), further confirming that CAI values were valid predictors of gene expression levels in K. lactis.

To explore whether the target sites of IR correlated with highly expressed genes, the CAI values of the genes in
which the insertion of pFA6a-KANMX occurred either within the ORF or in their 5’-promoter region were determined and compared with the genome global CAI value distribution. When the insertion target site corresponded to two different promoter regions, both ORFs were included in the calculation. The mean CAI values of the ORFs with intragenic insertions (0.204, 0.156 SD) and ORFs with insertions in the promoter region (0.256, 0.179 SD) were close to the global genome mean (0.206, 0.13 SD) and significantly lower than the mean value of the ribosomal protein genes (0.798, 0.156 SD). Hence, IR target sites did not correlate with highly expressed genes.

**Ploidy or a rad52 mutation did not change the target site preference for IGRs**

The IR-target site preference in IGRs and rDNA could be an artifact due to the use of a haploid strain in these experiments. Insertions within ORFs could result in lethality more often than insertions into IGRs or rDNA. Furthermore, it was possible that spontaneous DSBs occurred with equal frequency throughout the genome, but that DSBs in ORFs were preferentially repaired by the less error prone HR pathway. To test these possibilities we examined IR target sites in a rad52/rad52 diploid strain, a strain defective for HR and insensitive to recessive lethal insertions. If either of the above hypotheses were valid, then we would expect an altered target site distribution in the rad52/rad52 diploid strain. Forty-six targets sites were investigated using an inverse PCR approach. The insertion intensity varied somewhat between chromosomes, probably due to the smaller sample size (Table 3). Significantly, 31 insertions were in IGRs and only 12 were in ORFs (Table 4). The insertion intensity into IGRs was 5.6-fold (28.7/5.1) higher than into ORFs in the wild-type haploid strain and 6.4-fold (10.2/1.6) in the rad52/rad52 diploid strain. In addition, the rad52/rad52 diploid strain maintained the target site preference for transcriptional regulatory regions (Figure 4B). For insertions into rDNA, however, the rad52/rad52 diploid strain no longer showed a preference for this region. The insertion intensity into the rDNA loci was similar to the insertion intensity into the non rDNA portion of chromosome IV (Table 4).

**IR at ectopic DSBs**

It is well known that chromatin is a dynamic structure and that histone modification patterns differ across the genome. The observed preference of IR target sites within IGRs corresponding to promoter regions prompted us to examine whether the NHEJ machinery was more efficient at repairing DSBs in these regions compared to coding regions. Alternatively, more spontaneous DSBs may arise in promoters than in ORFs. To address whether IR could be directed to DSBs and whether the NHEJ efficiency was dependent on the genomic location (ORF versus promoter region) we introduced ectopic I-SceI recognition sites either within or upstream of ORF E00462g, generating strains SAY648 and PMY34, respectively (Figure 5A). Based on sequence conservation with *S.cerevisiae* we named this ORF VMR1. VMR1 encodes a predicted multiple drug resistance pump. Insertion of a strong promoter upstream of VMR1 results in strains resistant to high levels of G418 (E. Barsoum and S. U. Åström, unpublished data), suggesting that Vmr1 may rid cells of this compound. The strains were transformed with either pCXJ18 or pCXJ18 containing the I-SCEI gene controlled by the galactose-inducible GAL1 promoter. The effect of the expression of the endonuclease was examined on DNA-blots and on glucose and galactose containing plates (Figure 5B and C).

The I-SceI endonuclease cut the two sites with similar efficiency in vivo; ~11% efficiency was observed after 3 h in galactose containing medium in both PMY34 and SAY684 (Figure 5B). Both strains containing an ectopic I-SceI site and expressing the I-SceI endonuclease grew slowly on galactose compared to glucose (Figure 5C). However, plating efficiency was not severely affected probably because the DSBs were efficiently repaired. Endonuclease expression in a strain lacking the I-SceI site did not affect growth. In addition, SAY648 and PMY34 containing pCXJ18 alone had no growth defect on galactose. Both HR and NHEJ could repair the I-SceI induced break, since strains lacking either Rad52 (SAY686) or Nej1 (SAY685) and containing the I-SceI site in the VMR1 promoter, were sensitive to increased endonuclease expression. This sensitivity was seen as lowered plating efficiency on galactose plates. It should be noted that the GAL1 promoter was leaky in *K.lactis* (data not shown). This explained the reduced plating efficiency of the rad52 strain on glucose plates.

