



HAL
open science

Mécanismes de réparations d'une cassure double-brin et résection au sein d'un microsatellite humain

Wilhelm Vaysse-Zinkhöfer

► **To cite this version:**

Wilhelm Vaysse-Zinkhöfer. Mécanismes de réparations d'une cassure double-brin et résection au sein d'un microsatellite humain. Génétique humaine. Sorbonne Université, 2021. Français. NNT : 2021SORUS477 . tel-03719442

HAL Id: tel-03719442

<https://theses.hal.science/tel-03719442>

Submitted on 11 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT
SORBONNE UNIVERSITE

École doctorale Complexité du Vivant

UMR3525/Instabilités naturelles et synthétiques des génomes (INSG)

**Mécanismes de réparation d'une cassure
double-brin et résection au sein d'un
microsatellite humain**

Wilhelm Vaysse-Zinkhöfer

Dirigé par Guy-Franck Richard

Soutenue le 16 avril 2021

Mme. Anabelle DECOTTIGNIES

Rapporteure

M. Stéphane MARCAND

Rapporteur

Mme. Anne PLESSIS

Examinatrice

M. Bernard LOPEZ

Examinateur

Mme. Sophie LOUVET-VALLEE

Examinatrice

M. Guy-Franck RICHARD

Directeur de thèse

THESE DE DOCTORAT
SORBONNE UNIVERSITE

École doctorale Complexité du Vivant

UMR3525/Instabilités naturelles et synthétiques des génomes (INSG)

**Mécanismes de réparation d'une cassure
double-brin et résection au sein d'un
microsatellite humain**

Wilhelm Vaysse-Zinkhöfer

Dirigé par Guy-Franck Richard

Soutenue le 16 avril 2021

Mme. Anabelle DECOTTIGNIES

Rapporteure

M. Stéphane MARCAND

Rapporteur

Mme. Anne PLESSIS

Examinatrice

M. Bernard LOPEZ

Examinateur

Mme. Sophie LOUVET-VALLEE

Examinatrice

M. Guy-Franck RICHARD

Directeur de thèse

J'aimerais remercier Anabelle Decottignies et Stéphane Marcand, rapporteurs de cette thèse, ainsi que Anne Plessis, Sophie Louvet-Vallee et Bernard Lopez, examinateurs, d'avoir accepté de faire partie de mon jury de thèse et d'évaluer mes travaux. Encore merci à Bernard Lopez pour son suivi lors de mes comités de thèse.

Pour ces années de découvertes, de curiosités et d'apprentissage, je remercie Guy-Franck Richard. Merci de cette chance que tu m'as offerte, de cette confiance que tu m'accordais et qui se traduisait par une liberté dans laquelle j'ai pu m'épanouir. Merci pour tes combats administratifs qui m'ont permis de commencer et de conclure ces travaux. Merci de tes conseils, de tes corrections, de ton aide et de ton soutien. Je pense être devenu un peu plus grand grâce à toi, merci de m'avoir donné cette chance de devenir qui je suis.

Merci David Viterbo, pour la bonne humeur que tu sembles toujours porter avec toi, pour tes conseils toujours justes, ton humour et ton empathie qui ont éclairé plus d'une fois des journées sombres. Merci de ton aide, tout au long de mes travaux, sans laquelle j'aurais plusieurs fois été perdu. Merci de ta présence dans les moments difficiles.

Je te remercie Saliha de ton soutien lorsque tu ne me connaissais à peine. Je serais toujours reconnaissant de ces moments. Merci de m'avoir aidé à porter une partie des moments difficiles. Merci pour les rires et le courage que tu m'as donné.

Stéphane, nos débats du vendredi soir étaient toujours un *must*, ton aide en statistiques et en informatique s'est avérée cruciale. Qui aurait pu mieux égayer le bureau de sa présence ? Merci encore.

Merci à toi Laureline, pour ta gentillesse et ta passion sans limite.

Je remercie Lucie Poggi, Olivia Frenoy, Thomas Damagnez, Charles Monod-Broca, Estelle Dacheux, Simon Malesys pour toutes les bières, les rires et les bonnes choses. Merci Frédérique Moyrand et Jean-Yves Coppee d'avoir été de parfait voisin de bureau rendant agréable toutes ces années de travaux

Merci à Frank Feuerbach de s'être rendu disponible pour mes nombreuses questions et son intérêt dans mes travaux.

Merci à toi Christelle Lenormand pour ta bonne humeur sans faille et ton énergie sans fin, de la signature de mon contrat *in extremis* un 28 décembre jusqu'à aujourd'hui.

J'aimerais remercier tous les membres du 4^e étage Fernbach pour les bons moments passés. J'aimerais spécialement remercier le travail de Lucia Oreus pour son aide permanente dans l'activité de l'étage ainsi que les membres du l'unité Dynamique du génome.

Merci à mes amis.

Merci à toi Jan, mon frère, dont je ne saurais être plus fier.

Finalement merci à vous Papa et Maman, sans qui rien n'aurait été possible. Pour votre soutien, votre confiance, vos encouragements et votre amour. Merci.

LISTE DES ABREVIATIONS

NOR : Nucleolar organizer regions
CDB : Cassure double-brin
STR : Short tandem repeat
SSR : Short sequence repeat
SSM : Slipped strand mispairing
DM1 : Dystrophie myotonique de type 1
NER : Nucleotide excision repair
GG-NER : Global genome NER
TC-NER : Transcription coupled NER
BER : Base excision repair
SN-BER : Single-nucleotide BER
LP-BER : Long patch BER
MMR : Mismatch repair
CPD : Cyclobutane pyrimidine dimer
6-4PP : 6-4 Pyrimidine-Pyrimidone
RFC : Replication factor C
PCNA : Proliferating cell nuclear antigen
ROS : Reactive oxygen species
CFS : Common fragile site
RFS : Rare fragile site
CRISPR : Clustered Regularly Interspaced Short Palindromic Repeats
TALEN : Transcription activator-like effector nucleases
ZFN : Zinc finger nuclease
SSA : Single-strand annealing
NHEJ : Non-homologous end joining
MMEJ : Micro-homology mediated end joining
DSBR : Double-strand break repair
SDSA : Synthesis dependent strand annealing
BIR : Break induced replication

INDEX DES FIGURES ET TABLEAUX

FIGURE 1 : REPRESENTATION SCHEMATIQUE DES NATURES DES SEQUENCES COMPOSANT LES GENOMES EUCARYOTES.	8
FIGURE 2 : REPRESENTATION SCHEMATIQUE DE L'IMPACT DES EXPANSIONS DE TRIPLETS CTG SUR LE METABOLISME ARN DE LA CELLULE.....	10
FIGURE 3 : REPRESENTATION SCHEMATIQUE DES POSSIBLES STRUCTURES SECONDAIRES ADOPTÉES PAR LES TRINUCLEOTIDES (CTG) _N , (CCG) _N , (GAA) _N A DIFFÉRENTS LOCUS <i>IN VIVO</i>	14
FIGURE 4 : REPRESENTATION SCHEMATIQUE DU GLISSEMENT A LA REPLICATION (<i>SLIPPED-STRAND MISPAIRING</i> (SSM)).....	17
FIGURE 5 : REPRESENTATION DE LA FOURCHE DE REPLICATION AU NIVEAU D'UNE REPÉTITION CAG/CTG.	21
FIGURE 6 : REPRESENTATION SCHEMATIQUE DE LA FOURCHE DE REPLICATION PASSANT AU SEIN D'UNE REGION CAG/CTG.	23
FIGURE 7 : SCHEMA DES DIFFÉRENTS MECANISMES DE REPARATION PAR LE « MISMATCH REPAIR ».....	26
FIGURE 8 : HOMOLOGUES DES PROTEINES DE REPARATION DE LA LEVURE <i>SACCHAROMYCES CEREVISIAE</i> ET D' <i>HOMO SAPIENS</i>	35
FIGURE 9 : SCHEMA DES DIFFÉRENTS MECANISMES DE REPARATION PAR LE « NUCLEOTIDE EXCISION REPAIR ».	36
FIGURE 10 : SCHEMA DES DIFFÉRENTS MECANISMES DE REPARATION PAR LE <i>BASE EXCISION REPAIR</i>	44
FIGURE 11 : STRUCTURES CRISTALLINES DES QUATRE FAMILLES DE NUCLEASES.	59
FIGURE 12 : SCHEMA REPRESENTANT LES MECANISMES DE REPARATION DU <i>NON-HOMOLOGOUS END JOINING</i> (NHEJ), DU <i>MICRO-HOMOLOGY MEDIATED END JOINING</i> (MMEJ) ET DU <i>SINGLE STRAND ANNEALING</i> (SSA).....	66
FIGURE 13 : MODELE DE RESECTION D'UNE CASSURE DOUBLE-BRIN.....	72
FIGURE 14 : SCHEMA DES MECANISMES DE REPARATION PAR RECOMBINAISON HOMOLOGUE.	81
FIGURE 15 : REPRESENTATION SCHEMATIQUE DU MECANISME DE GLISSEMENT LORS DE LA RECOMBINAISON.	90
FIGURE 16 : REPRESENTATION SCHEMATIQUE DU SYSTEME TALEN.....	91
FIGURE 17 : REPRESENTATION SCHEMATIQUE DE L'INTERACTION ENTRE LE COMPLEXE MRX, TEL1 ET LE COMPLEXE RAPI-RIF1-RIF2.	96
FIGURE 18 : MODELE DE CASSETTE POUVANT SERVIR A L'ETUDE DES EXPANSIONS DANS DES CELLULES HUMAINES.	141
FIGURE 19 : MODELE DE L'INTERACTION ENTRE LE COMPLEXE TELOMERIQUE RIF1-RIF2, MRX ET UNE CASSURE DOUBLE-BRIN LOCALISE AU SEIN DE SEQUENCES REPETEES OU NON REPETEES.....	143

Table des matières

LISTE DES ABREVIATIONS.....	2
INDEX DES FIGURES ET TABLEAUX	3
INTRODUCTION	7
1 MICROSATELLITES	7
1.1 HISTORIQUE	7
1.2 MICROSATELLITES, DEFINITION ET PROPRIETES	10
2 FACTEURS DE L'INSTABILITE DES MICROSATELLITES	12
2.1 STRUCTURES SECONDAIRES DES REPETITIONS DE DINUCLEOTIDES	12
2.2 STRUCTURES SECONDAIRES DES REPETITIONS DE TRINUCLEOTIDES	13
2.2.1 ÉPINGLES IMPARFAITES	14
2.2.2 TRIPLES HELICES.....	15
2.2.3 G-QUADRUPLEX	15
2.2.4 LES REPETITIONS DE TRINUCLEOTIDES NON STRUCTUREES	16
2.3 GLISSEMENT LORS DE LA REPLICATION (SSM).....	16
2.3.1 GLISSEMENT LORS DE LA REPLICATION ET PURETE DES REPETITIONS	18
2.4 REPLICATION ET INSTABILITE	19
2.4.1 FOURCHE DE REPLICATION ET INSTABILITE	19
2.5 ORIENTATION DES MICROSATELLITES ET INSTABILITE	22
2.6 MECANISMES DE REPARATION DES LESIONS A L'ADN ET INSTABILITE	24
2.6.1 EFFET DU MMR SUR LA STABILITE DU GENOME	24
2.6.2 « NUCLEOTIDE EXCISION REPAIR » (NER)	35
2.6.3 BASE EXCISION REPAIR (BER)	42
2.7 ENVIRONNEMENT CHROMATINIEN ET INSTABILITE DES MICROSATELLITES	47
2.7.1 ENVIRONNEMENT CHROMATINIEN DES REPETITIONS DE TRINUCLEOTIDES CGG/CCG	48
2.7.2 ENVIRONNEMENT CHROMATINIEN DES REPETITIONS DE MICROSATELLITES GAA/TTC.....	49
2.7.3 ENVIRONNEMENT CHROMATINIEN DES REPETITIONS DE MICROSATELLITES CAG/CTG	49
3 SITES FRAGILES, CANCER ET MICROSATELLITES	51
4 LES NUCLEASES	52
4.1 LES MEGANUCLEASES	53
4.2 NUCLEASES A DOIGTS DE ZINC (ZFN)	55
4.3 LES TALENS	57
4.4 CRISPR-CAS.....	59
4.4.1 ADAPTATION	61
4.4.2 EXPRESSION.....	62
4.4.3 INTERFERENCE.....	62
4.4.4 CRISPR-Cas9	63
5 MECANISMES DE REPARATION DES CASSURES DOUBLE-BRIN ET REPETITIONS.....	64
5.1 LES DIFFERENTS MECANISMES DE REPARATION	65

5.1.1	NHEJ CHEZ L'HUMAIN	66
5.1.2	MICRO-HOMOLOGY MEDIATED END JOINING (MMEJ).....	68
5.1.3	RECOMBINAISON HOMOLOGUE	69
5.1.4	ÉTAPES TARDIVES DE LA RESECTION	75
5.1.5	MECANISMES DEPENDANTS DE <i>RAD51</i> ET <i>RAD52</i>	79
5.2	RECOMBINAISON HOMOLOGUE ET INSTABILITE DES REPETITIONS EN TANDEM	87
5.3	REPARATION D'UNE CASSURE DOUBLE-BRIN AU SEIN D'UNE REPETITION DE TRINUCLEOTIDES CAG/CTG.....	90
6	INTERACTIONS DE LA MACHINERIE TELOMERIQUE AVEC LA RESECTION.....	93
	PROBLEMATIQUE DE MON TRAVAIL DE THESE.....	97
	RESULTATS	98
	INTRODUCTION	101
	MATERIALS AND METHODS	104
	YEAST STRAINS AND PLASMIDS.....	104
	TALEN INDUCTION.....	105
	DSB ANALYSIS AND QUANTIFICATION.....	106
	TRINUCLEOTIDE REPEAT LENGTH ANALYSIS	107
	ANALYSIS OF DSB END RESECTION.....	108
	RESULTS	109
	A DOUBLE-STRAND BREAK INDUCED WITHIN CTG TRINUCLEOTIDE REPEATS REQUIRES THE <i>MRE11P</i> NUCLEASE ACTIVITIES AND <i>SAE2P</i> TO BE PROCESSED	109
	LONG RANGE RESECTION: <i>SGS1</i> , <i>DNA2</i> AND <i>EXO1</i>	110
	INVOLVEMENT OF THE TELOMERIC END PROTECTION COMPLEX IN CTG DSB REPAIR.....	111
	DSB RESECTION IS ABOLISHED IN <i>SAE2Δ</i> AND <i>MRE11</i> MUTANTS	112
	RESECTION AT CTG TRINUCLEOTIDE REPEATS IS INCREASED IN THE ABSENCE OF <i>RIF1P</i> OR <i>RIF2P</i>	113
	CELL SURVIVAL DEPENDS ON <i>SAE2</i> AND <i>MRE11</i>	113
	TALEN EXPRESSION DOES NOT INDUCE OTHER MUTATIONS	114
	DISCUSSION.....	115
	DIFFERENCES IN <i>MRE11</i> MUTATIONS IN CTG REPEAT RESECTION AND REPAIR	115
	CTG TRINUCLEOTIDE REPEATS DO NOT FORM CLEAN DSB ENDS	115
	LONG RANGE RESECTION	116
	INVOLVEMENT OF TELOMERIC PROTECTION COMPLEX IN CTG END RESECTION	117
	DISCUSSION & PERSPECTIVES	136
	REFERENCES	145
	ANNEXES.....	162

INTRODUCTION

1 Microsatellites

1.1 Historique

Regardez la composition en séquence de votre génome eucaryote préféré (dans mon cas *Panthera uncia*) et vous observerez que ce dernier est composé de deux types de séquences, les séquences uniques du génome et les séquences répétées. Les séquences uniques du génome concernent la majorité des gènes et des séquences régulatrices non codantes. Les séquences répétées du génome (retrouvées en plusieurs copies), sont-elles même divisées en deux familles. On distingue d'une part les séquences répétées dites « dispersées » (les transposons (classe I et II), les gènes à ARNt et les gènes paralogues) et d'autre part les séquences répétées en « tandem » (ADN satellites, ADN ribosomiques, télomères et les gènes paralogues en tandem) (figure 1). Historiquement, ces séquences répétées ont été mises en évidence après des expériences de dénaturation-renaturation, bien avant l'apparition des moyens de séquençage modernes. Lorsqu'une solution contenant de l'ADN génomique dénaturé est placée dans des conditions favorables de renaturation, la rapidité à laquelle une séquence particulière pourra se renaturer sera proportionnelle au nombre de copies de cette même séquence dans le génome. Le paramètre C_{ot} , a été défini comme la valeur associant la concentration d'une molécule d'ADN et le temps nécessaire à sa réassociation à une température donnée. Le paramètre $C_{ot_{1/2}}$ est la valeur à laquelle la moitié des molécules d'une séquence d'ADN, pour un organisme donné, est renaturée. Trois types de séquences ont ainsi été mises en évidence, les séquences fortement répétées avec un C_{ot} très bas, les séquences uniques avec un C_{ot} très élevé et les séquences allant de moyennement répétées à faiblement répétées, associant des C_{ot} progressivement moins élevés ¹. La définition de ce paramètre a permis de mieux comprendre

la composition et l'évolution des génomes eucaryotes et a été la première étape vers la caractérisation des séquences répétées. Ainsi l'ADN de chaque génome peut être défini par cette valeur ².

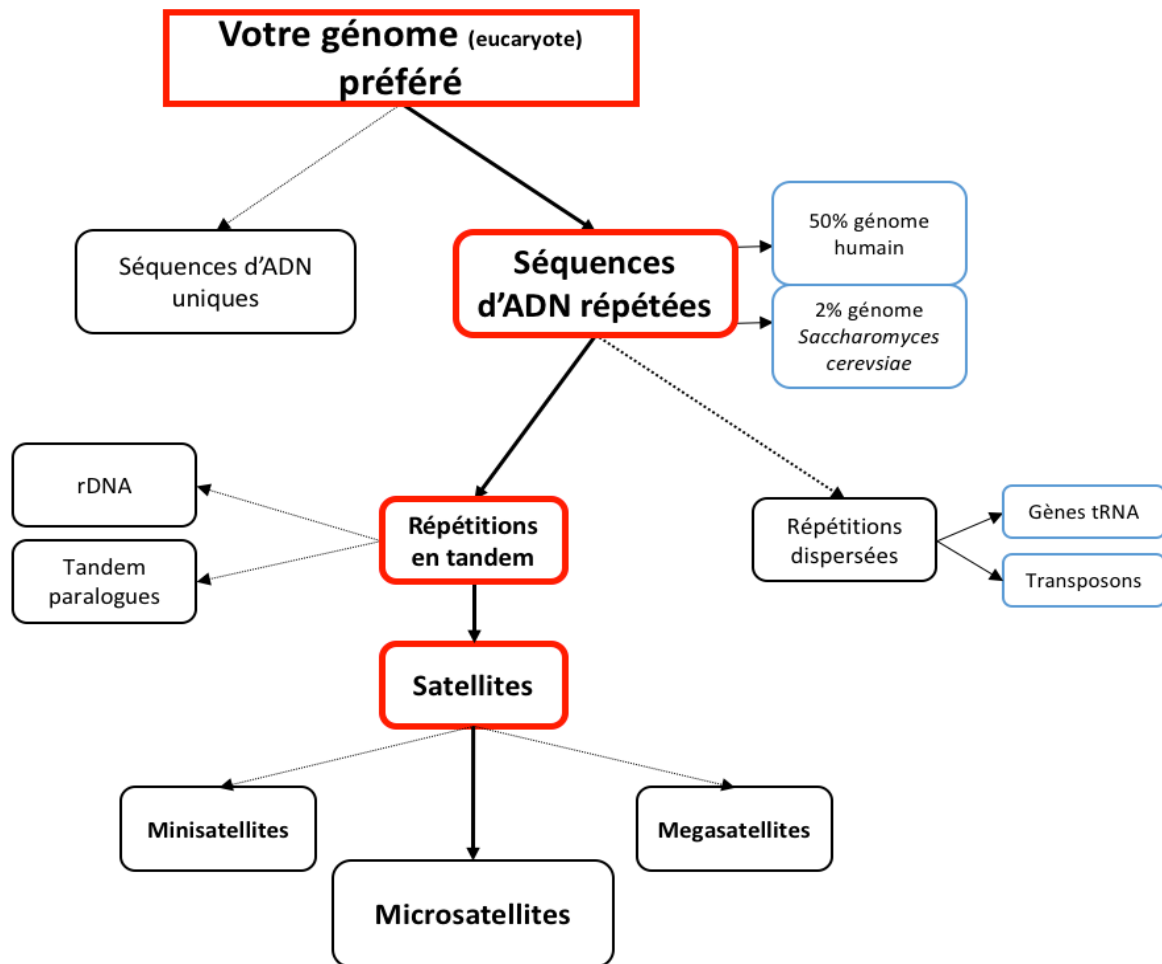


Figure 1 : Représentation schématique des natures des séquences composant les génomes eucaryotes.

Ces observations n'ont pas permis tout de suite de comprendre le rapport entre la taille d'un génome et la quantité de matériel génétique retrouvé dans les cellules. Il a été associé à cette quantité de matériel génétique, pour un organisme donné, la « valeur C ». Les premières observations montraient une différence de quantité de matériel génétique entre les organismes et ne montraient aucune relation entre la complexité d'un organisme et la quantité de d'ADN dans son noyau ³. Ces observations ont conduit au paradoxe de la « valeur C » ⁴. La réponse définitive est venue du séquençage complet des premiers eucaryotes ⁵⁻⁷. Tandis qu'une partie

des génomes était constituée d'une proportion importante, quoique variable, de séquences répétées ($\approx 50\%$ du génome humain, $\approx 2\%$ pour *S. cerevisiae*), l'ordre de grandeur des gènes restait le même entre les eucaryotes (6,000 gènes pour *S. cerevisiae* contre 20,000 à 25,000 pour l'humain), malgré une différence d'un facteur 200 entre les tailles des génomes de l'humain et de la levure. Longtemps considéré comme « Junk DNA » (ADN poubelle) ⁸, les implications de ces régions dans la régulation de l'expression génétique sont aujourd'hui reconnues et étudiées. On retrouve une fonction des séquences répétées dans l'évolution des génomes et la spéciation ⁹, la formation de la majorité de l'hétérochromatine, centromère, télomère et « NOR »¹ ou encore la régulation de l'expression génique par l'interaction entre les nucléosomes, les éléments répétés et les promoteurs ¹⁰. Les répétitions sont aussi impliquées dans plus de 40 pathologies neurologiques héréditaires humaines avec une surreprésentation des microsatellites à répétitions de trinuécléotides. On retrouve parmi ces maladies le syndrome du X fragile, la maladie de Huntington, la dystrophie myotonique de type 1 (DM1) ou l'ataxie de Friedreich ^{11,12}. La DM1 est une maladie héréditaire multisystémique caractérisée par une faiblesse musculaire progressive et une myotonie ¹³. L'apparition et la gravité de la maladie sont toutes deux positivement liées au nombre de répétitions trinuécléotidiques CTG. Les répétitions sont situées dans la partie 3'UTR du gène DMPK, qui code pour la protéine kinase de la dystrophie myotonique. Un individu sain possède entre 5 et 37 répétitions, tandis qu'entre 38 et 49 répétitions est considéré comme un état prémutationnel. Les individus présentant 50 répétitions et plus sont invariablement symptomatiques ¹⁴. L'ARN muté contenant des CUG se lie et séquestre MBLN1, inhibant sa fonction de facteur d'épissage de l'ARN ^{15,16} (figure 2). Les souris knock-out MBNL1 donnent le même phénotype que le trouble DM1 chez l'humain ¹⁷. Aucun traitement pour la DM1 n'existe à ce jour.

¹ (« *Nucleolar organiser regions* », qui participent à la formation du nucléole, la plus grande structure du noyau des cellules eucaryotes et centre de la biosynthèse des ribosomes (revu dans Biscotti, Harrison, 2015)

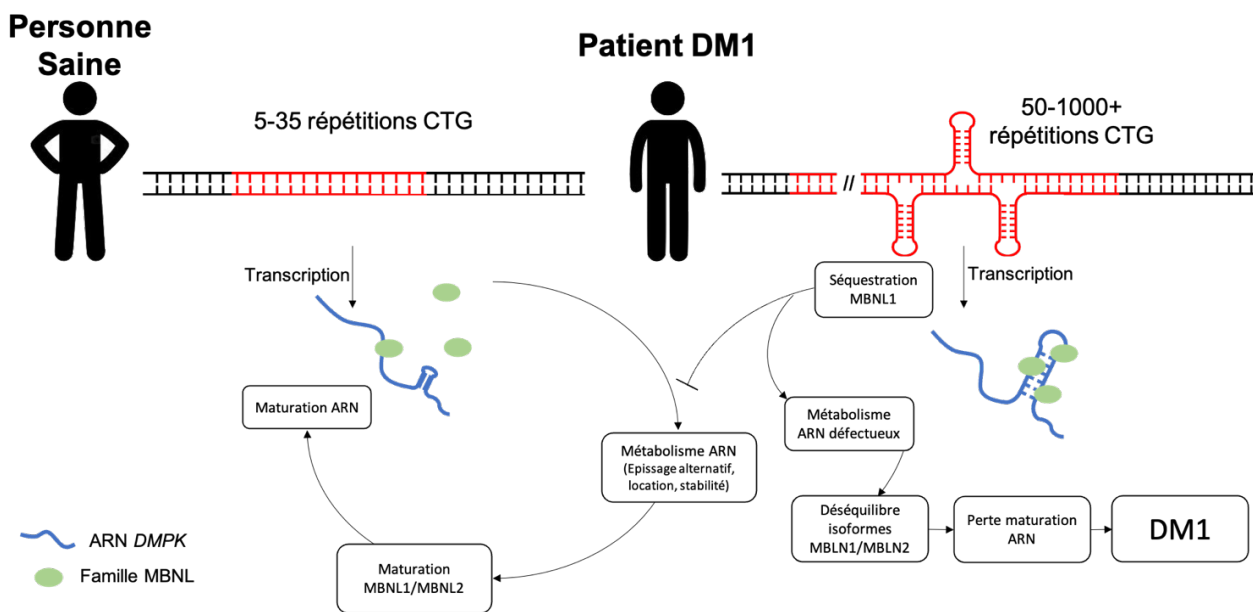


Figure 2 : Représentation schématique de l'impact des expansions de triplets CTG sur le métabolisme ARN de la cellule. Les protéines de la famille MBNL sont nécessaires au métabolisme ARN de la cellule notamment dans l'épissage alternatif des membres de la famille MBNL dont les différentes isoformes jouent un rôle critique dans le métabolisme ARN de la cellule. En présence d'un grand nombre de répétitions, une structure secondaire en épingle à cheveux séquestre une grande quantité des protéines MBNL1 conduisant à un défaut d'épissage et un déséquilibre des isoformes des protéines MBLN entraînant une perte de la maturation correcte des ARNs, responsable du développement de la DM1.

Certaines de ces répétitions sont aussi des sites fragiles, enclins à de fréquentes cassures double-brin (CDB) dès lors que la machinerie de réplication est ralentie ou altérée ¹⁸. Certains de ces sites fragiles sont associés à des translocations chromosomiques détectées dans certains adénocarcinomes, myélomes, leucémies et lymphomes ^{19,20}. De plus une instabilité des microsatellites (MSI) est présente dans 15% des cancers colorectaux et des cancers HNPCC ²¹. Dans cette introduction nous nous intéresserons à deux types de répétitions, d'abord aux répétitions de trinuécléotides puis aux répétitions télomériques.

1.2 Microsatellites, définition et propriétés

Les microsatellites, également appelés *Short Tandem Repeats (STR)* ou *Short Sequence Repeats (SSR)* sont des régions du génome composées de répétitions en tandem. Bien qu'il ne semble

pas y avoir de consensus dans la littérature, le motif de répétition est composé de 1-9 paires de bases. Les microsatellites sont considérés comme parfaits lorsque l'identité du motif de base se conserve et imparfait lorsque des interruptions du motif sont observées. Chez les eucaryotes ils sont généralement longs de quelques dizaines, jusqu'à une centaine de paires de bases. Ce sont les répétitions les plus abondantes du génome humain couvrant jusqu'à 3% de sa longueur, avec une fréquence de 0,5 par kilobase d'ADN ⁷. Chez la levure *Saccharomyces cerevisiae*, on retrouve environ 1800 di-, tri- et tétranucléotides dans un génome de 12Mb ²². Tous les microsatellites sont sous-représentés dans les régions codantes à l'exception des microsatellites à trinuécléotides et hexanucléotides car leur changement de taille ne produit pas de décalage du cadre de lecture. On retrouve, chez la levure, des répétitions de trinuécléotides dans les régions codantes avec une fréquence légèrement supérieure aux régions intergéniques du génome ²³. Chez *Homo sapiens* cette fréquence est doublée ²⁴. On retrouve des répétitions de trinuécléotides dans les 5'UTR et 3'UTR des gènes, ainsi que dans les exons et les introns ²⁵. Chez la levure, les régions subtélomériques (à l'exception du télomère droit du chromosome X) et centromériques (11 sur 16) contiennent peu de répétitions ²⁶.

La distribution et la taille moyenne des microsatellites, quant à elle, varie d'un organisme à l'autre ²³. Les microsatellites font partie des séquences d'ADN les plus variables du génome et sont hautement instables. Il a été montré que certains microsatellites sont capables de former des structures secondaires *in vitro* et probablement *in vivo* ²⁷. Contrairement aux régions uniques, le polymorphisme de ces séquences vient principalement de la variation en taille, par insertion ou délétion d'un nombre entier de motifs de base, avec une fréquence d'environ 10^{-5} à 10^{-7} pour les répétitions de di- tri- et tétranucléotides d'une taille allant de 15 à 36 répétitions. En comparaison, les séquences uniques du génome mutent à une fréquence de 10^{-9} par nucléotide par génération ²⁸.

Cette instabilité innée des microsatellites est particulièrement problématique dans le cas des

répétitions de trinuécléotides chez *Homo sapiens* (figure 2). La nature des séquences répétées permet à différents mécanismes d'amplifier de façon massive ces séquences et de conduire aux développements de maladies neurodégénératives graves, comme nous allons maintenant le voir.

2 Facteurs de l'instabilité des microsatellites

Plusieurs modèles ont été proposés afin d'expliquer l'expansion des séquences répétées. Le premier modèle fait intervenir la capacité de ces régions, par la nature des séquences les composant, à former des structures secondaires. Cette capacité à former des structures a été démontrée *in vitro* et laisse suggérer qu'elles en sont probablement capables *in vivo*. Ces structures capables de déstabiliser les fourches de réplication seraient à l'origine de cassures nécessitant l'intervention des mécanismes de réparation. Nous allons voir dans cette partie que la formation de structures secondaires, la réparation d'une cassure double brin, le « *gap repair* », le « *base excision repair* », la réplication en phase S ou n'importe quel mécanisme impliquant de la synthèse d'ADN *de novo*, est susceptible d'entraîner une expansion ou une contraction des répétitions ²⁹. Les différents mécanismes de réparations intervenant dans la réparation de ces cassures seront détaillés dans une autre partie de cette introduction.

2.1 Structures secondaires des répétitions de dinuécléotides

Une conséquence d'avoir un même motif répété le long d'un segment d'ADN est la possibilité d'être mal apparié sur le brin complémentaire, puisque rien ne distingue une unité de répétition d'une autre. Ce type de mésappariement permet la formation de structures secondaires. Cela a été historiquement démontré lors d'expériences *in vitro* sur des microsatellites à dinuécléotides. La nature des nucléotides composant la répétition influence le type de structure pouvant être généré ainsi que la capacité de la cellule à se débarrasser de ces structures. Une séquence d(TC)_n, est capable de former une triple hélice avec la séquence d(GA)_n du brin

complémentaire lorsqu'un des deux brins est sous forme d'ADN simple brin. La modification d'un carbone de la structure de l'adénine, de la guanine ainsi que l'ajout de la protéine *SSB* d'*E. coli* permet d'empêcher la formation de ces structures. La dynamique de fixation des protéines, *SSB*, *scRAD51*, *hRAD51*, *RECA* (fixant l'ADN simple brin (ADNsb) et non l'ADN double brin (ADNdb), a été mesurée sur les séquences d(CA)_n, d(CT)_n, d(GT)_n, d(CA)_n, d(GA)_n, d(AT)_n et d(GC)_n. Les d(AT)_n et d(GC)_n sont capables de former des « épingles » parfaites et sont très peu fixés par les protéines testées, suggérant que ces structures existent. Il est intéressant d'observer que 25% des d(AT)_n sont néanmoins fixés par les *SSB*, montrant que ces structures peuvent être éliminées par la cellule, impliquant que les séquences sont de nouveau accessibles aux protéines se fixant à l'ADN simple brin. Les séquences d(GA)_n pouvant former des triples hélices sont aussi incapables de lier des protéines de fixation à l'ADNsb^{30,31}.

2.2 Structures secondaires des répétitions de trinuécléotides

Plusieurs types de répétitions de trinuécléotides sont capables de former des structures secondaires et trois sont particulièrement intéressantes. Les répétitions CAG/CTG, CCG/CGG et GAA/TTC sont liés à un grand nombre de maladies neurodégénératives avec plus de 12 maladies pour les seules répétitions CAG/CTG. Les structures formées par ces répétitions sont des épingles imparfaites pouvant prendre deux configurations différentes en fonction du nombre pair ou impair de répétitions au sein de la séquence, des triples hélices ou encore des G-quadruplex.

nucléotides si le nombre de répétitions est impair (figure 3).^{32,33}. Un minimum de 4 répétitions est nécessaire pour former une structure en épingle³⁴.

2.2.2 Triples hélices

Les répétitions GAA/TTC peuvent former un type de structure secondaire en triple hélice. Celle-ci consiste en la formation de liaisons Hoogsteen en plus des liaisons Watson-Crick classiques, pour former une structure où le brin GAA riche en purines s'intercale au sein d'une épingle formée par le brin complémentaire TTC riche en pyrimidines (figure 3)^{34,35}.

2.2.3 G-quadruplex

Le brin riche en G des répétitions CGG est capable de former des G-quadruplex, structures plus complexes et très stables où des plateaux de G4 (ou G-quartet) sont connectés cycliquement par des liaisons hydrogènes. La présence de cations monovalents (Na⁺, K⁺) est nécessaire à la formation de ces structures³⁶. Il a été montré que ces structures jouent un rôle dans les processus de transcription, réplication, stabilité du génome et dans la régulation épigénétique³⁷. On les pensait jusqu'à récemment être le point d'ouverture des origines de réplifications chez l'humain. Il est fascinant de constater qu'il a récemment été montré que les origines de réplifications sont en réalité liées aux promoteurs et non au G-quadruplex³⁸. On retrouve des G-quadruplex au sein des télomères humains (la séquence télomérique est composée de répétition hexanucléotidiques (T)2A(G)3). Une dérégulation de la maintenance des G4 au sein des télomères peut entraîner certains types de cancers. Nécessaire au recrutement de la télomérase chez l'humain, la préservation de l'intégrité des télomères passe par le dépliement des G4 via l'activité de l'hélicase RTEL1, dont la mutation empêche le déroulement des structures G4 et déstabilise les télomères,³⁷. L'hélicase Pif1 joue aussi un rôle dans le déroulement des

télomères, son inactivation entraîne une forte augmentation de l'instabilité du locus minisatellite humain CEB1, capable de former des structures G4 *in vitro* ³⁹.

2.2.4 Les répétitions de trinuécléotides non structurées

On retrouve une corrélation entre la capacité d'une séquence répétée à former des structures secondaires et les expansions associées au développement d'une maladie. Les trinuécléotides ne formant pas de structures secondaires *in vitro*, tel que les répétitions de trinuécléotides GAT/ATC, ne sont associées à aucune maladie humaine pour l'instant. La formation de structures secondaires sur les répétitions de trinuécléotides restantes, TAG/CTA, ATA/TAT et TTG/CAA semble peu probable. Les dernières répétitions de trinuécléotides GAG/CTC et GTG/CAC, sont riches en guanine qui sont capables de stabiliser des structures en épingles à cheveux ou des quadruplex, voire des triples hélices, laissant supposer que si elles subissaient un phénomène d'expansion, elles pourraient causer des maladies ³². La présence de ces structures secondaires à une implication certaine dans les processus d'expansions observés. Différents modèles ont été proposés afin de lier les deux événements et d'en comprendre les mécanismes.

2.3 Glissement lors de la réplication (SSM)

Historiquement, le premier modèle proposé a été celui du « *Slipped-strand mispairing* » ou SSM. Le SSM implique une dénaturation locale et un mauvais réappariement (*mismatch*) des deux brins d'ADN complémentaires suite à un glissement de l'un des deux brins le long de la séquence répétée. Ce mécanisme correspond à l'appariement de l'extrémité 3' libre du brin néo-synthétisé avec la séquence complémentaire répétée en aval ou en amont de son site d'hybridation original. Si le brin néo-synthétisé s'hybride en aval, une boucle d'ADN se

formera et conduira à une expansion de la répétition par insertion d'un motif de répétitions. Si le brin néo-synthétisé s'hybride en amont, la boucle se forme sur le brin matrice et conduit à une contraction par délétion d'un motif de répétitions (figure 4). La séquence répétée étant son propre substrat d'instabilité, plus elle contiendra de motifs répétés, plus la probabilité de modification de la séquence sera élevée ⁴⁰. La fréquence d'élongation d'une répétition causée par le SSM diminue cependant avec l'augmentation de taille du motif de répétition ⁴¹. La mutation des gènes du système de réparation des mésappariements (*mismatch repair*) (dont les mécanismes liés à l'instabilité seront développés plus loin) *PMS1*, *MSH2* ou *MLH1* de *S. cerevisiae* augmente l'instabilité des microsatellites d'un facteur 100-700. Le mécanisme du SSM semble être un moteur d'instabilité pouvant ajouter 1-2 motifs répétés à une répétition de dinucléotides ⁴². En revanche, l'activité de correction des erreurs des ADN polymérases affecte peu la stabilité des séquences répétées en tandem ^{42,43}.

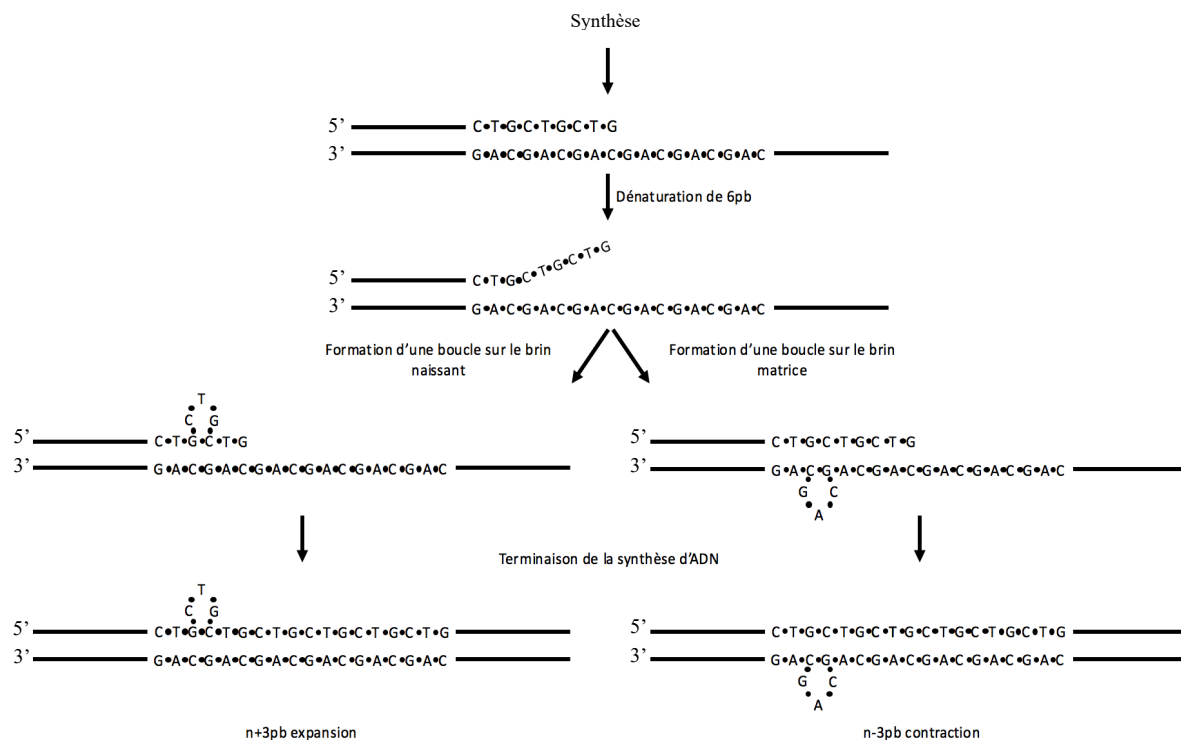


Figure 4 : Représentation schématique du glissement à la réplication (*Slipped-Strand Mispairing* (SSM)). Par mésappariement du brin naissant et renaturation accompagnée de mésappariement sur le brin naissant ou matrice, une boucle d'ADN peut se former. La formation d'une boucle sur le brin naissant conduit à une expansion du nombre de répétitions tandis que la formation d'une boucle sur le brin matrice conduit à une contraction.

2.3.1 Glissement lors de la réplication et pureté des répétitions

Nous avons vu que la stabilité d'une séquence répétée est directement liée à sa longueur. Une séquence répétée ne présentant aucune interruption de son motif de répétition est considérée comme parfaite. La substitution d'une base ou d'un motif entier le long de la séquence est considérée comme une interruption et influence de façon différente la stabilité de la séquence en fonction de sa position en 5', 3' ou central. Nous avons vu dans la partie précédente qu'un glissement lors de la réplication peut conduire à la formation de structures secondaires pouvant mener à des contractions ou des expansions. Une impureté le long de cette séquence peut influencer le glissement du brin d'ADN et la formation de structures secondaires ou affecter la stabilité des structures formées. Une séquence répétée doit porter en moyenne 35 à 40 triplets pour subir les phénomènes d'expansion et voir apparaître les différents symptômes associés. Il existe naturellement 1-3 interruptions CAT/ATG au sein des séquences (CAG/CTG)_n responsable de l'ataxie spinocérébelleuse de type 1 et 1 à 3 interruptions AGG/CCT au sein des séquences (CGG/CCG)_n responsable du syndrome de l'X fragile, ces interruptions empêchent la formation d'un segment d'ADN répété trop long qui pourrait déclencher les expansions et la maladie. On peut prendre l'exemple du plus long segment d'ADN répété SCA1 stablement transmis, qui portait 44 répétitions avec une configuration 12*1*12*1*14 (les * représentent une interruption). Bien que supérieur à la limite de stabilité, les interruptions empêchent la présence d'une séquence CAG trop longue ⁴⁴⁻⁴⁶. Une séquence (CAG/CTG)₃₉ parfaite formera des structures secondaires sur environ 40% de sa séquence tandis qu'une séquence portant des interruptions dans la configuration 9*9*9*9 ne forme des structures que sur environ 23% de sa séquence ⁴⁶. Une séquence (CAG/CTG)₁₇ parfaite forme des structures sur 27% de sa séquence ce qui reste supérieur à une plus grande séquence portant des interruptions. A l'inverse, la perte de ces interruptions doit augmenter l'instabilité des séquences et augmenter les risques d'expansion ⁴⁶. Une interruption placée en 5' de la séquence répétée, stabilise celle-ci d'un

facteur 90- fois plus important par rapport à une séquence répétée parfaite. Placé en 3' de la séquence répétée la stabilisation n'est que d'un facteur 2,4 et monte jusqu'à 16 quand l'interruption est centrale à la séquence. Ces interruptions sont conservées après la réplication de la séquence ⁴⁷. L'effet stabilisateur des interruptions est indépendant de la machinerie de réparation des mésappariements. L'inactivation de *MSH2*, *MSH3* ou *PMS1* entraîne une augmentation des événements de petites insertions ou délétions d'une ou deux paires de bases mais aucun impact sur la stabilisation des séquences n'est observé ⁴⁸.

2.4 Réplication et instabilité

Des expansions de répétitions CTG sont observées pré- et post-méiotiquement dans les lignées germinales de patients atteints de DM1 ⁴⁹. Cette observation démontre que la réplication de l'ADN et les différents mécanismes de réparations liés ou non à la réplication, peuvent entraîner des erreurs augmentant l'instabilité des microsatellites. Cette observation est confirmée par les nombreuses analyses des processus de réplication et des protéines impliquées.

2.4.1 Fourche de réplication et instabilité

La fourche de réplication est composée d'un ensemble de protéines nécessaires à la reproduction fidèle de l'information portées par les deux brins. Trois polymérases sont nécessaires à la réplication de l'ADN eucaryote, les polymérases α , δ et ϵ (figure 5). La polymérase α possède deux sous-unités primases, nécessaires pour l'amorçage et le départ de la réplication, elle possède une faible processivité et ne possède pas d'activité exonucléase 3'-5', ce qui l'empêche de corriger d'éventuelles erreurs de polymérisation. La polymérase δ , responsable de l'élongation des fragments d'Okazaki, est faiblement processive en l'absence du complexe tripartite *PCNA* et possède une activité exonucléase 3'-5'. Elle joue un rôle dans plusieurs mécanismes de réparation que nous ne détaillerons pas. La polymérase ϵ est

hautement processive même en l'absence de *PCNA* et possède une activité exonucléase 3'-5'^{50,51}. La mutation des gènes de la réplication, *PCNA*, *RFC* ou *POL32* (figure 5) augmente l'instabilité des microsatellites. *PCNA* est une molécule homotrimérique en forme d'anneau (*clamp*), chargé sur l'ADN par *RFC*, qui reste fixée jusqu'à être déchargée. L'interaction des protéines de réplication avec *PCNA* augmente leur durée d'interaction avec l'ADN. *RFC*, forme un complexe avec *PCNA* et le charge sur l'ADN (*clamp loader*) amorcé par *Polα*. *POL32* est une sous-unité de la polymérase δ et joue un rôle dans son interaction avec *PCNA*, une mutation de *POL32* diminuant la processivité de *Polδ*^{52,53}. *RAD27* (*FEN-1* chez l'humain) est une endonucléase à *flap* 5' possédant une activité exonucléase 5'-3' et appartenant à la superfamille *RAD2* (figure 5). Elle est nécessaire à la maturation des fragments d'okazaki et intervient dans le « *base excision repair* ». Une absence de *RAD27* entraîne chez la levure (et chez *Homo sapiens*) expansions et contractions des séquences répétées en tandem. Une seule copie du gène est nécessaire pour garder 70 répétitions CAG/CTG stables et de faibles expansions sont observées de 70 à 130 répétitions. Au-delà de 155 répétitions, la délétion d'une seule copie du gène suffit à entraîner des expansions importantes, montrant que *RAD27* est critique pour préserver la stabilité des longues répétitions^{54,55}. Il est montré que les structures secondaires en épingles formées sur le *flap*-5' peuvent inhiber l'activité de *RAD27*, qui peut cliver avec la même efficacité qu'une séquence contrôle des répétitions (CTG)₅₋₁₀ tout en étant incapable de cliver une séquence (CTG)₂₀⁵⁶. L'interaction de *RAD27* avec *PCNA* augmente son efficacité d'un facteur 30 et permet de résoudre plus efficacement des structures secondaires en épingle de petites tailles dont la base est composée de moins de 18 nucléotides⁵⁶. La létalité de la double mutation *rad27Δ rad52Δ* (*RAD52* est le gène central de la recombinaison homologue) indiquent qu'un *flap*-5' non clivé peut entraîner des cassures double-brin⁵⁷ et générer encore plus d'instabilité. L'inactivation de n'importe quel gène de la réparation par recombinaison homologue peut de même entraîner des cassures double-brin et augmenter l'instabilité des

microsatellites. La ligase *CDC9* termine la réplication en liant les fragments d'Okazaki clivés par Rad27 (figure 5). Sa surexpression entraîne une augmentation de l'instabilité des microsatellites sur le brin retardé lors de la réplication. Cette augmentation d'instabilité, indépendante de la fonction ligase de *CDC9*, est liée à son interaction avec *PCNA*. *RAD27* et *CDC9* forment chacun un complexe exclusif avec *PCNA*, un déséquilibre de la quantité de *CDC9* empêche la formation de complexes *RAD27-PCNA* et augmente l'instabilité des triplets

58.

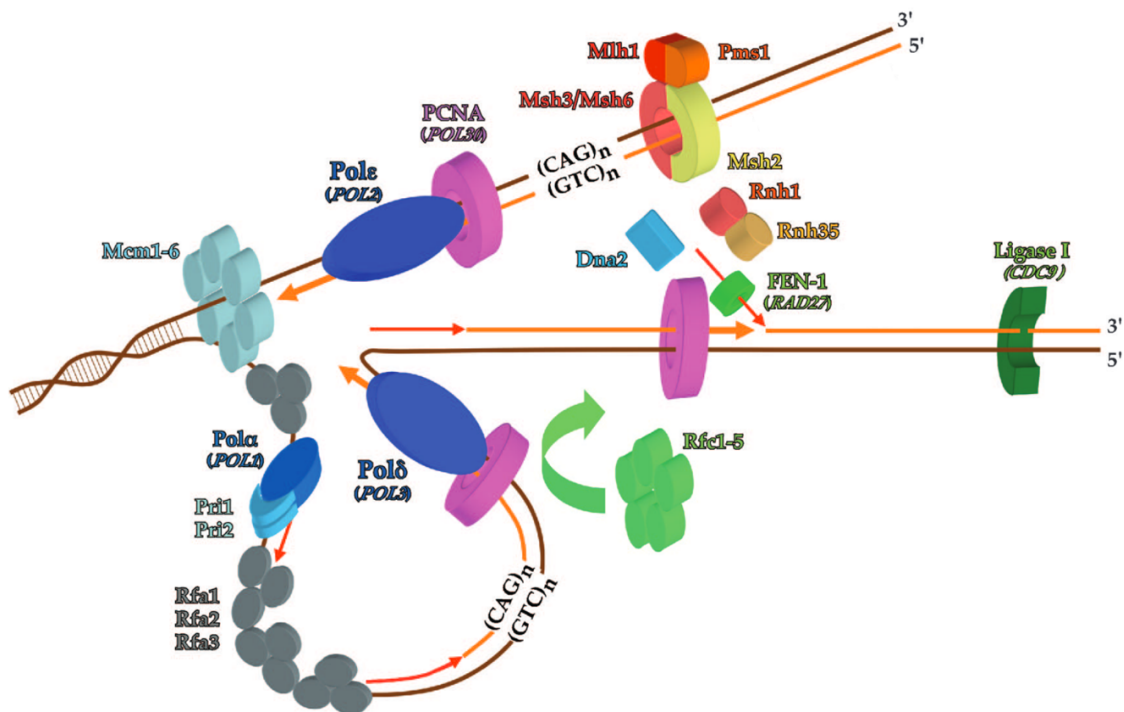


Figure 5 : Représentation de la fourche de réplication au niveau d'une répétition CAG/CTG. Le complexe MCM, hétérohexamère, ouvre l'ADN pour permettre la progression de la synthèse d'ADN sur le brin continu et discontinu. La polymérase α amorce l'ADN des deux brins avec une amorce ARN générant une extrémité 3' libre permettant aux polymérases ϵ et δ de synthétiser respectivement l'ADN du brin continu et discontinu. Les polymérases sont associées au facteur de processivité PCNA chargé sur l'ADN par le complexe RFC. Les fragments d'Okazaki du brin discontinu doivent être maturés et l'amorce ARN éliminée par l'action de la flap-5' endonucléase Rad27, l'exonucléase Dna2 et le complexe Rnase H. Les fragments d'Okazaki sont finalement liés entre eux par la ligase Cdc9. Les protéines du MMR (*mismatch repair*) sondent l'ADN et corrigent les éventuelles erreurs de réplication. Richard, Kerrest and Dujon, 2008.

Il est possible d'étudier les effets de l'altération de la fourche de réplication par traitements avec différentes drogues tel que l'aphidicoline (inhibe l'activité des polymérase α , δ et ϵ), la mimosine (inhibe l'initiation de la réplication) et l'emetine (inhibe l'élongation du brin tardif). L'inhibition totale de l'initiation de la fourche de réplication, par traitement à la mimosine, n'entraîne pas d'augmentation de l'instabilité des répétitions comparé à la condition sans traitement. L'instabilité des répétitions après inhibition des polymérase α , ϵ et δ , responsables de la polymérisation du brin précoce et du brin tardif, est augmentée avec des expansions trois fois plus importantes dans les souches traitées par rapport aux souches contrôles. L'émétine entraîne une amplification du nombre de répétitions et contrairement aux autres types de traitements, cause une hétérogénéité de la taille des expansions entre les cellules d'une même population ⁵⁹. L'analyse par gel bidimensionnel d'ADN de la réplication d'une séquence répétée (CGG)_n ou (CCG)_n montre une corrélation positive entre l'intensité du ralentissement de la fourche de réplication et la taille de la séquence répétée. L'effet du ralentissement de la fourche à l'intérieur de répétitions (CTG)_n est moins marqué que les répétitions (CGG)_n probablement à cause de la nature des structures secondaires formées par ces séquences, montrant que les triplets bloquent transitoirement la fourche de réplication ⁵².

Les observations de ces résultats montrent aussi que la nature de la séquence répétée ainsi que son orientation jouent un rôle prépondérant dans sa stabilité.

2.5 Orientation des microsatellites et instabilité

Il est devenu clair dès 1995 que l'orientation des répétitions par rapport au sens de la transcription ou aux fourches de réplication avait un impact sur leur stabilité ⁶⁰. Freudenreich et Zakian montrent que l'insertion d'une séquence (CAG/CTG)₁₃₀ dans les deux orientations (respectivement orientations I et II), n'influence pas la stabilité des séquences de la même

manière. Dans l'orientation I le brin CTG est orienté de façon opposée à la progression de la fourche de réplication, c'est à dire synthétisé sur le brin tardif.

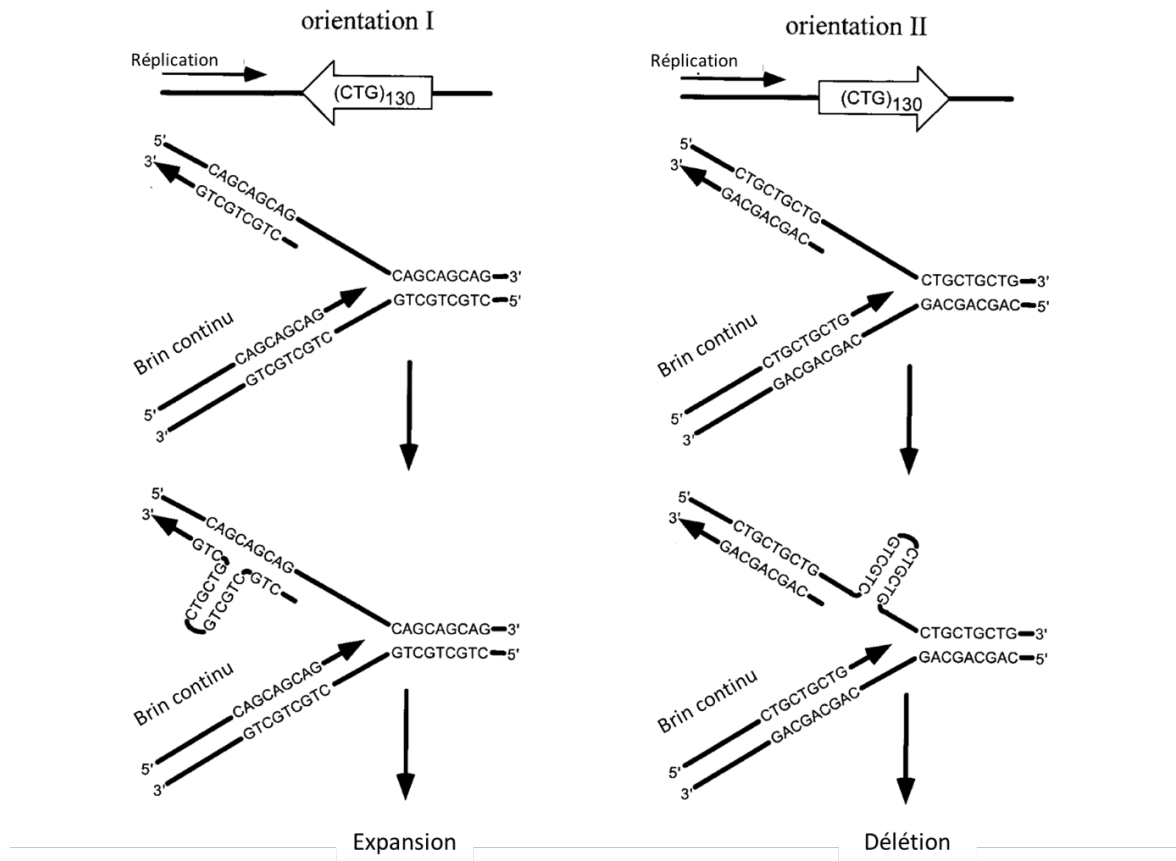


Figure 6 : Représentation schématique de la fourche de réplication passant au sein d'une région CAG/CTG. En orientation I le brin CTG est répliqué sur le brin tardif sur lequel se forme naturellement des séquences d'ADN simple brin au sein des répétitions. Comme vu précédemment dans cette introduction, la formation d'un brin simple au sein d'une séquence répétée peut conduire à la formation de structure secondaire (voir partie sur le glissement lors de la réplication). En orientation II le brin CTG est répliqué sur le brin continu, plus stable. Le brin CAG répliqué sur le brin tardif est moins apte à former des structures secondaires comme vu précédemment dans cette introduction. Tiré de Freudenreich, Zakian, 1996.

En orientation II le brin CTG est orienté dans le même sens que la fourche de réplication et sera synthétisé sur le brin précoce. En orientation II, les $(CTG)_n$ situés sur le brin matrice du brin tardif, sont capables de former des structures secondaires lorsque la séquence devient simple brin.

Nous avons déjà discuté de la plus grande stabilité des structures CTG comparées aux structures CAG, ce qui explique sans doute pourquoi en orientation I très peu d'instabilité est observée (figure 6). Il est intéressant de noter que la transcription des gènes en aval ou en amont des CAG/CTG n'est pas affectée par la présence de la séquence répétée ^{61,62}.

2.6 Mécanismes de réparation des lésions à l'ADN et instabilité

Il est évident que les structures secondaires formées ou les lésions qui se forment suite aux passages de fourches de réplication dans les séquences répétées doivent être réparées. Dans cette partie seront introduits les différents mécanismes de réparations et leurs rôles dans l'instabilité des séquences répétées tandis que les mécanismes moléculaires (qui seront néanmoins brièvement introduit dans cette partie) de ces mêmes mécanismes seront détaillés dans un chapitre consacré.

2.6.1 Effet du MMR sur la stabilité du génome

Le MMR (*Mismatch repair*) est un mécanisme de réparation impliqué dans un ensemble de fonctions métaboliques et retrouvé chez tous les organismes vivants. D'abord mis en évidence dans son rôle contre l'accumulation de mésappariements et la fixation de mutation dans le génome, il promeut la stabilité génomique (sauf à certains locus comme nous le verrons dans cette partie). Ainsi, les principaux substrats du MMR sont les substitutions d'une base et les petites insertions/délétions mais aussi les bases abimées par des dommages oxydatifs ⁶³. La perte du MMR conduit à une hausse du nombre de mutations spontanées et on observe notamment dans des familles portant des mutations dans les gènes *MSH2*, *MLH1*, *PML2* et/ou *PMS1* l'apparition de cancers colorectaux non polypeux héréditaire (HNPCC). Ces mutations s'accompagnent d'une instabilité des microsatellites qui sera détaillée plus tard ⁶⁴. Dans cette

partie seront détaillés les mécanismes du MMR dans la reconnaissance et la réparation des mésappariements (figure 7) ainsi que le rôle du MMR dans la stabilité des microsatellites.

2.6.1.1 Mécanisme du MMR

Le système MMR a été étudié de façon extensive dans les organismes procaryotes, notamment chez *Escherichia coli*. Le MMR de *E. coli* dépend des protéines MutS, MutL et MutH qui sont indispensables à la reconnaissance de la lésion et à l'initiation de la réparation. La première étape de la reconnaissance de la lésion se fait par MutS, qui fonctionne en homodimère. Une fois fixée sur le mésappariement, MutL (qui s'homodimérise de la même façon que MutS), interagit physiquement avec MutS et augmente sa capacité à reconnaître les mésappariements. MutS et MutL possèdent toutes les deux une activité ATPase indispensable au fonctionnement du MMR. MutL joue un rôle fondamental dans l'assemblage et la stabilisation du complexe MMR chez les bactéries. Une perte de l'activité ATPase de MutL conduit à une perte de la formation du complexe. Il a aussi été montré que MutL interagit avec l'ADN polymérase III, nécessaire pour la terminaison de la réparation par le MMR. Finalement MutL permet le recrutement de MutH au niveau du complexe et son activation.

Chez *E. coli* les séquences dGATC sont normalement méthylées sauf brièvement après la réplication sur le brin néosynthétisé. L'absence de méthylation sur le nouveau brin permet à MutH de différencier les deux séquences et d'induire une coupure simple brin au niveau du brin néosynthétisé au sein d'une séquence dGATC non méthylée. Ce *nick*² permet à des exonucléases tel que Exo1 (5'-3' exonucléase) de digérer le fragment portant le

² Cassure simple brin

La perte d'expression de *MSH2* augmente drastiquement l'accumulation de mutations dans le génome, tandis que la perte de *MSH6* augmente sensiblement l'accumulation des mutations par substitution et affecte faiblement les mutations par insertion/délétion. La mutation de *MSH3* entraîne l'accumulation d'insertions et de délétions supérieures à 2 paires de bases avec un biais 6 fois supérieurs pour les délétions⁶⁸. On observe une incidence plus importante de la perte de *MSH2* et *MSH6* par rapport à *MSH3* sur les réarrangements après recombinaison homologue. Ces résultats sont consistants avec la formation d'hétéroduplex contenant des mésappariements base-base et non pas des structures formées par des insertions/délétions, reconnues par MutS β ⁶⁹. La double perte d'expression de *MSH3* et *MSH6* conduit à une perte totale de la réparation par le MMR et présente un phénotype semblable à celui du mutant simple *MSH2*⁶⁸. Grâce à ces observations, nous savons que le complexe MutS α reconnaît les substitutions de bases uniques tandis que les insertions/délétions peuvent être reconnues à la fois par MutS α et MutS β . La cellule préfère néanmoins utiliser MutS β pour la reconnaissance des insertions/délétions (avec un facteur 5- à 10 fois supérieur à celui de MutS α)⁶⁸⁻⁷⁰. Des homologues de MutL sont aussi retrouvés chez *S. cerevisiae* : *MLH1*, *PMS1* (homologue de *PMS2* chez l'humain), *MLH2* (ne joue aucun rôle dans le MMR) et *MLH3*. Mlh1 et Pms1 forment le complexe (MutL α) qui interagit avec les complexes MutS α et MutS β . De la même manière Mlh1 et Mlh3 forment le complexe (MutL γ) qui n'interagit qu'avec MutS β (la perte de *MLH3/MSH6* augmente le nombre d'insertions/délétions). MutL α , via l'activité de Pms1/Pms2 joue le rôle d'endonucléase (dépendant de son activité ATPase) capable d'induire une cassure simple brin.

Chez *Homo sapiens* (et les mammifères), le modèle de réparation par le MMR fait intervenir MutS α / MutS β , MutL α , l'exonucléase 5'-3' *EXO1*, RPA (Protéine de fixation à l'ADN simple brin), RFC (charge PCNA sur l'ADN), le complexe PCNA (complexe tripartite interagissant avec des protéines de la fourche de réplication augmentant leur processivité), la polymérase δ

ainsi que l'ADN ligase. De façon intéressante, le complexe PCNA n'intervient que dans une orientation lors de la réparation. En effet la réparation par le MMR est dépendante d'une cassure simple brin (*nick*) déjà existant en 5' ou en 3' de la lésion. Lors d'une réparation dépendante d'un *nick* en 5' le complexe PCNA n'est pas nécessaire à la réparation et ne fait intervenir que les complexes MutS α / MutS β , MutL α , l'exonucléase 1, RPA, la polymérase δ ainsi que l'ADN ligase ^{65,71}. La réparation dépendante d'un *nick* en 5' de la cassure est donc initiée par la reconnaissance de la lésion par le complexe MutS α / MutS β . Un changement de conformation du complexe par hydrolyse de molécules d'ATP lui permet de stimuler l'exonucléase 1 afin de retirer le fragment portant la lésion. Une fois le fragment digéré par l'exonucléase 1, le fragment d'ADN complémentaire simple brin est recouvert par le complexe RPA protégeant l'ADN simple brin. La fixation de RPA au fragment simple brin inhibe l'activité de l'exonucléase 1 en la retirant de l'ADN ainsi que MutS α . En absence de lésion MutS α /MutS β et MutL α inhibent à leur tour l'activité de l'exonucléase 1. MutL α augmente la capacité de MutS α /MutS β à reconnaître les lésions mais n'intervient pas directement dans la réparation ⁷¹. La surexpression de la protéine p21 qui se fixe fortement sur le complexe PCNA, inhibe la réparation par le MMR lorsqu'elle est dépendante d'un *nick* situé en 3' de la lésion mais pas lorsque le *nick* est situé en 5' de la lésion. Cette observation indiquait que plusieurs mécanismes existaient en parallèle et étaient dépendants de l'orientation à la lésion ⁶⁵. La réparation orientée en 3' de la lésion est, contrairement à celle en 5', dépendante de RFC et de PCNA. Après la reconnaissance de la lésion réalisée par MutS α / MutS β et MutL α . Le complexe RFC et PCNA activent l'activité endonucléase latente de MutL α qui induit un *nick* en aval de la lésion créant une nouvelle extrémité 5' permettant à l'exonucléase 1 d'hydrolyser le fragment portant la lésion. La réparation est finalement terminée par l'ADN polymérase δ et l'ADN ligase I ^{72,73} (Figure 7).

Le MMR est aussi impliqué dans les mécanismes de réparation par recombinaison homologue (voir chapitre 5). Lors de l'appariement de deux séquences, l'homologie n'est pas toujours parfaite et des mésappariements peuvent apparaître, comme c'est le cas pour deux séquences d'allèles différents du même gène. La réparation de ces mésappariements est identique à la voie canonique du MMR si ce n'est qu'elle est couplée à la machinerie de réplication durant la phase S. Un autre exemple de l'importance du MMR dans la réparation par recombinaison homologue est son intervention dans le mécanisme du *Single-Strand Annealing* (SSA), qui répare une cassure double brin par association de séquence homologue de part et d'autre de la cassure. L'appariement des séquences homologues de chaque côté de la cassure génère des extrémités 3' flottantes qui sont spécifiquement ciblées par les protéines de MMR puis clivées par les *Flap*-endonucléases Rad1/Rad10 chez la levure et *ERCC1/XPF* chez l'humain, appartenant au mécanisme du NER ⁷⁴.

2.6.1.2 MMR et instabilité des microsatellites

Un plasmide transformé dans une souche de *E. coli* MMR+, portant une répétition de trinuécléotides CAG/CTG subit des événements de contraction et d'expansion dépendants de l'orientation des triplets à la fourche de réplication, bien que les grandes contractions soient favorisées. La réplication des répétitions « CTG » sur le brin tardif amène à deux fois plus d'évènements de contraction que si elles sont positionnées sur le brin précoce ⁷⁵. Une souche de *E. coli* déficiente dans le MMR (MMR-), subit une augmentation de l'instabilité dans les deux orientations de transcription, causée par l'insertion ou la délétion d'une seule unité de triplet. En contraste avec les souches MMR+ l'orientation à la fourche n'influence pas l'instabilité dans les souches MMR- ⁷⁵.

2.6.1.2.1 MMR et instabilité des microsatellites chez *Saccharomyces cerevisiae*

Nous avons déjà discuté de l'effet de l'orientation des répétitions CAG/CTG sur la stabilité de la séquence. Si la séquence « CTG » est sur le brin matrice du brin tardif, on observe une plus forte instabilité des répétitions CAG/CTG. La perte du MMR dans ces souches augmente l'instabilité des répétitions dans les deux orientations. On observe une augmentation des événements d'expansion et de contraction par l'ajout ou la délétion d'un seul triplet, ce phénomène n'est jamais observé dans une souche sauvage, ou les contractions se produisent à une fréquence bien supérieure aux contractions^{53,62}. C'est un effet déstabilisateur des séquences qui est observé dans la levure lors de la suppression du MMR. La fréquence de contraction des séquences instables reste la même. Ces observations semblent valider le modèle du « SSM » présenté dans un chapitre précédent sur la stabilité des répétitions chez les micro-organismes^{53,76}. Le complexe du MMR formé des protéines Mlh1-Pms1 (Pms2 chez l'Humain) (MutL α) chez la levure et les mammifères, ainsi que le complexe formé des protéines Msh2-Msh6 (MutS α) (substitution de bases et réparation de petites boucles d'ADN) sont impliqués dans le ralentissement des fourches de réplication le long des séquences répétées dans les deux orientations⁷⁷. Une surexpression de Msh2 entraîne une forte augmentation du nombre de contractions dans les deux orientations, contrastant avec l'absence d'effet de la délétion de *MSH2* sur les contractions. La perte d'activité nucléase de Msh2 donne le même phénotype qu'une souche *msh2 Δ* ⁷⁷. La mutation du gène *MSH3* augmente aussi sensiblement l'instabilité des microsatellites, bien qu'à un niveau moindre de celui de la perte d'expression de *MSH2* tandis que la mutation de *MSH6* n'entraîne qu'une faible augmentation de l'instabilité.

2.6.1.2.2 MMR et instabilité des microsatellites chez *Mus musculus*

La souris *Mus musculus* est un modèle très utilisé dans l'étude des mécanismes de stabilité des répétitions et de nombreuses lignées ont été générées afin de pouvoir étudier les maladies liées aux expansions. Dans des souris portant le gène DMPK humain ayant subi une expansion de plus de 300 triplets, une transmission d'expansions à la descendance se produit dans 100% et 62,3% des cas pour les souris mâles et femelles respectivement. La simple perte d'une copie fonctionnelle de *MSH2*, protéine centrale et nécessaire à la formation de complexes du MMR chez les eucaryotes avec *MSH3* et *MSH6*, entraîne une diminution de la transmission de ces expansions (100% à 75% et 62,3% à 37,9% respectivement). La perte complète de copies fonctionnelles de *MSH2* diminue très fortement la transmission d'expansion chez le mâle et mène à une perte complète de ces événements chez la femelle (12,1% et 0% respectivement). Au contraire, des phénomènes de contractions massives sont observés pour les deux sexes, bien que plus important pour les femelles⁷⁸⁻⁸⁰. De façon intéressante, la perte de l'activité ATPase de Msh2, conduit au même phénotype dans les cellules germinales, mais n'induit pas les contractions observées dans les souris *Msh2*^{-/-}⁸¹. Comme chez la levure, cette observation met en évidence l'importance d'un MMR fonctionnel et que la capacité de Msh2p à se fixer à l'ADN ne suffit pas à conserver le phénotype sauvage. Les expansions ou les contractions ne sont pas des événements limités seulement à la réplication puisque l'instabilité est retrouvée dans des tissus somatiques peu ou non répliqués, tels que le cœur ou le cerveau^{78,81,82}. Les complexes *MSH2-MSH3* (MutS β), *MLH1-PMS2* (MutL α) et *MLH1-MLH3* (MutL γ), jouent un rôle prépondérant dans l'instabilité des microsatellites. La perte d'activité de ces protéines conduit à une perte totale des mécanismes d'expansion chez l'humain. Une instabilité des microsatellites (MSI) causée par un MMR déficient, est retrouvée dans 10-15% des cancers colorectaux (CRC)⁸³.

L'importance du MMR sur l'instabilité des répétitions varie donc entre les organismes. Chez la levure ce sont majoritairement des événements de contractions qui se produisent naturellement, ces événements sont peu affectés par le MMR. Cela contraste avec les observations faites chez l'humain ou la souris, où ce sont des expansions qui sont observées avec une fréquence bien supérieure aux contractions. La perte complète de ces expansions en l'absence du MMR montre l'importance de ce mécanisme dans l'instabilité. Aucune étude n'a encore été capable de montrer par quel mécanisme le MMR favorisait les expansions chez l'humain. Il est intéressant de constater que malgré le biais favorisant les contractions chez la levure, elle est longtemps restée l'organisme modèle favorisé dans l'étude de l'instabilité des répétitions. Bien que le modèle humain ne permette pas encore de pouvoir étudier avec la même aisance la biologie des séquences répétées, il est à mon sens nécessaire de développer des systèmes humains afin de pouvoir valider et confirmer les résultats déjà trouvés chez la levure ainsi que de pouvoir étudier plus en détail des mécanismes qui concernent des maladies humaines. Je propose à la fin de cette introduction un système expérimental visant à pouvoir étudier ces mécanismes chez *Homo sapiens*.

Catégorie	Fonction	<i>S. cerevisiae</i>	Mammifère	NHEJ	MMEJ	HR	MMR	BER	NER
Reconnaissance de la lésion	Complexe, fixation aux extrémités de la cassure, en complexe avec les DNA-PKcs	HDF1-HDF2	Ku70-Ku80						
	Polymérase poly-ADP ribose	-	PARP1						
	Complexe, reconnaissance des insertions et substitutions de bases (MutS α) ainsi que des petites insertions ou délétions (MutS α / MutS β)	MutS α (MSH2/MSH6) MutS β (MSH2/MSH3)	MutS α (MSH2/MSH6) MutS β (MSH2/MSH3)						
	Complexe, reconnaissance de l'ADN simple brin et protection de la digestion par des nucléases ; Stimule l'activité catalytique d'enzymes de réparations	RPA	RPA						
	Lient les bases endommagées par les UVs	Rad4 ; Rad14	XPC ; XPA						
Reconnaissance de la lésion /Recombinaison homologue	Complexe, fixation aux extrémités de la cassure, connecte les deux extrémités	Mre11-Rad50-Xrs2 (MRX)	Mre11-Rad50-Nbs1 (MRN)						
Recombinaison homologue	Protéine centre de la recombinaison homologue, permet la recherche d'homologie et l'appariement des séquences homologues	Rad51	Rad51						
	Nécessaire à la formation des nucléo-filaments de <i>Rad51</i> et à l'appariement des séquences homologues	Rad52	Rad52						
Nucléase	Endonucléase, exonucléase 3'-5'	Mre11	Mre11						
	Sous-unité du complexe MRX/N, régule activité nucléase de <i>Mre11</i>	Rad50	Rad50						
	Sous-unité du complexe MRX/N, nécessaire à la localisation nucléaire du complexe	Xrs2	Nbs1						
	Activateur de l'activité nucléase de <i>Mre11</i> , cycle dépendant	Sae2	CtIP						
	Famille <i>Rad2/XPG</i> Flap endonucléase, exonucléase 5'-3'	Exo1	Exo1						
	Endonucléase, exonucléase 5'-3'	-	Artemis						
	Activité nucléase stimulée par PCNA/RFC	Pms1-Mlh1	Pms2-Mlh1						
	Incise la lésion en 5' ; joue un rôle dans la réparation par SSA et MMEJ	Rad1-Rad10	XPF-ERCC1						
	Incise la lésion en 3'	Rad2	XPG						
	Résolvase intervenant dans la résolution des jonctions de Holliday	Yen1	GEN1						
	Résolvase intervenant dans la résolution des jonctions de Holliday	Mus81-Mms4	Mus81-EME1						
	Flap-endonucléase 5' nécessaire pour la maturation des fragments d'Okazaki	Rad27	FEN1						

Catégorie	Fonction	<i>S. cerevisiae</i>	Mammifère	NHEJ	MMEJ	HR	MMR	BER	NER
Nucléase/Hélicase	Exonucléase et endonucléase. Possède une activité hélicase faible	Dna2	Dna2	Red	Green	Green	Red	Red	Red
Hélicase	Hélicase importante de la réparation par recombinaison homologue par réparation ou en méiose	Sgs1	BLM/WRN	Red	Red	Green	Red	Red	Red
	Hélicases 3'-5' et 5'-3' Appartiennent aux complexe TFIIH	Rad25 ; Rad3	XPB ; XPD	Red	Red	Red	Red	Red	Green
Polymérase	Sous unité de Pol δ ; Impliqué dans l'élongation des chromosomes et la majorité des mécanismes de réparations	Pol32 (Pol δ)	Pol3D (Pol δ)	Red	Green	Green	Green	Green	Green
	Synthèse d'ADN lors de la réparation par le NHEJ ou le NER	Pol4 (Pol β)	Pol β	Green	Red	Red	Red	Green	Red
	Synthèse d'ADN lors de la réparation par le NHEJ ou le BER	-	Pol μ	Green	Red	Red	Red	Red	Green
	Synthèse d'ADN lors de la réparation par le NHEJ ou le NER	-	Pol λ	Green	Red	Red	Red	Green	Red
	Impliqué dans l'élongation des chromosomes et la majorité des mécanismes de réparations	Pol2	Pole	Red	Red	Green	Green	Green	Green
Support polymérases	Rôle dans le support de la synthèse d'ADN lors de la réplication et de la réparation ; Stimule l'incision du brin endommagé dans le MMR	PCNA-RFC	PCNA-RFC	Red	Red	Red	Green	Red	Red
Ligase	Sous-unité catalytique de la ligase IV	Dnl4	Lig4	Green	Red	Red	Red	Red	Red
	Support fonctionnel de la Ligase IV	Lif1	Xrcc4	Green	Red	Red	Red	Red	Red
	Support fonctionnel de la Ligase IV	Nej1/Lif2	Xlf/Cernunnos	Green	Red	Red	Red	Red	Red
	Sous unité catalytique de la ligase I	Cdc9	Lig1	Red	Green	Grey	Green	Red	Green
	Sous unité catalytique de la ligase III	-	Lig3	Red	Green	Red	Red	Green	Red
	Support fonctionnel de la ligase III	-	Xrcc1	Red	Green	Grey	Red	Red	Red

Catégorie	Fonction	<i>S. cerevisiae</i>	Mammifère	NHEJ	MMEJ	HR	MMR	BER	NER
Kinase	Protéine kinase recruté au cassure double brin par Ku	–	DNA-PKcs	Verte	Rouge	Rouge	Rouge	Rouge	Rouge
	Protéine recruté aux cassures par le complexe MRX, régule positivement l'activité de MRX	Tel1	ATM	Verte	Verte	Verte	Rouge	Rouge	Rouge
Maintenance télomérique	En complexe avec <i>Rap1</i> et <i>Rif2</i> ; Régule l'addition de télomères ; inhibe la fixation du complexe MRX et <i>Tel1</i> aux télomères	Rif1	–	Noir	Noir	Verte	Noir	Noir	Noir
	En complexe avec <i>Rap1</i> et <i>Rif1</i> ; Régule l'addition de télomères ; inhibe la fixation du complexe MRX et <i>Tel1</i> aux télomères	Rif2	–	Noir	Noir	Verte	Noir	Noir	Noir

Figure 8 : Homologues des protéines de réparation de la levure *Saccharomyces cerevisiae* et d'*Homo sapiens*. Les complexes protéiques sont surlignés en jaune. Une case verte signifie que la protéine ou le complexe d'intérêt joue un rôle dans le mécanisme concerné tandis qu'une case rouge signifie que la protéine ou le complexe d'intérêt ne joue aucun rôle connu dans le mécanisme concerné. Une case grise signifie que le rôle de la protéine ou du complexe étudié n'a pas été étudié dans le mécanisme d'intérêt. Une case noire correspond à une absence totale de la protéine dans le mécanisme d'intérêt (N/A).

84-91

2.6.2 « *Nucleotide Excision Repair* » (NER)

Le *Nucleotide Excision Repair* ou NER est principalement un mécanisme de réparation de lésions causées par les UVs, bien que le NER soit aussi impliqué dans la réparation d'autres lésions chimiques. Les principaux dommages causés par les UVs sont la formation de dimères de pyrimidines (*Cyclobutanes pyrimidines dimers* ou CPDs) et des photo-produits 6-4 (6-4-PP). La perte de fonction du NER est impliquée dans les syndromes de Cockayne et de *Xeroderma pigmentosum* qui se traduisent par une sensibilité extrême aux UVs, un développement fréquent de cancers de la peau, des retards mentaux et de croissance. Dans cette partie seront d'abord présentés les deux mécanismes du NER (Figure 8) puis le rôle du NER dans l'instabilité des régions répétées.

2.6.2.1 Mécanismes du NER

Chez les eucaryotes, le NER fonctionne grâce à deux mécanismes distincts. Le premier, le *Global Genome NER* ou GG-NER peut intervenir à n'importe quel endroit du génome à l'exception des séquences de gènes transcrits, qui sont pris en charge par le deuxième mécanisme du NER, plus rapide, le *Transcription-coupled NER* ou TC-NER. Dans les deux cas les substrats de ces mécanismes de réparation sont des lésions à l'ADN, tel que des dimères de thymines (CPD) ou des photo-produits 6-4 (6-4PP) causées par des UVs ou des agents chimiques induisant des modifications de bases tels que des agents méthylant créant des O⁶-methylguanine. De manière générale le NER excise un fragment de 25 à 30 paires de bases portant la lésion⁹², avant de laisser la place aux polymérase δ ou ϵ et enfin à l'ADN ligase 1 qui terminera la réparation.

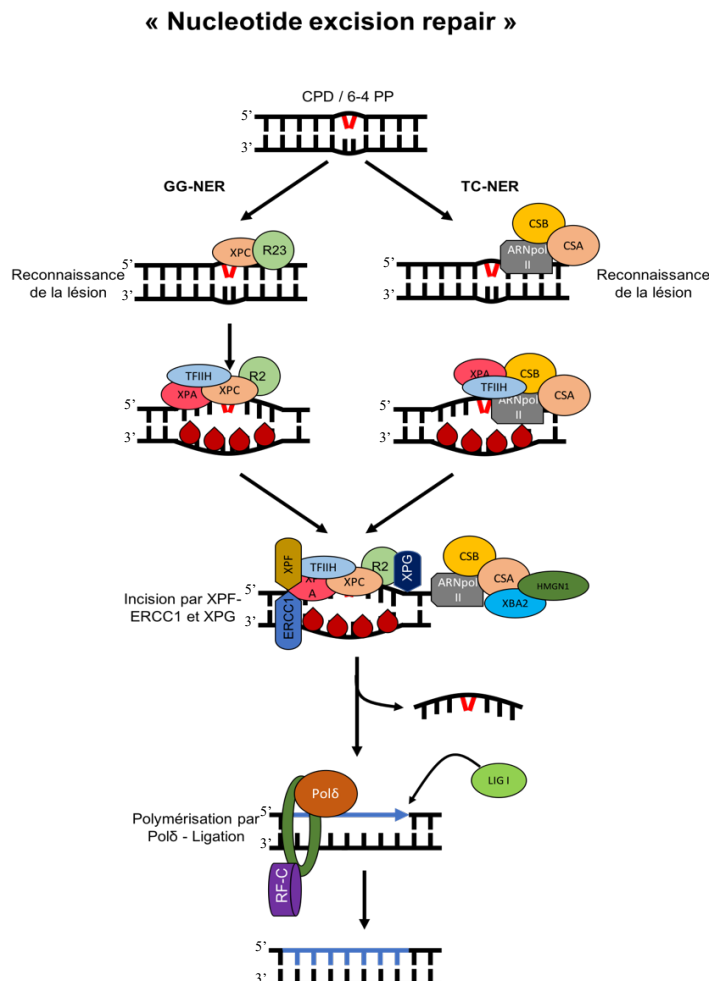


Figure 9 : Schéma des différents mécanismes de réparation par le « Nucleotide excision repair ». Les lésions sont repérées par XPC-RAD23 en l'absence de transcription (GG-NER). Dans les régions en cours de transcription (TC-NER), c'est l'arrêt de l'ARN polymérase II qui déclenche le recrutement de la machinerie de réparation. XPC-RAD23 recrute les facteurs TFIIH (portant les hélicases XPB et XPD), XPA, XPF-ERCC1 et XPG. XPF et XPG vont générer une cassure simple brin de part et d'autre de la lésion permettant d'inciser le fragment portant la lésion. Lors du TC-NER c'est CSB qui va permettre le recrutement de TFIIH, XPA (via XBA2) et RPA. Le reste de la réparation au niveau de la bulle de transcription se déroule comme dans le GG-NER. Une fois le fragment excisé, l'ADN polymérase δ qui synthétise le fragment manquant et l'ADN ligase I qui vient terminer la réparation.

Chez l'humain, le GG-NER fait intervenir plus de 30 protéines différentes, comprenant les protéines *XPA* à *XPG* dont la perte de fonction entraînent le développement du syndrome de *Xeroderma pigmentosum*, syndrome de développement de tumeur de la peau, retard de développement et de croissance. C'est le facteur d'activation hétérodimérique *XPC-RAD23B* après reconnaissance d'une lésion, qui déclenche la réparation, en formant le NER-C1 (NER-Complexe 1) lors de sa fixation à la lésion. *XPC* se fixe de façon non-spécifique à l'ADN et scanne les différentes régions à la recherche de lésions⁹³. Le NER-C1 est ensuite fixé par *TFIIH* (Transcription Factor II H). Composé entre autres des protéines *XPB* et *XPD*, ayant une activité hélicase (respectivement 3'-5' et 5'-3') et ATPase nécessaires à la réparation⁹⁴, sa fixation sur *XPC-RAD23B* forme le NER-C2 indépendamment de la présence d'ATP. La fixation des facteurs suivants ainsi que l'activation du système est néanmoins dépendante de la présence d'ATP. Les protéines *XPA*, *XPG* et *RPA* se fixent ensuite sur le NER-C2. L'addition de *XPA*, qui stabilise le complexe général, forme le NER-C3 puis le NER-C4 lors de l'addition des *RPA* pour finalement former le NER-C5 avec l'endonucléase *XPG*. Finalement lors de la fixation de la deuxième endonucléase *XPF-ERCC1* le NER-C6 pré-incision est formé. Les endonucléases *XPG* et *XPF-ERCC1* vont induire une cassure simple brin autour de la lésion, en 3' et 5' respectivement, avant de l'exciser. Tandis que les *RPA* protègent le simple brin, l'ADN polymérase δ ou ϵ viennent synthétiser la séquence manquante (en présence de *PCNA* et *RFC*) et c'est finalement l'ADN ligase *XRCC1* qui terminera la réparation⁹⁵ (Figure 8). Ce système dépend cependant des *UV-DDB* (*UV-Damaged DNA binding protein*) qui possèdent une forte affinité pour des lésions non déformantes et empaquetées au sein de la chromatine, afin de reconnaître les CPDs⁹⁶. Les CPDs déforment peu la structure de l'ADN et échappent plus efficacement aux systèmes de surveillance du génome face aux dommages. Reconnaissance et réparation semblent couplés à la transcription des gènes avec un enrichissement de la réparation au site d'initiation de la transcription (TSS) jusque dans la séquence du gène. Cette réparation

peut durer 24 à 48h après l'initiation de la réparation contre 4h pour les 6-4PP. Au contraire, le brin complémentaire au brin transcrit (brin non-transcrit) ne présente aucune activité de réparation après 24h et ne montre de manière générale qu'un faible signal de détection de la réparation (confirmant que la réparation des CPDs est majoritairement couplée à la transcription). Les 6-4PPs, qui déforment bien plus la structure de l'ADN, sont rapidement reconnus par les facteurs du GG-NER et montrent un schéma de réparation différent, comparé à celui des CPDs, avec un signal de réparation similaire entre le brin transcrit et le brin non-transcrit ainsi qu'une cinétique de réparation bien plus rapide. Bien que la réparation montre un niveau d'activité similaire entre le brin transcrit et le brin non-transcrit, on observe malgré tout un enrichissement légèrement plus important du signal de réparation sur le brin transcrit dû à l'état chromatinien ouvert autour des TSS ainsi que par l'action synergique du GG-NER et TC-NER ⁹⁷. La cellule utilise donc le deuxième mécanisme de réparation du NER couplé à la transcription afin de réparer certaines lésions ou afin de rapidement redémarrer la transcription de gènes portant une lésion (figure 9).

L'activation du TC-NER est dépendante de l'ARN polymérase II, dont l'arrêt au niveau d'une lésion lors de la transcription déclenche la cascade de réparation. La présence de ce système de reconnaissance durant la transcription élimine la nécessité d'avoir des facteurs, tels que *XPC* ou *UV-DDB*, qui reconnaissent la lésion. La protéine *CSB*, une hélicase ATP dépendante appartenant à la famille SNF2, interagit faiblement avec l'ARN polymérase II lors de la transcription. Si la transcription est stoppée, *CSB* (aussi appelé *ERCC6*) interagit plus fortement avec l'ARN Pol II et permet le recrutement de la protéine *CSA*, complexe intégrant une E3 ubiquitine ligase. C'est via *XPG* que *CSB* est activé, indiquant que *XPG* joue aussi un rôle dans la reconnaissance des lésions par le TC-NER ⁹⁸. *CSA* va à son tour permettre de recruter les facteurs *XAB2* (*XPA binding protein 2*) et *HMGNI* (*High Mobility Group Nucleosome binding domain containing protein 1*). Comme *XPC* n'est pas requis pour la réparation par le TC-NER,

c'est *CSB* qui servira d'antenne permettant le recrutement de *TFIIH* à la lésion. *TFIIH* active l'activité cryptique nucléase de l'ARN polymérase II, lui permettant de digérer le fragment naissant d'ARN et de reculer le long du segment d'ADN en cours de réplication (Figure 8). Ce mécanisme permet de libérer le site de lésion et le rend accessible à la machine de réparation classique du GG-NER. Les facteurs remodelant la chromatine recrutée par *CSA* (*HMGNI*, *p300*, *CAF-1*) ainsi que *TFIIS* (dont le rôle va être de cliver l'ARN dans le site actif de l'ARN polymérase afin de recadrer l'ARN naissant), permettent de défaire les nucléosomes s'étant reformés en aval de l'ARN polymérase II (*HMGNI*), d'ouvrir la chromatine (*p300*) et finalement de reformer les nucléosomes après la réparation (*CAF1*)⁹⁹.

Chez la levure, les protéines impliquées dans le NER sont connues et sont réparties en deux catégories. La première regroupe les gènes *RAD1*, *RAD2*, *RAD3*, *RAD4*, *RAD10*, *RAD14*, et *RAD25* tandis que le deuxième groupe comporte les gènes *RAD7*, *RAD16*, *RAD23* et *MMS19*. Des mutations dans le premier groupe entraînent une forte sensibilité aux UVs et aux agents provoquant des lésions aux bases de l'ADN, tandis que des mutations dans le second groupe n'entraînent qu'une sensibilité partielle.

A l'exception des gènes *RAD7*, *RAD16* et *MMS19*, les homologues de tous les gènes cités sont connus chez l'humain. *RAD4* est néanmoins plus important que son homologue humain *XPC*, dont l'activité est nécessaire à la fois pour le GG-NER et le TC-NER tandis que chez *Homo sapiens*, l'activité de *XPC* n'est nécessaire que dans la réparation par GG-NER¹⁰⁰. La lésion est reconnue et fixée par *RAD14/XPA*, *RAD4/XPC-RAD23*, *RPA*, *TFIIH* (*RAD3/XPD*, *RAD25/XPB*), *RAD1/XPF*, *RD10/ERCC1*, *RAD2/XPG*, *RAD26/CSB*, *RAD28/CSA*. La cassure est reconnue par Rad4-Rad23 en association avec Rad7-Rad16 qui augmente l'affinité de Rad4 pour la lésion. Finalement Rad23 permet le recrutement de Rad14-Rad1-Rad10 et de Rad2-*TFIIH* en parallèle⁸⁹.

2.6.2.2 NER et instabilité des microsatellites

Comme tout mécanisme de réparation impliquant de la synthèse d'ADN *de novo*, ou sensible à la formation de structure secondaire, le NER peut conduire à une perturbation de la stabilité des régions répétées. L'étude d'une séquence à répétitions de trinuécléotides CAG/CTG insérée au sein d'un intron d'un minigène (*hypoxanthine guanine phosphoribosyl transferase* (HPRT)) transcrit chez l'humain, montre une augmentation du nombre de contractions de la séquence avec le temps. La perte d'expression du gène *XPA* diminue le nombre de contractions comparé à une cellule normale. Au contraire, la perte d'expression de *XPC* n'entraîne pas de modification de la stabilité des répétitions. On observe le même effet lors de la perte d'expression de *MSH2-MSH3* (MutS α) (voir sous chapitre sur le MMR). Chez des souris modèles de la maladie de Huntington (expansion de CAG au sein du gène de l'Huntingtine), la perte d'expression partielle ou totale de *XPC* n'entraîne pas non plus de changement de stabilité de la séquence répétée^{101,102}. Comme il a été discuté dans la partie concernant des mécanismes du NER, *XPC* est indispensable (chez l'humain) au GG-NER mais pas au TC-NER tandis que *XPA* est indispensable aux deux mécanismes. Cette observation confirme que c'est la réparation d'une séquence répétée, couplée à sa transcription, qui influence sa stabilité. La double inhibition de *XPA* et de *MSH2* ne modifie pas la fréquence de contraction comparée à l'inhibition simple de chaque gène. Ce résultat indique une interaction du NER et du MMR dans la réparation des lésions durant la transcription, jouant en parallèle sur l'instabilité des régions répétées transcrites¹⁰². La perte d'expression de *CSB*, protéine jouant un rôle fondamental dans la réparation par TC-NER (voir sous chapitre précédent), entraîne une diminution de la transmission d'expansion chez des femelles souris âgées (absente chez de plus jeunes souris) et une augmentation du nombre d'allèles contractés. Chez les souris mâles, la transmission n'est pas affectée et ce sont majoritairement des expansions qui sont observées. Il est intéressant de constater que dans certains tissus (foie, reins, rates, testicules) la perte

d'expression de *CSB* diminue le nombre d'expansions somatiques chez les souris mâles. Cette augmentation de la stabilité n'est pas due à la diminution de l'expression des gènes qui entraîne la perte du TC-NER ¹⁰³. De même, la perte d'expression de *XPG*, *TFIIIS* ou *ERCC1* (en complexe avec *XPF*), entraîne une baisse des événements de contractions liés à la transcription des répétitions de trinuécléotides CAG/CTG dans des cellules humaines. Comme pour le MMR on constate une influence de l'âge et du sexe des souris sur les mécanismes de réparation et sur la stabilité des séquences répétées. L'évolution des systèmes de réparation avec l'âge n'est pas si surprenante, quand on sait aujourd'hui les effets qu'entraîne le vieillissement sur l'ensemble des processus métaboliques (y compris les systèmes de réparation) par l'accumulation de mutations délétères. De nombreuses théories existantes essayent d'expliquer pourquoi les organismes vieillissent et meurent ¹⁰⁴. Ce sujet, bien qu'intéressant, nous éloigne trop du sujet de ma thèse et ne sera pas développé.

La dynamique d'induction des cassures simple brin par *XPF-ERCC1* et *XPG*, en 5' et 3' respectivement de la lésion a été étudiée. Les résultats montrent que *XPF-ERCC1* en complexe induisent une cassure avant *XPG*, le *flap* 5' simple brin généré est potentiellement capable de se refermer sur lui-même et de former une structure secondaire en épingle à cheveux si *XPG* n'induit pas de cassure à son tour. L'activité de *XPF* est cruciale pour la terminaison de la réparation par le NER. L'inactivation de *XPF* ne permet pas, par exemple, la co-localisation de *PCNA* aux sites de dommages. La présence de *XPG*, mais pas son activité catalytique, est suffisante pour déclencher une réparation partielle ^{105,106}. Il est cependant difficile d'imaginer que ce mécanisme soit responsable des grandes expansions observées chez l'humain ou la souris quand le *flap* 5' généré fait de 10 à 20 nucléotides. Pour justifier le TC-NER comme moteur des expansions, il faudrait que l'activité de *XPG* soit inhibée et/ou que le nombre d'évènements d'incision générant des *flaps* 5' simple brin, soit très important. Des arrêts fréquents de l'ARN polymérase II dans des répétitions CAG/CTG ou CGG/CCG montre que

les répétitions entraînent naturellement une fréquence élevée de pauses de transcription ¹⁰⁷.

Il est intéressant d'imaginer la dynamique entre MMR et NER au vu de ces résultats. Si la génération d'une cassure simple brin peut conduire à la formation d'une structure secondaire en épingle à cheveux, alors un défaut de réparation par le NER peut devenir substrat du MMR. La réciproque est vraie puisque la présence des facteurs de réparations du MMR tels que MutS α /MutS β , RFC ou PCNA, peuvent alors qu'ils tentent de réparer un mésappariement, bloquer l'ARN polymérase II et déclencher le TC-NER. Des travaux cités dans cette partie montrent d'ailleurs la communication entre ces deux systèmes de réparation ¹⁰².