To explore if IR in the VMR1 ORF and in the VMR1 promoter region was equally efficient, a linearized LEU2 plasmid (pPMB18) with the break close to the ADH1 and TEF2 promoters and the plasmid expressing the endonuclease were co-introduced into SAY648 and PMY34 (Figure 5A). The resulting integration events were analyzed by PCR using the primers indicated in Figure 5A. The results showed that there was efficient targeting of pPMB18 to both ectopic DSBs. No significant difference in targeting efficiency between the strains was observed (Figure 5D). In addition, there was a perfect correlation between G418 resistance and isolates that the PCR analysis showed had an insertion of pPMB18 into the VMR1 promoter. Targeted integration required both the I-SceI site and the endonuclease, since none of the Leu+ transformants from a strain (CK213-4c) lacking the I-SceI site was G418 resistant. No I-SceI site targeted events were observed when pCXJ18 was co-introduced with pPMB18 into both SAY648 and PMY34. Targeted integration was also dependent on the NHEJ pathway since an isogenic nej1 mutant strain did not render any integrative transformants (data not shown). These results demonstrated that IR could be efficiently targeted to a specific locus by introducing a DSB. Moreover, when ectopic DSBs were generated using an endonuclease, a promoter region and an ORF were equally good targets for IR.

**DISCUSSION**

In this study we showed that IR depended entirely on the NHEJ pathway and present very strong evidence that IR occurred at spontaneous mitotic DSB in milk yeast. In contrast to *S.cerevisiae*, *K.lactis* performed IR very efficiently and therefore constituted an excellent model organism to analyze the molecular mechanisms underlying NHEJ. Indeed, IR in *K.lactis* was 1000-fold more efficient than in *S.cerevisiae* (Table 2). The molecular requirements for NHEJ were...
identical in both organisms. *K. lactis* mutant strains lacking *Nej1*, *Ku80*, *Lig4*, *Mre11* and *Rad50* were unable to rejoin a linearized episomal plasmid or integrate a non-homologous DNA molecule into the genome (Figure 1A and Table 2).

In similarity with mammalian cells, the *K. lactis* NHEJ pathway was capable of repairing linearized plasmids containing complementary, blunt and non-complementary ends (Figure 1B). Importantly, it was highly efficient at rejoining a plasmid linearized with restriction endonucleases generating a 5′ and a 3′ overhang. This observation may partly explain the efficient IR observed in *K. lactis*. The capability of efficient ligation of blunt and noncohesive ends is also a...
feature of S. pombe and mammalian NHEJ (17,20,61,62). However, in S. cerevisiae efficient NHEJ only takes place with complementary overhangs (54). K. lactis must be able to form a stable NHEJ-complex capable of aligning the two ends without the additional energy provided by base pairing between the severed ends. By analyzing the repair events from PstI–HindIII double digested plasmids, we found that the 3’ PSS was not completely removed in most repair events (Figure 1C). Provided that DNA polymerases only synthesize DNA in the 5’–3’ direction, two alternative models for the molecular mechanism underlying the repair of 3’ PSS-5’ PSS can be envisaged. First, a single stranded DNA ligation event takes place followed by fill-in DNA synthesis. Second, an alignment event stabilizing the interaction between the ends takes place, followed by fill-DNA synthesis and dsDNA ligation. In both cases DNA synthesis is primed by the recessed strand of the 5’ PSS end. Among eukaryotic DNA ligases only a viral ligase has been reported to possess ssDNA ligase activity (63) and there is no evidence of such an activity for DNA ligase IV. Therefore, we favor the latter model, in which milk yeast NHEJ can efficiently align the non-complementary DNA ends. The same model has been proposed for NHEJ of mismatched non-complementary ends in human lymphoblasts (61).

Another interesting difference between both yeasts was that NHEJ was not regulated by cell-type in K. lactis and that NEJ1 transcription was not repressed by the a1/a2 heterodimer (Figure 1D). In contrast to S. cerevisiae, K. lactis vegetative growth is mostly restricted to the haplophase, since diploids are rather unstable and spontaneously enter meiosis. The unstable nature of the diplophase may have prevented the evolution of cell-type regulation of NHEJ in K. lactis, given that milk yeast cells seldom linger in the diploid state. The NEJ1 gene was nevertheless essential for NHEJ in K. lactis, indicating that Nej1 did not evolve in S. cerevisiae as a function to impart cell-type regulation of NHEJ. We suggest that NHEJ required the NEJ1 gene in a common ancestor, and that Saccharomyces sensu stricto species later acquired a specific regulation of this gene. We cannot exclude, however, that the NEJ1 gene was regulated by cell-type in the common ancestor and that K. lactis lost this regulation.

In nature, IR must repair DNA lesions in the context of highly condensed chromatin fibers. Although plasmid rejoining assays have provided useful information regarding NHEJ mechanisms, it is important to note that the substrate for NHEJ in this assay is unlikely to be chromatin. The fact that K. lactis very efficiently integrates a non-homologous linear DNA fragment in the genome allowed us to further study the IR events taking place genome-wide and in chromatin. We showed that IR entirely depended upon the NHEJ pathway (Table 2). This observation stands in contrast to the situation in mouse ES-cells where NHEJ proteins are not required for many illegitimate integration events (32). In mammalian cells, there is a poorly characterized second pathway for NHEJ that is independent of the previously characterized NHEJ proteins (64). We could not observe IR events in strains compromised for NHEJ. Therefore, if there is such an alternative pathway in K. lactis, it must be extremely inefficient. The construction of K. lactis mutant strains is tedious due to the high IR efficiency. We found that K. lactis strains lacking Nej1, Yku80 or Lig4 were useful tools for improving gene targeting. Using strains compromised for NHEJ led to the recovery of integrants that were invariably targeted to the homologous locus (data not shown). The use of yku80 mutant strains for gene targeting purposes has also been proposed by others (65). Given that yku80 mutations in S. cerevisiae show pleiotropic phenotypes, including strain-dependent temperature sensitivity and telomere shortening (66), we suggest that nej1 or lig4 mutations may be a better choice for large-scale gene targeting projects. We have evidence that yku80 mutant K. lactis strains are partly compromised for telomere capping (S. D. Carter and S. U. Åström, unpublished data).