2.6.3 *Base Excision Repair (BER)*

Le mécanisme du *Base Excision Repair* ou BER, est le principal mécanisme de réparation des dommages oxydatifs causés à l'ADN. Les *reactive oxygen species* (ROS) sont des produits du métabolisme mitochondrial, incluant les superoxydes, peroxyde d'hydrogène, et les radicaux hydroxyles, qui génèrent la majorité des dommages à l'ADN avec plus de 100 bases modifiées répertoriées dans différents organismes modèles ¹⁰⁸. Les guanines sont les bases les plus sensibles à l'oxydation et forment notamment des 8-oxo-guanines (8-oxo-G). On retrouve à titre d'exemple pour les autres bases des 8-oxo-adénines, les cytosines transformées en uracile ou la dihydro 5-6 thymine (DHT). La modification d'une base entraîne généralement la modification de ses propriétés d'interaction avec les autres bases. Ces modifications peuvent alors entraîner la mutation par transition ou transversion de la base modifiée après répliation du génome. Ces mutations peuvent conduire aux développements de maladies telles que des cancers. On retrouve aussi, chez des patients atteints de maladies neurodégénératives telles que les maladies d'Alzheimer ou de Huntington, une altération du BER qui se traduit par une augmentation du nombre de 8-oxo-guanines. Tout comme le NER, le BER est constitué de deux mécanismes qui seront détaillés plus loin (figure 10). Le rôle du BER dans l'instabilité des

microsatellites sera ensuite développé dans la deuxième partie de ce sous-chapitre.

2.6.3.1 Mécanismes du BER

L'ajout d'un groupe oxo (un atome d'oxygène doublement lié à un atome de carbone) en C8 de la molécule et d'un hydrogène en N7 donne la forme oxydée la plus courante de la guanine et porte le nom de 8-oxoG. Cette forme très stable ne perturbe pas la réplication (à l'exception de l'ADN polymérase I). Cette guanine modifiée a extensivement été étudiée en raison de son implication dans le développement de nombreuses maladies (voir l'introduction de cette sous-partie). Il a été montré que les différentes polymérases de réparation ou jouant un rôle dans la réplication (Polymérases δ , κ , β , λ , et γ) ajoutent un dATP ou un dCTP en face d'une 8-oxoG. En effet la modification de la structure du 8-oxoG lui permet d'interagir via des liaisons de Hoogsteen et de former un mésappariement avec une cytosine ou une adénine en fonction de la conformation en *anti* ou en *syn* de la molécule¹⁰⁹. La stabilité de l'association 8-oxoG avec une adénine permet de muter une paire de bases C:G \rightarrow A:T par transversion après un cycle de réplication. La forme nucléotidique de la guanine est aussi capable d'être modifiée *via* des lésions oxydatives (formant une 8-oxoGTP) permettant d'augmenter le nombre d'incorporations de 8-oxoG dans le génome et d'avoir des transversions A:T \rightarrow C:G.

« Base Excision Repair »

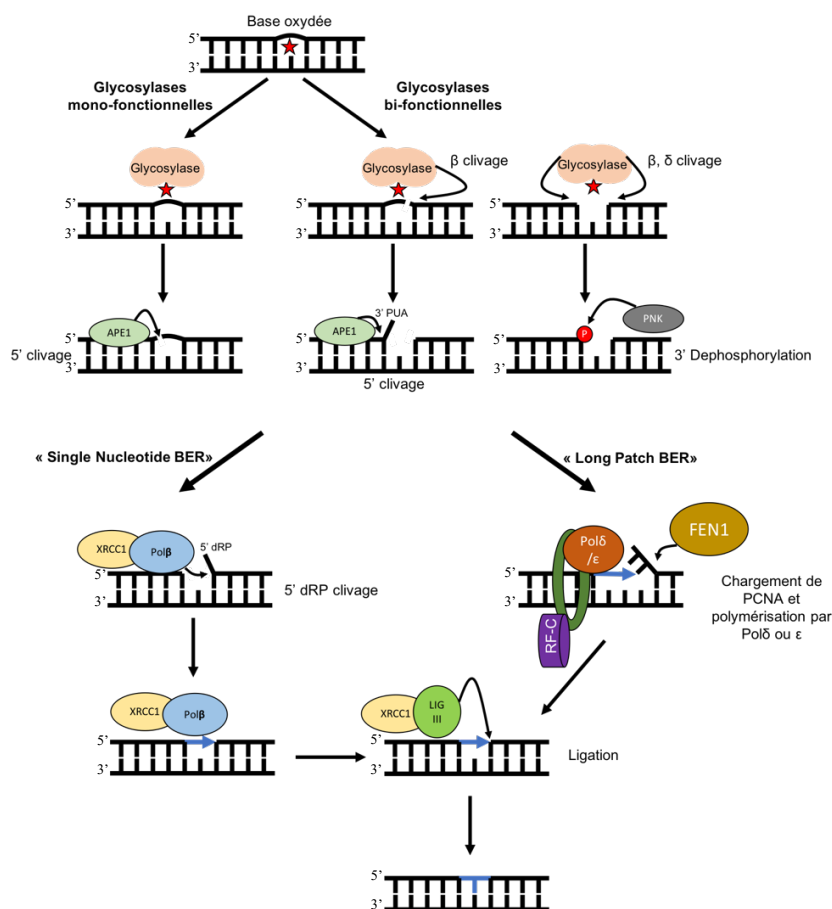


Figure 10 : Schéma des différents mécanismes de réparation par le Base excision repair. L'étape initiale de reconnaissance des lésions est réalisée par une glycosylases mono- (MF) ou bi-fonctionnelle (BF). Une glycosylase MF clive la jonction N-glycosidique de la base lésée. Le site abasique est clivé en 5' par l'AP endonucléase, générant un 5' desoxyribosephosphate (5' dRP). Une glycosylase BF peut cliver la liaison phosphodiester en 3' du site (β-élimination) ou de cliver le site en 3' et en 5' (β/δ élimination) générant un 5' PUA ou une extrémité 3'-P respectivement. La réparation par SN-BER nécessite de cliver le 5' dRP (si nécessaire) par la polymérase β via son activité lyase. C'est cette même polymérase qui ajoute la base manquante. La réparation par LP-BER fait intervenir RFC et PCNA qui permettent de charger les Pol δ/ε ou β. Le brin néo-synthétisé déplace le brin adjacent créant un substrat pour la Flap-5' endonucléase FEN-1. C'est l'ADN ligase I qui termine la réparation en comblant le gap.

Il existe deux mécanismes de réparation par le BER. Le premier mécanisme, *single nucleotide BER* (SN-BER), permet d'éliminer un nucléotide simple à travers l'activité de glycosylases et de la polymérase β (Figure 9). La réparation peut débuter par une glycosylase simple qui va repérer les lésions et cliver la jonction N-glycosylique entre la base et le desoxyribose formant des sites apuriques/aprimidiques (AP). Ce site sera clivé en 5' par une AP endonucléase 1 (APE1) laissant une extrémité 3'-OH libre et une extrémité 5' desoxyribosephosphate (dRP). La polymérase β, qui joue un rôle actif dans la réparation par le BER, utilise cette cassure simple brin (*nick*) afin de premièrement retirer le site AP, *via* son activité lyase, puis de combler le vide *via* son activité polymérase. C'est la ligase III avec l'aide de XRCC1 qui termine la réparation en refermant le *nick*¹¹⁰.

La réparation peut aussi être conduite par une glycosylase bi-fonctionnelle, capable de repérer la lésion, d'éliminer la base et de cliver la liaison phosphodiester en 3' du site (β -élimination) ou de cliver le site en 3' et en 5' (β/δ élimination). Si la coupure est réalisée uniquement en 3', l'aldéhyde α - β -3' insaturé restant sera éliminé par *APE1* qui clive le site en 5', la réparation se termine par le mécanisme classique. Lors d'un clivage en 3' et 5' de la base abimée, le phosphate en 3' est éliminé par la polynucleotide kinase (PNK) qui laisse un groupe hydroxyl à l'extrémité 3' ¹¹¹. Le sucre d'un dNTP peut aussi être oxydé ou réduit, bloquant l'activité lyase de la polymérase β ¹¹². Le *Long patch BER* (LP-BER) permet de réparer ces dommages en faisant intervenir un ensemble de polymérases et de facteurs appartenant à la réplication tel que l'ADN polymérase δ et/ou β et/ou ϵ , *FEN1*, PCNA ou encore l'ADN ligase I ¹¹³ (figure 10). C'est un court fragment de 2 à 15 nucléotides qui sera excisé, contrairement à la base unique retirée dans le SN-BER. Dans ce contexte, la lésion ne sera pas éliminée par la polymérase β . Le nick réalisé en 5' par *APE1* permet le déplacement du brin portant la lésion par la polymérisation d'un nouveau brin créant une extrémité 5' sortante (ou *flap* 5'). Cet intermédiaire de réparation est un substrat de *FEN1* qui va cliver l'extrémité sortante. Si la polymérase β est nécessaire à la réparation par le BER, l'activité de *FEN1* est critique dans la réparation par le LP-BER, stimulant l'activité de la polymérase β et le déplacement du brin ¹¹⁴. Les polymérases impliquées dans le déplacement du brin portant la lésion sont les polymérases δ et/ou β et/ou ϵ . Le chargement des différentes polymérases sur le brin d'ADN requiert l'intervention de PCNA et de RFC. Une fois le nouveau brin synthétisé, la ligase III termine la réparation (figure 10).

Il est intéressant de noter qu'il existe une interaction physique entre le NER et le BER via *CSB* et *APE1*. *CSB* n'est pas nécessaire à l'activation de *APE1* mais stimule fortement sa capacité à cliver les sites abasiques. L'interaction entre les deux protéines est plus forte au sein d'une bulle

de transcription, rappelant le rôle de *CSB* dans le TC-NER ¹¹⁵. Des interactions entre *XPG* et certaines glycosylases ou *APE1* ont aussi été mises en évidence. On retrouve aussi une interaction entre le BER et le MMR sous la forme d'interactions entre les complexes MutS α ou MutL α et certaines glycosylases ¹¹⁶.

Ces observations montrent à quel point les différents systèmes de réparation communiquent entre eux afin de pouvoir repérer et éliminer le plus grand nombre de lésions spécifiques à chaque mécanisme. Dans notre étude sur l'instabilité des régions répétées, il est intéressant de se demander à quel point cette communication influence la stabilité des microsatellites.

2.6.3.2 BER et instabilité des microsatellites

Il a été observé chez des patients atteints de la maladie d'Alzheimer (AD) une augmentation du nombre de lésions dues aux dommages oxydatifs avec une diminution de la quantité de la glycosylase OGG1 par cellule ¹¹⁷. OGG1 est une glycosylase bi-fonctionnelle qui possède la plus grande affinité pour les lésions 8-oxoG. Une lignée de souris (souris R6/1) ayant intégré l'exon 1 de l'allèle du gène de la Huntingtine, portant des répétitions CAG/CTG celles que retrouvées chez des patients malades, a servi à l'étude du BER dans les mécanismes d'instabilité des régions répétées. Il est naturellement observé chez cette souris une corrélation entre l'âge, l'accumulation de dommages oxydatifs et l'expansion des répétitions de trinuécléotides CAG. Cependant la perte d'expression de OGG1 diminue de 70% le nombre d'évènements d'expansions observés dans ces lignées. Cette observation n'est pas réalisée après la perte d'autres glycosylases et ce malgré des redondances de fonctions et de substrats ¹¹⁸. Les évènements d'expansion liés à OGG1 sont dépendants des séquences ayant dépassé le seuil pathologique d'instabilité et capables de former des structures secondaires. L'ajout *in vitro*, à une séquence portant des répétitions de trinuécléotides CAG/CTG, de *OGG1*, *APE1*, *POL β* et de la ligase I provoque des évènements d'expansions de plusieurs nucléotides, qu'on ne

retrouve pas lors de la réparation d'une séquence aléatoire (le produit majeur de la réparation comporte des séquences ayant incorporé un nucléotide supplémentaire dû à l'activité de *Polβ*)¹¹⁹. L'absence d'un de ces acteurs clés de la réparation bloque les expansions. Ces additions de nucléotides sont dues à une faible activité de déplacement de brin de la *polβ*¹²⁰. Nous avons vu que lors de la réparation par le LP-BER, un mécanisme de déplacement de brin lors de la synthèse du nouveau brin par les polymérase δ , ϵ ou β . Le *flap 5'* résultant est (comme dans les autres mécanismes vus dans cette introduction) capable de former une structure secondaire. Une structure secondaire formée sur un *flap 5'* empêche son clivage par FEN-1⁵⁶. Ainsi lors d'une réparation par LP-BER dans les répétitions, la coordination entre l'activité polymérase δ , ϵ ou β et FEN-1 favorise l'expansion des répétitions de trinucleotides CAG/CTG¹²¹. En l'absence de réparation par le BER, cette structure deviendra un substrat du MMR montrant encore une fois les connections entre les différents mécanismes de réparation de la cellule, le MMR pouvant être à l'origine de la stabilisation de la structure en épingle à cheveux formée. La polymérase β favorise aussi la stabilisation des structures en épingle à cheveux en ajoutant des nucléotides aux extrémités 3' de ces structures. Ce mécanisme empêche à la fois la polymérase δ , capable de retirer les épingles, de réparer la séquence, et offre une extrémité 3'-OH libre à Pol δ pour redémarrer la synthèse d'ADN fixant la structure dans la séquence¹²².

2.7 Environnement chromatinien et instabilité des microsatellites

Comme l'ADN n'est pas une simple molécule non structurée²⁷, il est nécessaire de se poser la question de l'impact des expansions (ou contractions) sur l'environnement chromatinien des répétitions. Plusieurs études montrent des modifications de la structure chromatinienne, de la méthylation ou des histones autour des microsatellites à répétitions CAG/CTG, CCG/CGG ou GAA/TTC. Ces études réalisées chez la levure ont aussi tenté d'expliquer les phénomènes d'expansions observés chez l'humain. Nous avons vu dans un chapitre précédent que les

répétitions dans la levure avaient naturellement tendance à subir des événements de contractions avec une fréquence bien supérieure aux événements d'expansions. L'utilisation du modèle levure s'est montrée important et nécessaire dans l'étude des mécanismes d'instabilités des répétitions. J'exprime cependant une réserve quant à l'utilisation de ce modèle dans l'étude des mécanismes d'expansions. Il est possible de modifier une souche de levure afin de se rapprocher au maximum des phénotypes observés chez l'humain ou la souris *via* la perte d'expression de *SRS2*¹²³. Il est à noter cependant que ce phénotype augmente fortement à la fois le taux d'expansions (8,1% d'augmentation des expansions) et de contractions (58,1% d'augmentation des contractions) qui ne correspond pas au phénotype humain. D'autre part, l'influence de l'âge et du sexe des patients joue clairement un rôle dans les expansions et la transmission à la descendance. Cet aspect est complètement perdu lors de l'étude chez la levure (voir chapitres précédents).

2.7.1 Environnement chromatinien des répétitions de trinuécléotides

CGG/CCG

L'expansion des séquences CGG, impliquée dans le syndrome de l'X fragile, cause une hyperméthylation³ des histones le long de la séquence et des régions flanquantes, inhibant l'expression du gène *FRMI*, causant la maladie¹²⁴. Il a été retrouvé chez certains patients une expansion CGG au locus *FRMI* n'entraînant pas d'hyperméthylation. Le niveau d'expression de *FRMI* est similaire à une personne ne portant pas d'expansion CGG au locus *FRMI*, il est cependant observé une forte diminution de la traduction des ARNm¹²⁵. La technique du 4C (*chromosome conformation capture on chip*) qui permet d'étudier les interactions d'un locus avec le reste du génome (*one vs all*) a permis de montrer qu'un locus *FMR1* portant des

³ Augmentation des marques de méthylation de la lysine 9 des histones H3 (H3K9me)

expansions de CGG/CCG subit une modification de la conformation de sa chromatine et interagit différemment avec le reste du génome ¹²⁶.

2.7.2 Environnement chromatinien des répétitions de microsatellites

GAA/TTC

La même augmentation des marques H3K9me et la même diminution des marques d'acétylation des histones H3 et H4 dans les expansions de GAA sont observées au sein et autour des répétitions ¹²⁷. Ces marqueurs ne sont pas retrouvés dans une séquence homologue n'ayant pas subi d'évènement d'expansion ¹²⁸. Une diminution des marques de méthylation et une augmentation des marques d'acétylation permettent d'augmenter le niveau de transcription d'un locus FRDA (ataxie de Friedreich) portant des répétitions microsatellites GAA. Ces modifications de marques épigénétiques s'accompagnent d'une décompaction du locus, augmentant l'accessibilité de la région ¹²⁹.

2.7.3 Environnement chromatinien des répétitions de microsatellites

CAG/CTG

On retrouve les marqueurs de méthylation dans le cas d'une séquence CTG ayant subi une expansion au locus *DMPK*. Accompagné d'une baisse de marqueur d'acétylation des histones H3, ce phénomène traduit une hétérochromatinisation qui s'accompagne d'une perte d'euchromatinisation (acétylation) des séquences ¹²⁷. Des études sur l'acétylation des histones H4 semblent montrer leur implication dans l'inhibition des expansions « CAG » chez la levure. Cela serait notamment dû à la sous-unité catalytique *esal* du complexe NUA4 (Histone Acetyl Transferase ou HAT) qui acétyle les 4 lysines de l'histone H4 ainsi que les lysines de l'histone H2A. La perte de l'activité de cette sous-unité entraîne une augmentation du nombre

d'évènements d'expansions. On retrouve ce même phénotype d'expansion dans des mutants *hat1Δ*, qui acétylent les histones néo-synthétisés H4. La perte de fonction de la dé-acétylase *rpd3* entraîne une diminution des évènements de contraction sans montrer de différence quant à la fréquence d'expansion. La perte d'activité de la déacétylase *hda1* augmente fortement les évènements d'expansion. Dans des levures dont les histones H4 portent une mutation simulant une acétylation constante des lysines, une élévation de la fréquence des expansions est observée. Un résultat montrant une acétylation transitoire des histones au sein des répétitions durant la phase S montre l'importance des cycles de méthylation et d'acétylation des histones, en fonction des phases du cycle cellulaire, dans la stabilité des répétitions. Ces évènements sont probablement connectés au passage des fourches de réplication ou des dommages à l'ADN liés à la nature des répétitions ¹³⁰. Dans des cellules humaines, la perte d'expression de la méthyltransférase Dmmt1 conduit à un doublement de la fréquence des évènements de contractions d'une répétition CAG/CTG ¹³¹. Les répétitions au locus SCA1 de la souris, responsable de l'ataxie spinocérébelleuse de type 1 chez l'Humain, tendent à se contracter lors de la transmission parentale avec des contractions plus importantes dans la transmission maternelle, la transmission paternelle semblant plus stable avec peu de variation du nombre de répétitions. La perte d'expression d'une copie du gène *DNMT1* conduit à une augmentation des évènements des expansions lors de la transmission parentale avec des évènements d'expansions plus importants lors de la transmission paternelle ¹³¹. La conformation de la chromatine aux locus HTT ou DMPK, portant respectivement des répétitions de microsatellites CAG et CTG, ne montre aucune modification après expansion. Ces résultats indiquent que la conformation en 3 dimensions de la chromatine n'influence pas la stabilité des répétitions, contrairement aux marques épigénétiques ¹²⁶.

3 Sites fragiles, cancer et microsatellites

L'implication des expansions de microsatellites est aujourd'hui reconnue dans plus d'une quarantaine de maladies neurodégénératives. Certains de ces microsatellites sont aussi des sites fragiles sensibles aux cassures double-brin, aux délétions, expansions ou translocations retrouvées dans certains cancers. De plus, une instabilité des microsatellites est observée dans 15% des cancers colorectaux et impliquée dans le déclenchement de certains cancers colorectaux héréditaire non-polypeux (HNPCC). Dans ce chapitre nous discuterons de l'implication de l'instabilité des microsatellites dans la biologie des sites fragiles et du développement de certains cancers.

Les sites fragiles (SF) sont des locus chromosomiques sensibles aux drogues induisant des ralentissements de fourches de réplication/de la synthèse de l'ADN et conduisant à l'apparition de cassures simple et double-brin lors de la phase S. Ils sont qualifiés de communs ou de rares en fonction de leur proportion dans la population et ont été retrouvés et documentés chez de nombreux mammifères. Ils sont impliqués dans des réarrangements chromosomiques tels que des translocations, des échanges de chromatides sœurs, des insertions de virus oncogéniques, des délétions ou expansions qu'on retrouve dans certains cancers tel que des adénocarcinomes, myélomes, leucémies et lymphomes ^{19,20}. L'expression d'un site fragile correspond à l'expression de ces phénotypes.

Les sites fragiles communs (CFS) sont retrouvés dans tous les individus et sont notamment sensibles aux traitements inhibant l'ADN polymérase (traitement aphidicoline). On les retrouve répliqués tardivement lors du cycle cellulaire, avec une réplication qui déborde sur la phase G2 ¹³². Ce phénotype est observé, par exemple, lors de l'induction du site fragile commun FRA3B avec de l'aphidicoline, inhibant l'activité des polymérases α , δ et ϵ , où le locus n'est pas encore totalement répliqué lors de l'entrée en G2 de la cellule ^{133,134}. Le laboratoire de Michelle

Debatisse a montré que la fragilité de *FRA3B* est spécifique d'un type de tissu ¹³⁴. Les sites fragiles rares (RFS) sont causés par l'expansion de microsattellites (ou minisatellites), dont les structures secondaires vont gêner la réplication. Le groupe sensible aux folates est le groupe majeur des RFS causé par l'expansion des répétitions trinuécléotides CGG, impliqué dans l'inhibition du gène *FRM1* et le syndrome du X fragile (FRAXA), ou du gène *FRM2* (FRAXE). D'autres groupes de RFS (FRA16B et FRA10B) non sensible aux folates sont sensibles à la distamycin A et à la bromodeoxyuridine (BrdU). Au total on distingue trois groupes de RFS, un groupe sensible aux folates (FRA11B, FRA16A, FRAXA, FRAXE, FRAXF), un groupe sensible à la distamycin A (FRA16B, FRA17A, FRA8E, FRA11I,) et un troisième groupe sensible au BrdU (FRA10B, FRA12C) ¹⁸. La fragilité de ces séquences semble proportionnelle à leur taille. L'insertion de répétitions CAG/CTG dans un chromosome synthétique de levure (YAC) se comporte comme un site fragile en présence d'hydroxyurée (HU), drogue qui déplete la réserve en dNTPs de la cellule, qui conduit à un arrêt des fourches de réplication. Plus la séquence CAG/CTG insérée est longue, plus le nombre d'évènements de cassures observées est important ^{54,135}.

4 Les nucléases

Comme nous avons pu le constater tout au long de cette introduction, la génération d'une cassure double-brin au sein d'une séquence répétée augmente fortement l'instabilité de cette dernière. Que cette cassure conduise à une expansion de la séquence, qui déclenchera une maladie neurodégénérative, ou à la tumorigénèse de la cellule, la compréhension des mécanismes de réparation de la cassure double-brin est nécessaire dans le développement de traitements. Ainsi dans mon sujet de thèse, la réparation d'une cassure double-brin unique générée par une nucléase est au centre de mon projet.

Au cours des deux dernières décennies, de nombreux outils ont été développés dans le but

d'éditer le génome. Des méthodes classiques de recombinaison homologue ¹³⁶ aux systèmes d'endonucléases modernes, la biotechnologie offre aujourd'hui la possibilité de générer à un locus précis et unique, une cassure double-brin de l'ADN. L'édition du génome permet à la fois la disruption d'un gène précis, pour son étude ou dans un but de thérapie génique et l'étude de la réparation des cassures double-brin. Dans ce chapitre sont présentées les différentes familles de nucléases qui ont été développées dans le but d'étudier ou de générer une cassure double-brin et notre choix afin de générer la cassure double-brin qui est étudiée dans mon projet de thèse.

4.1 Les Méganucléases

Ces endonucléases utilisées depuis les années 90 peuvent être divisées en cinq familles basées sur leurs séquences et les motifs de structures. On distingue les familles LAGLIDADG, GIY-YIG, HNH, à boîte His-Cys et PD-(D/E)XK, qui représentent des motifs d'acides aminés conservés par les membres de chaque famille. Leurs larges sites de reconnaissance (12-45pb) leur ont valu le surnom de « Méga » ¹³⁷ (figure 11). La famille la plus étudiée est celle des Méganucléases à domaine LAGLIDADG, retrouvée dans tout le vivant. Capable de s'exciser librement de leur séquence hôte ADN (endonucléase) ou protéique (intéine) sans perturber la biologie du gène ou protéine d'insertion, et de cliver spécifiquement la jonction exon-exon de la séquence cible de l'hôte dans laquelle la séquence codant la nucléase s'insère. La nucléase codée par une ORF (*open reading frame*, cadre de lecture susceptible d'être traduit en protéine) sur un élément mobile, dans un intron ou une intéine (intron protéique), est capable d'induire une cassure double-brin à une séquence cible complémentaire. Cette cassure sera réparée par recombinaison homologue avec la séquence de la nucléase conduisant à la duplication de l'intron ou de l'intéine ^{138,139}. Cette activité endonucléase dirigée contre leur propre intron, leur a donné le nom d'*homing endonuclease*. Le motif LAGLIDADG a été montré comme essentiel

dans l'activité enzymatique, la dimérisation (dans le cas du groupe I, voir suite du chapitre), le repliement et la reconnaissance de la séquence cible de la nucléase. Deux groupes peuvent être distingués : le premier groupe ne contient qu'un seul motif LAGLIDADG auquel appartient la nucléase I-CreI et un deuxième groupe portant deux motifs, auquel appartient la nucléase I-SceI qui fut la première découverte au sein d'un intron mitochondrial de la levure *Saccharomyces cerevisiae*^{140,141}.

I-CreI fonctionne en homodimère et reconnaît un motif pseudo-palindromique 5'-CAAACGTCGTGAGACAGTTTG -3' de 22 paires de bases. Des délétions ou des substitutions dans la séquence sont tolérées par la nucléase qui peut encore se fixer et couper¹⁴². La modification d'acides aminés ou de domaines a permis de développer de nouvelles I-CreI capable de reconnaître de nouveaux sites de coupure¹⁴³. I-SceI possède deux domaines LAGLIDADG et est active en tant que monomère. Son site de reconnaissance 5'-TAGGGATAACAGGGTAAT-3' fait 18 paires de bases. Avec cette découverte venait la possibilité de pouvoir induire une cassure double-brin à un site spécifique du génome, la taille importante du site de reconnaissance des Méganucléases rendant la spécificité de la coupure importante. Évidemment, la difficulté d'utilisation réside aussi dans ce site de reconnaissance. Pour induire une cassure, il faut avoir inséré au préalable la séquence cible de la nucléase au locus d'intérêt. Des variations de la nucléase ont été générées par Arnould et Pâques *in vitro*, capables de reconnaître et de couper de nouvelles séquences, par mutations du site de reconnaissance à l'ADN de I-CreI. La formation de complexe hétérodimérique avec des I-CreI modifiées augmente encore le nombre de possibilités¹⁴³. Ces modifications longues et laborieuses rendent compliquées l'utilisation des Méganucléases dans des expériences de cassures à fréquence élevée et dans des régions aussi complexes que les microsatellites. De plus, l'efficacité des différentes Méganucléases est sensible aux marques chromatiniennes portées par le locus d'intérêt et montrant des variations importantes d'efficacité d'un locus à un

autre *in vivo*¹⁴⁴ rendant d'autant plus compliquée leur utilisation dans des régions soumises à des variations importantes de méthylations comme les répétitions de trinuécléotides CTG.

4.2 Nucléases à Doigts de Zinc (ZFN)

Les *Zinc Finger Nucleases* (ZFN) sont des nucléases artificielles portant des domaines à doigts de zinc (ZF) ainsi que le domaine catalytique de la nucléase FokI¹⁴⁵ (figure 11). Les domaines à doigts de zinc ont d'abord été mis en évidence dans la protéine « transcription factor III A » ou TFIIIA (une protéine à doigts de zinc (ZFP)) nécessaire à l'initiation de la transcription de l'ARN ribosomal 5S sous-unité du ribosome par l'ARN polymérase III, purifié à partir d'extrait d'ovocyte de *Xenopus laevis*¹⁴⁶. TFIIIA interagit avec la sous unité ribosomique 5S et régule sa transcription en interagissant aussi avec l'ARN polymérase III. On retrouve dans son organisation trois larges domaines, avec deux domaines d'interaction à l'ADN et un domaine d'interaction avec l'ARN polymérase III¹⁴⁶. La séquence de la protéine met en évidence neuf motifs de répétitions de 30 acides aminés le long de la protéine. On retrouve aux mêmes positions dans chaque motif les acides aminés Y-C-C-F-L-H-H (T=tyrosine ; C=cystéine ; F=phénylalanine ; L=leucine ; H=histidine). Les deux cystéines et les deux histidines, qui sont les ligands biologiques les plus communs du zinc, sont les éléments les mieux conservés de chaque motif et permettent à chacun des 9 motifs de fixer une molécule de zinc nécessaire à la stabilisation de la structure qui adopte la forme d'une boucle (portant deux feuilletts β et une hélice α « $\beta\beta\alpha$ ») (figure 11). Cette structure en boucle ressemble à un doigt d'où le nom de ces protéines est tiré¹⁴⁶. Chaque doigt de zinc reconnaît un motif de 3-4pb et interagit avec l'ADN via l'hélice α du doigt de zinc. La modification des acides aminés en -1 à +6 au sein de l'hélice α (au sein de la structure $\beta\beta\alpha$), entraîne un changement de la séquence reconnue par le ZF. Chaque ZF influence la spécificité des ZFs voisins, rendant compliqué l'élaboration d'une ZFP ciblant une séquence d'intérêt. Différentes techniques ont été développées afin de

simplifier le processus d'élaboration des ZFP et ne seront pas développées dans cette introduction, mais ont permis de rendre plus accessible la fabrication de ZFP spécifiques^{146,147}.

Il est évident qu'un système permettant de reconnaître spécifiquement une séquence d'ADN est intéressant dans le développement de nucléases induisant spécifiquement une cassure double-brin. Ainsi un système de ZFP hybride, constitué de plusieurs ZFP (ayant une spécificité de séquence variable) et du domaine catalytique de la nucléase FokI a vu le jour. Il suffit d'assembler alors des ZFP connues afin de construire une nucléase capable de reconnaître et de cliver un site spécifique dans le génome¹⁴⁵. Afin d'induire une cassure double-brin, la nucléase FokI qui ne possède qu'un site catalytique, doit se dimériser au niveau du site de coupure. Deux sites de reconnaissance sont donc nécessaires à l'induction d'une cassure double brin et l'efficacité de la cassure d'une ZFN est largement dépendante de sa concentration⁴^{148,149}. Les premières ZFN ont été testées *in vivo* dans des cellules de souris ou humaines. Bien que fonctionnelles, ces dernières sont toxiques après traitement prolongé¹⁵⁰. Cette toxicité, provenant de la génération de cassures double-brin non spécifiques, causées par la formation d'homodimères de ZFN, peut être réduite par l'augmentation du nombre de ZF le long de la protéine afin d'augmenter la spécificité de la ZFN, par la création de ZFN plus précise ou par la modification du site catalytique de FokI afin d'éliminer les interactions homodimériques, qui diminue le nombre de cellules présentant des coupures *off-target*⁵ par deux^{151,152}.

Bien que cette technologie semblait difficilement applicable aux répétitions au vu de la faible spécificité de ces nucléases, il a malgré tout été développé de nouvelles ZFN capable d'induire des cassures double-brins au sein de régions répétées¹⁵³. Ces ZFN restent assez peu spécifiques (l'induction d'une cassure nécessite un bras reconnaissant les CAG et un bras reconnaissant les CTG, chaque bras étant malgré tout capable d'induire une cassure double-brin par

⁴ Corrélation entre la concentration et l'efficacité de coupure de la ZFN

⁵ Coupure générée en dehors du locus ciblé

homodimérisation). Une étude du laboratoire de Michael Leffak essaye de montrer le développement d'un couple hétérodimérique de ZFN spécifiques, incapable de former des homodimères et induisant spécifiquement des cassures double-brin au sein de microsatellites. L'incubation d'un seul bras des ZFN fait néanmoins apparaître des bandes, correspondant à des produits de digestion après incubation, que les auteurs justifient par la génération de cassures simple-brin au niveau des boucles simple-brin des structures secondaires se formant sur la séquence répétée et prouvant ainsi l'existence *in vivo* de ces structures. Ces conclusions restent à interprétation puisqu'aucune explication sur le mécanisme de fixation d'une ZFN sur une épingle à cheveux imparfaite n'est considérée. D'autres contrôles seraient nécessaires afin d'éliminer la possibilité d'homodimérisation et la spécificité de reconnaissance des ZFN pour les structures secondaires. On pourrait imaginer tester des ZFN similaires dans des répétitions ne formant pas de structures secondaires ¹⁵⁴.

4.3 Les TALENs

En 2009 fut découverte une famille de protéines, dans le genre *Xanthomonas* des pathogènes de plantes. Ces protéines sont capables de se fixer au promoteur d'un gène de l'hôte et d'activer sa transcription, afin de permettre la colonisation de la plante. Ces protéines, appelées *Transcription Activator-Like Effector* ou TALE ont permis par la suite de développer une nouvelle famille de nucléases, les TALEN (*TALE Nuclease*) (figure 11). Une TALE est constituée d'un domaine d'adressage au noyau (NLS), d'un domaine activateur de la transcription et d'un domaine central de 30 répétitions en tandem, de 33- à 35- acides aminés, invariantes à l'exception de deux acides aminés adjacents qui portent le nom de *repeat variable diresidue* (RVD) en position 12 et 13 du motif répété. Ces RVD jouent un rôle central dans le mécanisme de reconnaissance de la TALE à son site de fixation. La variation des RVD forme un code qui a été déchiffré et qui permet de déterminer à quelle paire de RVD est associé chaque

base, HD = C ; NG = T ; NI = A ; NS = A, C, G, or T ; NN = A ou G ; et IG =T. ^{155,156}. La construction des TALENs utilise l'avantage que procure ce code afin de construire un domaine TALE capable de reconnaître un locus précis, fusionné au domaine catalytique de la nucléase Fok1 ((figure 11). Comme dans le cas des ZFN, l'utilisation de deux TALE-Fok1 permet d'induire une cassure double-brin à un locus précis. Les TALENs ont d'ailleurs énormément bénéficié de l'expérience acquise lors de la construction des ZFN dans leur mis au point ¹⁵⁷. Contrairement aux ZFN, le domaine de reconnaissance de l'ADN des TALENs est beaucoup plus large (un domaine de 33-35 acides aminés reconnaît une base, via les RVD, tandis qu'un domaine ZF d'une ZFN reconnaît 3-4 paires de bases), mais aussi plus précis et n'est pas affecté par un domaine voisin, rendant prédictible l'usage du code TALE et la construction des domaines. Le domaine nucléase Fok1 doit comme les ZFN se dimériser afin de pouvoir induire la cassure. Deux bras de TALENs sont donc nécessaires à l'induction d'une cassure double-brin. L'efficacité des TALENs est similaire à celle des ZFN avec une toxicité moins importante ^{147,157}. Les points négatifs de l'utilisation des TALENs résident dans leur construction qui même si plus simple à mettre au point que les ZFN, est difficile à réaliser en raison de la nature des séquences répétées formant les domaines de la TALEN. La taille importante de ces dernières (le gène codant pour une TALEN est trois fois supérieur en taille à celui codant pour une ZFN) rend difficile la transfection de cellules de mammifère. Dans mon projet de thèse, une TALEN a été spécifiquement développée par le laboratoire afin d'induire une cassure double-brin unique, au sein de répétitions CTG. Le choix de la TALEN sera développé ci-dessous.

Finalement, alors que la technologie TALEN émergeait, une technologie concurrente aujourd'hui célèbre et ubiquitaire voyait le jour : la technologie CRISPR-Cas (figure 11).

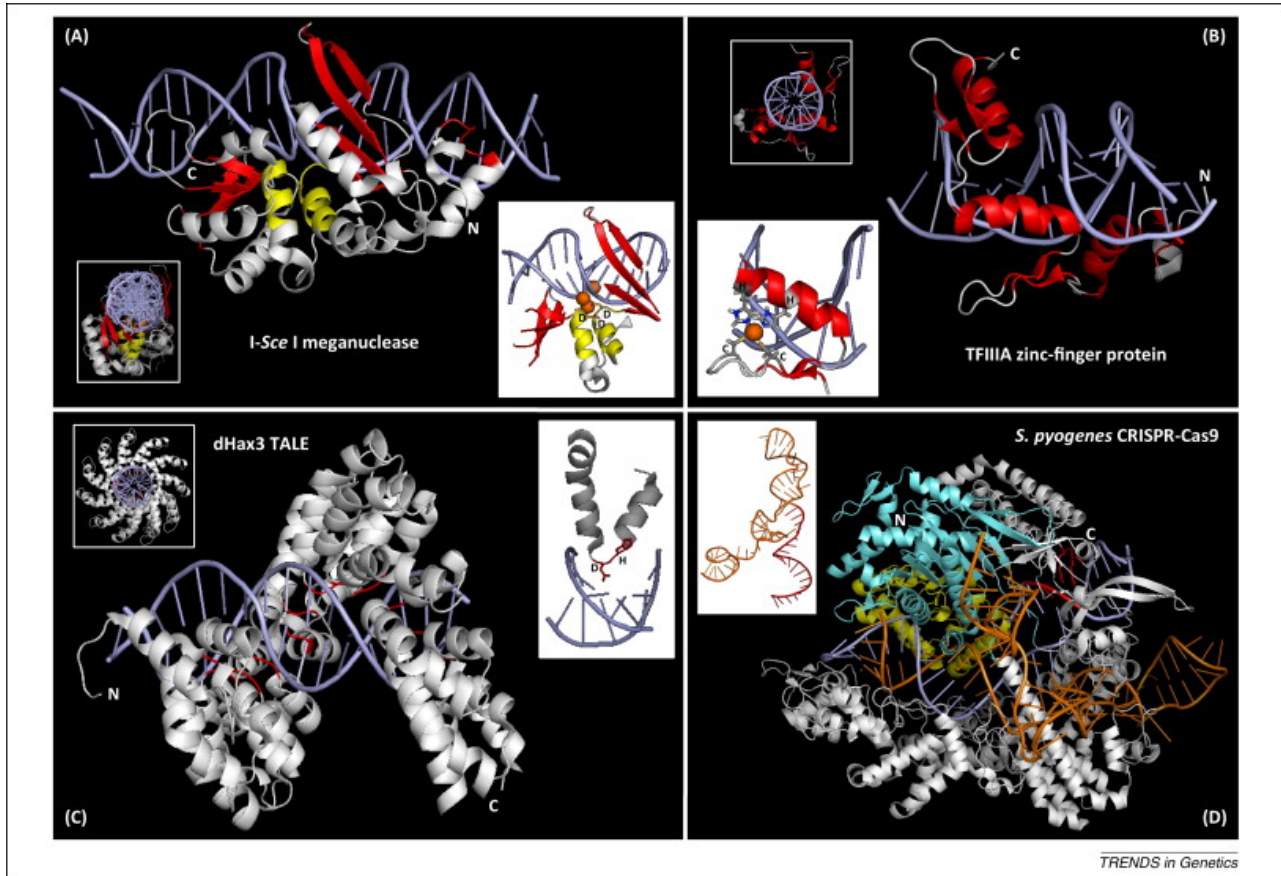


Figure 11 : Structures cristallines des quatre familles de nucléases. Les représentations insérées dans les carrés blancs représentent la vue perpendiculaire du complexe protéine-ADN le long de l'axe de l'ADN. (A) Structure cristalline de la méganucléase I-Sce I sur son site de fixation à l'ADN. Les deux motifs LAGLIDADG sont représentés en jaune. (B) Structure NMR des trois doigts de zinc de la protéine à doigts de zinc TFIIIA, facteur de transcription de *Xenopus laevis*, attaché au gène de l'ARN 5S. Chaque doigt est représenté en rouge. (C) Structure cristalline de la TALE dHax3 attaché à son site de reconnaissance. Les RVD des différents motifs répétés de la TALE sont représentés en rouge. (D) Structure cristalline de la nucléase Cas9 de *Streptococcus pyogenes* en complexe avec une molécule guide d'ARN.

4.4 CRISPR-Cas

Les CRISPRs (*Clustered Regularly Interspaced Palindromic Repeats*), ont commencé à attirer l'attention des chercheurs au milieu des années 2000. Initialement découvertes par hasard dans la région 3' du gène *IAP* de *Escherichia coli*¹⁵⁸, il a plus tard été montré que ces séquences sont présentes dans de très nombreuses familles de bactéries et archées tandis qu'elles sont

complètement absentes des organismes eucaryotes. Comme leur nom l'indique, les CRISPRs sont de courtes séquences d'ADN répétées, séparées par de courtes séquences d'ADN uniques (espaceurs ou *spacers*). Les segments répétés font de 21 à 37 paires de bases de long et les espaceurs sont de taille similaire. Dans les régions flanquant les CRISPRs, on retrouve invariablement une longue séquence d'ADN répétée (LR) de plusieurs centaines de paires de bases riches en AT qui n'est pas retrouvée ailleurs dans le génome. Quatre gènes *CAS* (appelé *cas1* à *cas4*) furent d'abord mis en évidence et toujours associés au locus CRISPR. *Cas3* fut montré comme une hypothétique hélicase appartenant à la superfamille 2 et *Cas4* comme une exonucléase appartenant à la famille RecB. La présence d'une hélicase et d'une nucléase laissait déjà supposer le rôle fonctionnel des locus CRISPR dans le métabolisme de l'ADN de la cellule. Dans les organismes possédant plusieurs locus CRISPR avec la même séquence répétée, on ne retrouve les gènes *CAS* qu'à un seul des locus ¹⁵⁹. Il a été, peu de temps après, mis en évidence que les espaceurs avaient une origine extrachromosomique et venaient pour la plupart de phages ou de plasmides. Une corrélation entre la résistance d'une bactérie à l'infection par un phage et le nombre d'espaceurs ayant une homologie de séquence avec ce même phage a été observée, indiquant que le système CRISPR-Cas pouvait jouer un rôle dans la protection des bactéries face à l'invasion de la cellule par du matériel génétique d'un organisme étranger ¹⁶⁰. Cette hypothèse fut rapidement vérifiée lors d'expériences visant à rendre des lignées de bactéries industrielles insensibles aux phages. Au contact du phage, de nouveaux espaceurs sont intégrés au locus CRISPR correspondant à des éléments du génome du phage invasif. Ces insertions sont polarisées à une extrémité du locus CRISPR et rendent les souches résistantes à l'infection ¹⁶¹.

La découverte de la synthèse d'ARN par le locus CRISPR a permis de comprendre par quel mécanisme la cellule se défend face à l'invasion par un phage ou un plasmide d'un organisme étranger. Ce sont par exemple 8 gènes *CAS* associés au locus CRISPR, qui seront découverts

chez *Escherichia coli*. Ces gènes sont : Les gènes *CASABCDE*, le gène *CAS1* (a été montré comme codant une intégrase), *CAS2* (code une endoribonucléase) et le gène *CAS3* (code une nucléase/hélicase). Les cinq protéines CasABCDE forment un complexe appelé « Cascade ». Ce complexe est nécessaire à la maturation d'un ARN transcrit à partir du locus CRISPR (crRNA), d'environ 57 paires de bases et comprenant un espaceur entouré de séquences répétées en 5' et 3' nécessaires pour l'interaction avec le complexe « Cascade ». CasE est, de façon surprenante, la seule sous-unité nécessaire du complexe pour la maturation des « crRNA ». La résistance des bactéries est au maximum en présence de *Cas3*, du complexe *CasABCDE* et lorsque les espaceurs du locus CRISPR ciblent le template (ADN) des ARNm envahissant la cellule ¹⁶².

La classification des CRISPRs évolue rapidement au fur et à mesure des découvertes de nouveaux systèmes dans différents organismes. Des classifications ont donc été proposées en fonction de l'architecture du locus, de sa composition en gènes et d'études phylogénétiques ^{163,164}. La plus récente réalisée par ces auteurs date de 2020 ¹⁶⁵. De façon générale, le mécanisme se déroule comme suit : (I) intégration d'une séquence d'ADN de l'organisme invasif au locus CRISPR sous forme d'espaceurs (adaptation) ; (II) Synthèse d'un crRNA qui doit être maturé (expression) ; (III) Clivage de la séquence complémentaire du crRNA par les protéines *Cas* (interférence).

4.4.1 Adaptation

L'étape d'adaptation, dans laquelle Cas1 et Cas2 jouent un rôle fondamental en permettant l'insertion d'un fragment d'ADN du pathogène au locus CRISPR (avec l'intervention de Cas4), est très conservée. Le choix de la séquence qui deviendra un espaceur dépend entièrement du motif flanquant le futur espaceur (proto-espaceur) et qui porte le nom de PAM (*Proto-spacers Adjacent Motifs*). Le choix de la PAM dépend du système CRISPR-Cas et est dépendant des

protéines Cas1 et Cas2 ^{162,166}.

4.4.2 Expression

La classification par Makarova et Koonin définit 2 classes de CRISPR. Le système de classe I fait intervenir un ensemble de protéines Cas dans la formation du complexe portant le crRNA (module effecteur) jusqu'à la reconnaissance de la cible et son clivage (par exemple avec le complexe cascade pour les systèmes de type I). A l'inverse, le système de classe II ne fait intervenir qu'une seule protéine cas à multidomaines, capable de catalyser toutes les étapes, depuis la maturation du crRNA jusqu'au clivage de la séquence cible ¹⁶⁵. Le système CRISPR-Cas de classe I (auquel *Escherichia coli* appartient) fait intervenir Cas6 pour la maturation du crRNA, les type I, III et IV appartenant à la classe I. Le système CRISPR-Cas de classe II fait intervenir la RNase III, Cas12 ou Cas13 pour les types II (auquel appartient le système CRISPR-Cas9 qui sera développé dans ce chapitre), V et VI respectivement ¹⁶⁵.

4.4.3 Interférence

L'interférence correspond à l'étape où le crRNA emmène les protéines et/ou complexes (en fonction de la classe et du type de CRISPR) responsables du clivage de la séquence cible du crRNA. Il est intéressant de voir apparaître un nouveau type de CRISPR dans la classification (type VI) capable de reconnaître et de cliver directement un ARN cible du pathogène à la place d'une séquence d'ADN ¹⁶⁵. Cas9, Cas12 et Cas13 sont respectivement les protéines majeures du module effecteur responsable de la reconnaissance, de la fixation à la séquence cible et de son clivage pour les types II, V et VI de classe 2.

4.4.4 CRISPR-Cas9

Dans le système CRISPR-Cas de classe II, on retrouve le célèbre système CRISPR-Cas9 de la bactérie *Streptococcus pyogenes*. Comme pour les autres types de la classe II, seule la protéine Cas9 est suffisante pour à la fois terminer la maturation du crRNA et induire une cassure double-brin au locus ciblé. La maturation du crRNA dépend d'un « *trans-encoded small RNA* » ou tracrRNA, de la RNase III et de la protéine Cas9. Chez les types V et VI de la classe II, la RNase III n'est pas nécessaire, ce sont les protéines Cas12 et Cas13 respectivement qui termineront la maturation du crRNA.

Le tracrRNA est une spécificité des systèmes de type II. Il est transcrit sur le brin opposé à celui portant l'unité espaceur-répétition et présente une homologie totale avec la séquence du crRNA. Lors de l'interférence, le crRNA guide le système à la séquence ADN cible du pathogène afin de la cliver^{163-165,167}. Rapidement il fut montré par le laboratoire d'Emmanuelle Charpentier que seule l'extrémité 5' du crRNA (interagissant avec l'ADN) et la région du tracrRNA interagissant avec le crRNA, étaient nécessaires à la reconnaissance du complexe par Cas9 et à l'induction de la cassure au locus cible. Un ARN chimère (ou *simple guide RNA (sgRNA)*) ne comprenant que ces régions (en fusionnant l'extrémité 3' du crRNA à l'extrémité 5' du tracrRNA, imitant le complexe double ARN) fut développé et se montra capable d'amener Cas9 au locus ciblé et d'induire une cassure double brin¹⁶⁸. La présence d'une PAM à proximité du site de cassure reste néanmoins nécessaire. Le motif PAM varie en fonction des différents systèmes CRISPR mais reste une séquence adjacente au locus de quelques nucléotides de long. Dans le cas du système CRISPR-Cas9, le motif PAM consensus est « NGG ». Ce système CRISPR-Cas9 permet d'induire une cassure double-brin à n'importe quel locus simplement en modifiant la séquence cible sur l'ARN guide chimère (sgRNA) et à condition d'une PAM « NGG » soit accessible de façon immédiatement adjacente à la séquence cible.

5 Mécanismes de réparation des cassures double-brin et répétitions

Par quels mécanismes une cassure double-brin générée au sein d'un microsatellite à trinuécléotides CAG/CTG est réparée, fut la question à l'origine de mon sujet de thèse. Une cassure double-brin est une lésion hautement toxique que la cellule doit absolument réparer sous peine de mort, de réarrangement chromosomique ou de perte d'information génétique pouvant mener au développement de maladies tel que le cancer. C'est environ 50 cassures double-brin qui se produisent par cycle cellulaire, par cellule, chez l'humain ¹⁶⁹. Ces lésions peuvent être causées par des facteurs extérieurs environnementaux tels que les radiations (UV ou ionisante), ou encore des produits chimiques. Elles peuvent aussi être causées par le métabolisme des cellules lors d'arrêt de fourche de réplication, des radicaux oxygénés libres ou des nucléases. Il existe deux mécanismes de réparation pouvant intervenir lors de la génération d'une cassure double-brin. En fonction des différentes phases du cycle cellulaire, ces mécanismes seront amenés préférentiellement à réparer ces cassures. En phase G1 la réparation par *Non Homologous End Joining* (NHEJ) (figure 12) sera préférentiellement utilisée. Ce mécanisme permet de relier les deux extrémités de la cassure protégeant l'intégrité du génome en empêchant l'apparition d'aberrations chromosomiques ¹⁷⁰. En effet une perte de Ku ou Xrcc4/Lig IV conduit à une forte augmentation du nombre de réarrangement chromosomique impliquant sont importance dans la stabilité du génome ^{171-173,174}. Historiquement montré comme un mécanisme favorisant les erreurs de réparation, en raison de nombreuses études ne faisant aucunes distinctions entre le NHEJ et les mécanismes de l'Alt-NHEJ, il est aujourd'hui reconnu comme un mécanisme précis et adaptatif ¹⁷³. Durant les phases S et G2, lorsque le génome a été entièrement répliqué, la cellule peut utiliser la recombinaison homologue (HR) afin de réparer les cassures. Ce mécanisme nécessite l'envahissement d'un brin d'ADN médié par *RAD51* sur la chromatide sœur ou le chromosome homologue. Le *Single-Strand Annealing* (SSA) et l'*alt-End Joining* (alt-EJ) ou *Micro-homology End Joining* (MMEJ) sont deux

mécanismes de réparations des cassures double-brin, utilisant des séquences d'homologie de part et d'autre de la cassure afin de la réparer, mais conduisant à une perte d'information génétique entre les séquences répétées et/ou favorisent les évènements de translocations interchromosomiques. Dans ce chapitre seront d'abord détaillés les mécanismes généraux de réparation des cassures double-brins ainsi que leurs rôles dans la réparation d'une cassure double-brin au sein d'une région répétée. Dans une deuxième partie, leurs rôles dans l'instabilité des régions répétées seront décrits de façon plus approfondie.

5.1 Les différents mécanismes de réparation

Des études récentes tendent à montrer que c'est PARP-1 qui serait à l'origine de la reconnaissance des cassures double-brin et du choix entre le NHEJ et la recombinaison homologue (HR) ^{175,176}. PARP-1 génère des poly(ADP-ribose) (PAR) qui permettent de relâcher la structure chromatinienne facilitant l'accès des différentes protéines de la machinerie de réplication à la cassure ainsi que leur recrutement ¹⁷⁵. La résection est la première étape obligatoire des mécanismes de réparation par recombinaison homologue où la digestion des extrémités 5' de la cassure permet de libérer des extrémités 3' sortantes nécessaires à la recombinaison ¹⁷⁷. Les étapes initiales de la résection sont réalisées par Mre11, Rad50 et Xrs2, formant le complexe MRX de la levure *S. cerevisiae* et MRN chez l'humain où Xrs2 est remplacée par son homologue humain Nbs1. Mre11 dispose de deux activités nucléases nécessaires à la réparation et qui seront détaillées plus tard. Rad50 possède une activité ATPase et permet de dimériser le complexe à la cassure, rapprochant les extrémités et facilitant la réparation. Xrs2 stabilise et régule le complexe. Le complexe MRX/N possède un site de reconnaissance des PAR permettant à PARP-1 de négativement réguler l'activité du complexe MRX à la cassure ¹⁷⁵. D'autre part, PARP-1 interagit avec *DNA-PK* (DNA dependent protein kinase) et *Ku70-Ku80*, favorisant ainsi l'initiation du NHEJ.

5.1.1 NHEJ chez l'humain

Le *Non-Homologous End Joining* (NHEJ) permet de lier deux extrémités d'une cassure sans avoir besoin de séquence homologue de part et d'autre de celle-ci. Il est favorisé lors des phases G1 en absence de séquence homologue (chromatide sœur ou chromosome homologue) permettant la recombinaison homologue (figure 12).

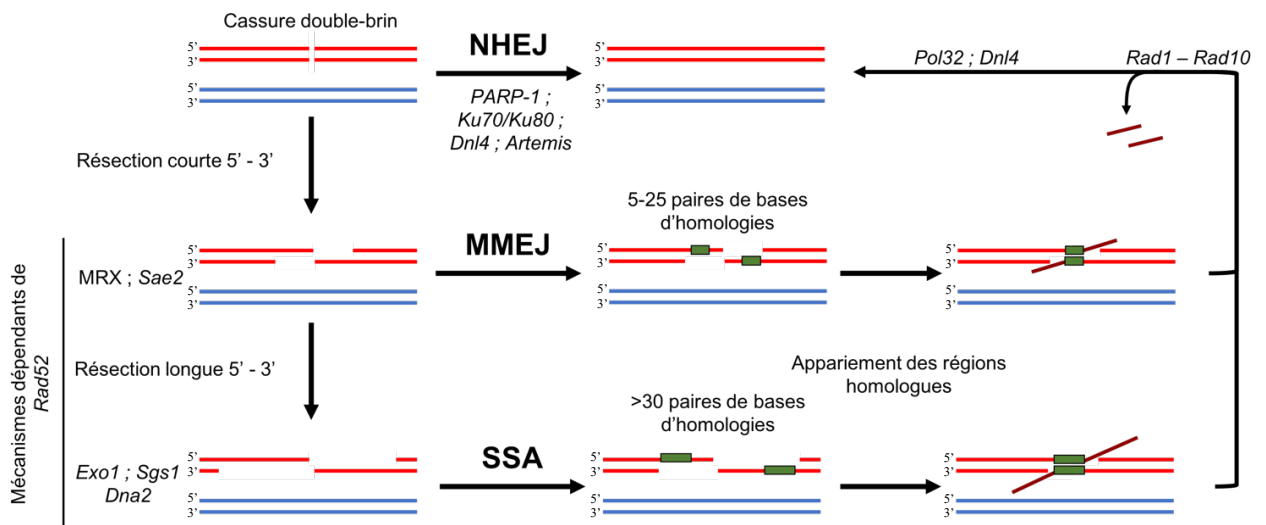


Figure 12 : Schéma représentant les mécanismes de réparation du *Non-Homologous End Joining* (NHEJ), du *Micro-homology Mediated End Joining* (MMEJ) et du *Single Strand Annealing* (SSA). Le NHEJ est indépendant des mécanismes de recombinaison et dépend de la fixation du complexe Ku et DNA-Pk afin d'amorcer le recrutement et l'activation de la nucléase Artemis (qui permet la digestion des extrémités de la cassure afin de rendre plus favorable la ligation des extrémités) et de la ligase IV qui permet de refermer la cassure. Le MMEJ et le SSA sont tous deux dépendants de *Rad52*. En effet ces deux mécanismes dépendent de la présence de région d'homologie de part et d'autre de la cassure. Les régions d'homologies libérées par la résection des extrémités 5' (par le complexe *MRX-Sae2*, *Sgs1*, *Dna2* et *Exo1*) sont appariées via l'action de *Rad52*. Les flaps 3' générés par l'appariement des séquences homologues sont digérés par *Rad1*-*Rad10*. Ce sont la polymérase δ et la ligase IV qui terminent la réparation.

Ce sont les protéines *Ku70* et *Ku80* en complexe (Ku) qui vont se fixer sur les extrémités de la cassure et servir d'antennes pour le recrutement des autres protéines de réparation par le NHEJ. L'importance du complexe Ku a été mise en évidence lorsque la perte de la protéine *yKu80* et du NHEJ rend les cellules incapables de réparer une cassure double-brin chez la levure *S. cerevisiae*¹⁷⁸. Il est intéressant de noter que la présence du complexe MRX chez la levure, est

nécessaire pour le fonctionnement du NHEJ, les mutants *mre11*, *xrs2* ou *rad50* présentant le même phénotype qu'une souche *yku70* ou *yku80* mutante¹⁷⁹. C'est Xrs2, par son domaine FHA (*Fork-Head Associated*⁶), qui permet d'éliminer Ku des extrémités des cassures double-brin et de terminer la réparation ainsi que d'interagir avec Lif1 du complexe ligase DNL4. Chez l'humain une troncation de la protéine Nbs1 (homologue de Xrs2) entraînant la perte du domaine FHA, conduit au développement du syndrome de cassure de Nijmegen¹⁸⁰.

5.1.1.1 Réparation dépendante des nucléases

La *DNA-dependent protein kinase catalytic subunit* ou DNA-PKcs possède une forte affinité pour le complexe Ku *via* un domaine d'interaction spécifique en N-terminal¹⁸¹. Une fois recrutée au niveau des cassures, elle forme le complexe DNA-PK avec Ku. Le complexe DNA-PK est une sérine/thréonine kinase qui joue un rôle central dans le NHEJ, servant de pont entre les deux extrémités de la cassure¹⁸² et active un ensemble de protéines nécessaires à la réparation *via* son activité kinase (composition du complexe revue dans¹⁸³). De manière intéressante, en présence de Ku, DNA-PK peut se fixer sur n'importe quelle extrémité, mais l'activation de son activité kinase dépend des extrémités sur lequel le complexe se fixe¹⁸⁴. Une cassure double-brin peut présenter des extrémités incompatibles qui devront être résectées afin d'être réparées. Chez l'humain, c'est la nucléase Artemis qui va venir digérer les extrémités de la cassure, aucun homologue n'existe chez la levure. DNA-PK forme un complexe avec Artemis qui sera activé par phosphorylation¹⁸⁵. Artemis est une exonucléase avec une polarité 5'-3' et ne possède que peu d'affinité pour l'ADN double brin (pas d'activité endonucléase) en absence de DNA-PK. Dès lors qu'Artemis se retrouve associé à DNA-PK son activité est modifiée et présente une activité endonucléase capable de digérer des extrémités sortantes 5' (laisse une extrémité franche) ou 3' (laisse une extrémité sortante de 4 paires de bases). DNA-

⁶ Domaine de reconnaissance des phosphopeptides

PK confère de plus une activité nucléase capable d'ouvrir les structures en épingles à cheveux à Artemis¹⁸⁵. Artemis sera capable de reconnaître les régions de transitions double-brin vers simple brin et de cliver l'ADN à leurs jonctions¹⁸⁶. Les extrémités digérées par Artemis peuvent devenir un substrat du complexe *XRCC4/ligase IV* (*XRCC4* stimule l'activité de la *ligase IV* correspond à *LIFI* chez la levure¹⁸⁷, la ligase IV étant codée par le gène *DNL4*¹⁸⁸. *XLF* et *PAXX* sont deux acteurs intervenant dans la ligation en favorisant l'activité de la ligase IV et présentent tous les deux une homologie structurale avec *XRCC4*. *PAXX* favorise la ligation des extrémités franches tandis que *XLF* intervient dans la ligation des extrémités sortantes 5' ou 3' (revu dans¹⁸⁹). Lorsqu'une extrémité 3' doit être digérée puis réparée par le NHEJ, la polymérase μ génère un fragment de 4 paires de bases homologue à l'extrémité 3' sortante laissée par Artemis. Une fois les deux séquences appariées la polymérase λ synthétise les fragments manquants de part et d'autre de la région homologue appariée¹⁸⁹. L'information portée par les extrémités sortantes digérées par Artemis est perdue. Il faut noter que le NHEJ n'est pas fondamentalement mutagène, mais peut le devenir en fonction de la nature des extrémités de la cassure-double brin¹⁷³.

Bien que les protéines centrales au NHEJ soient très conservées chez les eucaryotes, il existe des différences entre l'humain et la levure. Le complexe MRX est indispensable à la réparation par le NHEJ chez *S. cerevisiae*¹⁷⁹ tandis que l'absence du complexe MRN n'empêche pas la réparation chez l'humain, bien que moins efficace¹⁹⁰. Si le complexe MRX est indispensable chez la levure, c'est par sa fonction de pont entre les deux extrémités de la cassure, assuré par DNA-PK chez l'humain¹⁸². L'homologie de séquence de DNA-PK avec Tel1 et Mec1 suggère que la fonction de DNA-PK dans la levure est reprise par ces protéines¹⁹¹.

5.1.2 *Micro-homology Mediated End Joining (MMEJ)*

Le MMEJ est un mécanisme favorisant les erreurs et repose sur de courtes séquences

d'homologies de 5 à 25 paires de bases situé de chaque côté de la cassure double brin. De plus ce mécanisme est indépendant de *RAD52* (figure 12) ^{192,193}. Chez l'humain le MMEJ est nécessaire au *class-switch recombination* (CSR), qui consiste en la maturation des IgM en IgG, IgE ou IgA ¹⁷¹. Bien que les protéines centrales du NHEJ soient dispensables pour le bon déroulement du MMEJ, la présence de Sae2/CtIP et du complexe MRX/N est indispensable ¹⁹³.

5.1.3 Recombinaison homologue

La recombinaison homologue est le deuxième grand mécanisme par lequel la cellule va initier la réparation d'une de ses cassures double-brin. Longtemps considérée sans erreurs, puisque la recombinaison fait appel à une séquence homologue pour la réparation, il a été montré qu'elle pouvait être source d'instabilité et de mutations. Lors de l'envahissement de la séquence homologue, quelques hétérologies sont tolérées (mésappariements) par *RAD51* ¹⁹⁴. L'information génétique pourra donc être modifiée sur la séquence réparée, par conversion génique, si ces hétéroduplex ne sont pas reconnus et réparés par la cellule ¹⁹⁵. De la même façon, il est possible d'intégrer une séquence qui n'existait pas à ce locus ou encore de réarranger les allèles par *crossing-over* ¹⁹⁶. La conversion génique peut également être source d'instabilité dans les régions répétées du génome et ce mécanisme sera détaillé plus tard. Cependant, la recombinaison homologue reste un mécanisme indispensable à la ségrégation méiotique des chromosomes et par la formation de *crossing-over* à la diversité génétique des organismes sexués.

Il existe plusieurs mécanismes de réparation utilisant la recombinaison homologue. Tous utilisent chez la levure le médiateur *RAD52*, dont l'homologue fonctionnel (mais non structural) chez l'humain est BRCA2 ¹⁹⁷. On distingue ainsi les mécanismes de recombinaison homologue dépendants de *RAD51* tel que le *double-strand break repair* (DSBR), le *synthesis dependent strand annealing* (SDSA) et le *break induced replication* (BIR), et des mécanismes

indépendants de *RAD51* mais dépendants de *RAD52* tel que le *single-strand annealing* (SSA). Finalement certains mécanismes sont indépendants de *RAD51* et *RAD52* tel que le *microhomology end-joining* (MMEJ), qui fait appel -comme son nom l'indique- à des microhomologies et qui est apparenté au Non-Homologous End Joining. Bien que ces différents mécanismes aient chacun leurs spécificités moléculaires, une étape de résection de la cassure double-brin, fabriquant des extrémités 3'-OH simple brin est dans tous les cas la première étape de tous ces mécanismes.

5.1.3.1 Initiation de la résection

La majorité des études visant à comprendre les mécanismes de résection ont été réalisées dans la levure *Saccharomyces cerevisiae*. La résection est l'étape initiale visant à dégager des extrémités 3' sortantes nécessaires pour la recherche d'homologie et/ou l'appariement à une séquence homologue. Elle est initiée par le complexe MRX-Sae2 chez *Saccharomyces cerevisiae* (MRN chez *Homo sapiens*). Ce complexe est composé des protéines Mre11, Rad50, Xrs2 et Sae2 (figure 13).

La protéine Mre11 possède, *in vitro*, à la fois une activité exonucléase 3'-5'¹⁹⁸, qui lui permet de résecter les extrémités 3' avec une efficacité 50 fois supérieure pour les extrémités franches et entrantes que sortantes. Cette activité est dépendante de la présence de manganèse et est fortement stimulée par la présence de Rad50¹⁹⁸. Le complexe Mre11-Rad50 est orthologue au complexe SbcC-SbcD de la bactérie *E. coli*, qui possède à la fois une activité exonucléase double-brin et endonucléase simple brin. Bien que SbcD porte les domaines catalytiques de l'activité nucléase, son activité exonucléase dépend de la présence de SbcC¹⁹⁹. Le rôle principal du complexe SbcCD est d'éliminer les structures secondaires de type épingle à cheveux formées lors de la réplication du génome bactérien²⁰⁰.

L'activité endonucléase de Mre11 est nécessaire à la réparation des cassures double-brin méiotiques réalisées par Spo11, des cassures dont les extrémités sont bloquées par des protéines tel que Ku ou des topoisomérases, ou encore par des structures secondaires empêchant l'accès aux autres nucléases impliquées dans la résection (Exo1 et Dna2) aux extrémités de la cassure²⁰¹⁻²⁰⁵. L'activité endonucléase de Mre11 est dépendante de l'hydrolyse d'ATP par Rad50, de la présence de manganèse ainsi que de Sae2 qui joue un rôle critique dans la stimulation de l'activité endonucléase de Mre11²⁰⁶ (figure 13). En présence d'ATP le complexe MRX adopte une conformation fermée qui favorise la fixation à l'ADN et la liaison des deux extrémités de la cassure²⁰⁷. Après hydrolyse de l'ATP le complexe MRX adopte une conformation ouverte qui favorise la résection²⁰⁷ (figure 13). De plus il a été montré que la conformation fermée de MRX est nécessaire à l'activation de Tel1 par le complexe²⁰⁸.

L'élimination des blocages des extrémités d'une cassure est observée *in vitro* uniquement lorsque Sae2 est présent avec un complexe MRX fonctionnel et en présence de blocage²⁰⁶. *In vitro* le complexe MRX génère un simple *nick* 15-20 nucléotides en amont des extrémités d'une cassure double-brin bloquées par un complexe streptavidine-biotine²⁰⁶. *In vivo* l'ADN est incisé plus loin en amont des extrémités, jusqu'à 300 nucléotides lors de la résection de cassures générées en méiose²⁰⁹. Le *nick* ainsi généré permet à Mre11 d'utiliser son activité exonucléase 3'-5' afin de dégrader le fragment entre le site de cassure simple brin et l'extrémité de la cassure²⁰⁹. Une perte d'une des deux activités catalytiques de Mre11 bloque complètement la résection (phénotype observé avec les mutations D56A/N et H125A/N).

Mre11 porte 5 domaines phosphoestérase responsables de son activité nucléase, localisés en N-terminal de la protéine et un domaine d'interaction à l'ADN au niveau C-terminal²¹⁰. Il a été montré que plusieurs mutations au sein du domaine N-terminal étaient capables de perturber l'activité nucléase de Mre11 sans affecter sa capacité à se fixer sur une séquence d'ADN²¹¹. Les trois mutations les plus étudiées affectant l'activité nucléase de Mre11, sont les mutations

ponctuelles D16A, D56A/N et H125A/N dans les motifs phosphoestérase I, II et III respectivement. La hiérarchie de la résistance aux dommages par irradiation de ces mutants est résumée dans une étude du laboratoire de Lorraine Symington : les mutations *mre11D56N* et *mre11H125N* présentent la même sensibilité aux radiations et sont moins sensibles que les mutations *mre11D16N*, *mre11D56A* et *mre11H125A*. *mre11D16A* est la mutation qui déstabilise le plus l'activité de Mre11 mais reste moins sensible qu'un mutant nul *mre11Δ*. Les mutations au sein du motif I avec le phénotype de sensibilité aux radiations le plus important, déstabilisent la formation du complexe MRX²¹¹. La mutation H59S abolit l'activité exonucléase de la protéine en affectant peu l'activité endonucléase²⁰⁹ ce qui a permis de mieux comprendre le rôle de chacune des deux activités.

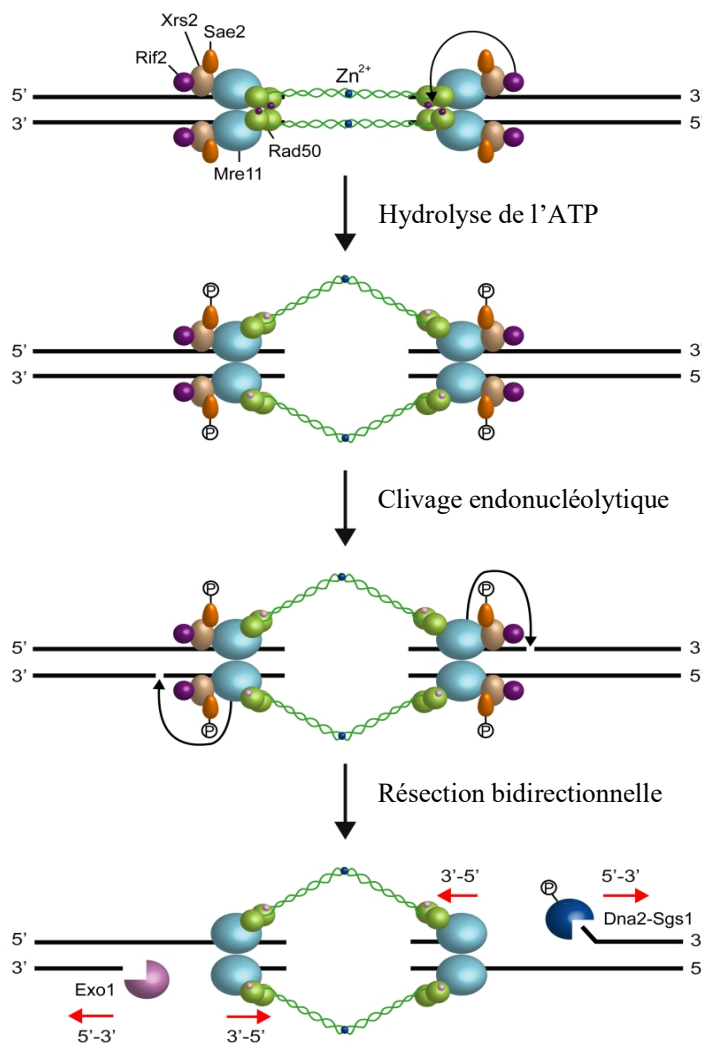


Figure 13 : modèle de résection d'une cassure double-brin. Le complexe MRX est recruté à la cassure, lié à l'ATP (point violet). Sous cette forme Rad50 inhibe l'activité nucléase de Mre11. Après hydrolyse de l'ATP en ADP (point rose) par Rad50, le site catalytique de l'activité endonucléase de Mre11 devient apparent et génère une cassure en amont de l'extrémité de la cassure après activation par Sae2. Rif2 est montré interagissant avec le complexe MRX et est proposé comme favorisant l'hydrolyse de l'ATP par Rad50 par les auteurs de la publication. Sae2 est activé après phosphorylation par Cdk1 ce qui permet de limiter la résection aux phases S et G2 du cycle cellulaire. L'incision générée par Mre11 permet aux nucléases Exo1 et Dna2 (en tandem avec Sgs1) de résecter l'ADN dans la direction 5'-3'. Finalement le complexe MRX dégrade l'ADN dans la direction 3'-5', vers l'extrémité de la cassure double-brin, via son activité exonucléase. Tiré de Gobbi et al. 2016.

Le rôle de la protéine Rad50 est à la fois structurel et fonctionnel. Rad50 interagit avec Mre11 et permet la dimérisation de deux complexes MRX via une autre molécule Rad50 par son domaine *hook*²¹². Rad50 possède une activité ATPase nécessaire pour le fonctionnement de Mre11. Deux domaines de liaisons à l'ATP sont retrouvés en position N-terminale et C-terminale de Rad50. Les domaines « *Walker A* » et « *Walker B* » s'assemblent afin de former un domaine catalytique fonctionnel. L'inactivation du domaine « *Walker A* » par insertion de mutations (K40R, K40A, K40E), entraîne la perte de l'activité ATPase de Rad50, empêchant l'interaction du complexe MRX avec l'ADN et inhibant l'activité nucléase de Mre11²¹³. L'hydrolyse de l'ATP par Rad50 active le domaine nucléase de Mre11 et permet la résection²¹⁴ (figure 13). L'inactivation de *RAD50* dans un mutant nul conduit à une destruction totale de l'assemblage du complexe MRX et une absence totale de résection, se traduisant par une forte sensibilité aux agents endommageant l'ADN à la fois dans les régions non répétées et répétées du génome^{214,215}.

Xrs2 (homologue fonctionnel, mais non structurel de Nbs1 chez l'humain) est nécessaire à la stabilisation du complexe. Son l'absence entraîne une perte de la localisation nucléaire de MRX. L'ajout d'un tag NLS (signal d'adressage au noyau) sur Mre11 restaure le phénotype sauvage en absence de *XRS2/NBS1*^{216,217}. De plus Xrs2 est nécessaire au recrutement de Tell1, PI(3)K kinase, dont l'orthologue chez les mammifères est ATM, au niveau de la cassure ou aux télomères par son domaine C-terminal²¹⁸.

Les modèles actuels suggèrent que le complexe MRX recrute Tell1 au niveau de la cassure entraînant son activation par autophosphorylation stimulée par la présence de protéines bloquant les extrémités de la cassure^{219,220}. Tell1 phosphoryle à son tour le complexe MRX qui initie la résection. Les mutations *mre11Δ*, *rad50Δ* ou *xrs2Δ* conduisent à un raccourcissement des télomères similaire à celui d'une souche *tell1Δ* suggérant que ces protéines interviennent dans le même mécanisme de régulation de la taille des extrémités chromosomiques²²¹. Bien

que le complexe MRX joue un rôle épistatique sur Tel1 lors de la réparation de dommages à l'ADN, il est intéressant de constater que ce n'est pas le cas lors de la régulation de l'élongation des télomères en présence d'une forme constitutivement active de Tel1 ²²². Cependant les mécanismes généraux de recrutement et d'activation de Tel1 par le complexe MRX aux télomères restent similaires.

5.1.3.2 *SAE2*

Le complexe MRX fonctionne en tandem avec la protéine Sae2 (*CtIP* chez l'humain, considérée comme la « quatrième protéine » du complexe MRX) qui interagit via son domaine C-terminal et active l'activité endonucléase de Mre11 en plus de stimuler son activité exonucléase ^{206,223} (figure 13). Cette interaction est dépendante de la présence d'ADN formant un complexe MRX-Sae2-ADN ²²³. La perte d'expression de *SAE2* entraîne un phénotype de perte totale de la résection similaire à celui d'une souche *rad50Δ* ²²⁴. Une forme phosphorylée de Sae2 est détectable périodiquement durant les phases S, G2 et M du cycle cellulaire ainsi qu'en présence de dommages à l'ADN ²²⁵. L'inhibition de l'expression de Cdc28 ou la mutation de la sérine 267 de Sae2, motif de phosphorylation CDK, entraîne une forte augmentation de la sensibilité aux dommages à l'ADN de la cellule ²²⁶. La phosphorylation de la sérine 267 par Cdc28 est nécessaire au bon déroulement de la résection ²²⁶.

En présence de facteurs endommageant l'ADN tel que le MMS⁷, l'hydroxyurée (HU)⁸ ou la bleomycine⁹, il est possible d'observer la forme phosphorylée de Sae2, indépendamment du déroulement du cycle cellulaire, après synchronisation en phase G1 de cellules de levures par le facteur α . La perte d'expression de *MEC1* ou de *TEL1* (sérine/thréonine PI(3)K kinases,

⁷ Agent alkylant entraînant des dommages à l'ADN par modification de la guanine en 7-méthylguanine et de l'adénine en 3-méthyladénine, responsable de mésappariements et de blocages de réplication

⁸ Ralentit la réplication de l'ADN en diminuant le *pool* de nucléotides

⁹ Agent radiomimétique, entraîne la génération de cassures simple brin à faible concentration et de cassures double-brin à forte concentration

appartenant au mécanisme de réponse aux dommages de l'ADN, respectivement homologues de ATR et de ATM) entraîne une diminution de la phosphorylation de Sae2 tandis que la double inactivation des gènes *MEC1* et *TEL1* entraîne la perte totale de la phosphorylation de Sae2²²⁵. Une baisse de la phosphorylation est aussi détectée dans un mutant *rad50Δ*. La double mutation *rad50Δ tel1Δ* entraîne une diminution de la phosphorylation de Sae2 similaire aux simples mutations *tel1Δ* et *rad50Δ* tandis que la double mutation *mec1Δ rad50Δ* conduit à une perte totale de la phosphorylation indiquant que Rad50 et Tel1 font partie du même mécanisme de phosphorylation de Sae2 tandis que Mec1 appartient à une voie métabolique différente²²⁵.

Sae2 se dimérise via un motif LKEX4E(V/L) porté en N-terminal de la protéine, ce motif étant critique pour son recrutement à la cassure et pour le bon déroulement de la résection. Le domaine de dimérisation est aussi critique pour l'activation par phosphorylation de Sae2 par *TEL1*²²⁷⁻²²⁹. Le laboratoire de Tanya Paull a proposé que Sae2, en plus de stimuler l'activité exonucléase de Mre11, posséderait une activité endonucléase propre, stimulée par le complexe MRX et nécessaire à l'élimination des structures en épingles à cheveux pouvant bloquer les extrémités des cassures double-brin²²³. Sae2 serait donc en réalité directement responsable de l'activité endonucléase du complexe MRX. Ces résultats font encore débat aujourd'hui²⁰⁶ et ont été une des questions de mon projet de thèse.

5.1.4 Etapes tardives de la résection

Les exonucléases 5'-3' Exo1 et Dna2 (en tandem avec Sgs1) viennent réaliser une deuxième étape de résection, plus longue. Le complexe RPA interagit avec les extrémités simple brin afin de les protéger. Cette deuxième étape de résection libère les longues extrémités nécessaires à la recherche d'homologie.

5.1.4.1 EXO1

Exo1 est une exonucléase 5'-3', appartenant à la famille des nucléases XPG avec Rad2 (homologue de XPG chez la levure), Rad27/FEN-1 et Yen1/GEN1. Elle est aussi capable de cliver des flap 5' ²³⁰⁻²³². Exo1 intervient dans le MMR en interagissant avec *MSH2*. Les mutations *msh2Δ*, *mlh1Δ*, et *pms1Δ* sont épistatiques sur *exo1Δ*, indiquant que ces gènes fonctionnent bien dans la même voie métabolique ²³³. Exo1 a été rapidement montrée comme étant impliquée dans les événements de recombinaison mitotique et méiotique ^{234,235}.

La protéine Exo1 est directement recrutée au niveau de la cassure par le complexe MRX indépendamment de l'activité nucléase du complexe ²⁰³. Bien qu'il ait été montré qu'elle avait une préférence pour les extrémités 5' laissées par le complexe MRX ²³⁶, en absence du complexe MRX et de Ku fixé aux extrémités de la cassure, qui inhibe l'activité d'Exo1, cette dernière est capable d'initier seule la résection ^{203,237}. Le complexe MRX, en éliminant Ku des extrémités de la cassure, permet à Exo1 d'engager la résection ²⁰³ (figure 13). Le même rôle inhibiteur de Ku sur l'activité nucléase d'Exo1 est aussi observé aux télomères ²³⁸.

Le complexe RPA protège les extrémités simple brin sortantes de la dégradation par Exo1 tout en stimulant son activité de résection sur l'extrémité 5' ²³⁶. En présence de RPA et du complexe MRX-Sae2, la stimulation de l'activité exonucléase d'Exo1 est encore plus importante et est indépendante de l'activité nucléase de Mre11 ^{236,239}. La double mutation *exo1Δ mre11Δ* augmente la sensibilité aux radiations ainsi que les retards de croissance cellulaire tandis que la double mutation *exo1Δ mre11H125N* entraîne une augmentation de la sensibilité aux radiations, mais pas de retard de croissance ²⁴⁰. La surexpression d'Exo1 est capable de partiellement compenser la sensibilité aux radiations ainsi que de partiellement restaurer le retard dans le changement de type sexuel chez la levure, induit par la mutation *mre11Δ*. Ces résultats suggèrent qu'Exo1 et Mre11 ont une activité partiellement redondante sur la résection

mais que d'autres nucléases interviennent dans la réparation ²⁴⁰. La perte d'expression du gène *DNA2* induit des défauts de résection comparables au simple mutant *sgs1Δ* et au double mutant *sgs1Δ dna2Δ pif1-m2* ²⁴¹. Ces résultats suggèrent que Sgs1 et Dna2 agissent au sein du même mécanisme de résection. Une souche portant la mutation *exo1Δ*, avec le gène *DNA2* sous contrôle d'un promoteur inductible conduit à une perte totale de la résection longue similaire à un double mutant *exo1Δ sgs1Δ*. Ces résultats montrent qu'Exo1 et Dna2 ont une activité redondante pendant la résection ²⁴¹.

5.1.4.2 *SGS1*

Les hélicases RecQ de la superfamille SF2, conservée dans tout le vivant, comprennent les hélicases RecQ chez *Escherichia coli*, Sgs1 chez *Saccharomyces cerevisiae* ²⁴² et 5 représentants chez l'humain avec notamment *BLM* et *WRN*. La perte de ces hélicases conduit à des pathologies liées au cancer et au vieillissement ²⁴³. Leurs structures se ressemblent et possèdent toutes un domaine hélicase (portant une boîte DEAH) qui permet d'ouvrir l'ADN (ou l'ARN) d'une manière dépendante de l'ATP (la famille des hélicases DEAH peut hydrolyser tous les dNTPs) ²⁴⁴. Il est intéressant de noter que Sgs1 (et son homologue BLM) intervient à la fois dans la résection initiant les différents mécanismes de réparation par recombinaison homologue ainsi que dans la résolution des jonctions de Holliday qui sont des intermédiaires de réparation du mécanisme du *double-strand break repair* (DSBR) ²⁴⁵. Des expériences d'immunoprécipitation suggèrent que c'est le complexe MRX qui recrute Sgs1 au niveau des cassures via une interaction avec Mre11 ²⁴⁶.

Sgs1 fonctionne en complexe avec les protéines Rmi1 et Top3. Sgs1 ouvre la double hélice d'ADN tandis que Top3 (une topoisomérase) retire les tensions sur l'ADN en complexe avec Rmi1 qui reconnaît certaines structures spécifiques de l'ADN tel que des jonctions de Holliday ou l'ADN simple brin ^{247,248}. Chacun des éléments du complexe Sgs1-Top3-Rmi1 stabilise le

complexe²⁴⁸. Sgs1 est capable d'ouvrir des séquences portant des extrémités 5' ou 3' sortantes ainsi que des extrémités franches (figure 14). Malgré tout, ce sont les jonctions de Holliday qui sont le plus efficacement résolues par Sgs1²⁴⁹. La présence de RPA stimule l'activité de Sgs1 et favorise l'ouverture des séquences avec une polarité 3'-5'. L'absence de RPA entraîne une forte diminution de l'activité de Sgs1²⁴⁹.

Sgs1 fonctionne en tandem avec la nucléase Dna2, dont l'activité et l'affinité à l'ADN sont stimulées à la fois par le complexe Top3-Rmi1 et le complexe MRX²⁵⁰. Ainsi en l'absence d'Exo1, Sgs1 est nécessaire pour mener à bien la résection, et la perte d'expression des gènes *EXO1* et *SGS1* entraîne une perte complète de la résection. La perte d'expression de *EXO1* ou de *SGS1* et *DNA2* n'entraîne pas de diminution de la résection en comparaison d'une souche sauvage. Ces résultats indiquent qu'Exo1 et Sgs1-Dna2 interviennent dans deux voies métaboliques distinctes réalisant la résection^{236,251}.

5.1.4.3 *DNA2*

La protéine Dna2 est importante dans un ensemble de mécanismes et processus métaboliques. On la retrouve dans la réparation des cassures double-brin de l'ADN (figure 12), la maintenance télomérique ainsi que dans la réplication de l'ADN en phase S chez *S. cerevisiae*. Etant un gène essentiel, les souches *dna2Δ* ne sont pas viables²⁵². Dna2 possède à la fois une activité hélicase, exo- et endonucléase avec une préférence pour les extrémités libres 5' ou 3' et peut aussi digérer les *flaps* 5' ou 3' avec une préférence pour les *flaps* 3'²⁵³. Il est cependant observé que Dna2 est capable de cliver avec la même efficacité les séquences portant à la fois un flap 5' et un flap 3'. Ces résultats indiquent que Dna2 est capable de reconnaître certaines structures particulières de l'ADN²⁵³. L'activité hélicase de Dna2 est faible et ne permet pas d'ouvrir un duplex d'ADN au-delà de quelques nucléotides²⁵⁴. C'est en réalité l'activité nucléase de Dna2 qui bloque son activité hélicase. Une perte de l'activité nucléase ou une absence d'ATP (nécessaire au

fonctionnement du domaine catalytique nucléase) entraîne une augmentation de l'activité hélicase^{254,255}. Les séquences présentant une extrémité 5' sortante sont les substrats favoris de l'activité hélicase de Dna2 indiquant que son activité hélicase présente une polarité 5'-3'²⁵⁵.

RPA joue aussi un rôle dans la discrimination du brin digéré par Dna2. En absence de RPA, Dna2 digère les extrémités 5' et 3' sortantes de façon équivalente. En présence de RPA, les extrémités 3' sortantes sont protégées de la dégradation mais pas les extrémités 5' sortantes. Ces résultats expliquent le biais de digestion observé lors de la résection par Dna2. La nucléase ne digère cependant que les extrémités simple brin et ne se montre pas capable d'ouvrir l'ADN en absence de Sgs1²⁵⁰. De plus en absence de RPA, Dna2 n'est pas recruté au niveau de la cassure²⁵⁶.

Le recrutement de Dna2 à la cassure dépend de la présence physique du complexe MRX. Les mutations *mre11Δ* ou *rad50Δ* entraînant une perte du recrutement de Dna2 à la cassure, tandis que la mutation nucléase déficiente *mre11-3* n'entraîne pas de diminution du recrutement de Dna2 à la cassure²⁰³. Le complexe MRX et Dna2 présentent une activité redondante sur la résection d'extrémité propre^{203,237}. Le modèle suggéré par Cejka et Kowalczykowski propose qu'après son recrutement par le complexe MRX, Sgs1 ouvre l'ADN et les séquences simples brins générées se recouvrent de RPA. Dna2 par interaction avec Sgs1, dégrade l'extrémité 5', son activité étant stimulée par RPA²⁵⁰.

5.1.5 Mécanismes dépendants de *RAD51* et *RAD52*

Une fois les extrémités 3' sortantes libérées par la résection, la cellule peut engager différents mécanismes de réparation par recombinaison homologe, afin de réparer la cassure double brin. Ces mécanismes sont le *double-strand break repair* (DSBR) classique, le *synthesis-dependent strand annealing* (SDSA) ainsi que le *break-induced replication* (BIR). L'acteur partagé par

tous ces mécanismes est *RAD51*. Rad51 envahit les extrémités simple brin 3' sortantes laissées par la résection et catalyse, de manière dépendante de l'ATP, l'appariement de séquences homologues en présence de RPA ²⁵⁷. La formation de nucléofilaments de Rad51 est dépendante de la présence de Rad52 qui stimule le remplacement de RPA par Rad51 sur l'ADN simple brin ^{258,259}. Il est intéressant de noter que Rad52 interagit directement avec RPA. De plus RPA stimule à la fois l'activité enzymatique de Rad51 et de Rad52 ²⁶⁰. La perte d'expression de *RAD52* entraîne une perte de la formation des foci Rad51 co-localisant avec les cassures. A l'inverse, dans un mutant *rad51Δ*, on observe toujours des foci Rad52 aux cassures. Ces foci Rad52 ne sont pas capables de quitter la cassure en absence de Rad51 ²⁶¹.

On divise en trois phases les étapes communes aux mécanismes de réparation par recombinaison homologue. L'étape pré-synaptique correspond à la formation de nucléofilaments de Rad51 le long des séquences d'ADN simple brin. La synapse est formée par l'ouverture de l'ADN double brin portant la séquence d'homologie nécessaire à la réparation. L'ouverture de la séquence et l'envahissement par le brin portant Rad51 correspondent à la formation d'une *D-loop* et à l'étape de formation de la synapse. Une *D-loop* est un intermédiaire de réparation dans lequel deux brins complémentaires de l'ADN donneur sont déshybridés sur plusieurs bases et stabilisés par un troisième brin invasif provenant du locus endommagé. Finalement, la synthèse d'ADN démarre à partir de l'extrémité 3' du brin invasif qui sert d'amorce initiant la dernière phase post-synaptique (revu dans ²⁶²). Une fois l'étape post-synaptique amorcée différents mécanismes de réparation sont possibles.

5.1.5.1 Réparation d'une cassure double-brin par conversion génique

La conversion génique est la forme la plus courante de recombinaison homologue initiée après cassure double-brin. Elle correspond à l'échange de matériel génétique non réciproque entre une séquence homologue à la région portant la cassure double-brin. Dans le modèle du *double-*

strand break repair (DSBR) proposé par Szostak, Orr-Weaver, Rothstein et Stahl ²⁶³ il y a bientôt 40 ans, la synthèse d'ADN initiée à partir de l'extrémité 3' du brin invasif se poursuit jusqu'à ce que des séquences d'homologie avec la deuxième extrémité de la cassure apparaissent. L'extrémité invasive est alors capturée par l'autre extrémité 5' de la cassure. Au fur et à mesure de la synthèse d'ADN la *D-loop* s'agrandit jusqu'à ce que l'extrémité 3' non invasive de la cassure double-brin s'apparie avec la séquence simple brin de la *D-loop* et initie à son tour la synthèse d'ADN (figure 14). Il en résulte la formation de deux jonctions de Holliday ²⁶³.

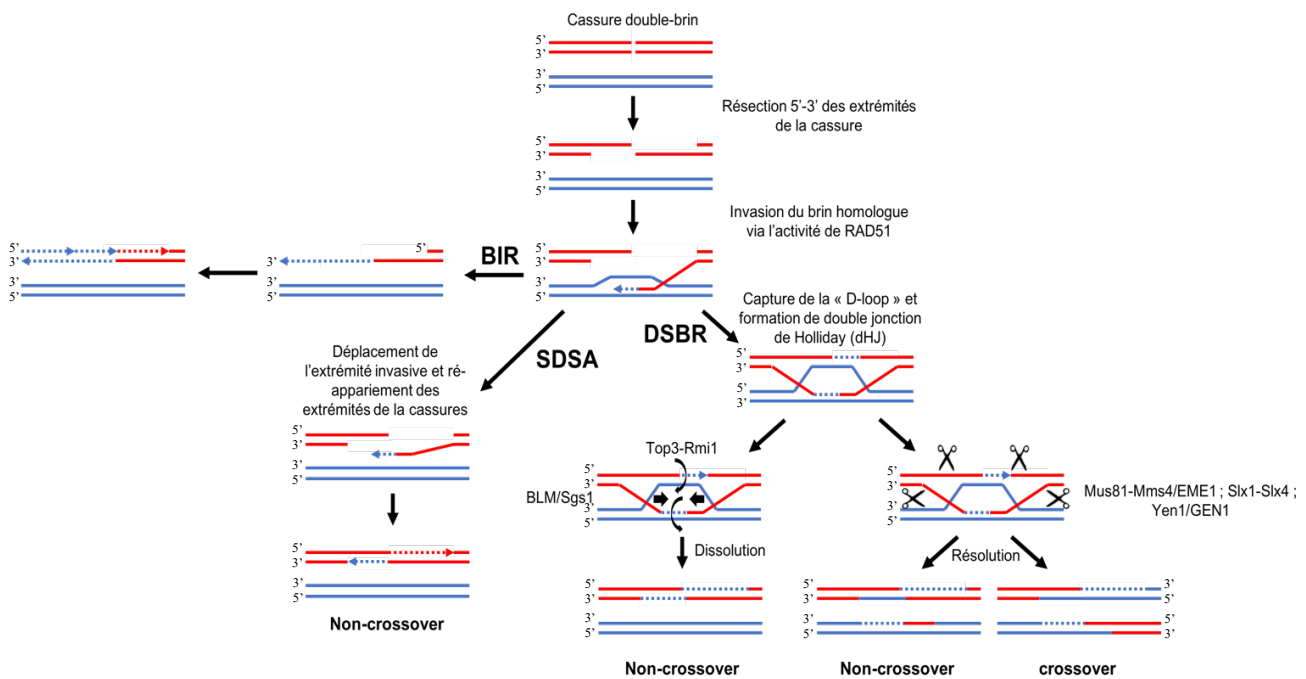


Figure 14 : Schéma des mécanismes de réparation par recombinaison homologue. Une première étape de résection génère des extrémités 3' libres sortantes. Rad51 peut lier ces extrémités 3' et initier l'envahissement et le déplacement d'un brin d'ADN homologue, formant une *D-loop* et une première jonction de Holliday. Dans le DSBR (*Double-strand break repair*), le brin apparié servira d'amorce à la synthèse sur le brin donneur et formera une deuxième jonction de Holliday (dHJ) lors de la fermeture de la séquence. L'autre extrémité de la cassure interagit avec la *D-loop* afin de synthétiser le fragment d'ADN manquant. La réparation se termine par la résolution ou la dissolution des dHJ, qui peut aboutir sur un *crossover* (CO) ou sans *crossover* (NCO). La dissolution, qui se termine toujours par un NCO, fait intervenir à la fois les hélicases BLM/Sgs1 qui vont rapprocher les jonctions tandis que le complexe Top3-Rmi1 déroule l'ADN entre les jonctions. La résolution des dHJ se fait par le clivage nucléolytique des jonctions par les résolvases Mus81-Mms4/EME1 ; Slx1-Slx4 ; Yen1/GEN1. Une coupure sur le brin interne d'une jonction et sur le brin externe de l'autre jonction aboutit sur un CO, une coupure sur le même brin de chaque jonction aboutit sur un NCO. Dans le SDSA (*Synthesis-dependant strand annealing*) l'extrémité ayant envahi le brin donneur est déplacée et ré-appariée à l'autre extrémité de la cassure. Le BIR (*Break-induced replication*) consiste en la formation d'une fourche de réplication après invasion du brin donneur. La synthèse se poursuivra jusqu'à l'extrémité du chromosome et aboutit à une perte d'hétérozygotie (LOH) sur le chromosome portant la cassure.

Une jonction de Holliday, un intermédiaire de la recombinaison homologue, est une jonction mobile entre deux molécules d'ADN qui doivent être dissociées afin de terminer la réparation. Cette jonction tire son nom éponyme de Robin Holliday qui, le premier, suggéra dans les années 1960 une telle structure en tant qu'intermédiaire de recombinaison menant aux *crossing-over*. La preuve physique de l'existence d'une double jonction de Holliday méiotique sera démontrée des décennies plus tard par Tony Schwacha et Nancy Kleckner ²⁶⁴.

Ces jonctions peuvent être dissociées *via* l'activité de résolvas. Les résolvas chez la levure sont Yen1, Mus81-Mms4 et Slx1-Slx4 qui appartiennent à la famille des nucléases *RAD2/XPG*, avec une spécificité pour les jonctions de Holliday (on ne retrouve pas d'activité au niveau des jonctions par d'autres membres de la même famille comme Exo1 ou Rad27) ²⁶⁵. Dans le modèle classique de Szostak et collaborateurs, la résolution des jonctions peut conduire à un *crossing-over* ou à un *non crossing-over* (figure 14) ²⁶³, en fonction des coupures générées par les résolvas Yen1, Mus81-Mms4 et Slx1-Slx4.