Exploring the genomic target sites for IR revealed a globally random distribution (Figure 2) and an absence of both microhomology and consensus Top1 sites (Figure 3). Interestingly, an obvious over-representation of IGRs and the rDNA locus compared to ORFs was observed (Table 4). The analysis of the IGR events showed a striking preference for 5’ promoter regions compared to 3’ non-coding regions (Figure 4A). However, no correlation with transcriptional activity could be made (Figure 4B). The insertions into the rDNA did not show a hot spot for a particular region of the rDNA repeats (data not shown), which would be expected if a particular feature of rDNA such as a replication fork block sequence would facilitate the increased level of IR. Loci with two divergently directed promoters and rDNA are expected to experience negative super-coiling that is relieved by topoisomerase activities. In S. cerevisiae, there are no reports that IR occurs preferentially in IGRs, but a preference for the rDNA locus is observed after human Top1 over expression (67). Similar studies in Candida glabrata also revealed a preference for IGRs, absence of Top1 sites and limited presence of micro-homology (68). However, in this study a smaller number of IR-events were analyzed, and the transformants were partly selected for their inability to adhere to human epithelial cells. This selection may have biased the genome wide IR target site distribution.

The IR target site distribution in a diploid rad52/rad52 strain was similar to the distribution in the wild-type haploid strain with respect to the preference for IGRs and excess of 5’ promoter regions (Table 4 and Figure 4A). These results ruled out the possibility that recessive lethal insertions significantly biased the observed distribution in haploid wild-type cells. In addition, we analyzed 32 independent wild-type diploid strains with random insertions of the pFA6a-KANMX plasmid using tetrad analysis. We found only two strains (6%) in which the insertions co-segregated with lethality. Furthermore, a compromised HR pathway did not change the target site preference for IGRs, showing that DSBs arising in ORFs were not preferentially repaired by HR. However, in the rad52/rad52 diploid strain, the IR target site preference for rDNA was lost, suggesting that HR somehow facilitated high levels of IR in the rDNA loci. To firmly establish this notion, a larger number of IR target sites from the rad52/rad52 diploid strain should be analyzed.

The observation that IR could be targeted very efficiently to an ectopically induced DSB demonstrated that spontaneous genomic DSBs likely constituted the target sites for illegitimate integration events. A model for the observed target site preference for IGRs and rDNA was that DSBs arise more frequently in IGRs and rDNA compared to ORFs. Alternatively, the NHEJ-proteins could be more efficiently recruited to...
these loci as a result of specific protein–protein interactions or a generally more accessible chromatin structure. Our results showed that I-SceI induced breaks in a promoter region or in an ORF were equally efficient targets for IR, strongly supporting the first model. In bacteria, it has also been proposed an IR pathway mediated by DSBs followed by end-joining (69).

Studies of the location of meiotic DSBs in *S. cerevisiae* demonstrate that most hotspots are intergenic rather than intragenic (70,71). In addition, among 20 intergenic DSBs representing meiotic hotspots, 13 were between the 5′ ends of two genes and only two were between the 3′ ends of two genes (72). Our tentative mapping of mitotic DSBs thus showed similarities with the location of meiotic DSBs in *S. cerevisiae*, suggesting that chromosomal sites prone to experience DSBs are similar during both vegetative growth and meiosis.

In summary, our data strongly suggest that mitotic DSBs arise in IGRs ~6-fold more frequently than in ORFs. To our knowledge, this report is the first comprehensive attempt to study the locations of mitotic DSBs in a eukaryote. We propose that spontaneous mitotic DSBs can be mapped by analyzing sites of IR integration events, providing a molecular tool for understanding how these DSBs arise. Since mitotic DSBs contribute to genomic instability in all organisms, these findings should be of significant interest.

ACKNOWLEDGEMENTS

The authors thank Jan-Olov Persson for help with statistical calculations and Monique Bolotin for generously supplying *K. lactis* DNA sequences prior to the release of the genome. The authors acknowledge Wu Gang and Stephen Freeland for the gift of the plasmid expressing I-SceI. This work was supported by grants from the Swedish Cancer Council (621-2004-1942) to S.U.Å. Funding to pay the Open Access publication charges for this article was provided by the Swedish Research Council.

Conflict of interest statement. None declared.

REFERENCES