Il est intéressant de noter que Yen1 préfère couper les jonctions de façon symétrique (sans *crossing-over*) tandis que Mus81-Mms4 et Slx1-Slx4 les coupent de manière asymétrique (avec *crossing-over*) ²⁶⁵. Les jonctions peuvent aussi être dissoutes au travers de l'activité des hélicases Sgs1-Top3-Rmi1, Srs2 ou Mph1, sans *crossing-over* ²⁶⁶ (figure 14).

Les *crossing-overs* apparaissant suite à la résolution des jonctions de Holliday par les résolvas Yen1, Mus81-Mms4 et Slx1-Slx4 sont un mécanisme indépendant des *crossing-overs* générés lors de la méiose et sont même dispensables à la résolution de la majorité des intermédiaires de recombinaison ²⁶⁷. En absence de Yen1, Mus81-Mms4 et Slx1-Slx4, bien que la résolution des intermédiaires de recombinaison soit altérée, seul une modeste diminution des *crossing-over* est observée. Les protéines Sgs1, Mlh1-Mlh3 et Exo1 semblent être les facteurs *pro-crossover* lors de la résolution des intermédiaires de recombinaison en méiose ²⁶⁷.

Les modèles les plus récents montrent que les *crossing-over* peuvent émerger de deux mécanismes distincts. Le premier mécanisme que nous venons de détailler repose sur la résolution symétrique ou asymétrique par les résolvases des jonctions de Holliday. Le deuxième mécanisme repose sur l'intégrité du complexe synaptonemal (CS) se déroulant durant la prophase I de méiose. Des cassures double-bris générées de manière non aléatoire le long des chromosomes permettent l'assemblage des chromosomes homologues après résection et invasion des séquences homologues complémentaires ²⁶⁸. Les protéines du complexe ZMM (chez la levure Zip1; Zip2; Zip3; Zip4; Spo16; Mer3; Msh4-Msh5 et Pph3, revu dans ²⁶⁸) sont responsables de la résolution des jonctions en Holliday en *crossing-over* en stabilisant l'invasion de la séquence homologue par l'extrémité 3' de la cassure double-brin, facilitent la recapture du brin invasif par l'extrémité de la cassure, permettant d'engager la résolution vers un *crossing-over* ²⁶⁹. En absence des membres du ZMM on observe une diminution de 80% des événements de *crossing-overs* ^{268,269}, ce qui démontre l'importance de ZMM dans la régulation des *crossing-overs* à la méiose. Les mécanismes favorisant le *crossing-over* par l'intermédiaire du ZMM ne sont pas connus.

Le modèle proposé par Szostak, Orr-Weaver, Rothstein et Stahl ²⁶³, s'il a servi de paradigme pendant près de 40 ans, ne permet néanmoins pas d'expliquer totalement la répartition des événements de *crossing-overs* observé *in vivo* ²⁷⁰. Sous le modèle classique de réparation des cassures double-brin, la résolution des jonctions de Holliday devrait aboutir à une répartition de part et d'autre de la cassure double-brin des hétéroduplex d'ADN. Il a cependant été observé des hétéroduplex d'un côté seulement de la cassure ²⁷¹. D'autre part, le modèle classique prédit que le point d'échange des *crossing-over* doit systématiquement être localisé au niveau des hétéroduplex d'ADN, eux même localisés à la cassure. Le séquençage des sites des cassures induites par Spo11 montre que 11% des événements de *crossing-over* observés ne correspondent pas à un site de cassure ^{270,272}. L'analyse de résolution des intermédiaires de

recombinaison montre que la majorité des événements de résolution avec ou sans *crossing-over* présentent une symétrie des hétéroduplex après résolution. Ces observations suggèrent que les jonctions de Holliday sont dynamiques et capables de migrer le long de l'ADN ²⁷⁰.

5.1.5.2 *Synthesis-Dependent Strand Annealing (SDSA)*

Historiquement, le SDSA a été postulé comme un mécanisme alternatif de recombinaison homologue en 1994 par Nassif et collaborateurs en étudiant les événements de réparation de cassures double-brin chez la drosophile, suite à l'excision d'un élément P ²⁷³. Peu de temps après, ce mécanisme fut proposé pour expliquer les réarrangements d'ADN ribosomique 5S chez la drosophile ²⁷⁴. L'existence du SDSA fut formellement démontrée chez la levure *S. cerevisiae* par Pâques et Haber, en utilisant un élégant mécanisme de réparation d'une cassure double-brin. Un gène *LEU2* tronqué dans sa partie centrale ne pouvait être réparé qu'en allant recopier deux informations complémentaires présentes sur deux chromosomes différents, excluant de fait la possibilité de générer un *crossing-over* et l'obligation de dérouler au moins l'un des brins néosynthétisés lors de la réparation pour l'hybrider avec son complémentaire ²⁷⁵. Ce mécanisme a été montré comme le mécanisme principal générant des expansions et des contractions de répétitions en tandem, que ce soit chez l'humain ou dans la levure, comme il sera expliqué en détail plus loin ²⁷⁶.

D'un point de vue strictement mécanistique, une fois la synthèse d'ADN initiée (étape post-synaptique) la *D-loop* peut se dissocier et l'extrémité 3' du brin invasif recapturé par l'extrémité 5' de la cassure (figure 14). Le mécanisme du *Synthesis-Dependant Strand Annealing (SDSA)* est un mécanisme de conversion génique conservatif qui permet de réparer une cassure double brin sans générer de *crossing-over*. Les hélicases Sgs1-Top3-Rmi1, Srs2 ou Mph1 favorisent le SDSA en ouvrant la première *D-loop* formée lors de la synapse ^{277,278}. Le SDSA est responsable de la majorité des événements de recombinaison sans *crossing-over*, l'ensemble

des évènements de *crossing-overs* en méiose étant dépendants des ZMMs^{268,269}.

5.1.5.3 *Break-Induced Replication (BIR)*

Contrairement aux autres mécanismes de réparation des cassures double-brin où les deux extrémités de la cassure vont envahir simultanément une séquence d'ADN homologue, le *break-induced replication* (BIR) est initié lorsqu'une seule extrémité de la cassure est disponible²⁷⁹. Cela peut se produire lorsqu'une seule extrémité est physiquement capable d'envahir une séquence homologue, lors de l'invasion de deux séquences homologues localisées en *trans*, par chaque extrémité de la cassure ou encore lors d'une cassure due à un arrêt de fourche de réplication (revu dans²⁸⁰). Comme la majorité des mécanismes de réparation par recombinaison, la réparation commence par la résection extensive 5'-3' de la cassure, laissant une extrémité 3'-OH sortante, suivie de la formation de nucléofilaments de Rad51 catalysée par Rad52²⁷⁹ (figure 14). La *D-loop* formée par l'invasion de la séquence complémentaire par l'extrémité 3' de la cassure, formant une jonction de Holliday, se déplace le long de la séquence matrice²⁸¹. La synthèse d'ADN est asynchrone avec une accumulation d'ADN simple brin durant la synthèse du brin invasif qui servira de matrice à l'autre extrémité de la cassure²⁸¹. L'efficacité du BIR dépend de la distance entre la séquence homologue envahie par l'extrémité 3' de la cassure et le plus proche télomère, ainsi que de la présence des polymérasés α , ϵ et δ (notamment de la sous-unité Pol32)^{282,283}. Le laboratoire de Lorraine Symington a montré que le BIR était dépendant de *RAD51*²⁷⁹ alors que ce mécanisme avait initialement été décrit par le laboratoire de Jim Haber comme un processus indépendant de *RAD51*²⁸⁴. Ainsi deux types de BIR peuvent exister, dépendants ou pas de la recombinase Rad51. Le BIR dépendant de *RAD51* nécessite une grande région d'homologie du côté télomère proximal de la cassure et est une alternative à la conversion génique en absence de séquence d'homologie des deux côtés du site de cassure²⁸⁵. De plus il est dépendant de Rad54, Rad55 et

Rad57 et est indépendant du complexe MRX ²⁷⁹. Le BIR indépendant de RAD51, observé par le laboratoire de Haber, est indépendant de Rad54, Rad55 et Rad57 mais nécessite la présence de Rad9, Rdh54 et du complexe MRX ²⁸⁶. Le même laboratoire a plus tard montré que ce processus reposait sur de petites séquences d'homologies de part et d'autre de la cassure, et était dépendant de *RAD50* (et probablement du complexe MRX). Les auteurs proposent un mécanisme associant le BIR au SSA entre deux séquences répétées inversées qui permettent la réparation par BIR en absence de Rad51 ²⁸⁷, mécanisme déjà observé par le laboratoire de Lorraine Symington ^{280,288,289}. Il a également été montré que dans un cas bien particulier, le BIR pouvait utiliser une séquence de 200 nucléotides proche d'une origine de réplication pour envahir le chromosome donneur en l'absence de la recombinaise Rad51 ²⁹⁰.

Il faut noter que le BIR est un mécanisme peu vraisemblable pour expliquer les expansions et contractions de répétitions de trinuécléotides lors de la réparation d'une cassure double-brin dans une répétition de triplets, car celle-ci est indépendante du gène *POL32*, indispensable au BIR ²¹⁵.

5.1.5.4 Le *Single-Strand Annealing*, un mécanisme de recombinaison dépendant de *RAD52*, indépendant de *RAD51*

Le *Single-Strand Annealing* (SSA) (figure 12) permet la réparation d'une cassure double brin par l'hybridation de deux régions homologues de part et d'autre de la cassure. La région comprise entre les deux séquences homologues est perdue par délétion ^{195,291} et rend donc la réparation par SSA mutagène. Ce mécanisme est préféré à la conversion génique dans les régions répétées ²⁹¹. Il est de plus dépendant de la résection des extrémités de la cassure qui libère des extrémités 3' simple brin sur lesquelles des régions homologues pourront s'hybrider ^{195,291}. Les mutations *rad50Δ* ou *xrs2Δ* diminuent l'efficacité de réparation par le SSA ²⁹².

La perte d'expression de *RAD52* diminue fortement le taux de réparation par SSA mais peut malgré tout se dérouler en présence de grandes régions d'homologie^{291,293}. Bien que ce mécanisme dépende de la présence de séquences homologues il est indépendant de *RAD51* et même stimulé par la mutation *rad51Δ*^{288,291,292}. Rad52 qui forme un complexe avec l'ADN simple brin, stimule l'alignement des régions homologues exposées par la résection de part et d'autre de la cassure²⁹⁴. L'activité de Rad52 est stimulée par la présence de RPA et ce de façon espèce spécifique^{10 260}. L'alignement des séquences homologues génère des *flaps* 3' qui seront des substrats du complexe Rad1-Rad10 (*XPF-ERCC1* chez l'humain)²⁹².

5.2 Recombinaison homologue et instabilité des répétitions en tandem

On a longtemps cru que les changements de taille des minisatellites¹¹ était dûs à des *crossing-over* inégaux se produisant entre deux allèles lors de la recombinaison méiotique. Comme vu précédemment, la conversion génique associée à la réparation d'une cassure double-brin peut être associée à un *crossing-over* mais ce n'est pas toujours le cas²⁷⁰ (figure 14). L'étude de la conversion génique associée à l'instabilité de plusieurs minisatellites, dans la levure et chez l'humain, a montré sans ambiguïté possible que celle-ci était indépendante des *crossing-over*²⁷⁶.

Dans *S. cerevisiae*, il a été observé des événements de conversion génique associés à des échanges non réciproques de matériel génétique au locus répété *CUPI*, conduisant à des expansions ou à des contractions de grandes répétitions en tandem²⁹⁸. L'étude du minisatellite hypervariable humain MS32 par le laboratoire de Alec Jeffreys, a montré que 74% des mutations observées dans des spermatozoïdes étaient des expansions²⁹⁹. Les expansions ou contractions sont polarisées préférentiellement d'un côté des répétitions et les expansions

¹⁰ Aucune interaction observée entre Rad52 et l'ADN en présence de hRPA ou de SSB mais Rad52 est capable de se fixer à l'ADN en présence de yRPA

¹¹ Les minisatellites sont des répétitions en tandem dont le motif de base est long de plus de 9 nucléotides.

observées sont dues dans la majorité des cas à des échanges d'information génétique inter- et intra-alléliques ²⁹⁹. L'absence d'échanges des marqueurs flanquant les répétitions dans la majorité des cas observés indique que ces échanges sont issus de recombinaison indépendante des *crossing-over*. Certains évènements présentent néanmoins des duplications intra-alléliques pouvant correspondre à des échanges de matériel après *crossing-over* ²⁹⁹. Les auteurs proposaient alors que des facteurs extérieurs aux répétitions influençaient leur stabilité et la polarité des évènements de conversion génique. L'analyse des évènements de *crossovers* au locus MS32 montre qu'un *hotspot* de recombinaison méiotique, en amont des répétitions, favorise l'instabilité et explique la polarité mutationnelle observée ³⁰⁰. La recombinaison partant du point de cassure localisé sur le *hotspot*, la conversion génique observée diminue à mesure qu'on s'éloigne du point de cassure. Ces résultats expliquent l'instabilité du minisatellite MS32 chez l'humain, et le fait que celle-ci soit plus importante en méiose qu'en mitose lorsque le même minisatellite fut intégré dans le génome de *S. cerevisiae* ³⁰¹.

Lorsque le minisatellite humain CEB1 est inséré près du locus *ARG4*, connu comme étant un *hotspot* de recombinaison méiotique chez la levure *S. cerevisiae*, l'instabilité de CEB1 augmente lors de la méiose ²⁰¹. L'inactivation de *SPO11*, responsable de la génération des cassures double-brin en méiose, fait disparaître cette instabilité. L'insertion de CEB1 dans une région qui ne subit pas de cassure double-brin stabilise aussi les répétitions, démontrant que la stabilité du minisatellite dépend de la génération d'une cassure double-brin à sa proximité ²⁰¹.

Le SDSA est un modèle de réparation des cassures double-brin dépendant, entre autres, de l'hélicase Srs2. Le SDSA permet de copier l'information génétique d'un brin homologue d'ADN, après recombinaison, sans échange des séquences flanquantes. Les deux brins de la molécule d'ADN cassée envahissent la molécule donneuse et subissent une expansion ou une contraction tandis que la molécule donneuse ne subit aucune modification. Un système expérimental mis au point par Pâques et Haber, a permis de mettre en évidence directement de

tels évènements, en empêchant les *crossing-over*²⁷⁵. Des expériences ultérieures de conversion génique de microsatellites CTG dans lesquelles les *crossing over* étaient rendus possibles a montré qu'ils ne se produisaient pas à une fréquence détectable par PCR³⁰². L'envahissement de la séquence donneuse d'information par le brin receveur peut être sujet au SSM (*Slipped Strand Mismatching*) ainsi que le réappariement des deux brins receveurs. La déstabilisation d'une répétition de trinuécléotides (CTG)₃₉ causée par la recombinaison homologue est 800 fois plus importante que durant la réplication en phase S²⁷⁶.

Ce mécanisme de glissement lors de la recombinaison a été montré comme induisant des contractions et des expansions d'une répétition de trinuécléotides CAG/CTG dans la levure *S. cerevisiae*. Ces évènements sont sous le contrôle du complexe MRX, suggérant un rôle dans la résolution des structures secondaires associées à ces répétitions, vraisemblablement lors du ré-envahissement du brin néosynthétisé contenant de telles structures (figure 15²⁷⁶).

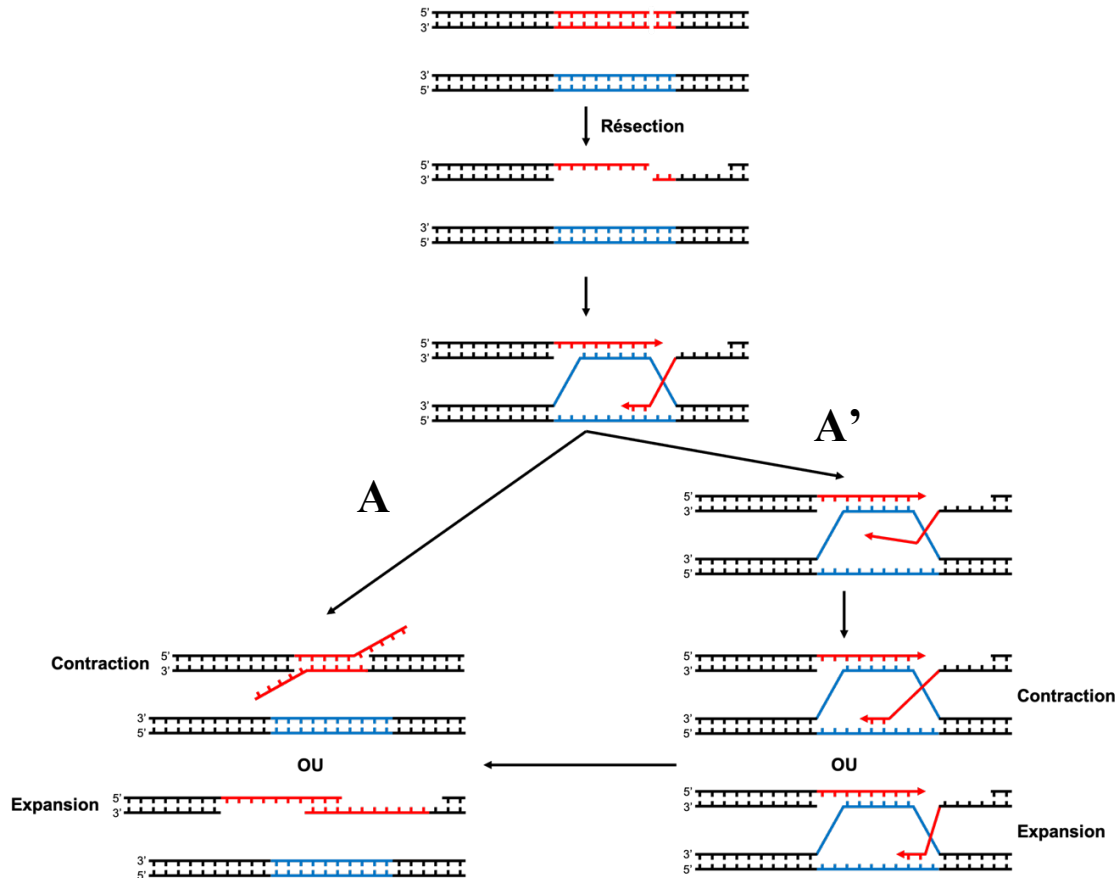


Figure 15 : Représentation schématique du mécanisme de glissement lors de la recombinaison. (A) Les deux brins néosynthétisés peuvent se dérouler et s'apparier de façon décalée, générant des expansions et des contractions de la répétition (gauche). (A') La nature de la séquence répétée peut la conduire à ré-envahir la molécule donneuse de façon décalée, pouvant conduire à une expansion ou une contraction (droite).

Notons que dans toutes ces observations, que ce soit celles des minisatellites humains ou bien des expériences dans la levure, la cassure double-brin est faite à l'extérieur de la région répétée. Dans mes expériences, la cassure est faite à l'intérieur de celle-ci, changeant vraisemblablement la dynamique de sa réparation ainsi que les mécanismes pouvant être utilisés par la cellule pour la réparer.

5.3 Réparation d'une cassure double-brin au sein d'une répétition de trinuécléotides CAG/CTG

Par quels mécanismes une cassure double-brin générée au sein d'une région répétée et

Les mutations *rad50Δ* et *sae2Δ* conduisent à la perte de la résection et de la réparation complète des cassures au sein de la région répétée tandis que les mutations *pol32Δ*, *rad51Δ* et *dln4Δ* n'ont aucun effet détectable sur la réparation, suggérant une absence d'implication des mécanismes de réparation du NHEJ, BIR ou des mécanismes dépendant de Rad51 (DSBR, SDSA). De façon intéressante, la perte d'expression de *RAD52* conduit à une très forte diminution du pourcentage de cassure détectable par Southern blot ²¹⁵. La cassure est néanmoins détectable par PCR en utilisant le protocole développé pour séquencer les télomères ²⁹⁷. Il a ainsi été montré que la cassure se produisait 1-4 triplets avant la fin de la région répétée. Ces résultats suggèrent que c'est le SSA, mécanisme dépendant de Rad52 mais indépendant de Rad51 (voir sous chapitre 5.1.5.4), qui est favorisé dans la réparation des microsatellites CAG/CTG. L'induction d'une cassure double-brin entraîne la contraction du microsatellite pendant la réparation de la cassure, de façon extrêmement efficace, car presque 100% des colonies analysées montrent une contraction de la répétition ²⁹⁶.

C'est à partir de ces résultats que mon projet de recherche a démarré. J'ai souhaité déterminer le rôle et l'importance de chacun des acteurs intervenant dans la résection d'une cassure double brin dans une répétition de triplets CAG/CTG. L'objectif de cette thèse est d'étudier le rôle de plusieurs gènes (*MRE11*, *EXO1*, *SGS1*, *DNA2*, *SAE2*, *RIF1*, *RIF2*), impliqués dans la résection et la réparation des cassures double-brin chez la levure. Le but est de comprendre en détail les mécanismes à l'œuvre dans la résection et la réparation d'une cassure double-brin dans une région d'ADN répétée et structurée, question qui n'avait jamais été abordée dans le passé. D'autre part, comprendre les mécanismes à l'œuvre dans un eucaryote modèle nous guidera sur la voie permettant d'utiliser une approche similaire chez l'humain. Ainsi à partir d'une TALEN, élaborée dans le but d'introduire une cassure double-brin au sein d'un microsatellite (CTG)n inséré dans le génome de la levure *S. cerevisiae* ²⁹⁶, des cassures double-brin ont été générées au sein de mutants des gènes impliqués dans la résection et la réparation de cette cassure.

6 Interactions de la machinerie télomérique avec la résection

Une des idées de mon projet a été l'étude de l'implication des gènes *RIF1* et *RIF2* dans la réparation des cassures double-brin au sein des répétitions de trinuécléotides CTG. Je me suis rapidement intéressé aux mécanismes qui pourraient expliquer les grandes expansions chez les patients atteints de DM1 (ou d'autres maladies neurodégénératives causées par des expansions) en parallèle de mon projet initial.

Les modèles proposés pour expliquer les expansions (*i.e* glissement de brins, structures secondaires, insertions par les mécanismes de réparations, *etc.*) sont capables d'expliquer des expansions de quelques nucléotides mais jamais les expansions de plusieurs centaines voire milliers de bases qui sont transmis à la descendance de patients. Mon raisonnement a d'abord été de réfléchir par quel mécanisme une grande séquence d'ADN pouvait être générée au sein de la région portant les répétitions. Seule une polymérase était à mon sens capable d'accomplir une telle tâche. Comme nous avons pu le voir au cours de cette introduction, le rôle des polymérases de phase S dans l'instabilité des répétitions est déjà connu et les mutants de ces enzymes ne génèrent pas de grandes expansions.

Cela m'a conduit à imaginer un rôle de la télomérase (et de la machinerie de maintenance télomérique) dans l'expansion des régions répétées, au vu de travaux préalables sur la capacité de la télomérase à s'hybrider, dans un contexte favorable, sur des séquences ressemblant aux télomères^{303,304}. La télomérase de *S.cerevisiae* est composée d'une matrice ARN (TER) *TLC1* (constituée des répétitions C₁₋₃A) ainsi que d'une transcriptase inverse (TERT). *TLC1* est nécessaire à la réplication des télomères et son absence conduit à leur raccourcissement³⁰⁵. La transcriptase inverse de *S. cerevisiae* est constituée de trois sous unités *EST1*, *EST2* et *EST3*^{304,306}. Est2 est la sous unité catalytique, sa quantité est dépendante de celle de *TLC1*. Est1 et Est3 sont nécessaires à la maintenance télomérique mais pas à l'élongation³⁰⁴.

Cette idée d'un rôle de la télomérase dans les expansions était renforcée par le fait qu'une addition de séquence télomérique avait été précédemment détectée au niveau d'une cassure double-brin en absence de Rqh1 (hélicase homologue de Sgs1) et Rad55 (impliqué dans la réparation des cassures double-brin, stabilise les nucléofilaments de Rad51), mais en présence d'Exo1 et Rad51³⁰⁷. De plus, bien que limité et dans le contexte *rad55Δ*, le complexe de résection MRX semble être impliqué dans la prévention de l'ajout de télomère³⁰⁷. D'autre part il a été montré que dans un contexte dépendant de Tel1¹², la télomérase était capable *in vivo* d'agrandir de très courtes séquences télomériques de moins de 125 paires de bases³⁰⁸.

De récents travaux du laboratoire, montrent que les cassures double-brin au sein de répétitions de trinuécléotides CAG/CTG sont réparées par un mécanisme indépendant de *RAD51*, le *Single-Strand Annealing* (SSA) et largement dépendant des étapes initiales de la résection²¹⁵. De plus il a été montré dans un autre système expérimental que la résection favorisait l'ajout de télomère au niveau de la cassure^{309,310}. Serait-il alors possible qu'au sein d'un mécanisme de réparation indépendant de *RAD51* et *RAD55* et entièrement dépendant de la résection tel que le SSA, la télomérase puisse ajouter des fragments de télomères à l'extrémité d'une répétition de trinuécléotides CAG/CTG ?

Deux types d'extrémités de l'ADN peuvent être rencontrées dans une cellule, les extrémités d'une cassure double-brin et les extrémités télomériques. Le consensus des séquences télomérique chez *S. cerevisiae* est C₁₋₃A/TG₁₋₃, formant ainsi une séquence répétée imparfaite. Sur celle-ci se fixe un ensemble de facteurs protégeant l'extrémité d'un chromosome de la dégradation ou de la recombinaison, garantissant ainsi leur intégrité et leur stabilité³⁰⁴. Les protéines formant le *capping*, qui protège les extrémités télomériques chez la levure *S. cerevisiae* sont les protéines Ku, le complexe Rap1-Rif1-Rif2 et le complexe CST (Cdc13-Stn1-

Ten1)^{311,312}. Rap1, Rif1 et Rif2, qui interagissent directement, régulent négativement l'élongation télomérique³¹³. Bien que l'interaction entre Rap1 et Rif2 soit faible, Rif2 lie deux Rap1 via ses domaines AAA+ et RBM^{314,315}. Le complexe CST joue un rôle de « RPA télomérique » protégeant les extrémités simple brin des télomères de la dégradation tandis que le complexe Rap1-Rif1-Rif2 et Ku bloquent l'accès au complexe MRX et à Exo1^{271,272,304}. En l'absence de ces facteurs, rien ne distingue en théorie une extrémité télomérique d'une cassure double-brin, pouvant conduire à des pertes ou des fusions de chromosomes.

Des études comparant l'interaction entre Rif1 et Rif2 au niveau de courtes séquences télomériques (TG₈₁) ou plus longues (TG₁₆₂) avec Tel1 et le complexe MRX montrent qu'elles régulent négativement le recrutement de Tel1 et du complexe MRX au niveau d'une cassure double-brin télomérique³¹⁰. La mutation *rif1*Δ et/ou *rif2*Δ augmente l'association du complexe MRX et de Tel1 au niveau des séquences télomériques. Les protéines Rif1 et Rif2 sont recrutées aux télomères en fonction de la longueur des séquences télomériques, plus celles-ci sont longues plus les deux protéines sont recrutées. Il fut proposé que l'inhibition du recrutement de Tel1 par Rif2 se joue par interaction avec le domaine C-terminal de Xrs2, domaine par lequel Tel1 interagit avec le complexe MRX, entraînant une compétition entre Tel1 et Rif2³¹⁰. Ainsi Rif1 et Rif2 empêchent le recrutement de Tel1 à la cassure, diminuant l'affinité du complexe MRX pour les cassures dans les régions télomériques TG₈₁³¹⁰. Finalement, Rap1 exclut le complexe MRX de la cassure en absence de Tel1 (figure 17)³¹⁰. Rif1 semble dépendant de Rif2 dans l'inhibition du recrutement de Tel1³¹⁰.

Les modèles actuels montrent cependant que l'inhibition du recrutement de Tel1 se fait via l'activité de Rad50 avec lequel il interagit directement en absence de Xrs2³¹⁴. L'interaction de Rif2 avec Rad50 induirait un changement conformationnel de Rad50, stimulant son activité ATPase³¹⁴. Le recrutement du complexe MRX au niveau des cassures ainsi que son affinité pour l'ADN dépendent de Tel1^{310,316}. Comme l'association du complexe MRX est dépendante

Une autre étude d'induction de cassure double-brin par HO suggère un rôle de Rif1 dans la régulation positive de la résection par le complexe MRX. En absence de Rif1, on observe une baisse de l'efficacité de résection par le complexe MRX³¹⁸. Les mutations *rif1Δ* *exo1Δ* et *rif1Δ* *sae2Δ* entraînent une diminution plus forte de l'efficacité de résection que les simples mutations *rif1Δ*, *exo1Δ* ou *sae2Δ*³¹⁸.

Problématique de mon travail de thèse

Le rôle et l'interaction du complexe Rap1-Rif1-Rif2 dans l'instabilité et la biologie des séquences répétées de type microsatellites n'ont jamais été auparavant adressés.

Bien que l'implication des mécanismes de réparation de cassures double-brin sur l'instabilité des microsatellites ait été développée dans cette introduction, le rôle des différentes protéines intervenant dans la réparation de ces cassures ainsi que leur rôle précis dans l'instabilité des microsatellites CTG restait encore à déterminer. J'ai essayé au cours de ma thèse de répondre à la problématique suivante. Par quels mécanismes moléculaires une cassure double-brin au sein d'un microsatellite CTG est résectée puis réparée ? Cette question a été posée en étudiant le rôle des protéines Mre11, Sae2, Exo1, Sgs1 et Dna2 dans leur fonction au sein des mécanismes de réparation et de résection d'une cassure double-brin unique au sein d'un microsatellite CTG.

Comme développé dans le chapitre 6 de cette introduction, j'ai souhaité poser une deuxième problématique durant mes travaux. La machinerie télomérique de *S. cerevisiae*, joue-t-elle un rôle dans l'instabilité et la réparation d'une cassure double-brin au sein d'un microsatellite CTG ? Ces questions m'ont poussé à construire des souches *rif1Δ*, *rif2Δ* et *rif1Δ rif2Δ* et à étudier l'effet de telles mutations sur la réparation d'un microsatellite CTG.

RESULTATS

**Repair of a double-strand break in a CTG trinucleotide repeat depends on short range
resection and on the telomeric *RIF1/RIF2* complex**

Wilhelm Vaysse-Zinkhöfer^{1,2}, David Viterbo¹ and Guy-Franck Richard¹

¹ Institut Pasteur, UMR3525 CNRS, 25 rue du Dr Roux, 75015 Paris, France

² Sorbonne Université, Collège doctoral, 75005 Paris, France

Corresponding author: Guy-Franck Richard

gfrichar@pasteur.fr

Tel: +33-1-45-68-84-36

Keywords: trinucleotide repeats, double-strand break repair, resection, telomere, secondary structure

Abstract

A large expansion of CTG trinucleotide repeats is responsible for myotonic dystrophy type 1 (DM1) in humans. Contracting the CTG tract below the pathological length could be envisioned as a possible gene therapy approach. We previously showed that expression of a TALE nuclease (TALEN) specifically designed to induce a double-strand break (DSB) within an expanded DM1 allele, led to frequent repeat contractions. Here, we demonstrate that efficient repair of this DSB requires proteins involved in the early steps of resection, *MRE11* and *SAE2*, but not in the later steps. The *mre11D56N* and *mre11H125N* alleles are as deficient in resecting and repairing the break as the *SAE2* deletion, and no additive effect was found in double mutants. This resection defect is specific of the CTG repeat tract, showing that the Mre11 nuclease activity is essential to initiate resection within a CTG trinucleotide repeat. We also show that both *RIF1* and *RIF2*, components of the telomeric end protection complex, repress resection of the CTG repeat tract but not of a DSB induced in a non-repeated region. This suggests an unexpected protective role of these telomeric proteins on a CTG repeat-containing DSB.

Introduction

Microsatellites are short tandem repeats whose base motif length is short (1-9pb). These repeats are found ubiquitously in all organisms but are particularly abundant in eukaryotic genomes, covering 3% of the human genome ¹. Some of these microsatellites are prone to expansion, sometimes massive, leading to numerous neurological diseases in humans such as fragile X syndrome, Huntington disease, myotonic dystrophy type 1 (DM1) or Friedreich Ataxia ². Some microsatellites were shown to form secondary structure *in vitro* and possibly *in vivo* ³⁻⁶. The mechanisms leading to massive microsatellite expansions in humans are still not known.

DM1 is an inherited multisystem disorder characterized by a progressive muscle weakness and myotonia ⁷. The onset and the severity of the disease are both positively linked with the CTG trinucleotide repeat number. The repeats are located within the 3'UTR of the DMPK gene, encoding the dystrophin myotonia protein kinase. A healthy individual has between 5 and 37 repeats, while between 38 and 49 repeats is considered a pre-mutational state. Individuals with 50 repeats and above are invariably symptomatic ⁸. The mutated CUG-containing RNA binds and sequesters *MBLN1*, inhibiting its function as a RNA splicing factor ^{9,10}. *MBNL1* knock-out mice gives the same phenotype as the DM1 disorder in humans ¹¹. No treatments for DM1 exists as of today.

A daughter that inherited her father DM1 chromosome did not show any clinical sign of the disease after a spontaneous contraction of the CTG trinucleotide repeat tract ¹². This suggested that contracting the repeat tract could reduce the most dramatic symptoms of the pathology. It was previously shown that the generation of a DSB between two short CTG repeat tracts, using the I-SceI meganuclease, led to contraction of the repeated region ¹³. With

the discovery of TALE Nucleases (TALEN) and of the CRISPR-Cas systems, it became possible to precisely edit complex genomes and generate DSB within microsatellites and expanded repeats. In previous work, a TALE nuclease (TALEN) specifically designed to induce a single DSB within an expanded (CTG)_n trinucleotide repeat, showed good efficacy at contracting this microsatellite^{14,15}. Similar experiments were conducted in cells deficient in several repair pathways, showing that *RAD51*, *POL32* and *DNL4* were dispensable for DSB resection and repair within a CTG trinucleotide repeat. Only *RAD50*, *SAE2* and *RAD52* were necessary, suggesting that cells were repairing DSB in trinucleotide repeats by a homology-driven single-strand annealing mechanism (SSA)¹⁶. Since SSA requires efficient DNA resection, we decided to study the effect of several mutants known to be involved in early or later steps of DNA resection, on the repair of an expanded CTG repeat tract¹⁷⁻²⁰. Resection of the double-strand break ends is critical to initiate repair by SSA and as such, the SSA pathway relies on Sae2 and the MRX complex²⁰⁻²³ as well as other nucleases such as Exo1 and Dna2 (working with Sgs1) to reveal homologous sequences on each side of the break to initiate annealing^{24,25}. The MRX-Sae2 complex and the Exo1, Dna2/Sgs1 nucleases acting in different steps of the resection²⁰. The activity of the MRX complex (Mre11-Rad50-Xrs2) is necessary for meiotic Spo11 induced double-strand break but not required for HO-induced double strand break^{26,27}.

The Mre11 nuclease possesses both endonuclease and exonuclease activities *in vitro*. Following a DSB, the endonuclease activity generates a single-strand nick upstream of the DSB end. The newly formed gap allows the protein, through its 3'-5' exonuclease, to digest the fragment between the nick and the DSB end^{28,29}. The loss of any of these activities leads to a complete loss of resection on "dirty" double-strand break extremities^{27,30}, while they are not necessary for clean double-strand break ends made by the HO meganuclease^{26,27}. Removal of

blockages from the ends of a break is observed in vitro only when Sae2 is present with a functional MRX complex and in the presence of blockage³¹. Since both activities are distinct, they can theoretically be selectively inhibited by point mutations³⁰ without hindering Mre11 abilities to form a complex with its partners, Rad50 and Xrs2. The three most studied mutations affecting Mre11 nuclease activity are the point mutations D16A, D56A/N and H125A/N in the phosphoesterase motifs I, II and III, respectively. Severity of IR sensibility showed that the D56N mutation was as impacted as the H125N mutation³⁰. To better understand Mre11 role in CTG repeat resection and repair, two strains were designed carrying either the *mre11D56N* or the H125N mutations to disrupt its nuclease activity. The endonuclease activity of Mre11 is dependent on the hydrolysis of ATP by Rad50, on the presence of manganese as well as on Sae2 which plays a critical role in the stimulation of the endonuclease activity of Mre11³¹.

The MRX complex functions in tandem with the Sae2 protein (CtIP in humans, considered the "fourth protein" of the MRX complex) which interacts via its C-terminal domain and activates the endonuclease activity of Mre11 in addition to stimulating its exonuclease activity^{31,32}. Loss of SAE2 expression results in a total loss of resection phenotype similar to that of a *rad50Δ* strain³³. It was proposed that Sae2 possesses its own nuclease activity, independent of Mre11³². We therefore decided to evaluate the possibility that Sae2 plays a role independently of Mre11 in CTG repeat DSB resection and repair. Long-range resection carried out by *EXO1*, *SGS1* and *DNA2* was also examined through the generation of knock-out *EXO1* strain and anchor-away *SGS1* and *DNA2* strains³⁴.

In another experimental system, it was shown that resection favored telomere addition at the break ^{35,36}. We asked if within a repair mechanism independent of RAD51 and RAD55 and entirely dependent on resection such as SSA, telomerase could add telomere fragments to the end of a CAG/CTG trinucleotide repeat. The Rap1-Rif1-Rif2 complex is known to negatively regulate telomere elongation by inhibiting the interaction of the MRX and Exo1 complex with telomere ends ³⁷. This Rif2- and Rap1-dependent inhibition mechanism increases the ATPase activity of Rad50, decreasing the ATP+ time of the MRX complex required for the interaction and activation of Tel1 (functional homolog of ATM) ³⁸⁻⁴¹. Former work showed that the Rif1 complex was playing a role in *Saccharomyces cerevisiae* resection by directly interacting with the MRX complex ⁴².

Analyzing *RIF1* and *RIF2* function at the CTG repeat DSB revealed an unsuspected role of both genes in inhibiting resection specifically within an expanded CTG trinucleotide repeat, and not in a non-repeated region.

Materials and Methods

Yeast strains and plasmids

SAE2 and *EXO1* were deleted in the GY6162-3D strain ⁴³ and replaced by the KANMX4 cassette ⁴⁴. The marker was amplified from the EUROSCARF deletion library, for each deletion, using primers located 0.5 kb (for *EXO1*) or 1 kb (for *SAE2*) upstream and downstream of the KANMX4 cassette. *RIF1* and *RIF2* were replaced respectively by the *HIS3* and *LEU2* auxotrophic markers in the GY6162-3D strain, using primers carrying 50 bp of homology to the *RIF1* or *RIF2* 5' and 3' UTR ⁴⁵. The *mre11D56N* and *mre11H125N* mutations were ordered from Twist Bioscience

as 760 bp DNA fragments containing the point mutation. These fragments were assembled by homologous recombination in *S. cerevisiae* with the remaining 1.5 kb of the *MRE11* gene into the pRS416 centromeric plasmid⁴⁶. Recombined plasmids were extracted from yeast and the reconstituted *mre11D56N* or *mre11H125N* were cloned into the integrative pRS406 plasmid. Recombinant plasmids were digested with *Cl*I and transformed into GFY6162-3D for integration at the *MRE11* locus. Pop-out of the plasmid was selected on 5-FOA plates⁴⁷. Presence of the *MRE11* mutations in pop-out clones was confirmed by PCR and Sanger sequencing (Eurofins).

The *SGS1*-RFB strain was built using the anchor away system³⁴ after crossing GFY6162-3D with the LMA3670 strain. *DNA2*-RFB was built after deletion and replacement of *TRP1* by the *HphMX6* cassette followed by insertion of 80 CTG triplets at the *SUP4* locus in the YDS43 strain⁴⁸.

All primers used in these constructions are shown in Supplementary Table S1 and yeast strains are described in Supplementary Table S2.

TALEN induction

Vectors pCMha182KN9996 and pCMha188KN16715 respectively carrying left and right TALEN arms under the control of the Tet_{OFF} promoter¹⁶, were transformed in GFY6161-3D and its mutant derivatives. Transformants were selected on synthetic SC-Ura⁻Trp medium supplemented with 10 µg/ml doxycycline. Vectors pCMha188KNL1 and pCMha182KNR1¹⁶, carrying respectively, the TALEN_{noCTG} left and right arm, were transformed in strains WVZ26, WVZ27 and WVZ29. Transformants were selected on synthetic SC-Ura⁻Trp medium supplemented with 10 µg/ml doxycycline. Yeast cells were grown overnight at 30°C in 100 ml liquid SC-Ura⁻Trp medium supplemented with 10 µg/ml doxycycline. The culture was divided

in two fractions, each containing $1.35 \cdot 10^9$ cells, and each one was washed two times in water to remove any trace of doxycycline. Each fraction was subsequently diluted to $9 \cdot 10^6$ cells/ml in fresh SC-Ura -Trp medium, and $10 \mu\text{g/ml}$ doxycycline was added in one of the cultures to repress TALEN induction. For Anchor Away strains, rapamycin was added to the medium ($2 \mu\text{g/ml}$ final concentration) with or without doxycycline.

Since it takes 14 hours at 30° for the TALEN to induce a DSB¹⁶, cells were collected at 0, 14, 16, 18, 20, 22, 24, and 48 hours after TALEN induction. For each time point, 2×10^8 cells were rapidly centrifuged, washed with water, and frozen in dry ice before DNA extraction. To determine viability after DSB induction, cells were plated 24 hours after TALEN induction on SC-Ura- Trp plates supplemented with doxycycline ($10 \mu\text{g/mL}$) for the TALEN-repressed culture and on SC-Ura -Trp plates for the TALEN-expressed culture. CFUs were counted after 3–5 days of growth at 30°C and survival was calculated as the ratio of CFUs on TALEN-expressed over TALEN-repressed plates.

DSB analysis and quantification

Total genomic DNA ($\sim 4 \mu\text{g}$) of cells collected at each time point was digested for 5 hours by EcoRV (20 U) (NEB) and analyzed by Southern blot using non-radioactive DIG probe. Labeling was performed in $100 \mu\text{l}$ volume reactions containing 50–60 ng genomic DNA, 1 M of each Su37/Su8 and Su39/Su40 primers in two separate reactions, $10 \mu\text{l}$ PCR DIG labeling mix (Roche 11 585 550 910), $10 \mu\text{l}$ PCR buffer and 5 units of Dream Taq polymerase (Thermo Fischer). The PCR program used started by an initial denaturation step at 98°C for 2 min, followed by 35 cycles of: 98°C denaturation for 30 s, 50°C annealing for 30 s, 72°C extension for 1 min. A final 10 min extension step at 72°C was added at the end. A $5 \mu\text{l}$ aliquot was

loaded on gel for quantification and approximately 250 ng of the probe was used for hybridization, after a 5 min denaturation step at 95°C ⁴⁹.

Hybridization of non-radioactive probes was performed overnight, at 65°C in a rotating tube (Hybaid hybridization oven, Thermo Scientific) in 250 mM sodium phosphate buffer (pH 7.4), 7% SDS, 1 mM EDTA. Two washes were performed in 20 mM sodium phosphate buffer (pH 7.4), 1% SDS, 1 mM EDTA. The membrane was blocked in 75 mM maleic acid pH 7.5, 200 mM NaCl, 5% non-fat dry milk powder (blocking buffer) for 1h at room temperature. The anti-digoxigenin antibody (Fab fragment, Roche 11 093 274 910) was diluted (1/10 000) in 50 ml blocking buffer supplemented with milk powder and incubated with the membrane for 30 min to 1 h at room temperature. Subsequently, two washes were performed in blocking buffer supplemented with 0.3% Tween 20 for 15 min each at room temperature. One quick final wash in 100 mM Tris-HCl (pH 9.5), 100 mM NaCl was performed for 5 min at room temperature before the chemiluminescent substrate (CSPD, Roche 11 655 884 001) was added to the membrane undiluted (1 ml/10 cm²). It was subsequently incubated during 5 min at room temperature, the CSPD was removed and a further 10 min incubation at 37°C was performed to enhance the chemoluminescent reaction, as recommended by the manufacturer ⁴⁹. Both DSB end signals were quantified and shown as a percentage of the total signal in the lane. In total 6 wild types, 4 *sae2Δ*, 7 *exo1Δ*, 6 *mre11H125N*, 3 *mre11D56N*, 3 *mre11D56N sae2Δ*, 3 *mre11D56N exo1Δ*, 6 *mre11D56N sae2Δ exo1Δ*, 5 *rif1Δ* with and without CTG, 2 *rif2Δ* with and without CTG and 2 *rif1Δ rif2Δ* with CTG clones were analysed.

Trinucleotide repeat length analysis

Colonies from each induced or repressed plates were picked, total genomic DNA was extracted, and ~4 μg was digested for 5 hours by SspI (20 U) (NEB) and analyzed by Southern

blot. As described previously (Viterbo et al., 2016) full length repeated tracts and contracted tracts were scored and recorded. The length of the tract was determined by comparisons against a specific CTG ladder with bands of known CTG repeats length. In each strain, repeat tract length was analyzed in 48 clones, by PCR amplification using primers su3 and su7 and Sanger sequencing using su2 (Eurofins).

Analysis of DSB end resection

Real-time qPCR assays using primer pairs flanking EcoRV sites respectively 0.81 kb and 2.94 kb away from the 3' end of the CTG repeat tract (VMS20/VMAS20 (right proximal) and VMS21/VMAS21 (right distal), and respectively 0.88 kb and 1.88 kb away from the 5' end of the CTG repeat tract (VMS22/VMAS22 (left proximal) and VMS23/VMAS23 (left distal), were used to quantify DSB end resection¹⁶. Another primer pair was used to amplify a region of chromosome X near the *ARG2* gene to serve as the DNA input control (JEM1f-JEM1r). Genomic DNA of cells collected at T = 0 hr, T = 18 hr, T = 24 hr, and T = 48 hr was split in two fractions; one was used for EcoRV digestion and the other one for a mock digestion in a final volume of 15 μ L. Samples were incubated for 5 hours at 37°C and then the enzyme was inactivated for 20 min at 80°C. DNA was subsequently diluted by adding 55 μ L of ice-cold water, and 4 μ L was used for each real-time PCR reaction in a final volume of 25 μ L. PCRs were performed with Absolute SYBR Green Fluorescein Mix (Thermo Scientific) in a Mastercycler S Realplex (Eppendorf) using the following program: 95°C for 15 min, 95°C for 15 s, 55°C for 30 s, and 72°C for 30 s repeated 40 times, followed by a 20-min melting curve. Reactions were performed in triplicate, and the mean value was used to determine the amount of resected DNA using the following formula: raw resection = $2/(1+2\Delta Ct)$ with $\Delta Ct = Ct_{EcoRV} - Ct_{mock}$. Relative resection values were calculated by dividing raw resection values by the percentage

of DSB quantified at the corresponding time point ⁵⁰. All t-tests were performed using the R package ⁵¹.

Results

A double-strand break induced within CTG trinucleotide repeats requires the Mre11p nuclease activities and Sae2p to be processed

In the present work, we used a TALEN designed to specifically induce a DSB within an expanded CTG trinucleotide repeat tract from the DMPK human locus. The repeats were inserted at the modified *SUP4* allele in *Saccharomyces cerevisiae* genome ^{13,15,16,43}. The TALEN is under the control of the Tet_{OFF} promoter ⁵². In previous work, it was shown that it takes 14 hours for the DSB to be detected by Southern blot ¹⁶. Therefore, the first time point was collected 14 hours after T0. When the TALEN is repressed no DSB is observed, but a smear can be seen, corresponding to the natural instability of CTG trinucleotide tracts in *S. cerevisiae* ^{53,54}. After TALEN expression, both ends of the DSB are visible (Figure 1A). The 5' end of the break appears as a smear due to repeat length heterogeneity during repeat tract contraction ¹⁶. The 3' end of the break appears as a sharper band since it contains only a few triplets. Loss of Mre11p nuclease activity leads to a significant two-fold increase of the 5' end signal, as compared to wild type, in the *mre11D56N* mutant (unilateral student's t-test p-value = 0.0002) (Figure 1B). The same observation was made for the *mre11H125N* strain (p-value = 0.006). We concluded that both phosphoesterase residues were essential for early resection step of a DSB made within an expanded CTG trinucleotide repeat. The D56N mutation seems to block DSB processing more efficiently than the H125N mutation (p-value = 0.01). The 3' end of the break

does not show the same defect as the 5' end since no significant difference between wild-type and mutant strains could be found (Figure 1B).

The Sae2 protein plays a critical role in early resection step and its absence results in a total loss of resection phenotype similar to that of a *rad50Δ* strain³³. Moreover Sae2 is directly interacting with Mre11 while activating its endonuclease activity and stimulating its exonuclease activity^{31,32}. Previous work suggests that Sae2, that in addition to activating the endonuclease activity of Mre11, Sae2 would play a direct role in the resection and removal of secondary structures blocking the double-strand break ends through its own endonuclease activity, stimulated by the MRX complex.³² Sae2 would therefore in fact be directly responsible for the endonuclease activity of the MRX complex. Our results shows that when compared to the wild-type strain, the *sae2Δ* mutant showed a significant increase of the 5' end signal after 24 hours, a result identical to the *mre11D56N* mutant (unilateral student's t-test p-value=0.0003) (Figure 1C). We expected that if Sae2p were carrying its own nuclease activity and if it were required to process the DSB, the double mutant phenotype would exhibit a greater defect. However, no additional effect of the *SAE2* deletion on the *mre11D56N* mutation was observed (Figure 1C).

Long range resection: *SGS1*, *DNA2* and *EXO1*

The *EXO1* deletion does not lead to any particular phenotype and no significant difference with the wild-type strain was observed (unilateral student's t-test p-value > 0.017 after Bonferroni correction) (Figure 2A). When *sae2Δ* and *mre11D56N* were introduced in the *exo1Δ* background, 5' end signal increased to the level seen with the *mre11D56N* or the *sae2Δ* single mutants, proving that *MRE11* was epistatic on *EXO1* (Figure 2B).

Involvement of the telomeric end protection complex in CTG DSB repair

The Rap1-Rif2-Rif2 complex has been shown to negatively regulate the recruitment of the MRX complex at breakage through a direct interaction of Rif2 with Xrs2 and Rad50^{36,38}. This inhibition is dependent on the presence of Rap1 and the effect of Rif1 on this inhibition is dependent on the presence of Rif2^{36,55}. A study on the interaction between the MRX complex and Rif2 shows that Rif2 negatively regulates the recruitment of MRX also to the site of HO induced intrachromosomal double-strand breaks^{36,56}. In order to determine if the telomere maintenance machinery could play a role on trinucleotide repeat processing, the same experiments were performed in *rif1Δ* and *rif2Δ* strains, as well as the double *rif1Δ rif2Δ* mutant. No effect was observed in the *rif1Δ*, *rif2Δ* or *rif1Δ rif2Δ* mutants, showing that these mutations had no effects on CTG repeats double-strand break repair (Figure 3A). No significant difference can be observed between the 5' and 3' ends accumulations of the break for the *rif2Δ* and *rif1Δ rif2Δ* mutants (Bilateral student's t-test p-value > 0.017 after Bonferroni correction, for *rif2Δ* and *rif1Δ rif2Δ* respectively) (Figure 3A).

As a control we inactivated *rif1Δ* and *rif2Δ* in an isogenic strain except for the presence of the CTG repeat tract inserted at *SUP4*¹⁶. In this strain with no CTG, the only way to repair the DSB is by non-homologous end-joining (NHEJ) or by single-strand annealing between LTR sequences located kilobases away on each side of the repeat tract^{13,57}. As compared to wild type, both mutants show an accumulation of broken molecules, on each end of the DSB 48 hours after TALEN induction (Figure 3B) (Unilateral student's t-test p-value < 0.05 for *rif1Δ* and unilateral student's t-test p-value < 0.01 for *rif2Δ*). We concluded that in *rif1Δ* and *rif2Δ* strains, a single TALEN-induced DSB in a non-repeated region is very inefficiently repaired by NHEJ.

DSB resection is abolished in *sae2Δ* and *mre11* mutants

Resection efficacy 24 hours and 48 hours after TALEN induction were determined by real-time qPCR of total genomic DNA, after digestion by the EcoRV endonuclease. Four EcoRV sites are located around the repeat tract, two of them 0.8 and 0.9 kb upstream and downstream of the repeat tract (proximal sites) and two others 1.8 and 2.9 kb upstream and downstream (distal sites) (Figure 4A). When resected, the single stranded DNA cannot be digested by EcoRV allowing to amplify the region located around the restriction site using dedicated primers. Comparison of cycle thresholds (Ct) between a non-digested and digested DNA fraction allows to determine the amount of raw resection at a particular site. This raw resection is then reported to the amount of 5' and 3' DSB ends as determined by Southern blot quantification, to obtain the true resection value at each EcoRV site⁵⁸.

In the wild type, resection is naturally hindered at the 5' repeated end of the break, whereas the non-repeated end is resected (Figure 4B). In the *sae2Δ*, D56N and H125N strains, there is an almost complete loss of resection of the repeated end, proximal to the break (left proximal) (unilateral student's t-test p-value < 0.01). In *sae2Δ*, D56N or H125N mutants, resection is reduced but not statistically different from wild type at the repeated end distal to the break (left distal). The non-repeated end shows no significant resection difference between the mutants and wild-type strains at the right proximal and right distal sites. In the *exo1Δ* mutant, no difference at either end of the break can be seen (Figure 4B). Results for the *mre11D56N sae2Δ*, *mre11D56N exo1Δ* double mutants and the *mre11D56N sae2Δ exo1Δ* triple mutant are all similar to the single *mre11D56N* mutant, confirming the epistasy of *MRE11* on other mutations (Supplementary Figure S1). These results put in evidence the critical role of the early resection machinery for DSB repair in CTG trinucleotide repeats.

Resection at CTG trinucleotide repeats is increased in the absence of Rif1p or Rif2p

Resection was restored to wild-type levels at left proximal and distal sites in the *rif1Δ*, *rif2Δ* and *rif1Δ rif2Δ* mutants (Figure 4C). Resection was also increased in *rif1Δ*, *rif2Δ* and *rif1Δ rif2Δ* strains as compared to wild type at the right proximal site, while at the right distal site only *rif1Δ* and the double mutant showed a significant increase in resection. The loss of *RIF2* seems to have a stronger effect on CTG repeat resection than the loss of its partner as was previously shown in telomeric regions or non-repeated DNA ^{36,55,56}

In the strain that contains no CTG triplet, the opposite effect was observed, DSB resection was significantly decreased in *rif1Δ*, *rif2Δ* and *rif1Δ rif2Δ* mutants, on both DSB ends (Figure 4D). Once again, the effect on resection was stronger with the *rif2Δ* mutation than its partner.

Cell survival depends on *SAE2* and *MRE11*

Approximately 60% of wild-type cells survived the DSB. This was not significantly different from cell survival in the *exo1Δ* mutant (Figure 5, left). All single and double mutants showed higher mortality rates and were all significantly different from wild type. In addition, *mre11D56N* and *mre11D56N sae2Δ* exhibit a slight but significant higher survival as compared to the *sae2Δ* mutant.

In *rif1Δ* and *rif2Δ* strains, survival is not significantly different from wild type (Figure 5, right). Without flanking homologies to help repairing the break, a strain without CTG is really sensitive to a single DSB and exhibits low viability. There are no significant differences between the *rif1Δ* and *rif2Δ* and the wild type strain in absence of CTG repeat.

TALEN expression does not induce other mutations

Southern blot analysis showed that CTG repeat contraction was very efficient in survivors following TALEN expression, confirming previous results obtained with the same TALEN^{15,16} (Supplementary Figure S2). Contraction occurred in 100% of the cases, accompanied by a small smear in approximately 10% of the survivors, testifying repeat length heterogeneity among cells of the same colony.

As expected, being deficient in DSB repair, the *mre11D56N* mutant showed significantly less contraction than the wild type as it was the case for the *mre11D56N* *exo1Δ* double mutant (exact Fisher's test p-value < 0.01 after Bonferroni correction). The *exo1Δ*, *mre11D56N* *sae2Δ* and triple mutants were not significantly different from the wild type strain (Supplementary Figure S2). Interestingly, *rif1Δ* and *rif2Δ* mutants did not show significantly higher CTG repeat contraction efficacy (exact Fisher's test p-value < 0.016 after Bonferroni correction). Both the *rif1Δ* and *rif1Δ* *rif2Δ* mutants showed increased smear formations, especially the *rif1Δ* strain when compared to the wild type and all other mutant strains (exact Fisher's test p-value < 0.00625 after Bonferroni correction). This suggests that DSB repair was differently processed in these strain backgrounds in which resection was more efficient.

DNA was extracted from surviving colonies and the *SUP4* locus was PCR amplified and sequenced. No other mutation was observed in any of the survivors, confirming previous results obtained with the same TALEN. No telomeric addition, nor small pieces of telomeric repeats could be seen in any of the 42 and 36 clones respectively of *rif1Δ* and *rif2Δ* survivors.

Discussion

Differences in *MRE11* mutations in CTG repeat resection and repair

In the present work, we investigated DSB-repair within a long CTG trinucleotide repeat. Mechanisms of repair of a repeated and structured DNA sequence were poorly studied in comparison with extensive studies focusing on unique genomic regions. Previous work could not distinguish the *mre11D56N* and H125N mutations (in phosphoesterase motifs II and III, respectively), their phenotypes being very similar. Both mutations show no defect in HO-induced DSB-repair or in telomere length in *S. cerevisiae*^{26,27}. However, both mutants are as defective in processing perfect hairpins as the null mutant, proving that nuclease activity was at least partly compromised⁵⁹. Our results show a significant difference between both mutations, D56N being more dramatically affected than H125N. These results could be explained by the discontinuous presence of several imperfect hairpins or other secondary structures along the repeated tract. In such instance, the Mre11p endonuclease activity would be required to successively remove each structure that would otherwise hinder the loading and processing of DSB ends by long range resection proteins. Such conditions would greatly enhance the differences in nuclease function between Mre11 mutants, allowing us to detect small phenotypical variation among them.

CTG trinucleotide repeats do not form clean DSB ends

It is known that *MRE11* is not required to process clean DSB end²⁶. Our results show that Mre11 is necessary to process breaks within CTG trinucleotide repeats, strongly suggesting that such DSB ends are blocked by secondary structures, hindering the recruitment of other proteins such as Exo1p or Dna2p involved in long range resection^{28,60,61}. The *in vivo* nature

of secondary structures within CTG microsatellites is still vividly discussed today ⁵, while *in vitro* evidences are well described ^{3,4}. Therefore, our results can be interpreted as another indirect proof of secondary structures forming within these repeats. Sae2 which is the fourth member of the MRX complex plays a critical role in the repair process of CTG repeat ¹⁶, activating the endonuclease activity of Mre11p as well as stimulating its exonuclease activity ^{31,32}. Loss of expression of *SAE2* led to the same phenotype as the *mre11D56N* and *mre11D56N sae2Δ* mutants with no significant differences, suggesting that Sae2 does not play a role independently from Mre11 and that it does not have any nuclease activity by itself on CTG repeats. Surprisingly, our results showed that the survival rate of the *sae2Δ* mutant was significantly lower than the *mre11D56N* and *mre11D56N sae2Δ* mutants but not different from the *mre11H125N* mutant. It is possible that the loss of Sae2p has an impact on DSB repair independently of nuclease activity, maybe in stabilizing the MRX complex once it is bound at the DSB. The second interpretation of these results would be the differences between the D56N and H125N mutation, the former one leading to a greater nuclease defect associated with increased accumulation of broken and unrepaired DNA molecules.

Long range resection

Long range resection is partially dependent on *EXO1* on HO clean double-strand breaks that do not require MRX activity ²⁶. Exo1p shows a strong preference for MRX generated DNA ends and have its nuclease activity stimulated by MRX. Our results demonstrate that knock-out of *EXO1* does not induce any phenotypical change in CTG repeats and was not significantly different from the wild type in resection efficacy and survival.

Involvement of telomeric protection complex in CTG end resection

In previous studies, induction of a HO DSB near an internal TG_n telomeric repeat recruited Rif1p, Rif2p and Rap1p at the DSB site. In *rif1Δ* and *rif2Δ* mutants, association of Mre11p with TG_{162} ends, following an HO DSB, is increased³⁶. Our results show a greater effect on resection of the *rif2Δ* mutation than the *rif1Δ* mutation in accordance with published work³⁶. This suggests that lack of Rif2p facilitates the accessibility of the MRX-Sae2p complex to the DSB site. In another work, *RIF1* was shown to increase cell survival in the presence of DNA damaging agents, in the absence of a functional MRX complex, suggesting a role for this gene in DSB repair. Loss of *RIF1* dramatically decreases repair by single-strand annealing over a long distance (25 kb), but not on a shorter one (4.6 kb), suggesting a long-range resection defect in this mutant. Molecular analysis of single-stranded DNA generated by a HO DSB showed that resection was defective in a *sae2Δ rif1Δ* double mutant, but not in a *rif1Δ* single mutant, showing that Rif1p was facilitating DNA end resection when the main activity promoted by the MRX-Sae2p complex was deficient⁴². In addition, Rif1p was shown by ChIP to bind to a HO cut site that was not close to a telomeric repeat, in contradiction to a previous publication³⁶. More recently the Rap1-Rif1-Rif2 complex was shown to be recruited near intrachromosomal DSB induced by HO in a Rap1 dependent fashion and seems to negatively regulate the MRX complex activity⁵⁵.

In our experiments, the DSB is made by a TALEN within a $(CTG)_{80}$ trinucleotide repeat. We found that the loss of *RIF1* or *RIF2* leads to increased DSB resection (Figure 4) while the survival rate is not affected (Figure 5). These results suggest that CTG repeats are protected from resection by the Rif1p-Rif2p complex, similarly to a telomeric end. Although a $(CTG)_{80}$ repeat is very different from a TG_{1-3} telomeric repeat, we propose that the Rif1p-Rif2p complex is attracted by the repeated nature of the CTG repeat (or maybe by its secondary structure),

protecting DSB ends from resection (Figure 6A). The effect is more pronounced at the 5' end of the break, since most of the CTG triplets are there, but is also detectable at the 3' end (Figure 4). In the absence of *RIF1* and/or *RIF2*, CTG trinucleotide repeat ends are not protected by the telomeric complex and are more accessible to MRX-Sae2p to initiate resection (Figure 6B).

In the case of the strain containing no CTG repeat, *rif1Δ* and *rif2Δ* mutants show the opposite effect, break resection is decreased (Figure 4D). The same phenotype was also observed in a *rif1Δ sae2Δ* mutant, suggesting that Rif1p helps resection of an HO DSB in the absence of the solid resection activity mediated by Sae2p⁴². These results are surprising, as it was recently shown that Rif2 negatively regulates the activity of the MRX complex within intrachromosomal double-strand breaks generated by HO⁵⁵.

It is possible that the TALEN interferes with both DSB ends more strongly than the Ho endonuclease, slightly decreasing resection efficacy. It would therefore require Rif1p-Rif2p to remove the nuclease in order to allow efficient resection by the MRX-Sae2p complex. We recently observed significant differences in DSB repair whether the break was induced by a TALEN or by *Streptococcus pyogenes* Cas9¹⁶. It is therefore possible that a DSB made by Ho or by a TALEN exhibit slightly different properties. Additional experiments will be required to test this hypothesis.

The discovery that the Rif1p-Rif2p complex specifically interacts with CTG trinucleotide repeat DSB ends is unexpected and suggests that such repeats may confuse the telomeric end protection machinery. The end-resection defect observed in CTG trinucleotide repeats may therefore not be directly caused by secondary structures, but rather by Rif1p-Rif2p binding on CTG repeats, blocking the access of the MRX-Sae2p complex. More experiments are now necessary to understand how this process is regulated and whether it is restricted to CTG

trinucleotide repeats or if it also affects other microsatellites forming alternative DNA structures.

Acknowledgements

W. V.-Z. was the recipient of a PhD fellowship from la Ligue Nationale Contre le Cancer and of a short-term fellowship from Fondation Hardy. Work in G.-F. Richard laboratory is generously supported by the Centre National de la Recherche Scientifique (CNRS) and by the Institut Pasteur. We are grateful to Frank Feuerbach for his kind advices on the Anchor Away system and the generous gift of the LMA3670 strain. We'd like to thank Eduardo Rocha for its continuous support and interest of this work.

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Orr, H. T. & Zoghbi, H. Y. Trinucleotide Repeat Disorders. *Annu. Rev. Neurosci.* **30**, 575–621 (2007).
3. Gacy, A. M., Goellner, G., Juranić, N., Macura, S. & McMurray, C. T. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* **81**, 533–540 (1995).
4. Mitas, M. Trinucleotide repeats associated with human disease. *Nucleic Acids Res.* **25**, 2245–2254 (1997).
5. Poggi, L. & Richard, G.-F. Alternative DNA Structures In Vivo: Molecular Evidence and Remaining Questions. *Microbiol. Mol. Biol. Rev. MMBR* **85**, (2021).
6. Richard, G.-F., Kerrest, A. & Dujon, B. Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes. *Microbiol. Mol. Biol. Rev. MMBR* **72**, 686–727 (2008).
7. Harper, P. *Myotonic Dystrophy*. (OUP Oxford, 2009).
8. Turner, C. & Hilton-Jones, D. The myotonic dystrophies: diagnosis and management. *J. Neurol. Neurosurg. Psychiatry* **81**, 358–367 (2010).
9. Mankodi, A. *et al.* Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.* **10**, 2165–2170 (2001).
10. Miller, J. W. *et al.* Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.* **19**, 4439–4448 (2000).
11. Kanadia, R. N. *et al.* Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3. *Gene Expr. Patterns GEP* **3**, 459–462 (2003).
12. O’Hoy, K. *et al.* Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. *Science* **259**, 809 (1993).
13. Richard, G.-F., Dujon, B. & Haber, J. E. Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats. *Mol. Gen. Genet. MGG* **261**, 871–882 (1999).
14. Richard, G.-F. Shortening trinucleotide repeats using highly specific endonucleases: a possible approach to gene therapy? *Trends Genet. TIG* **31**, 177–186 (2015).
15. Richard, G.-F. *et al.* Highly Specific Contractions of a Single CAG/CTG Trinucleotide Repeat by TALEN in Yeast. *PLOS ONE* **9**, e95611 (2014).
16. Mosbach, V., Poggi, L., Viterbo, D., Charpentier, M. & Richard, G.-F. TALEN-Induced Double-Strand Break Repair of CTG Trinucleotide Repeats. *Cell Rep.* **22**, 2146–2159 (2018).
17. Fishman-Lobell, J., Rudin, N. & Haber, J. E. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol. Cell. Biol.* **12**, 1292–1303 (1992).
18. Pâques, F. & Haber, J. E. Multiple Pathways of Recombination Induced by Double-Strand Breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).
19. Symington, L. S. End Resection at Double-Strand Breaks: Mechanism and Regulation. *Cold Spring Harb. Perspect. Biol.* **6**, (2014).
20. Zhu, Z., Chung, W.-H., Shim, E. Y., Lee, S. E. & Ira, G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double strand break ends. *Cell* **134**, 981–994 (2008).
21. Bennardo, N., Cheng, A., Huang, N. & Stark, J. M. Alternative-NHEJ Is a Mechanistically Distinct Pathway of Mammalian Chromosome Break Repair. *PLoS Genet.* **4**, (2008).
22. Bhargava, R., Onyango, D. O. & Stark, J. M. Regulation of Single Strand Annealing and

- its role in genome maintenance. *Trends Genet. TIG* **32**, 566–575 (2016).
23. Ivanov, E. L., Sugawara, N., Fishman-Lobell, J. & Haber, J. E. Genetic requirements for the single-strand annealing pathway of double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* **142**, 693–704 (1996).
 24. Mimitou, E. P. & Symington, L. S. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455**, (2008).
 25. You, Z. *et al.* CtIP Links DNA Double-Strand Break Sensing to Resection. *Mol. Cell* **36**, 954–969 (2009).
 26. Llorente, B. & Symington, L. S. The Mre11 Nuclease Is Not Required for 5' to 3' Resection at Multiple HO-Induced Double-Strand Breaks. *Mol. Cell. Biol.* **24**, 9682–9694 (2004).
 27. Moreau, S., Ferguson, J. R. & Symington, L. S. The nuclease activity of Mre11 is required for meiosis but not for mating type switching, end joining, or telomere maintenance. *Mol. Cell. Biol.* **19**, 556–566 (1999).
 28. Garcia, V., Phelps, S. E. L., Gray, S. & Neale, M. J. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature* **479**, 241–244 (2011).
 29. Paull, T. T. & Gellert, M. The 3' to 5' Exonuclease Activity of Mre11 Facilitates Repair of DNA Double-Strand Breaks. *Mol. Cell* **1**, 969–979 (1998).
 30. Krogh, B. O., Llorente, B., Lam, A. & Symington, L. S. Mutations in Mre11 Phosphoesterase Motif I That Impair *Saccharomyces cerevisiae* Mre11-Rad50-Xrs2 Complex Stability in Addition to Nuclease Activity. *Genetics* **171**, 1561–1570 (2005).
 31. Cannavo, E. & Cejka, P. Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. *Nature* **514**, 122–125 (2014).
 32. Lengsfeld, B. M., Rattray, A. J., Bhaskara, V., Ghirlando, R. & Paull, T. T. Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex. *Mol. Cell* **28**, 638–651 (2007).
 33. Prinz, S., Amon, A. & Klein, F. Isolation of COM1, a new gene required to complete meiotic double-strand break-induced recombination in *Saccharomyces cerevisiae*. *Genetics* **146**, 781–795 (1997).
 34. Haruki, H., Nishikawa, J. & Laemmli, U. K. The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes. *Mol. Cell* **31**, 925–932 (2008).
 35. Chung, W.-H., Zhu, Z., Papusha, A., Malkova, A. & Ira, G. Defective Resection at DNA Double-Strand Breaks Leads to De Novo Telomere Formation and Enhances Gene Targeting. *PLOS Genet.* **6**, e1000948 (2010).
 36. Hirano, Y., Fukunaga, K. & Sugimoto, K. Rif1 and Rif2 inhibit localization of Tel1 to DNA ends. *Mol. Cell* **33**, 312–322 (2009).
 37. Wotton, D. & Shore, D. A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in *Saccharomyces cerevisiae*. *Genes Dev.* **11**, 748–760 (1997).
 38. Hailemariam, S. *et al.* The telomere-binding protein Rif2 and ATP-bound Rad50 have opposing roles in the activation of yeast Tel1/ATM kinase. *J. Biol. Chem.* **294**, 18846–18852 (2019).
 39. Jessop, L., Allers, T. & Lichten, M. Infrequent Co-conversion of Markers Flanking a Meiotic Recombination Initiation Site in *Saccharomyces cerevisiae*. *Genetics* **169**, 1353–1367 (2005).
 40. Martini, E. *et al.* Genome-Wide Analysis of Heteroduplex DNA in Mismatch Repair-Deficient Yeast Cells Reveals Novel Properties of Meiotic Recombination Pathways. *PLOS Genet.* **7**, e1002305 (2011).

41. Wellinger, R. J. & Zakian, V. A. Everything You Ever Wanted to Know About *Saccharomyces cerevisiae* Telomeres: Beginning to End. *Genetics* **191**, 1073–1105 (2012).
42. Martina, M., Bonetti, D., Villa, M., Lucchini, G. & Longhese, M. P. *Saccharomyces cerevisiae* Rif1 cooperates with MRX-Sae2 in promoting DNA-end resection. *EMBO Rep.* **15**, 695–704 (2014).
43. Contractions and Expansions of CAG/CTG Trinucleotide Repeats occur during Ectopic Gene Conversion in Yeast, by a MUS81-independent Mechanism. *J. Mol. Biol.* **326**, 769–782 (2003).
44. Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci.* **78**, 6354–6358 (1981).
45. Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **21**, 3329–3330 (1993).
46. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
47. Boeke, J. D., Trueheart, J., Natsoulis, G. & Fink, G. R. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.* **154**, 164–175 (1987).
48. Kahli, M., Osmundson, J. S., Yeung, R. & Smith, D. J. Processing of eukaryotic Okazaki fragments by redundant nucleases can be uncoupled from ongoing DNA replication in vivo. *Nucleic Acids Res.* **47**, 1814–1822 (2019).
49. Viterbo, D. *et al.* A fast, sensitive and cost-effective method for nucleic acid detection using non-radioactive probes. *Biol. Methods Protoc.* **3**, (2018).
50. Mosbach, V., Poggi, L., Viterbo, D., Charpentier, M. & Richard, G.-F. TALEN-induced double-strand break repair of CTG trinucleotide repeats. *Cell Rep.* **22**, 2146–2159 (2018).
51. Millot, G. *Comprendre et réaliser les tests statistiques à l'aide de R: manuel de biostatistique.* (De Boeck, 2011).
52. Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766–1769 (1995).
53. Freudenreich, C. H., Stavenhagen, J. B. & Zakian, V. A. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell. Biol.* **17**, 2090–2098 (1997).
54. Viterbo, D., Michoud, G., Mosbach, V., Dujon, B. & Richard, G.-F. Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. *DNA Repair* **42**, 94–106 (2016).
55. Bonetti, D. *et al.* DNA binding modes influence Rap1 activity in the regulation of telomere length and MRX functions at DNA ends. *Nucleic Acids Res.* **48**, 2424–2441 (2020).
56. Cassani, C. *et al.* Tel1 and Rif2 Regulate MRX Functions in End-Tethering and Repair of DNA Double-Strand Breaks. *PLoS Biol.* **14**, (2016).
57. Mosbach, V. *et al.* Resection and repair of a Cas9 double-strand break at CTG trinucleotide repeats induces local and extensive chromosomal deletions. *PLoS Genet.* **16**, (2020).
58. Chen, H., Lisby, M. & Symington, L. S. RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Mol. Cell* **50**, (2013).
59. Lobachev, K. S., Gordenin, D. A. & Resnick, M. A. The Mre11 Complex Is Required for Repair of Hairpin-Capped Double-Strand Breaks and Prevention of Chromosome Rearrangements. *Cell* **108**, 183–193 (2002).
60. Cannavo, E., Cejka, P. & Kowalczykowski, S. C. Relationship of DNA degradation by

Saccharomyces cerevisiae Exonuclease 1 and its stimulation by RPA and Mre11-Rad50-Xrs2 to DNA end resection. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E1661–E1668 (2013).

61. Shim, E. Y. *et al.* Saccharomyces cerevisiae Mre11/Rad50/Xrs2 and Ku proteins regulate association of Exo1 and Dna2 with DNA breaks. *EMBO J.* **29**, 3370–3380 (2010).

Figure 1: Time course of TALEN induction in wild type, *mre11* mutants and *sae2* Δ strains

A: Schematic representation of the TALEN experimental system. The TALEN is represented by the scissors and the repeats by the dark pink square. Two EcoRV sites surrounds the repeated tract and are used to digest DNA before analyzing the sequences by Southern blot. The 5' and 3' non-radioactive probes used in the experiments are in red and blue respectively. **B:** Southern blots of time courses of TALEN induction in a wild-type strain. The nuclease is repressed (OFF) or induced (ON) in the presence or absence of doxycyclin. Signals visible on the Southern blot correspond to the intact uncut allele containing ~70-80 CTG triplets ("Uncut allele"), to the same allele following contraction to a smaller length ("Contracted allele"), a smear representing all possible CTG repeat lengths on the 5' end of the DSB ("5' ends (CTG)*n*") and to the 3' end of the DSB containing only a few CTG triplets ("3' end"). At T0, doxycycline is added ("DOX") or omitted ("-DOX") from the medium. The first time point is taken 14 hours afterwards. **C:** Southern blots of time courses of TALEN induction in *mre11* mutants. Quantification of 5' and 3' signals for wild type and mutants are shown on each graph, for each time point. Values are given in percentage on total signal in each lane. **D:** Southern blots of time courses of TALEN induction in *sae2* Δ and *mre11D56N* mutants. Same as B.

Figure 2: Time course of TALEN induction in *exo1* Δ mutants

A: Southern blot of time course in *exo1* Δ strain. **B:** Southern blot of time courses in *mre11D56N exo1* Δ and *mre11D56N exo1* $\Delta sae2$ Δ strains. Otherwise, same legend as Figure 1.

Figure 3: Time course of TALEN induction in *rif1Δ* and *rif2Δ* mutants

A: Southern blot of time courses in single and double mutants, in a strain containing a (CTG)₈₀ trinucleotide repeat. B: Southern blot of time courses in wild type, *rif1Δ* and *rif2Δ* mutants, in a strain containing no CTG trinucleotide repeat. Otherwise, same legend as Figure 1.

Figure 4: Analysis of DSB resection

A: Schematic representation of the four *EcoRV* restriction sites used to study resection by real-time qPCR. Each restriction site is surrounded by a couple of specific primers. The scissors represent the DSB site. B: Resection values in the WT, *sae2Δ*, *mre11D56N*, *mre11H125N* and *exo1Δ* strains. Each graph is shown under its corresponding proximal or distal restriction site. For calculation of relative resection values, see Materials & Methods. Asterisks: resection is significantly different from the wild type (unilateral student's t-test); *: p-value < 0,0127; ** p-value < 0,0025 after Bon feronni correction. C: Resection values in the wild type, *sae2Δ*, *mre11D56N*, *rif1Δ*, *rif2Δ* and *rif1Δ rif2Δ* strains. Results in *sae2Δ* and *mre11D56N* strain are shown to make comparisons easier but are identical to 4B. Asterisks: resection is significantly different from the wild type (unilateral student's t-test); *: p-value < 0,0127; ** p-value < 0,0025. A single asterisk on the *rif1Δ*, *rif2Δ* or *rif1Δ rif2Δ* means a significant difference with the WT in a unilateral student's t.test (pvalue < 0,017) and two asterisks a significant difference in a unilateral student's t.test (pvalue < 0,0033) after Bon feronni correction. D: Resection values in the wild type, *rif1Δ*, *rif2Δ* and *rif1Δ rif2Δ* mutants, in a strain containing no CTG. Results in the strain containing (CTG)₈₀ are shown to make comparisons easier but are identical to 4C. Black asterisks: resection is significantly different from the wild type containing a (CTG)₈₀ repeat tract (unilateral student's t-test); *: p-value < 0,017; ** p-value < 0,0033. Red

asterisks: resection is significantly different from the wild type containing no CTG repeat (unilateral student's t-test); *: p-value < 0,0253; ** p-value < 0,0050.

Figure 5: Cell viability following a TALEN-induced DSB

A: Survival of the wild type and resection mutants. Black asterisk: significant difference with wild type (unilateral student's t-test p-value < 0,01). Red asterisk: significant difference with with the *sae2* Δ strain. **B:** Survival of the wild type, *rif1* Δ and *rif2* Δ mutants in strains containing (CTG)₈₀ or no repeat.

Figure 6: Interaction of the MRX-Sae2p complex with Rif1p-Rif2p at a broken CTG trinucleotide repeat.

A: In a wild type cell, the Rap1p-Rif1-Rif2p complex is naturally attracted to CTG trinucleotide repeats, protecting it from processing by MRX-Sae2p. The 5' left end of the break that contains most of the CTG triplets is harder to access than the 3' right end, and resection is therefore less efficient and slower. **B:** In the absence of Rif1-Rif2p, the MRX-Sae2p complex readily accesses the DSB and resection is faster and more complete.

Supplementary Figure 1: Analysis of DSB Resection in *mre11D56N sae2* Δ , *mre11D56N exo1* Δ and *mre11D56N sae2* Δ *exo1* Δ

A: Schematic representation of the four *EcoRV* restriction sites used to study resection by real-time qPCR. Each restriction site is surrounded by a couple of specific primers. The scissors represent the DSB site. **B:** Resection values in the WT, *mre11D56N sae2* Δ , *mre11D56N exo1* Δ , and *mre11D56N sae2* Δ *exo1* Δ strains. Each graph is shown under its corresponding proximal or distal restriction site. For calculation of relative resection values, see Materials & Methods.

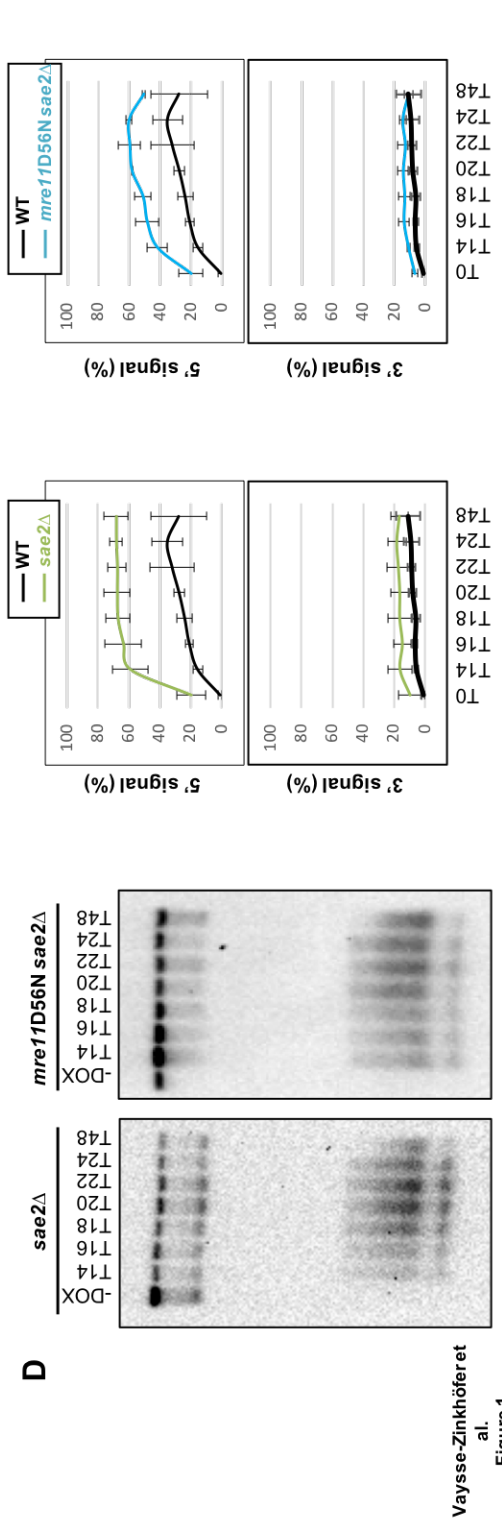
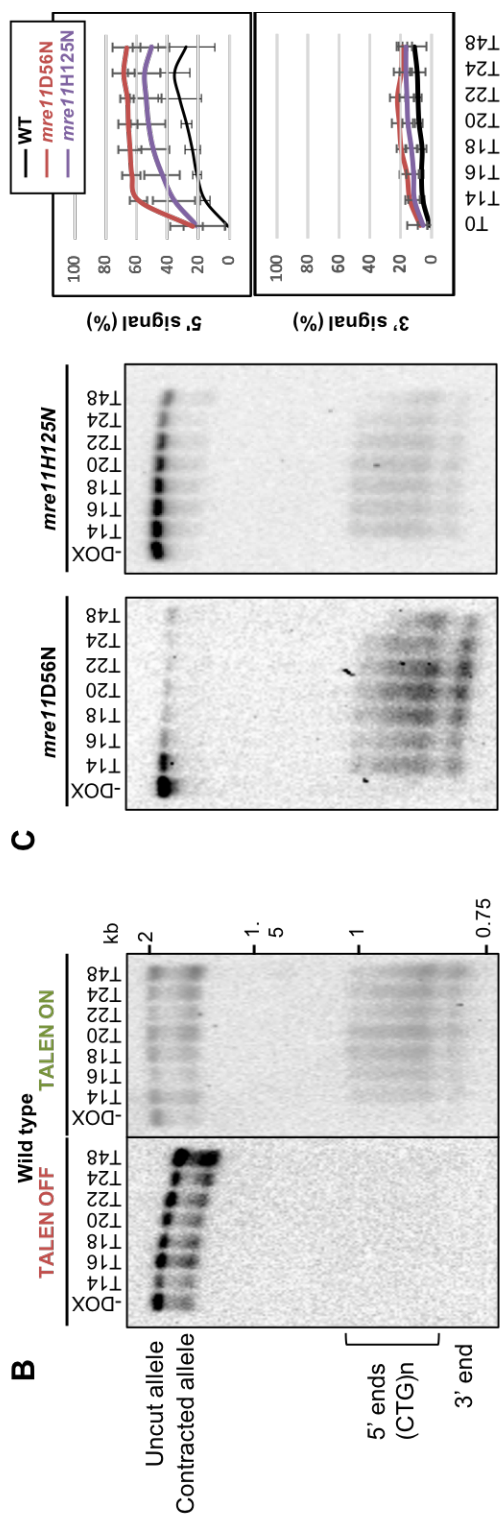
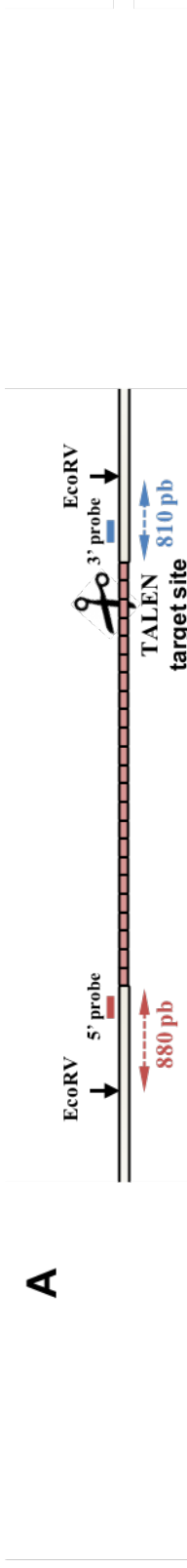
Asterisks: resection is significantly different from the wild type (unilateral student's t-test); *: p-value < 0,016; ** p-value < 0,0033 after Bon feronni correction.

Supplementary Figure 2: Southern blot analysis of CTG repeat length in cells after TALEN induction

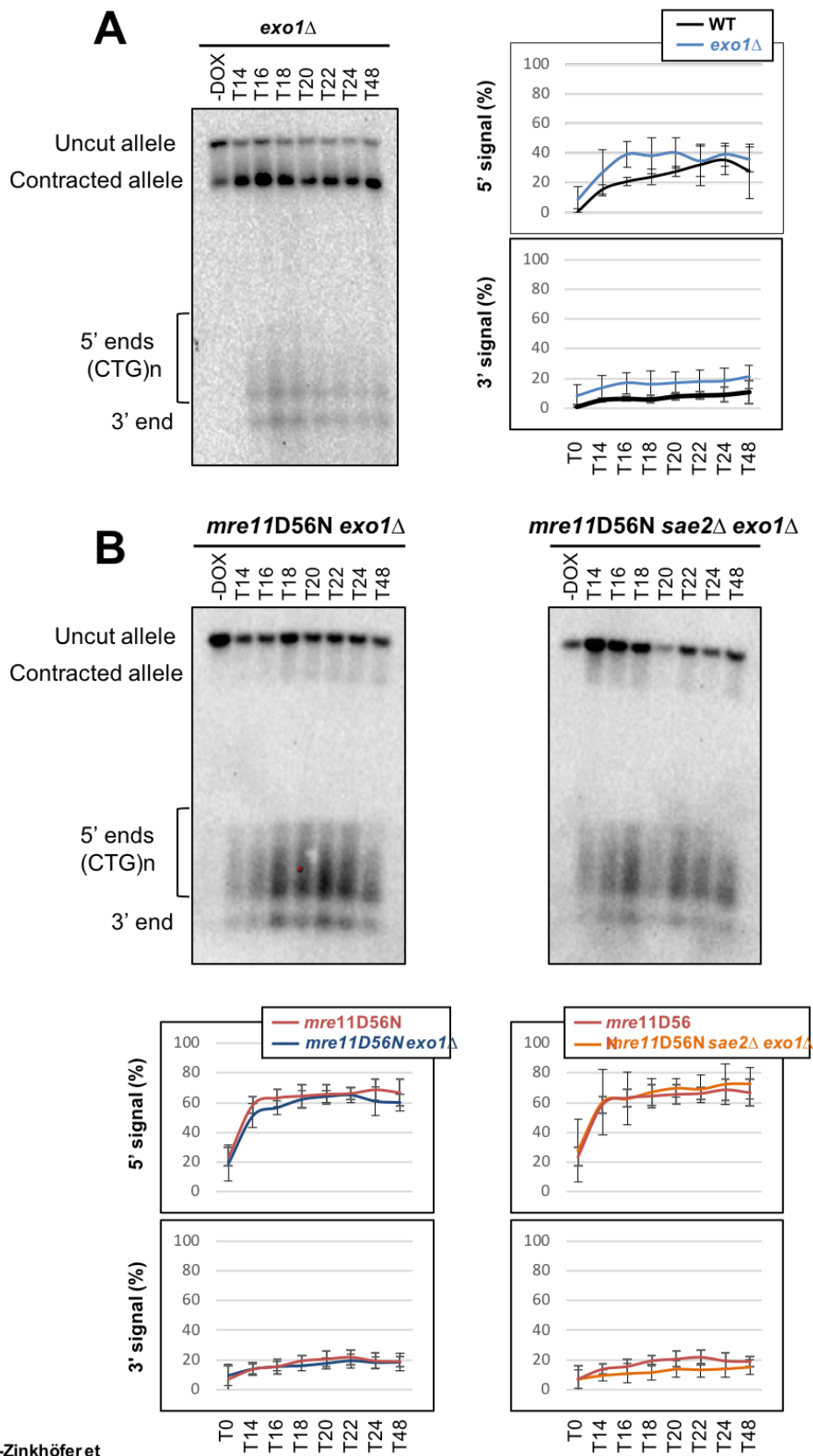
A: Examples of results obtained in different mutant strains. On each gel, a CTG ladder was run in parallel to digested genomic DNA ⁵⁴. The ladder hybridizes with the probe, for an easy determination of repeat tract length (CTG nbr). Red arrows point to clones in which the CTG repeat tract was not contracted. Red brackets indicate smears (see text). **B:** Comparison of CTG repeat tract length of different mutants with wild type. The number of clones analyzed in each strain are: *mre11D56N*: 39; *mre11H125N*: 42; *mre11D56N sae2Δ*: 47; *mre11D56N exo1Δ*: 48; *mre11D56N sae2Δ exo1Δ*: 45; *exo1Δ*: 39; *rif1Δ*: 42; *rif2Δ*: 36; *rif1Δ rif2Δ*: 45; WT: 45. The *mre11D56N* and *mre11D56N exo1Δ* mutants shows significantly less contraction than the wild type (exact Fisher's test p-value < 0.01 after Bonferroni correction). The *exo1Δ*, *mre11D56N sae2Δ* and triple mutants were not significantly different from the wild type strain. The *rif1Δ* and *rif2Δ* mutants did not show significantly higher CTG repeat contraction efficacy (exact Fisher's test p-value < 0.016 after Bonferroni correction). The *rif1Δ* and *rif1Δ rif2Δ* mutants showed increased smear formations (exact Fisher's test p-value < 0.00625 after Bonferroni correction).

Supplementary Table 1: Primers for strains construction

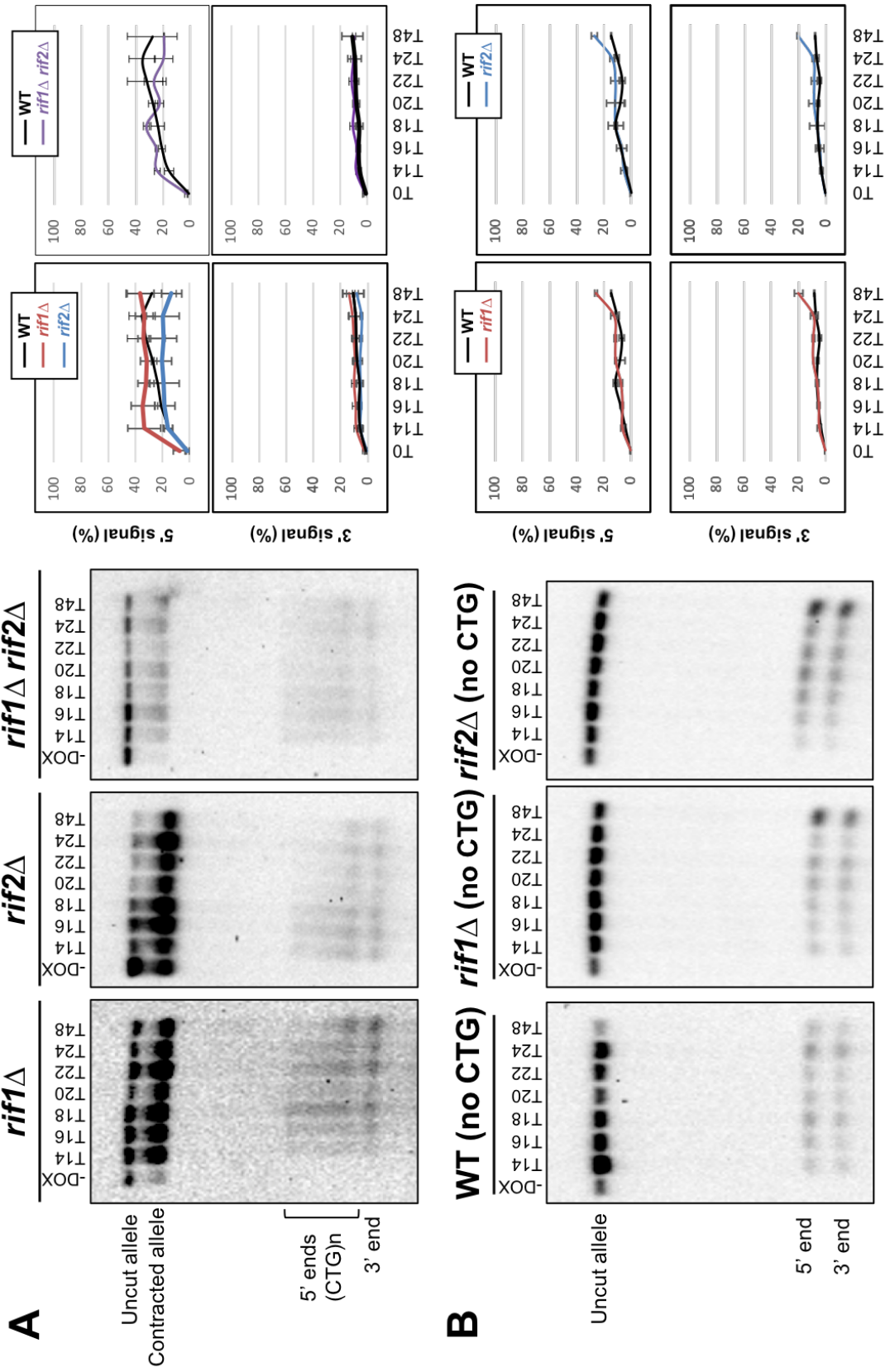
Supplementary Table 2: Strain names and genotypes



Vaysse-Zinkhöfer et al.
Figure 1

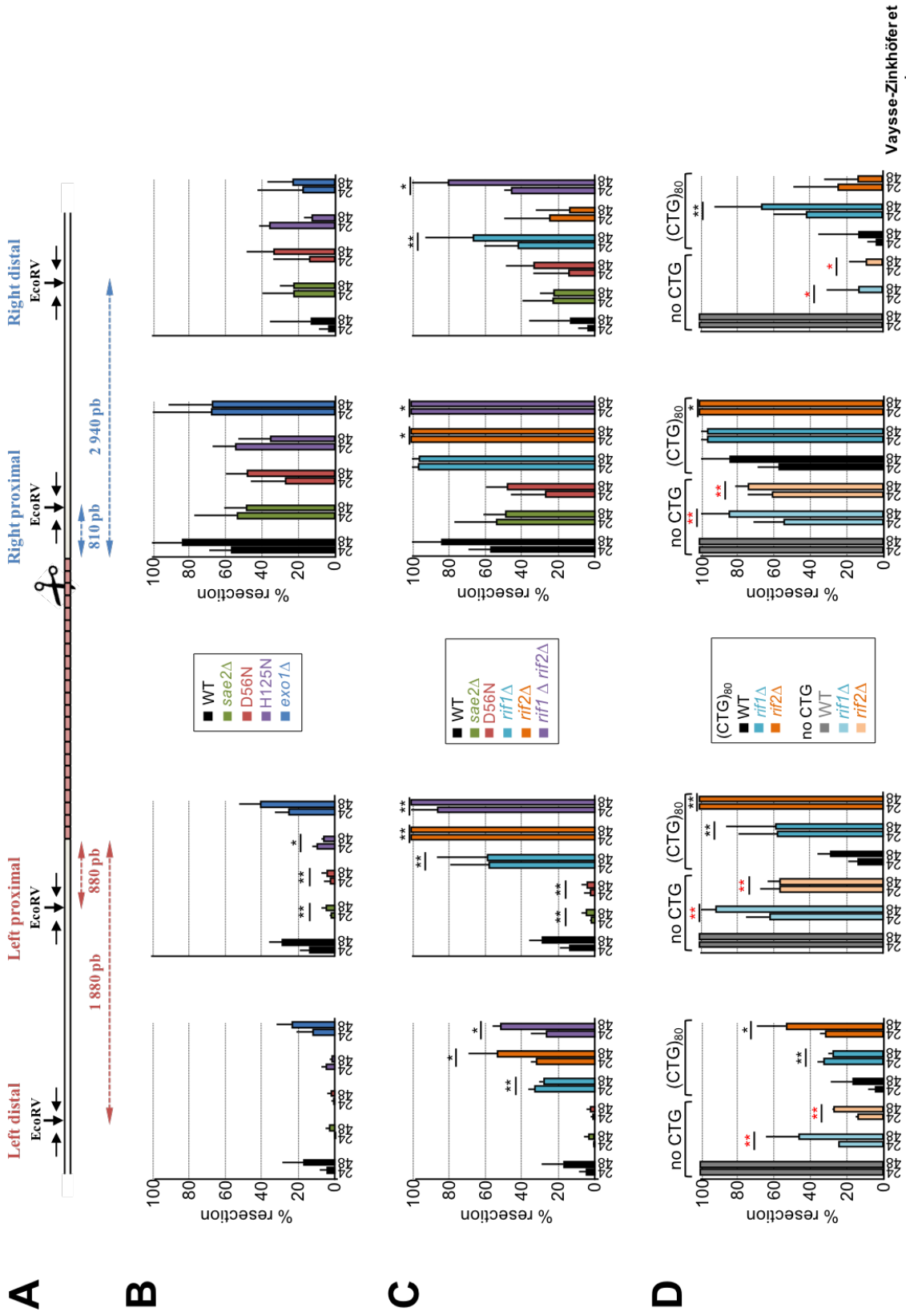


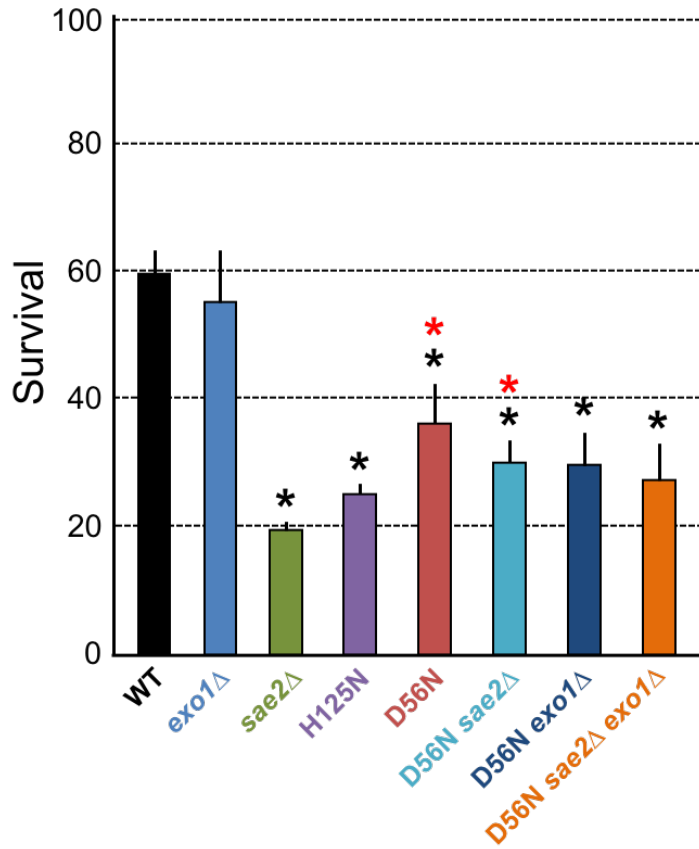
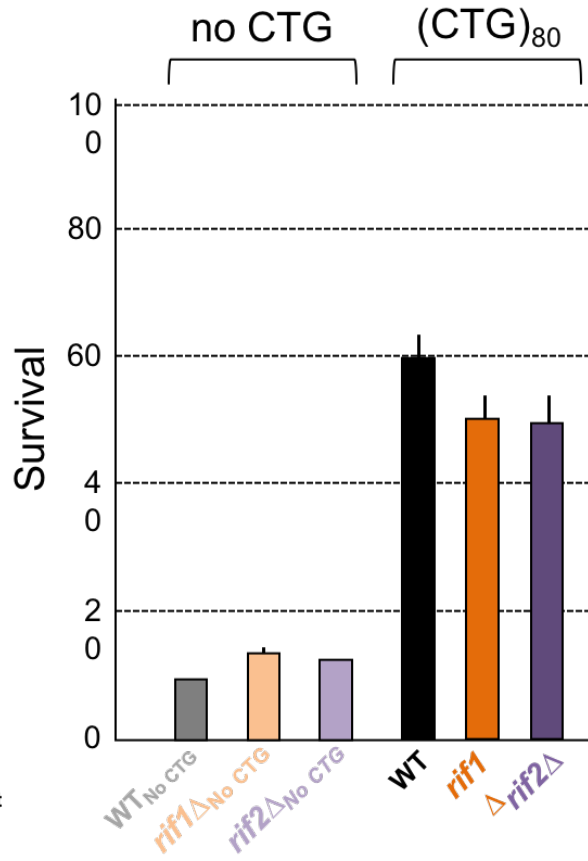
Vaysse-Zinkhöfer et al.
Figure 2



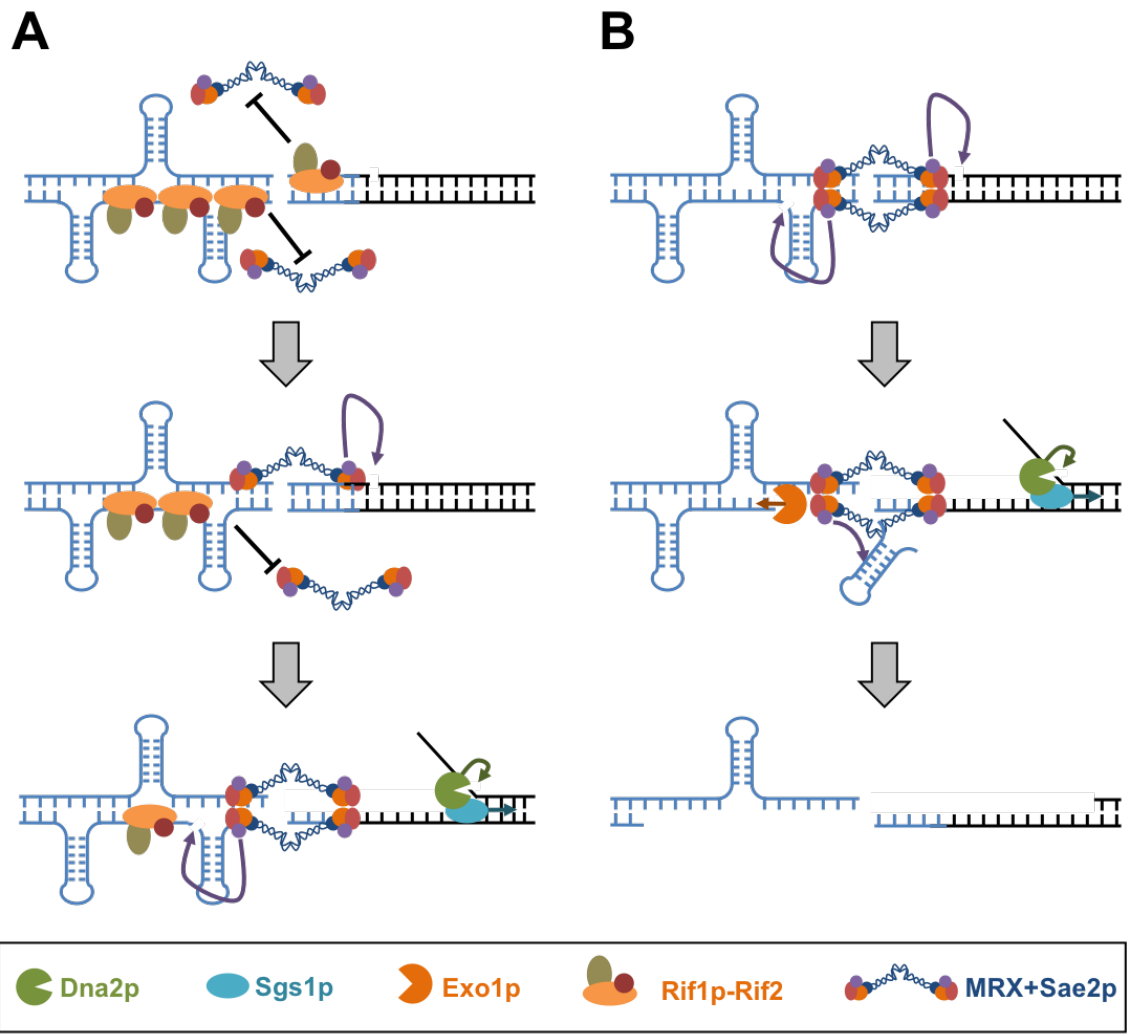
Vaysse-Zinkh fer et al.

Figure 3



A**B**

Vaysse-Zinkhöfer et al.
Figure 5



Vaysse-Zinkhøfer et al.
Figure 6

Supplementary Table 1 : Primers for strains construction

Primers' name	Description	Sequence
FW Exo1	Amplify the KANMX cassette with 500bp <i>EXO1</i> region homology in the EUROSCARF <i>exo1Δ</i> mutant	TGCATGGACAGGTTCCACAA
Rv Exo1		ACAGCAACGAGACGGTTCC
FW MRE11 P1	Amplify a MRE11 fragment carrying the point mutation D16A, D56N and H125N as well as the rest of the gene with no mutation. The plasmids were designed and sent by Twist Bioscience	GTAAAACGACGGCCAGT
RV MRE11 P1		CAGGAAACAGCTATGAC
FW RIF2leu2	Amplify <i>LEU2</i> for insertion at <i>RIF2</i> locus (50bp homology on each primer with the <i>RIF2</i> locus)	ATATGTACGAATATAGATATAAATACGAACGTGGTTAGTATATAGAGA CAGTCGGCTTGCTACCTTGCC
RV RIF2leu2		GAAGTAAAGACAGAGCTGGGATTATGATTCATTCCCGGTATAAATA TACCTCTAGCTAGAGTCGAGATCCCC
FW RIF1HIS3	Amplify <i>HIS3</i> for insertion at <i>RIF1</i> locus (50bp homology on each primer with the <i>RIF1</i> locus)	ACGTTATTACAGGATTGCCATTGCAAATCGTTTTTGTGGTCAATTTGC AACTCTTGGCCTCCTCTAG
RV RIF1HIS3		ATTATTGTAATTAATTTATTGCCATTTTGATCTATTCTACATACTAACAAT CGTTCAGAATGACACG
FW URA3pp	Amplify URA3 for insertion in pTRI612 carrying 80 CAG. This allow for genomic insertion of CAG and selection by the popin-popout method	ATCACGGAAGAGGGGACACGAGAAAACGAGGGGGGGCCCATGCCG CATCAGAGCAGATT
FW URA3pp		GGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTGCTCCTTAC GCATCTGTGCGG
FW SGS1 AA	Amplification of the FRB tag with 50bp homology with <i>SGS1</i> Nt region, in the pFA6-FRB-kan plasmid	ATGGGAGACGAGGTTTTAGAAATTACCGAGGCTACTACAGAGGAAGA AAGCGGATCCCCGGGTTAATTAA
RV SGS1 AA		TTGGCGAATGGTGTGCTAGTTATAAGTAACACTATTTATTTTTCTACTCT GAATTCGAGCTCGTTTAAAC
FW DNA2 AA	Amplification of the FRB tag with 50bp homology with <i>DNA2</i> Nt region, in the pFA6-FRB-kan plasmid	TCGTAAGCGATAAACCTATCATAAAGGAAATTCTACAAGAGTATGAAA GTCGGATCCCCGGGTTAATTAA
RV DNA2 AA		TATGCTGTGATAGCTTTCTGTTATGGAGAAGCTCTTCTTATCCCCTG GAATTCGAGCTCGTTTAAAC

Supplementary Table 2 : Strain names and genotypes

Name	Nuclear Genotype	MAT
WVZ 1-A	<i>mre11D56N ura3Δ851 leu2Δ1 his3Δ200 trp1Δ63 ade2-opal SUP4::CAG255</i>	α
WVZ 6-A	<i>exo1::KanMX ura3Δ851 leu2Δ1 his3Δ200 trp1Δ63 ade2-opal SUP4::CAG255</i>	α
WVZ 7	<i>mre11D56N sae2::HIS3 ura3Δ851 leu2Δ1 his3Δ200 trp1Δ63 ade2-opal SUP4::CAG255</i>	α
WVZ 8	<i>mre11H125N2 ura3Δ851 leu2Δ1 his3Δ200 trp1Δ63 ade2-opal SUP4::CAG255</i>	α
WVZ 10	<i>mre11D56N exo1::HIS3 ura3Δ851 leu2Δ1 his3Δ200 trp1Δ63 ade2-opal SUP4::CAG255</i>	α
WVZ 12	<i>mre11D56N sae2::HIS3 exo1::KanMX ura3Δ851 leu2Δ1 his3Δ200 trp1Δ63 ade2-opal SUP4::CAG255</i>	α
WVZ 14	<i>Tor1-1 fpr1Δlox RPL13a FKBP2::his3MX SUP4::CAG60</i>	α
WVZ 18	<i>rif1::HIS3 ura3D851 leu2D1 his3D200 trp1D63 ade2-opal SUP4::CAG255</i>	α
GFY6162-3D	<i>ura3D851 leu2Δ1 his3D200 trp1D63 ade2-opal SUP4::CAG255</i>	α
LMA3670	<i>Tor1-1 fpr1Δlox RPL13a FKBP2::his3MX Lys2 met15Δ0 ura3Δ0 leu2Δ0 his3Δ1</i>	a
WVZ 23	<i>rif2::LEU2 ura3D851 leu2Δ1 his3D200 trp1D63 ade2-opal SUP4::CAG255</i>	α
WVZ 24	<i>ura3D851 leu2Δ1 his3D200 trp1Δ63 ade2-opal SUP4::CAG30 SGS1-FRB</i>	α
WVZ 25	<i>rif1::HIS3 rif2::LEU2 ura3Δ851 leu2Δ1 his3Δ200 trp1Δ63 ade2-opal SUP4::CAG255</i>	α
WVZ 26	<i>Tor1-1 fpr1Δlox RPL13a FKBP2::his3MX SUP4::CAG255</i>	ND
WVZ 27	<i>rif2::LEU2 ura3D851 leu2D1 his3D200 trp1D63 ade2-opal</i>	ND
WVZ 30	<i>rif1::HIS3 ura3D851 leu2D1 his3D200 trp1D63 ade2-opal</i>	ND
WVZ 31	<i>rif1::HIS3 rif2::LEU2 ura3D851 leu2D1 his3D200 trp1D63 ade2-opal</i>	ND
WVZ 32	<i>SGS1-FRB KanMX Tor1-1 fpr1Δlox RPL13a FKBP2::his3MX SUP4::CAG255</i>	ND
YDS43	<i>RAD5 tor1-1 ::HIS3 fpr1 ::NatMX4 RPL13a-2xFKBP12::TRP1 DNA2-FRB</i>	a

DISCUSSION & PERSPECTIVES

Les différents mécanismes régissant la stabilité des répétitions de trinuécléotides sont complexes et leurs rôles respectifs pas toujours pas parfaitement déterminés. La nature des microsatellites et notamment des microsatellites à répétitions de CTG, qui ont été centraux à mes travaux de thèse, facilite la formation de structures secondaires menant aux expansions qui conduisent au développement de maladies neurodégénératives tel que la DM1¹². En 1993, la découverte d'une jeune femme qui après transmission d'un allèle paternel ayant subi une contraction des répétitions CTG³¹⁹, n'a pas développé les symptômes associés à la maladie, pose le problème suivant : est-il possible d'induire volontairement des contractions par génération de cassures au sein de répétitions afin de guérir des patients atteints de maladies liés aux expansions de répétitions ?

L'utilisation d'une TALEN s'est montrée très efficace pour induire des cassures et contractions au sein de répétitions CTG insérées chez la levure *S. cerevisiae*^{29,296} et *M. musculus*^{215,320}. Des expériences similaires dans des cellules humaines sont en cours. Ce modèle unique nous permet de suivre une cassure double-brin et sa réparation à la fois au sein d'une région répétée et vraisemblablement structurée²⁷, mais aussi au sein d'une région non répétée et non structurée (figure 15). Ce modèle a permis de déterminer qu'une cassure double-brin générée au sein de répétitions de triplets CTG était réparée par le mécanisme du *Single-Strand Annealing* (SSA)²¹⁵. Le SSA permet la réparation par hybridation de grandes homologies de part et d'autre de la cassure indépendamment de *RAD51* mais est dépendant de la résection, qui libère de grandes régions simple brin et génère de l'homologie permettant l'hybridation des deux extrémités de la cassure. L'interaction de la cassure avec l'ensemble des facteurs intervenant dans sa réparation est une façon de mieux comprendre comment ces derniers peuvent jouer un rôle dans la stabilité des répétitions. Ainsi, mes travaux se sont articulés autour de deux grands thèmes.

(I) Les mécanismes moléculaires et rôles des protéines impliquées dans la réparation d'une cassure double-brin au sein de répétitions CTG

Mes travaux ont permis de déterminer le rôle et les fonctions de plusieurs protéines intervenant directement dans la réparation par *Single-Strand Annealing* (SSA) et la résection des cassures double-brin des microsatellites CTG. Les mécanismes du SSA sont connus et les protéines impliquées ont été caractérisées. Mon travail s'est donc concentré sur l'étude de l'importance de l'activité nucléase de Mre11, appartenant au complexe MRX de la levure *S. cerevisiae*, de Sae2, Exo1, Sgs1 et Dna2.

Le complexe MRX intervient dans les étapes initiales de la résection en éliminant tout blocage des extrémités de la cassure, au travers de l'activité nucléase de Mre11. Mre11 qui possède à la fois une activité endonucléase et exonucléase¹⁹⁸ et dont l'activité est régulée par la protéine Sae2^{206,223} sont critiques dans la réparation d'une cassure générée par Spo11, nucléase responsable des cassures méiotiques conduisant à la recombinaison³²¹, mais n'est pas nécessaire dans la réparation de cassures liées à des radiations ou générées par des nucléases tel que HO ou I-SceI²³⁷. Mes travaux mettent en évidence l'importance critique de l'activité nucléase de Mre11 et de Sae2 sur la réparation d'une cassure double-brin au sein de répétitions CTG. Ces résultats suggèrent fortement la présence de structures secondaires qui bloquent l'accès aux extrémités de la cassure aux autres protéines de la résection telles que Exo1 et Dna2/Sgs1. RPA a été montré comme capable d'empêcher la formation de structures secondaires, favorisant la résection²⁵⁶. La surexpression de RPA dans un mutant D56N ou H125N portant des répétitions CTG pourrait empêcher la formation de structures secondaires et valider notre théorie.

La mutation H59S, caractérisée plus tardivement que les mutations D56N et H125N, inhibe complètement l'activité exonucléase de Mre11 tout en préservant son activité endonucléase ²⁰⁹ permettant d'étudier séparément chaque activité nucléases de Mre11. L'étude de cette mutation dans le contexte de la réparation d'une cassure double-brin au sein de répétitions CTG pourrait apporter une preuve indirecte de la présence de structures secondaires gênant la résection. L'étude de ce mutant pourrait aussi permettre une meilleure distinction de l'importance de chacune des deux activités nucléases de Mre11 durant la réparation d'une cassure double-brin.

Les expériences réalisées sur Exo1 ne montrent aucun effet de la perte de la protéine sur la réparation. Il est encore nécessaire d'élucider le rôle de Dna2/Sgs1 à la cassure avant de pouvoir déterminer si l'absence d'effets observés en absence d'Exo1 est due à une activité redondante de Dna2 sur la résection de la cassure ou si le complexe MRX est suffisant, pour résecter et éliminer assez de structure secondaire afin de libérer des séquences d'homologie nécessaires à la recombinaison et réparation par SSA.

Notre modèle de levure est un moyen simple et efficace d'appréhender les mécanismes de réparation et d'instabilité des répétitions chez *Homo sapiens*, mais ne reflète pas ces mécanismes dans leur intégralité. Le développement d'un modèle humain où il serait possible de contrôler la génération d'une cassure unique à un locus précis permettrait de grandes avancées dans la compréhension de l'instabilité des répétitions. Un modèle dans des cellules HEK293, développé par une ancienne étudiante du laboratoire, Lucie Poggi, permet de suivre des événements de recombinaison dans des cellules humaines après induction d'une nucléase (TALEN ou CRISPR-Cas) capable de reconnaître et d'induire une cassure au sein de différentes répétitions ³²⁰. Ce modèle est une première étape vers le développement d'un système permettant l'analyse du rôle des orthologues des protéines Exo1, Sae2, Mre11, Sgs1,

Dna2 dans la réparation et la stabilité des répétitions chez *Homo sapiens* et de manière générale de comparer les résultats obtenus dans la levure avec un modèle humain.

Une limitation du modèle levure est le biais d'instabilité des répétitions vers les contractions, qui ne reflète pas le biais vers les expansions observées chez l'Humain. Les mécanismes générant de l'instabilité nécessitent d'être étudiés dans un modèle humain qui devrait posséder les caractéristiques suivantes : isolations des événements de contractions afin de pouvoir étudier spécifiquement les expansions, observations simples et visuelles des événements d'expansions et finalement contre-sélection des mutations pouvant causer des faux positifs lors des observations. Je propose une construction pouvant servir d'outil à l'étude des expansions et des contractions chez l'humain. Les « *GA binding proteins* » ont été montrées comme pouvant jouer un rôle activateur ou répresseur sur les promoteurs ribosomiques rpL30 et rpS16 respectivement. Tirant avantage de cette particularité, deux gènes rapporteurs (i.e mCherry et GFP) sont placés en aval des promoteurs rpL30 et rpS16. Un nombre de répétitions à la limite de l'instabilité (36-50 triplets) sont placées directement en 3' ou 5' de la séquence codante GABP sans altérer la fonction de la protéine. Une contraction ou une absence d'instabilité des répétitions permettent l'expression correcte de GABP qui favorise l'expression du promoteur rpL30 et la production de GFP (figure 18). Un événement d'expansion conduit à la perte de fonction de la protéine GABP, favorisant l'expression du promoteur rpS16 et la production de mCherry (figure 18).

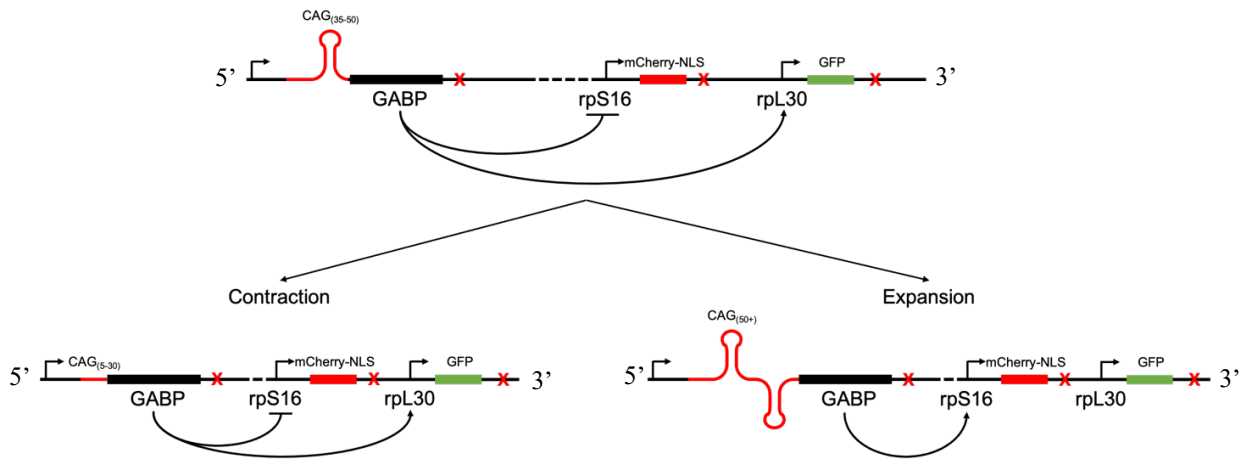


Figure 18 : modèle de cassette pouvant servir à l'étude des expansions dans des cellules humaines. Un nombre de répétition critique à la limite de l'instabilité (environ 35 répétitions) sont placées en amont de la séquence du gène GABP (*GA binding protein*). GABP est capable d'induire l'expression du promoteur ribosomique rpL30 tout en ayant un effet inhibiteur de l'expression du promoteur ribosomique rpS16 (Genuario et al, 1996). En présence d'un nombre de répétition à la limite de l'instabilité ou après contraction, un gène rapporteur (i.e mCherry) sous contrôle du promoteur rpS16 sera inhibé tandis qu'un gène rapporteur (i.e GFP) sous contrôle du rapporteur rpL30 sera transcrit.

(II) Interaction de la machinerie de maintenance télomérique avec la réparation d'une cassure double-brin au sein de répétitions CTG

Mon questionnement sur les origines des grandes expansions et le rôle que pouvait jouer la machinerie télomérique dans ce processus m'a permis de dégager un deuxième axe de réflexion et de recherche me demandant si les répétitions ne pouvaient pas être identifiées à tort comme des séquences télomériques par la télomérase et/ou la machinerie de maintenance télomérique (régulation de la longueur des télomères). Ces questions m'ont poussé à étudier le rôle de Rif1 et Rif2, qui associés à Rap1, forment un complexe télomérique qui régule la longueur des télomères en inhibant l'interaction du complexe MRX avec les extrémités télomériques^{310,316,317}.

Le complexe MRX est connu comme étant actif aux télomères et joue un rôle important dans l'activation de l'activité kinase de Tell1, critique dans la régulation de la longueur des télomères^{314 218-220,310,314,316,322}. L'activation de Tell1 par le complexe MRX aux télomères est

suffisante pour réguler la taille des extrémités télomériques. Le complexe télomérique Rap1-Rif1-Rif2 régule négativement le recrutement de Tel1 et du complexe MRX au niveau des télomères ainsi que des cassures intrachromosomiques^{310,317} montrant que ces complexes interagissent déjà dans la cellule. D'autres travaux montrent que Rif1 régule positivement la résection au niveau d'une cassure intrachromosomique, suggérant que l'interaction du complexe Rap1-Rif1-Rif2 avec le complexe MRX est plus complexe que les interactions observées aux télomères³¹⁸.

Mes travaux montrent que la perte d'expression des gènes *RIF1* et *RIF2* conduit à une augmentation de l'efficacité de la résection aux extrémités des cassures au sein de répétitions CTG, sans pour autant affecter la réparation. Des expériences similaires réalisées en l'absence des répétitions montrent que la perte d'expression de *RIF1* et *RIF2* diminue l'efficacité de réparation des cassures double-brin et la résection au sein d'une région non répétée et non structurée. Cela indique que le complexe télomérique formé de Rap1-Rif1-Rif2 joue un rôle différentiel en fonction de la nature de la séquence en cours de réparation. Je propose un modèle afin d'expliquer les différences observées en fonction du type de séquence dans laquelle une cassure double-brin est générée (figure 19). Il serait intéressant de déterminer l'épistasie des protéines du complexe MRX avec celle du complexe Rap1-Rif1-Rif2 et leur compétition sur la fixation des régions répétées. L'étude de la réparation d'une cassure au sein de différents types de répétitions permettrait de déterminer si cet effet est spécifique des répétitions CTG ou au contraire général à tous les types de répétitions. La comparaison de répétitions capable de former des structures secondaires et de répétitions connues comme ne formant pas de structures secondaires serait aussi un point d'étude à développer. Il serait finalement intéressant de comprendre les mécanismes impliqués dans la régulation positive ou négative de la résection du complexe Rap1-Rif1-Rif2 par interaction avec MRX et Tel1.

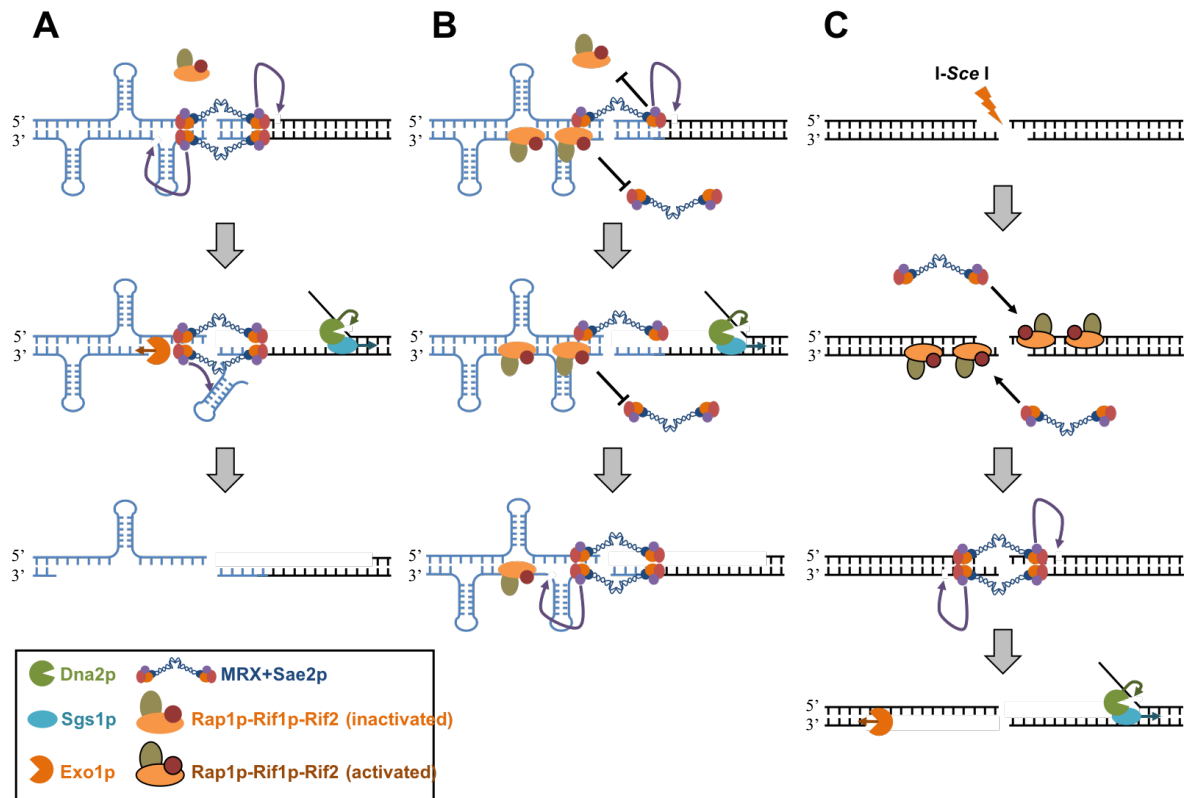


Figure 19 : Modèle de l'interaction entre le complexe télomérique Rif1-Rif2, MRX et une cassure double-brin localisé au sein de séquences répétées ou non répétées. (A) En absence de Rif1 ou Rif2, le complexe MRX se fixe sur les extrémités de la cassure et élimine les potentielles structures secondaires bloquant l'accès aux nucléases Exo1 et Dna2 responsables des étapes longues de la résection. (B) En présence du complexe Rif1-Rif2 au sein de séquence répétée, le complexe MRX entre en compétition pour la fixation aux extrémités de la cassure. Rif1-Rif2 inhibe la fixation du complexe MRX comme observé aux extrémités télomériques. (C) Au sein d'une séquence non répétée et non structurée, l'affinité du complexe Rif1-Rif2 pour les extrémités de la cassure est faible et le complexe MRX est capable d'initier les étapes initiales de la résection.

Il serait intéressant d'étudier l'interaction de la machinerie télomérique humaine avec la réparation des cassures double-brin au sein de différentes répétitions. Les télomères humains sont différents des télomères de levures, au niveau de leur séquence et de leur structure. Ces différences de structure font que Rif1 ne joue pas le même rôle chez la levure et chez *Homo sapiens*. De plus aucun orthologue de Rif2 n'existe chez l'humain ainsi deux questions peuvent se poser :

- (i) La machinerie télomérique humaine interagit-elle avec la réparation des cassures double-brin de l'ADN au sein de répétitions ?
- (ii) La télomérase joue-t-elle un rôle dans les expansions des répétitions ?

Afin de répondre à la première question des expériences de stabilité des répétitions en absence de gènes de maintenance télomérique ou au contraire en contexte de surexpression pourraient être autant d'indices de ces interactions. Un système d'induction d'une cassure unique au sein de répétitions doit être développé pour suivre le rôle exact de ces mutations sur la réparation et la stabilité des répétitions. De la même manière, la stabilité des répétitions pourrait être suivie dans une souche où le gène de la télomérase a été invalidé ou au contraire surexprimé. Le séquençage du locus portant les répétitions en cours de réparation, servirait à déterminer si la télomérase est capable de reconnaître les extrémités d'une cassure double-brin au sein de répétitions comme une extrémité télomérique et d'y ajouter des fragments de télomères.

REFERENCES

1. Peterson, D. G. *et al.* Integration of Cot Analysis, DNA Cloning, and High-Throughput Sequencing Facilitates Genome Characterization and Gene Discovery. *Genome Res.* **12**, 795–807 (2002).
2. Britten, R. J. & Kohne, D. E. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161**, 529–540 (1968).
3. Mirsky, A. E. & Ris, H. THE DESOXYRIBONUCLEIC ACID CONTENT OF ANIMAL CELLS AND ITS EVOLUTIONARY SIGNIFICANCE. *J. Gen. Physiol.* **34**, 451–462 (1951).
4. Thomas, C. A. The Genetic Organization of Chromosomes. *Annu. Rev. Genet.* **5**, 237–256 (1971).
5. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563–567 (1996).
6. C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
7. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
8. Palazzo, A. F. & Gregory, T. R. The Case for Junk DNA. *PLoS Genet.* **10**, e1004351 (2014).
9. Jurka, J., Kapitonov, V. V., Kohany, O. & Jurka, M. V. Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* **8**, 241–259 (2007).
10. Huda, A., Mariño-Ramírez, L., Landsman, D. & Jordan, I. K. Repetitive DNA elements, nucleosome binding and human gene expression. *Gene* **436**, 12–22 (2009).
11. Orr, H. T. & Zoghbi, H. Y. Trinucleotide Repeat Disorders. *Annu. Rev. Neurosci.* **30**, 575–621 (2007).
12. Paulson, H. Repeat expansion diseases. *Handb. Clin. Neurol.* **147**, 105–123 (2018).
13. Harper, P. *Myotonic Dystrophy*. (OUP Oxford, 2009).
14. Turner, C. & Hilton-Jones, D. The myotonic dystrophies: diagnosis and management. *J. Neurol. Neurosurg. Psychiatry* **81**, 358–367 (2010).
15. Mankodi, A. *et al.* Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.* **10**, 2165–2170 (2001).
16. Miller, J. W. *et al.* Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.* **19**, 4439–4448 (2000).
17. Kanadia, R. N. *et al.* Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3. *Gene Expr. Patterns GEP* **3**, 459–462 (2003).
18. Sutherland, G. R. & Richards, R. I. The molecular basis of fragile sites in human chromosomes. *Curr. Opin. Genet. Dev.* **5**, 323–327 (1995).
19. Popescu, N. C. Genetic alterations in cancer as a result of breakage at fragile sites. *Cancer Lett.* **192**, 1–17 (2003).
20. Richards, R. I. Fragile and unstable chromosomes in cancer: causes and consequences. *Trends Genet. TIG* **17**, 339–345 (2001).
21. Liu, B. *et al.* Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nat. Genet.* **9**, 48–55 (1995).
22. Malpertuy, A., Dujon, B. & Richard, G.-F. Analysis of Microsatellites in 13 Hemiascomycetous Yeast Species: Mechanisms Involved in Genome Dynamics. *J. Mol. Evol.* **56**, 730–741 (2003).
23. Tóth, G., Gáspári, Z. & Jurka, J. Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Res.* **10**, 967–981 (2000).
24. Subramanian, S., Mishra, R. K. & Singh, L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol.* **4**,

R13 (2003).

25. Wren, J. D. *et al.* Repeat Polymorphisms within Gene Regions: Phenotypic and Evolutionary Implications. *Am. J. Hum. Genet.* **67**, 345–356 (2000).
26. Richard, G. F. & Dujon, B. Trinucleotide repeats in yeast. *Res. Microbiol.* **148**, 731–744 (1997).
27. Poggi, L. & Richard, G.-F. Alternative DNA Structures In Vivo: Molecular Evidence and Remaining Questions. *Microbiol. Mol. Biol. Rev. MMBR* **85**, (2021).
28. Lynch, M. *et al.* A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9272–9277 (2008).
29. Richard, G.-F. Shortening trinucleotide repeats using highly specific endonucleases: a possible approach to gene therapy? *Trends Genet. TIG* **31**, 177–186 (2015).
30. Baran, N., Lapidot, A. & Manor, H. Formation of DNA triplexes accounts for arrests of DNA synthesis at d(TC). and d(GA)_n tracts. **5**.
31. Biet, E., Sun, J.-S. & Dutreix, M. Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure. *Nucleic Acids Res.* **27**, 596–600 (1999).
32. Mitas, M. Trinucleotide repeats associated with human disease. *Nucleic Acids Res.* **25**, 2245–2254 (1997).
33. Mitas, M., Yu, A., Dill, J. & Haworth, I. S. The trinucleotide repeat sequence d(CGG)₁₅ forms a heat-stable hairpin containing Gsyn. Ganti base pairs. *Biochemistry* **34**, 12803–12811 (1995).
34. Chi, L. M. & Lam, S. L. Structural roles of CTG repeats in slippage expansion during DNA replication. *Nucleic Acids Res.* **33**, 1604–1617 (2005).
35. Bidichandani, S. I., Ashizawa, T. & Patel, P. I. The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am. J. Hum. Genet.* **62**, 111–121 (1998).
36. Bhattacharyya, D., Mirihana Arachchilage, G. & Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **4**, (2016).
37. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **21**, 459–474 (2020).
38. Chen, Y.-H. *et al.* Transcription shapes DNA replication initiation and termination in human cells. *Nat. Struct. Mol. Biol.* **26**, 67–77 (2019).
39. Ribeyre, C. *et al.* The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo. *PLoS Genet.* **5**, (2009).
40. Levinson, G. & Gutman, G. A. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. *Nucleic Acids Res.* **15**, 5323–5338 (1987).
41. Wells, R. D. & Blair, J. E. Studies on polynucleotides: LXXI. Sedimentation and buoyant density studies of some DNA-like polymers with repeating nucleotide sequences. *J. Mol. Biol.* **27**, 273–288 (1967).
42. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276 (1993).
43. Kroutil, L. C., Register, K., Bebenek, K. & Kunkel, T. A. Exonucleolytic Proofreading during Replication of Repetitive DNA. *Biochemistry* **35**, 1046–1053 (1996).
44. Chung, M. *et al.* Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat. Genet.* **5**, 254–258 (1993).
45. Eichler, E. E. *et al.* Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat. Genet.* **8**, 88–94 (1994).

46. Pearson, C. E. *et al.* Interruptions in the Triplet Repeats of SCA1 and FRAXA Reduce the Propensity and Complexity of Slipped Strand DNA (S-DNA) Formation. *Biochemistry* **37**, 2701–2708 (1998).
47. Rolfsmeier, M. L. & Lahue, R. S. Stabilizing Effects of Interruptions on Trinucleotide Repeat Expansions in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **20**, 173–180 (2000).
48. Petes, T. D., Greenwell, P. W. & Dominska, M. Stabilization of Microsatellite Sequences by Variant Repeats in the Yeast *Saccharomyces cerevisiae*. *Genetics* **146**, 491–498 (1997).
49. Temmerman, N. D. *et al.* Intergenerational Instability of the Expanded CTG Repeat in the DMPK Gene: Studies in Human Gametes and Preimplantation Embryos. *Am. J. Hum. Genet.* **75**, 325–329 (2004).
50. Burgers, P. M. J. Eukaryotic DNA polymerases in DNA replication and DNA repair. *Chromosoma* **107**, 218–227 (1998).
51. Lehman, I. R. & Kaguni, L. S. DNA polymerase alpha. *J. Biol. Chem.* **264**, 4265–4268 (1989).
52. Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R. & Mirkin, S. M. Replication and Expansion of Trinucleotide Repeats in Yeast. *Mol. Cell. Biol.* **23**, 1349 (2003).
53. Schweitzer, J. K. & Livingston, D. M. Destabilization of CAG Trinucleotide Repeat Tracts by Mismatch Repair Mutations in Yeast. *Hum. Mol. Genet.* **6**, 349–355 (1997).
54. Freudenreich, C. H., Kantrow, S. M. & Zakian, V. A. Expansion and Length-Dependent Fragility of CTG Repeats in Yeast. *Science* **279**, 853–856 (1998).
55. Yang, J. & Freudenreich, C. H. Haploinsufficiency of Yeast FEN1 Causes Instability of Expanded CAG/CTG Tracts in a Length-Dependent Manner. *Gene* **393**, 110–115 (2007).
56. Henriksen, L. A., Tom, S., Liu, Y. & Bambara, R. A. Inhibition of Flap Endonuclease 1 by Flap Secondary Structure and Relevance to Repeat Sequence Expansion. *J. Biol. Chem.* **275**, 16420–16427 (2000).
57. Symington, L. Homologous recombination is required for the viability of rad27 mutants. *Nucleic Acids Res.* **26**, 5589–5595 (1998).
58. Subramanian, J., Vijayakumar, S., Tomkinson, A. E. & Arnheim, N. Genetic Instability Induced by Overexpression of DNA Ligase I in Budding Yeast. *Genetics* **171**, 427–441 (2005).
59. Yang, Z., Lau, R., Marcadier, J. L., Chitayat, D. & Pearson, C. E. Replication Inhibitors Modulate Instability of an Expanded Trinucleotide Repeat at the Myotonic Dystrophy Type 1 Disease Locus in Human Cells. *Am. J. Hum. Genet.* **73**, 1092–1105 (2003).
60. Kang, S., Jaworski, A., Ohshima, K. & Wells, R. D. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nat. Genet.* **10**, 213–218 (1995).
61. Freudenreich, C. H., Stavenhagen, J. B. & Zakian, V. A. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell. Biol.* **17**, 2090–2098 (1997).
62. Miret, J. J., Pessoa-Brandão, L. & Lahue, R. S. Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **95**, 12438–12443 (1998).
63. Russo, M. T. *et al.* The Oxidized Deoxynucleoside Triphosphate Pool Is a Significant Contributor to Genetic Instability in Mismatch Repair-Deficient Cells. *Mol. Cell. Biol.* **24**, 465–474 (2004).
64. Peltomaki, P. & Vasen, H. Mutations predisposing to hereditary nonpolyposis colorectal cancer: Database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. *Gastroenterology* **113**,

1146–1158 (1997).

65. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18**, 85–98 (2008).
66. Chi, N. W. & Kolodner, R. D. Purification and characterization of MSH1, a yeast mitochondrial protein that binds to DNA mismatches. *J. Biol. Chem.* **269**, 29984–29992 (1994).
67. Hunter, N. & Borts, R. H. Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis. *Genes Dev.* **11**, 1573–1582 (1997).
68. Johnson, R. E., Kovvali, G. K., Prakash, L. & Prakash, S. Requirement of the Yeast MSH3 and MSH6 Genes for MSH2-dependent Genomic Stability. *J. Biol. Chem.* **271**, 7285–7288 (1996).
69. Putnam, C. D., Hayes, T. K. & Kolodner, R. D. Specific pathways prevent duplication-mediated genome rearrangements. *Nature* **460**, 984–989 (2009).
70. Putnam, C. D. & Kolodner, R. D. Pathways and Mechanisms that Prevent Genome Instability in *Saccharomyces cerevisiae*. *Genetics* **206**, 1187–1225 (2017).
71. Genschel, J. & Modrich, P. Mechanism of 5'-Directed Excision in Human Mismatch Repair. *Mol. Cell* **12**, 1077–1086 (2003).
72. Kadyrov, F. A., Dzantiev, L., Constantin, N. & Modrich, P. Endonucleolytic Function of MutL α in Human Mismatch Repair. *Cell* **126**, 297–308 (2006).
73. Kadyrova, L. Y. & Kadyrov, F. A. Endonuclease activities of MutL α and its homologs in DNA mismatch repair. *DNA Repair* **38**, 42–49 (2016).
74. Tham, K.-C., Kanaar, R. & Lebbink, J. H. G. Mismatch repair and homeologous recombination. *DNA Repair* **38**, 75–83 (2016).
75. Schmidt, K. H., Abbott, C. M. & Leach, D. R. F. Two opposing effects of mismatch repair on CTG repeat instability in *Escherichia coli*. *Mol. Microbiol.* **35**, 463–471 (2000).
76. Parniewski, P., Jaworski, A., Wells, R. D. & Bowater, R. P. Length of CTG·CAG repeats determines the influence of mismatch repair on genetic instability | Edited by M. Yaniv. *J. Mol. Biol.* **299**, 865–874 (2000).
77. Viterbo, D., Michoud, G., Mosbach, V., Dujon, B. & Richard, G.-F. Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. *DNA Repair* **42**, 94–106 (2016).
78. Kovtun, I. V., Thornhill, A. R. & McMurray, C. T. Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations. *Hum. Mol. Genet.* **13**, 3057–3068 (2004).
79. Manley, K., Shirley, T. L., Flaherty, L. & Messer, A. Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.* **23**, 471–473 (1999).
80. Savouret, C. *et al.* CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J.* **22**, 2264–2273 (2003).
81. Tomé, S. *et al.* MSH2 ATPase Domain Mutation Affects CTG·CAG Repeat Instability in Transgenic Mice. *PLoS Genet.* **5**, (2009).
82. Lia, A.-S. *et al.* Somatic Instability of the CTG Repeat in Mice Transgenic for the Myotonic Dystrophy Region is Age Dependent But Not Correlated to the Relative Intertissue Transcription Levels and Proliferative Capacities. *Hum. Mol. Genet.* **7**, 1285–1291 (1998).
83. Boland, C. R. & Goel, A. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* **138**, 2073–2087.e3 (2010).
84. Compe, E. & Egly, J.-M. TFIIH: when transcription met DNA repair. *Nat. Rev. Mol. Cell Biol.* **13**, 343–354 (2012).
85. Cotterill, S. & Kearsley, S. Eukaryotic DNA Polymerases. in *eLS* (American Cancer Society, 2002). doi:10.1038/npg.els.0001045.

86. Doublíé, S. & Zahn, K. E. Structural insights into eukaryotic DNA replication. *Front. Microbiol.* **5**, (2014).
87. Fukui, K. DNA Mismatch Repair in Eukaryotes and Bacteria. *J. Nucleic Acids* **2010**, (2010).
88. Mattarocci, S., Hafner, L., Lezaja, A., Shyian, M. & Shore, D. Rif1: A Conserved Regulator of DNA Replication and Repair Hijacked by Telomeres in Yeasts. *Front. Genet.* **7**, (2016).
89. Prakash, S. & Prakash, L. Nucleotide excision repair in yeast. *Mutat. Res. Mol. Mech. Mutagen.* **451**, 13–24 (2000).
90. Sfeir, A. & Symington, L. S. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem. Sci.* **40**, 701–714 (2015).
91. Shuck, S. C., Short, E. A. & Turchi, J. J. Eukaryotic nucleotide excision repair, from understanding mechanisms to influencing biology. *Cell Res.* **18**, 64–72 (2008).
92. Huang, J. C., Svoboda, D. L., Reardon, J. T. & Sancar, A. Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5' and the 6th phosphodiester bond 3' to the photodimer. *Proc. Natl. Acad. Sci.* **89**, 3664–3668 (1992).
93. Hoogstraten, D. *et al.* Versatile DNA damage detection by the global genome nucleotide excision repair protein XPC. *J. Cell Sci.* **121**, 2850–2859 (2008).
94. Evans, E. Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J.* **16**, 6559–6573 (1997).
95. Riedl, T., Hanaoka, F. & Egly, J.-M. The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J.* **22**, 5293–5303 (2003).
96. Moser, J. *et al.* The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions. *DNA Repair* **4**, 571–582 (2005).
97. Adar, S., Hu, J., Lieb, J. D. & Sancar, A. Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2124–E2133 (2016).
98. Sarker, A. H. *et al.* Recognition of RNA Polymerase II and Transcription Bubbles by XPG, CSB, and TFIIH: Insights for Transcription-Coupled Repair and Cockayne Syndrome. *Mol. Cell* **20**, 187–198 (2005).
99. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* **9**, 958–970 (2008).
100. Miller, R. D., Prakash, L. & Prakash, S. Genetic control of excision of *Saccharomyces cerevisiae* interstrand DNA cross-links induced by psoralen plus near-UV light. *Mol. Cell Biol.* **2**, 939–948 (1982).
101. Dragileva, E. *et al.* Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol. Dis.* **33**, 37–47 (2009).
102. Lin, Y., Dion, V. & Wilson, J. H. Transcription promotes contraction of CAG repeat tracts in human cells. *Nat. Struct. Mol. Biol.* **13**, 179–180 (2006).
103. Zhao, X.-N. & Usdin, K. Gender and cell-type specific effects of the transcription coupled repair protein, ERCC6/CSB, on repeat expansion in a mouse model of the Fragile X-related disorders. *Hum. Mutat.* **35**, 341–349 (2014).
104. Maynard, S., Fang, E. F., Scheibye-Knudsen, M., Croteau, D. L. & Bohr, V. A. DNA Damage, DNA Repair, Aging, and Neurodegeneration. *Cold Spring Harb. Perspect. Med.* **5**, (2015).
105. McMurray, C. T. Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* **11**, 786–799 (2010).

106. Staresinic, L. *et al.* Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J.* **28**, 1111–1120 (2009).
107. Parsons, M. A., Sinden, R. R. & Izban, M. G. Transcriptional properties of RNA polymerase II within triplet repeat-containing DNA from the human myotonic dystrophy and fragile X loci. *J. Biol. Chem.* **273**, 26998–27008 (1998).
108. Cadet, J. & Wagner, J. R. DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).
109. Oda, Y. *et al.* NMR studies of a DNA containing 8-hydroxydeoxyguanosine. *Nucleic Acids Res.* **19**, 1407–1412 (1991).
110. Kubota, Y. *et al.* Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. *EMBO J.* **15**, 6662–6670 (1996).
111. Wallace, S. S. DNA Glycosylases Search for and Remove Oxidized DNA Bases. *Environ. Mol. Mutagen.* **54**, 691–704 (2013).
112. Dianov, G., Price, A. & Lindahl, T. Generation of single-nucleotide repair patches following excision of uracil residues from DNA. *Mol. Cell. Biol.* **12**, 1605–1612 (1992).
113. Klungland, A. Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1). *EMBO J.* **16**, 3341–3348 (1997).
114. Prasad, R., Dianov, G. L., Bohr, V. A. & Wilson, S. H. FEN1 stimulation of DNA polymerase beta mediates an excision step in mammalian long patch base excision repair. *J. Biol. Chem.* **275**, 4460–4466 (2000).
115. Wong, H.-K. *et al.* Cockayne syndrome B protein stimulates apurinic endonuclease 1 activity and protects against agents that introduce base excision repair intermediates. *Nucleic Acids Res.* **35**, 4103–4113 (2007).
116. Kovtun, I. V. & McMurray, C. T. Crosstalk of DNA glycosylases with pathways other than base excision repair. *DNA Repair* **6**, 517–529 (2007).
117. Sliwinska, A. *et al.* The levels of 7,8-dihydrodeoxyguanosine (8-oxoG) and 8-oxoguanine DNA glycosylase 1 (OGG1) – A potential diagnostic biomarkers of Alzheimer's disease. *J. Neurol. Sci.* **368**, 155–159 (2016).
118. Dizdaroglu, M. Substrate specificities and excision kinetics of DNA glycosylases involved in base-excision repair of oxidative DNA damage. *Mutat. Res. Mol. Mech. Mutagen.* **531**, 109–126 (2003).
119. Kovtun, I. V. *et al.* OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature* **447**, 447–452 (2007).
120. Hartenstine, M. J., Goodman, M. F. & Petruska, J. Weak strand displacement activity enables human DNA polymerase beta to expand CAG/CTG triplet repeats at strand breaks. *J. Biol. Chem.* **277**, 41379–41389 (2002).
121. Liu, Y. *et al.* Coordination between Polymerase β and FEN1 Can Modulate CAG Repeat Expansion. *J. Biol. Chem.* **284**, 28352–28366 (2009).
122. Chan, N. L. S. *et al.* Coordinated Processing of 3' Slipped (CAG)_n/(CTG)_n Hairpins by DNA Polymerases β and δ Preferentially Induces Repeat Expansions. *J. Biol. Chem.* **288**, 15015–15022 (2013).
123. Kerrest, A. *et al.* SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. *Nat. Struct. Mol. Biol.* **16**, 159–167 (2009).
124. Sutcliffe, J. S. *et al.* DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum. Mol. Genet.* **1**, 397–400 (1992).
125. Feng, Y. *et al.* Translational suppression by trinucleotide repeat expansion at FMR1. *Science* **268**, 731–734 (1995).
126. Ruiz Buendía, G. A. *et al.* Three-dimensional chromatin interactions remain stable

- upon CAG/CTG repeat expansion. *Sci. Adv.* **6**, (2020).
127. Greene, E., Mahishi, L., Entezam, A., Kumari, D. & Usdin, K. Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. *Nucleic Acids Res.* **35**, 3383–3390 (2007).
 128. Steinbach, P., Gläser, D., Vogel, W., Wolf, M. & Schwemmle, S. The DMPK Gene of Severely Affected Myotonic Dystrophy Patients Is Hypermethylated Proximal to the Largely Expanded CTG Repeat. *Am. J. Hum. Genet.* **62**, 278–285 (1998).
 129. Chan, P. K. *et al.* Heterochromatinization induced by GAA-repeat hyperexpansion in Friedreich's ataxia can be reduced upon HDAC inhibition by vitamin B3. *Hum. Mol. Genet.* **22**, 2662–2675 (2013).
 130. House, N. C. M., Yang, J. H., Walsh, S. C., Moy, J. M. & Freudenreich, C. H. NuA4 Initiates Dynamic Histone H4 Acetylation to Promote High-Fidelity Sister Chromatid Recombination at Postreplication Gaps. *Mol. Cell* **55**, 818–828 (2014).
 131. Dion, V., Lin, Y., Hubert, L., Waterland, R. A. & Wilson, J. H. Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Hum. Mol. Genet.* **17**, 1306–1317 (2008).
 132. Laird, C., Jaffe, E., Karpen, G., Lamb, M. & Nelson, R. Fragile sites in human chromosomes as regions of late-replicating DNA. *Trends Genet.* **3**, 274–281 (1987).
 133. Le Beau, M. M. *et al.* Replication of a Common Fragile Site, FRA3B, Occurs Late in S Phase and is Delayed Further Upon Induction: Implications for the Mechanism of Fragile Site Induction. *Hum. Mol. Genet.* **7**, 755–761 (1998).
 134. Letessier, A. *et al.* Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* **470**, 120–123 (2011).
 135. Callahan, J. L., Andrews, K. J., Zakian, V. A. & Freudenreich, C. H. Mutations in Yeast Replication Proteins That Increase CAG/CTG Expansions Also Increase Repeat Fragility. *Mol. Cell. Biol.* **23**, 7849–7860 (2003).
 136. Rothstein, R. J. One-step gene disruption in yeast. *Methods Enzymol.* **101**, 202–211 (1983).
 137. Colleaux, L., D'Auriol, L., Galibert, F. & Dujon, B. Recognition and cleavage site of the intron-encoded omega transposase. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6022–6026 (1988).
 138. Cooper, A. A. & Stevens, T. M. H. Protein splicing: Excision of intervening sequences at the protein level. *BioEssays* **15**, 667–674 (1993).
 139. Gogarten, J. P. & Hilario, E. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol. Biol.* **6**, 94 (2006).
 140. Dujon, B. Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the ω and rib-1 loci. *Cell* **20**, 185–197 (1980).
 141. Jacquier, A. & Dujon, B. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41**, 383–394 (1985).
 142. Argast, G. M., Stephens, K. M., Emond, M. J. & Monnat, R. J. I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. Edited by G. Smith. *J. Mol. Biol.* **280**, 345–353 (1998).
 143. Arnould, S. *et al.* Engineering of Large Numbers of Highly Specific Homing Endonucleases that Induce Recombination on Novel DNA Targets. *J. Mol. Biol.* **355**, 443–458 (2006).
 144. Daboussi, F. *et al.* Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. *Nucleic Acids Res.* **40**, 6367–6379 (2012).
 145. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci.* **93**, 1156–1160 (1996).
 146. Ginsberg, A. M., King, B. O. & Roeder, R. G. Xenopus 5S gene transcription factor,

- TFIIIA: Characterization of a cDNA clone and measurement of RNA levels throughout development. *Cell* **39**, 479–489 (1984).
147. Chandrasegaran, S. & Carroll, D. Origins of Programmable Nucleases for Genome Engineering. *J. Mol. Biol.* **428**, 963–989 (2016).
 148. Bitinaite, J., Wah, D. A., Aggarwal, A. K. & Schildkraut, I. FokI dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 10570–10575 (1998).
 149. Vanamee, É. S., Santagata, S. & Aggarwal, A. K. FokI requires two specific DNA sites for cleavage. Edited by T. Richmond. *J. Mol. Biol.* **309**, 69–78 (2001).
 150. Porteus, M. H. & Baltimore, D. Chimeric Nucleases Stimulate Gene Targeting in Human Cells. *Science* **300**, 763–763 (2003).
 151. Cornu, T. I. *et al.* DNA-binding Specificity Is a Major Determinant of the Activity and Toxicity of Zinc-finger Nucleases. *Mol. Ther.* **16**, 352–358 (2008).
 152. Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–785 (2007).
 153. Mittelman, D. *et al.* Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9607–9612 (2009).
 154. Liu, G., Chen, X., Bissler, J. J., Sinden, R. R. & Leffak, M. Replication dependent instability at (CTG)_n(CAG) repeat hairpins in human cells. *Nat. Chem. Biol.* **6**, 652–659 (2010).
 155. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509–1512 (2009).
 156. Schornack, S., Meyer, A., Römer, P., Jordan, T. & Lahaye, T. Gene-for-gene-mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. *J. Plant Physiol.* **163**, 256–272 (2006).
 157. Christian, M. *et al.* Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **186**, 757–761 (2010).
 158. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).
 159. Jansen, R., Embden, J. D. A. van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
 160. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
 161. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**, 1709–1712 (2007).
 162. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **321**, 960–964 (2008).
 163. Makarova, K. S. & Koonin, E. V. Annotation and Classification of CRISPR-Cas Systems. *Methods Mol. Biol. Clifton NJ* **1311**, 47–75 (2015).
 164. S. Makarova, K. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477 (2011).
 165. Makarova, K. S. *et al.* Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
 166. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
 167. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host

- factor RNase III. *Nature* **471**, 602–607 (2011).
168. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
169. Vilenchik, M. M. & Knudson, A. G. Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12871–12876 (2003).
170. Teoh, N. C. *et al.* Defective DNA strand break repair causes chromosomal instability and accelerates liver carcinogenesis in mice. *Hepatology* **47**, 2078–2088 (2008).
171. Yan, C. T. *et al.* IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449**, 478–482 (2007).
172. Boboila, C. *et al.* Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. *Proc. Natl. Acad. Sci.* **107**, 3034–3039 (2010).
173. Bétermier, M., Bertrand, P. & Lopez, B. S. Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? *PLoS Genet.* **10**, (2014).
174. Yan, C. T. *et al.* IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449**, 478–482 (2007).
175. Caron, M.-C. *et al.* Poly(ADP-ribose) polymerase-1 antagonizes DNA resection at double-strand breaks. *Nat. Commun.* **10**, 2954 (2019).
176. Pascal, J. M. The comings and goings of PARP-1 in response to DNA damage. *DNA Repair* **71**, 177–182 (2018).
177. Symington, L. S. & Gautier, J. Double-Strand Break End Resection and Repair Pathway Choice. *Annu. Rev. Genet.* **45**, 247–271 (2011).
178. LIANGt, F., ROMANIENKOt, P. J., WEAVERt, D. T., Jeggo, P. A. & Jasin, M. Chromosomal double-strand break repair in Ku80-deficient cells. *Proc Natl Acad Sci USA* **5** (1996).
179. Boulton, S. J. & Jackson, S. P. Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *EMBO J.* **17**, 1819–1828 (1998).
180. Iwasaki, D. *et al.* The MRX Complex Ensures NHEJ Fidelity through Multiple Pathways Including Xrs2-FHA–Dependent Tell1 Activation. *PLoS Genet.* **12**, (2016).
181. Jin, S., Kharbanda, S., Mayer, B., Kufe, D. & Weaver, D. T. Binding of Ku and c-Abl at the Kinase Homology Region of DNA-dependent Protein Kinase Catalytic Subunit. *J. Biol. Chem.* **272**, 24763–24766 (1997).
182. Spagnolo, L., Rivera-Calzada, A., Pearl, L. H. & Llorca, O. Three-Dimensional Structure of the Human DNA-PKcs/Ku70/Ku80 Complex Assembled on DNA and Its Implications for DNA DSB Repair. *Mol. Cell* **22**, 511–519 (2006).
183. Meek, K., Dang, V. & Lees-Miller, S. P. Chapter 2 DNA-PK: The Means to Justify the Ends? in *Advances in Immunology* vol. 99 33–58 (Academic Press, 2008).
184. Hammarsten, O., DeFazio, L. G. & Chu, G. Activation of DNA-dependent Protein Kinase by Single-stranded DNA Ends. *J. Biol. Chem.* **275**, 1541–1550 (2000).
185. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin Opening and Overhang Processing by an Artemis/DNA-Dependent Protein Kinase Complex in Nonhomologous End Joining and V(D)J Recombination. *Cell* **108**, 781–794 (2002).
186. Ma, Y., Schwarz, K. & Lieber, M. R. The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. *DNA Repair* **4**, 845–851 (2005).
187. Herrmann, G., Lindahl, T. & Schär, P. *Saccharomyces cerevisiae* LIF1: a function involved in DNA double-strand break repair related to mammalian XRCC4. *EMBO J.* **17**, 4188–4198 (1998).
188. Ma, Y. *et al.* A Biochemically Defined System for Mammalian Nonhomologous DNA

- End Joining. *Mol. Cell* **16**, 701–713 (2004).
189. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).
 190. Quennet, V., Beucher, A., Barton, O., Takeda, S. & Löbrich, M. CtIP and MRN promote non-homologous end-joining of etoposide-induced DNA double-strand breaks in G1. *Nucleic Acids Res.* **39**, 2144–2152 (2011).
 191. Jackson, S. P. DNA damage detection by DNA dependent protein kinase and related enzymes. *Cancer Surv.* **28**, 261–279 (1996).
 192. Welcker, A. J., de Montigny, J., Potier, S. & Souciet, J.-L. Involvement of Very Short DNA Tandem Repeats and the Influence of the RAD52 Gene on the Occurrence of Deletions in *Saccharomyces cerevisiae*. 9.
 193. Lee, K. & Lee, S. E. *Saccharomyces cerevisiae* Sae2- and Tel1-Dependent Single-Strand DNA Formation at DNA Break Promotes Microhomology-Mediated End Joining. *Genetics* **176**, 2003–2014 (2007).
 194. Holmes, V. F., Benjamin, K. R., Crisona, N. J. & Cozzarelli, N. R. Bypass of heterology during strand transfer by *Saccharomyces cerevisiae* Rad51 protein. *Nucleic Acids Res.* **29**, 5052–5057 (2001).
 195. Pâques, F. & Haber, J. E. Multiple Pathways of Recombination Induced by Double-Strand Breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).
 196. Guirouilh-Barbat, J., Lambert, S., Bertrand, P. & Lopez, B. S. Is homologous recombination really an error-free process? *Front. Genet.* **5**, (2014).
 197. Sung, P., Krejci, L., Van Komen, S. & Sehorn, M. G. Rad51 recombinase and recombination mediators. *J. Biol. Chem.* **278**, 42729–42732 (2003).
 198. Paull, T. T. & Gellert, M. The 3' to 5' Exonuclease Activity of Mre11 Facilitates Repair of DNA Double-Strand Breaks. *Mol. Cell* **1**, 969–979 (1998).
 199. Connelly, J. C. & Leach, D. R. F. The *sbcC* and *sbcD* genes of *Escherichia coli* encode a nuclease involved in palindrome inviability and genetic recombination. *Genes Cells* **1**, 285–291 (1996).
 200. Connelly, J. C. & Leach, D. R. F. Tethering on the brink: the evolutionarily conserved Mre11–Rad50 complex. *Trends Biochem. Sci.* **27**, 410–418 (2002).
 201. Debrauwère, H. *et al.* Meiotic instability of human minisatellite CEB1 in yeast requires DNA double-strand breaks. *Nat. Genet.* **23**, 367–371 (1999).
 202. Langerak, P., Mejia-Ramirez, E., Limbo, O. & Russell, P. Release of Ku and MRN from DNA Ends by Mre11 Nuclease Activity and Ctp1 Is Required for Homologous Recombination Repair of Double-Strand Breaks. *PLoS Genet.* **7**, e1002271 (2011).
 203. Shim, E. Y. *et al.* *Saccharomyces cerevisiae* Mre11/Rad50/Xrs2 and Ku proteins regulate association of Exo1 and Dna2 with DNA breaks. *EMBO J.* **29**, 3370–3380 (2010).
 204. Hartsuiker, E., Neale, M. J. & Carr, A. M. Distinct Requirements for the Rad32Mre11 Nuclease and Ctp1CtIP in the Removal of Covalently Bound Topoisomerase I and II from DNA. *Mol. Cell* **33**, 117–123 (2009).
 205. Lobachev, K. S., Gordenin, D. A. & Resnick, M. A. The Mre11 Complex Is Required for Repair of Hairpin-Capped Double-Strand Breaks and Prevention of Chromosome Rearrangements. *Cell* **108**, 183–193 (2002).
 206. Cannavo, E. & Cejka, P. Sae2 promotes dsDNA endonuclease activity within Mre11–Rad50–Xrs2 to resect DNA breaks. *Nature* **514**, 122–125 (2014).
 207. Gobbin, E., Cassani, C., Villa, M., Bonetti, D. & Longhese, M. P. Functions and regulation of the MRX complex at DNA double-strand breaks. *Microb. Cell* **3**, 329–337 (2016).
 208. Cassani, C. *et al.* The ATP-bound conformation of the Mre11–Rad50 complex is

- essential for Tel1/ATM activation. *Nucleic Acids Res.* **47**, 3550–3567 (2019).
209. Garcia, V., Phelps, S. E. L., Gray, S. & Neale, M. J. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature* **479**, 241–244 (2011).
210. Furuse, M. *et al.* Distinct roles of two separable *in vitro* activities of yeast Mre11 in mitotic and meiotic recombination. *EMBO J.* **17**, 6412–6425 (1998).
211. Krogh, B. O., Llorente, B., Lam, A. & Symington, L. S. Mutations in Mre11 Phosphoesterase Motif I That Impair *Saccharomyces cerevisiae* Mre11-Rad50-Xrs2 Complex Stability in Addition to Nuclease Activity. *Genetics* **171**, 1561–1570 (2005).
212. Wiltzius, J. J. W., Hohl, M., Fleming, J. C. & Petrini, J. H. J. The Rad50 hook domain is a critical determinant of Mre11 complex functions. *Nat. Struct. Mol. Biol.* **12**, 403–407 (2005).
213. Chen, L. *et al.* Effect of Amino Acid Substitutions in the Rad50 ATP Binding Domain on DNA Double Strand Break Repair in Yeast*. *J. Biol. Chem.* **280**, 2620–2627 (2005).
214. Symington, L. S. End Resection at Double-Strand Breaks: Mechanism and Regulation. *Cold Spring Harb. Perspect. Biol.* **6**, (2014).
215. Mosbach, V., Poggi, L., Viterbo, D., Charpentier, M. & Richard, G.-F. TALEN-Induced Double-Strand Break Repair of CTG Trinucleotide Repeats. *Cell Rep.* **22**, 2146–2159 (2018).
216. Carney, J. P. *et al.* The hMre11/hRad50 Protein Complex and Nijmegen Breakage Syndrome: Linkage of Double-Strand Break Repair to the Cellular DNA Damage Response. *Cell* **93**, 477–486 (1998).
217. Tsukamoto, Y., Mitsuoka, C., Terasawa, M., Ogawa, H. & Ogawa, T. Xrs2p Regulates Mre11p Translocation to the Nucleus and Plays a Role in Telomere Elongation and Meiotic Recombination. *Mol. Biol. Cell* **16**, 597–608 (2005).
218. Nakada, D., Matsumoto, K. & Sugimoto, K. ATM-related Tel1 associates with double-strand breaks through an Xrs2-dependent mechanism. *Genes Dev.* **17**, 1957–1962 (2003).
219. Fukunaga, K., Kwon, Y., Sung, P. & Sugimoto, K. Activation of Protein Kinase Tel1 through Recognition of Protein-Bound DNA Ends. *Mol. Cell. Biol.* **31**, 1959–1971 (2011).
220. Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
221. Ritchie, K. B. & Petes, T. D. The Mre11p/Rad50p/Xrs2p Complex and the Tel1p Function in a Single Pathway for Telomere Maintenance in Yeast. **5**.
222. Keener, R., Connelly, C. J. & Greider, C. W. Tel1 Activation by the MRX Complex Is Sufficient for Telomere Length Regulation but Not for the DNA Damage Response in *Saccharomyces cerevisiae*. *Genetics* **213**, 1271–1288 (2019).
223. Lengsfeld, B. M., Rattray, A. J., Bhaskara, V., Ghirlando, R. & Paull, T. T. Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex. *Mol. Cell* **28**, 638–651 (2007).
224. Prinz, S., Amon, A. & Klein, F. Isolation of COM1, a new gene required to complete meiotic double-strand break-induced recombination in *Saccharomyces cerevisiae*. *Genetics* **146**, 781–795 (1997).
225. Baroni, E., Viscardi, V., Cartagena-Lirola, H., Lucchini, G. & Longhese, M. P. The Functions of Budding Yeast Sae2 in the DNA Damage Response Require Mec1- and Tel1-Dependent Phosphorylation. *Mol. Cell. Biol.* **24**, 4151–4165 (2004).
226. Huertas, P., Cortés-Ledesma, F., Sartori, A. A., Aguilera, A. & Jackson, S. P. CDK targets Sae2 to control DNA-end resection and homologous recombination. *Nature* **455**, 689–692 (2008).
227. Li, S. *et al.* Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response. *Nature* **406**, 210–215 (2000).

228. Kim, H.-M. *et al.* Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J.* **27**, 2896–2906 (2008).
229. Wang, H. *et al.* CtIP Protein Dimerization Is Critical for Its Recruitment to Chromosomal DNA Double-stranded Breaks. *J. Biol. Chem.* **287**, 21471–21480 (2012).
230. Lieber, M. R. The FEN-1 family of structure-specific nucleases in eukaryotic dna replication, recombination and repair. *BioEssays* **19**, 233–240 (1997).
231. Tran, P. T., Erdeniz, N., Dudley, S. & Liskay, R. M. Characterization of nuclease-dependent functions of Exo1p in *Saccharomyces cerevisiae*. *DNA Repair* **1**, 895–912 (2002).
232. Tran, P. T., Erdeniz, N., Symington, L. S. & Liskay, R. M. EXO1-A multi-tasking eukaryotic nuclease. *DNA Repair* **3**, 1549–1559 (2004).
233. Tishkoff, D. X. *et al.* Identification and characterization of *Saccharomyces cerevisiae* EXO1, a gene encoding an exonuclease that interacts with MSH2. *Proc. Natl. Acad. Sci.* **94**, 7487–7492 (1997).
234. Kirkpatrick, D. T., Ferguson, J. R., Petes, T. D. & Symington, L. S. Decreased Meiotic Intergenic Recombination and Increased Meiosis I Nondisjunction in *exo1* Mutants of *Saccharomyces cerevisiae*. *Genetics* **156**, 1549–1557 (2000).
235. Fiorentini, P., Huang, K. N., Tishkoff, D. X., Kolodner, R. D. & Symington, L. S. Exonuclease I of *Saccharomyces cerevisiae* functions in mitotic recombination in vivo and in vitro. *Mol. Cell. Biol.* **17**, 2764–2773 (1997).
236. Cannavo, E., Cejka, P. & Kowalczykowski, S. C. Relationship of DNA degradation by *Saccharomyces cerevisiae* Exonuclease 1 and its stimulation by RPA and Mre11-Rad50-Xrs2 to DNA end resection. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E1661–E1668 (2013).
237. Llorente, B. & Symington, L. S. The Mre11 Nuclease Is Not Required for 5' to 3' Resection at Multiple HO-Induced Double-Strand Breaks. *Mol. Cell. Biol.* **24**, 9682–9694 (2004).
238. Maringele, L. & Lydall, D. EXO1-dependent single-stranded DNA at telomeres activates subsets of DNA damage and spindle checkpoint pathways in budding yeast *yku70Δ* mutants. *Genes Dev.* **16**, 1919–1933 (2002).
239. Nicolette, M. L. *et al.* Mre11–Rad50–Xrs2 and Sae2 promote 5' strand resection of DNA double-strand breaks. *Nat. Struct. Mol. Biol.* **17**, 1478–1485 (2010).
240. Moreau, S., Morgan, E. A. & Symington, L. S. Overlapping Functions of the *Saccharomyces cerevisiae* Mre11, Exo1 and Rad27 Nucleases in DNA Metabolism. *Genetics* **159**, 1423–1433 (2001).
241. Zhu, Z., Chung, W.-H., Shim, E. Y., Lee, S. E. & Ira, G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double strand break ends. *Cell* **134**, 981–994 (2008).
242. Gangloff, S., McDonald, J. P., Bendixen, C., Arthur, L. & Rothstein, R. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. *Mol. Cell. Biol.* **14**, 8391–8398 (1994).
243. Bernstein, K. A., Gangloff, S. & Rothstein, R. The RecQ DNA helicases in DNA Repair. *Annu. Rev. Genet.* **44**, 393–417 (2010).
244. Gilman, B., Tijerina, P. & Russell, R. Distinct RNA unwinding mechanisms of DEAD-box and DEAH-box RNA helicase proteins in remodeling structured RNAs and RNPs. *Biochem. Soc. Trans.* **45**, 1313–1321 (2017).
245. Wu, L. & Hickson, I. D. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* **426**, 870–874 (2003).
246. Niu, H. *et al.* Mechanism of the ATP-dependent DNA End Resection Machinery from *S. cerevisiae*. *Nature* **467**, 108–111 (2010).
247. Gangloff, S., McDONALD, J. P. & Rothstein, R. The Yeast Type I Topoisomerase Top3 Interacts with Sgs1, a DNA Helicase Homolog: a Potential Eukaryotic Reverse Gyrase. *MOL CELL BIOL* **14**, 8 (1994).

248. Mullen, J. R., Nallaseth, F. S., Lan, Y. Q., Slagle, C. E. & Brill, S. J. Yeast Rmi1/Nce4 Controls Genome Stability as a Subunit of the Sgs1-Top3 Complex. *Mol. Cell. Biol.* **25**, 4476–4487 (2005).
249. Cejka, P. & Kowalczykowski, S. C. The Full-length *Saccharomyces cerevisiae* Sgs1 Protein Is a Vigorous DNA Helicase That Preferentially Unwinds Holliday Junctions. *J. Biol. Chem.* **285**, 8290–8301 (2010).
250. Cejka, P. *et al.* DNA end resection by Dna2-Sgs1-RPA and its stimulation by Top3-Rmi1 and Mre11-Rad50-Xrs2. *Nature* **467**, 112–116 (2010).
251. Mimitou, E. P. & Symington, L. S. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455**, (2008).
252. Budd, M. E., Choe, W.-C. & Campbell, J. L. DNA2 Encodes a DNA Helicase Essential for Replication of Eukaryotic Chromosomes. *J. Biol. Chem.* **270**, 26766–26769 (1995).
253. Kim, J.-H. *et al.* Isolation of human Dna2 endonuclease and characterization of its enzymatic properties. *Nucleic Acids Res.* **34**, 1854–1864 (2006).
254. Bae, S.-H. *et al.* Dna2 of *Saccharomyces cerevisiae* Possesses a Single-stranded DNA-specific Endonuclease Activity That Is Able to Act on Double-stranded DNA in the Presence of ATP. *J. Biol. Chem.* **273**, 26880–26890 (1998).
255. Levikova, M., Klaue, D., Seidel, R. & Cejka, P. Nuclease activity of *Saccharomyces cerevisiae* Dna2 inhibits its potent DNA helicase activity. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E1992–E2001 (2013).
256. Chen, H., Lisby, M. & Symington, L. S. RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Mol. Cell* **50**, (2013).
257. Sung, P. Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* **265**, 1241–1243 (1994).
258. New, J. H., Sugiyama, T., Zaitseva, E. & Kowalczykowski, S. C. Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. *Nature* **391**, 407–410 (1998).
259. Sugiyama, T. & Kowalczykowski, S. C. Rad52 Protein Associates with Replication Protein A (RPA)-Single-stranded DNA to Accelerate Rad51-mediated Displacement of RPA and Presynaptic Complex Formation. *J. Biol. Chem.* **277**, 31663–31672 (2002).
260. Sugiyama, T., New, J. H. & Kowalczykowski, S. C. DNA annealing by Rad52 Protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6049–6054 (1998).
261. Miyazaki, T., Bressan, D. A., Shinohara, M., Haber, J. E. & Shinohara, A. In vivo assembly and disassembly of Rad51 and Rad52 complexes during double-strand break repair. *EMBO J.* **23**, 939–949 (2004).
262. Krejci, L., Altmannova, V., Spirek, M. & Zhao, X. Homologous recombination and its regulation. *Nucleic Acids Res.* **40**, 5795–5818 (2012).
263. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33**, 25–35 (1983).
264. Schwacha, A. & Kleckner, N. Identification of double holliday junctions as intermediates in meiotic recombination. *Cell* **83**, 783–791 (1995).
265. Ip, S. C. Y. *et al.* Identification of Holliday junction resolvases from humans and yeast. *Nature* **456**, 357–361 (2008).
266. Mitchel, K., Lehner, K. & Jinks-Robertson, S. Heteroduplex DNA Position Defines the Roles of the Sgs1, Srs2, and Mph1 Helicases in Promoting Distinct Recombination Outcomes. *PLoS Genet.* **9**, e1003340 (2013).
267. Zakharyevich, K., Tang, S., Ma, Y. & Hunter, N. Delineation of Joint Molecule Resolution Pathways in Meiosis Identifies a Crossover-Specific Resolvase. *Cell* **149**, 334–347

(2012).

268. Lynn, A., Soucek, R. & Börner, G. V. ZMM proteins during meiosis: Crossover artists at work. *Chromosome Res.* **15**, 591–605 (2007).
269. Börner, G. V., Kleckner, N. & Hunter, N. Crossover/Noncrossover Differentiation, Synaptonemal Complex Formation, and Regulatory Surveillance at the Leptotene/Zygotene Transition of Meiosis. *Cell* **117**, 29–45 (2004).
270. Marsolier-Kergoat, M.-C., Khan, M. M., Schott, J., Zhu, X. & Llorente, B. Mechanistic View and Genetic Control of DNA Recombination during Meiosis. *Mol. Cell* **70**, 9-20.e6 (2018).
271. Martini, E. *et al.* Genome-Wide Analysis of Heteroduplex DNA in Mismatch Repair-Deficient Yeast Cells Reveals Novel Properties of Meiotic Recombination Pathways. *PLoS Genet.* **7**, e1002305 (2011).
272. Jessop, L., Allers, T. & Lichten, M. Infrequent Co-conversion of Markers Flanking a Meiotic Recombination Initiation Site in *Saccharomyces cerevisiae*. *Genetics* **169**, 1353–1367 (2005).
273. Nassif, N., Penney, J., Pal, S., Engels, W. R. & Gloor, G. B. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol. Cell Biol.* **14**, 1613–1625 (1994).
274. Pâques, F., Bucheton, B. & Wegnez, M. Rearrangements Involving Repeated Sequences Within a P Element Preferentially Occur Between Units Close to the Transposon Extremities. *Genetics* **142**, 459–470 (1996).
275. Pâques, F., Leung, W.-Y. & Haber, J. E. Expansions and Contractions in a Tandem Repeat Induced by Double-Strand Break Repair. *Mol. Cell Biol.* **18**, 2045–2054 (1998).
276. Richard, G.-F. & Pâques, F. Mini- and microsatellite expansions: the recombination connection. *EMBO Rep.* **1**, 122–126 (2000).
277. Liu, J. *et al.* Srs2 promotes synthesis-dependent strand annealing by disrupting DNA polymerase δ -extending D-loops. *eLife* **6**, e22195 (2017).
278. Prakash, R. *et al.* Yeast Mph1 helicase dissociates Rad51-made D-loops: implications for crossover control in mitotic recombination. *Genes Dev.* **23**, 67–79 (2009).
279. Davis, A. P. & Symington, L. S. RAD51-Dependent Break-Induced Replication in Yeast. *Mol. Cell Biol.* **24**, 2344–2351 (2004).
280. Sakofsky, C. J. & Malkova, A. Break induced replication in eukaryotes: mechanisms, functions, and consequences. *Crit. Rev. Biochem. Mol. Biol.* **52**, 395–413 (2017).
281. Saini, N. *et al.* Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502**, 389–392 (2013).
282. Donnianni, R. A. & Symington, L. S. Break-induced replication occurs by conservative DNA synthesis. *Proc. Natl. Acad. Sci.* **110**, 13475–13480 (2013).
283. Lydeard, J. R., Jain, S., Yamaguchi, M. & Haber, J. E. Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* **448**, 820–823 (2007).
284. Malkova, A., Ivanov, E. L. & Haber, J. E. Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc. Natl. Acad. Sci.* **93**, 7131–7136 (1996).
285. Malkova, A., Naylor, M. L., Yamaguchi, M., Ira, G. & Haber, J. E. RAD51-Dependent Break-Induced Replication Differs in Kinetics and Checkpoint Responses from RAD51-Mediated Gene Conversion. *Mol. Cell Biol.* **25**, 933–944 (2005).
286. Signon, L., Malkova, A., Naylor, M. L., Klein, H. & Haber, J. E. Genetic Requirements for RAD51- and RAD54-Independent Break-Induced Replication Repair of a Chromosomal Double-Strand Break. *Mol. Cell Biol.* **21**, 2048–2056 (2001).
287. Ira, G. & Haber, J. E. Characterization of RAD51-Independent Break-Induced Replication That Acts Preferentially with Short Homologous Sequences. *Mol. Cell Biol.* **22**,

6384–6392 (2002).

288. Kang, L. E. & Symington, L. S. Aberrant Double-Strand Break Repair in rad51 Mutants of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **20**, 9162–9172 (2000).
289. Kramara, J., Osia, B. & Malkova, A. Break-Induced Replication: The Where, The Why, and The How. *Trends Genet.* **34**, 518–531 (2018).
290. Malkova, A. *et al.* RAD51-independent break-induced replication to repair a broken chromosome depends on a distant enhancer site. *Genes Dev.* **15**, 1055–1060 (2001).
291. Fishman-Lobell, J., Rudin, N. & Haber, J. E. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol. Cell. Biol.* **12**, 1292–1303 (1992).
292. Ivanov, E. L. & Haber, J. E. RAD1 and RAD10, but not other excision repair genes, are required for double-strand break-induced recombination in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 2245–2251 (1995).
293. Ozenberger, B. A. & Roeder, G. S. A unique pathway of double-strand break repair operates in tandemly repeated genes. *Mol. Cell. Biol.* **11**, 1222–1231 (1991).
294. Mortensen, U. H., Bendixen, C., Sunjevaric, I. & Rothstein, R. DNA strand annealing is promoted by the yeast Rad52 protein. *Proc. Natl. Acad. Sci.* **93**, 10729–10734 (1996).
295. Richard, G.-F., Dujon, B. & Haber, J. E. Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats. *Mol. Gen. Genet. MGG* **261**, 871–882 (1999).
296. Richard, G.-F. *et al.* Highly Specific Contractions of a Single CAG/CTG Trinucleotide Repeat by TALEN in Yeast. *PLOS ONE* **9**, e95611 (2014).
297. Förstemann, K., Höss, M. & Lingner, J. Telomerase-dependent repeat divergence at the 3' ends of yeast telomeres. *Nucleic Acids Res.* **28**, 2690–2694 (2000).
298. Welch, J. W., Maloney, D. H. & Fogel, S. Unequal crossing-over and gene conversion at the amplified CUP1 locus of yeast. *Mol. Gen. Genet. MGG* **222**, 304–310 (1990).
299. Jeffreys, A. J. *et al.* Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.* **6**, 136–145 (1994).
300. Jeffreys, A. J., Murray, J. & Neumann, R. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell* **2**, 267–273 (1998).
301. Appelgren, H., Cederberg, H. & Rannug, U. Mutations at the human minisatellite MS32 integrated in yeast occur with high frequency in meiosis and involve complex recombination events. *Mol. Gen. Genet. MGG* **256**, 7–17 (1997).
302. Contractions and Expansions of CAG/CTG Trinucleotide Repeats occur during Ectopic Gene Conversion in Yeast, by a MUS81-independent Mechanism. *J. Mol. Biol.* **326**, 769–782 (2003).
303. Kramer, K. M. & Haber, J. E. New telomeres in yeast are initiated with a highly selected subset of TG1-3 repeats. *Genes Dev.* **7**, 2345–2356 (1993).
304. Wellinger, R. J. & Zakian, V. A. Everything You Ever Wanted to Know About *Saccharomyces cerevisiae* Telomeres: Beginning to End. *Genetics* **191**, 1073–1105 (2012).
305. Singer, M. S. & Gottschling, D. E. TLC1: template RNA component of *Saccharomyces cerevisiae* telomerase. *Science* **266**, 404–409 (1994).
306. Lue, N. F. *Yeast Telomerases: Structure, Mechanisms and Regulation. Madame Curie Bioscience Database [Internet]* (Landes Bioscience, 2013).
307. Davé, A. *et al.* Homologous recombination repair intermediates promote efficient de novo telomere addition at DNA double-strand breaks. *Nucleic Acids Res.* **48**, 1271–1284 (2020).
308. Chang, M., Arneric, M. & Lingner, J. Telomerase repeat addition processivity is increased at critically short telomeres in a Tel1-dependent manner in *Saccharomyces*

- cerevisiae. *Genes Dev.* **21**, 2485–2494 (2007).
309. Chung, W.-H., Zhu, Z., Papusha, A., Malkova, A. & Ira, G. Defective Resection at DNA Double-Strand Breaks Leads to De Novo Telomere Formation and Enhances Gene Targeting. *PLoS Genet.* **6**, e1000948 (2010).
310. Hirano, Y., Fukunaga, K. & Sugimoto, K. Rif1 and Rif2 inhibit localization of Tel1 to DNA ends. *Mol. Cell* **33**, 312–322 (2009).
311. Dewar, J. M. & Lydall, D. Similarities and differences between ‘uncapped’ telomeres and DNA double-strand breaks. *Chromosoma* **121**, 117–130 (2012).
312. Lange, T. de. How Telomeres Solve the End-Protection Problem. *Science* **326**, 948–952 (2009).
313. Wotton, D. & Shore, D. A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in *Saccharomyces cerevisiae*. *Genes Dev.* **11**, 748–760 (1997).
314. Hailemariam, S. *et al.* The telomere-binding protein Rif2 and ATP-bound Rad50 have opposing roles in the activation of yeast Tel1/ATM kinase. *J. Biol. Chem.* **294**, 18846–18852 (2019).
315. Shi, T. *et al.* Rif1 and Rif2 shape telomere function and architecture through multivalent Rap1 interactions. *Cell* **153**, 1340–1353 (2013).
316. Cassani, C. *et al.* Tel1 and Rif2 Regulate MRX Functions in End-Tethering and Repair of DNA Double-Strand Breaks. *PLoS Biol.* **14**, (2016).
317. Bonetti, D. *et al.* DNA binding modes influence Rap1 activity in the regulation of telomere length and MRX functions at DNA ends. *Nucleic Acids Res.* **48**, 2424–2441 (2020).
318. Martina, M., Bonetti, D., Villa, M., Lucchini, G. & Longhese, M. P. *Saccharomyces cerevisiae* Rif1 cooperates with MRX-Sae2 in promoting DNA-end resection. *EMBO Rep.* **15**, 695–704 (2014).
319. O’Hoy, K. *et al.* Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. *Science* **259**, 809 (1993).
320. Poggi, L., Dumas, B. & Richard, G.-F. Monitoring Double-Strand Break Repair of Trinucleotide Repeats Using a Yeast Fluorescent Reporter Assay. *Methods Mol. Biol. Clifton NJ* **2056**, 113–120 (2020).
321. Moreau, S., Ferguson, J. R. & Symington, L. S. The nuclease activity of Mre11 is required for meiosis but not for mating type switching, end joining, or telomere maintenance. *Mol. Cell. Biol.* **19**, 556–566 (1999).
322. Ritchie, K. B., Mallory, J. C. & Petes, T. D. Interactions of TLC1 (Which Encodes the RNA Subunit of Telomerase), TEL1, and MEC1 in Regulating Telomere Length in the Yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 6065–6075 (1999).

ANNEXES

Annexe 1: Resection and repair of a Cas9 double-strand break at CTG trinucleotide repeats induces local and extensive chromosomal deletions

Article sur les mécanismes de réparation d'une cassure double-brin généré au sein de répétitions CTG par SpCas9. J'ai réalisé les expériences de qPCR de la figure 8


RESEARCH ARTICLE

Resection and repair of a Cas9 double-strand break at CTG trinucleotide repeats induces local and extensive chromosomal deletions

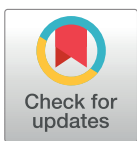
Valentine Mosbach¹[✉], David Viterbo¹[✉], Stéphane Descorps-Declère^{1,2}, Lucie Poggi^{1,3}, Wilhelm Vaysse-Zinkhöfer^{1,3}, Guy-Franck Richard¹^{*}

1 Institut Pasteur, CNRS, UMR3525, Paris, France, **2** Institut Pasteur, Center of Bioinformatics, Biostatistics and Integrative Biology (C3BI), Paris, France, **3** Sorbonne Universités, Collège doctoral, Paris, France

 These authors contributed equally to this work.

 Current address: Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), UMR7104 CNRS/Unistra, INSERM U1258, Illkirch, France

* gfrichar@pasteur.fr



 OPEN ACCESS

Citation: Mosbach V, Viterbo D, Descorps-Declère S, Poggi L, Vaysse-Zinkhöfer W, Richard G-F (2020) Resection and repair of a Cas9 double-strand break at CTG trinucleotide repeats induces local and extensive chromosomal deletions. *PLoS Genet* 16(7): e1008924. <https://doi.org/10.1371/journal.pgen.1008924>

Editor: Lorraine S. Symington, Columbia University, UNITED STATES

Received: December 6, 2019

Accepted: June 10, 2020

Published: July 16, 2020

Copyright: © 2020 Mosbach et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: V. M. was supported by Fondation Guy Nicolas and Fondation Hardy. W. V.-Z. is the recipient of a PhD fellowship from la Ligue Nationale Contre le Cancer. This work was generously supported by the Institut Pasteur and by the Centre National de la Recherche Scientifique (CNRS). The funders had no role in study design,

Abstract

Microsatellites are short tandem repeats, ubiquitous in all eukaryotes and represent ~2% of the human genome. Among them, trinucleotide repeats are responsible for more than two dozen neurological and developmental disorders. Targeting microsatellites with dedicated DNA endonucleases could become a viable option for patients affected with dramatic neurodegenerative disorders. Here, we used the *Streptococcus pyogenes* Cas9 to induce a double-strand break within the expanded CTG repeat involved in myotonic dystrophy type 1, integrated in a yeast chromosome. Repair of this double-strand break generated unexpected large chromosomal deletions around the repeat tract. These deletions depended on *RAD50*, *RAD52*, *DNL4* and *SAE2*, and both non-homologous end-joining and single-strand annealing pathways were involved. Resection and repair of the double-strand break (DSB) were totally abolished in a *rad50Δ* strain, whereas they were impaired in a *sae2Δ* mutant, only on the DSB end containing most of the repeat tract. This observation demonstrates that Sae2 plays significant different roles in resecting a DSB end containing a repeated and structured sequence as compared to a non-repeated DSB end. In addition, we also discovered that gene conversion was less efficient when the DSB could be repaired using a homologous template, suggesting that the trinucleotide repeat may interfere with gene conversion too. Altogether, these data show that *SpCas9* may not be the best choice when inducing a double-strand break at or near a microsatellite, especially in mammalian genomes that contain many more dispersed repeated elements than the yeast genome.

Author summary

With the discovery of highly specific DNA endonucleases such as TALEN and CRISPR-Cas systems, gene editing has become an attractive approach to address genetic disorders. Myotonic dystrophy type 1 (Steinert disease) is due to a large expansion of a CTG trinucleotide

data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare no competing interest.

repeat in the DMPK gene. At the present time, despite numerous therapeutic attempts, this dramatic neurodegenerative disorder still has no cure. In the present work, we tried to use the Cas9 endonuclease to induce a double-strand break within the expanded CTG repeat of the DMPK gene integrated in the yeast genome. Surprisingly, this break induced chromosomal deletions around the repeat tract. These deletions were local and involved non-homologous joining of the two DNA ends, or more extensive involving homologous recombination between repeated elements upstream and downstream the break. Using yeast genetics, we investigated the genetic requirements for these deletions and found that the triplet repeat tract altered the capacity of the repair machinery to faithfully repair the double-strand break. These results have implications for future gene therapy approaches in human patients.

Introduction

Microsatellites are short tandem repeats ubiquitously found in all eukaryotic genomes sequenced so far [1]. Altogether, they cover ~2% of the human genome, a figure similar to the whole protein-coding sequence [2]. Naturally prone to frequent repeat length polymorphism, some microsatellites are also prone to large expansions that lead to human neurological or developmental disorders, such as trinucleotide repeats involved in Huntington disease, myotonic dystrophy type 1 (Steinert disease), fragile X syndrome or Friedreich ataxia [3]. These expansion-prone microsatellites share the common property to form secondary DNA structures *in vitro* [4] and genetic evidence suggest that similar structures may also form *in vivo* [5,6], transiently stalling replication fork progression [7–11]. Among those, CCG/CGG trinucleotide repeats are fragile sites in human cells, forming frequent double-strand breaks when the replication machinery is slowed down or impaired [12]. Similarly, CAG/CTG and CCG/CGG microsatellites are also fragile sites in *Saccharomyces cerevisiae* cells [13,14]. Therefore, microsatellite abundance and the natural fragility of some of them make these repeated sequences perfect targets to generate chromosomal rearrangements potentially leading to cancer.

Double-strand break (DSB) repair mechanisms have been studied for decades in model organisms as well as in human cells and led to the identification of the main genes involved in this process [15]. Many of these advances were made possible by the use of highly specific DNA endonucleases, such as the meganucleases I-SceI or HO [16–18]. Other frequently used methods involved ionizing radiation making genome-wide DSBs [19]. However, the fate of a single double-strand break within a repeated and structured DNA sequence has never been addressed, until recently. In a former work, we used a TALE Nuclease (TALEN) to induce a unique DSB into a long CTG trinucleotide repeat integrated into a *S. cerevisiae* chromosome. TALENs are made of the fusion between a Transcription Activator-like Effector, a *Xanthomonas* family of modular transcription activators (TALE), and the FokI nuclease domain. We showed that 100% of yeast cells in which the TALEN was expressed exhibited a large contraction of the repeat tract, going from an initial length of ~80 CTG triplets to less than 35. *POL32*, *DNL4* and *RAD51* were shown to play no detectable role in repairing this DSB. On the contrary, *RAD50*, *RAD52* and *SAE2* were required for proper repair of the DSB, and a functional Sae2 protein was found to be essential for efficient DSB resection, suggesting that repeat contraction occurred by a single-strand annealing (SSA) process, involving preliminary resection of the break, followed by annealing of the two DSB ends carrying the repeat tract [20,21].

In the present work, we used the *Streptococcus pyogenes* Cas9 endonuclease (*SpCas9*) to induce a DSB within the same long CTG trinucleotide repeat integrated in the yeast genome. The break was made at the 3' end of the repeat tract (Fig 1A), using a guide RNA that targets

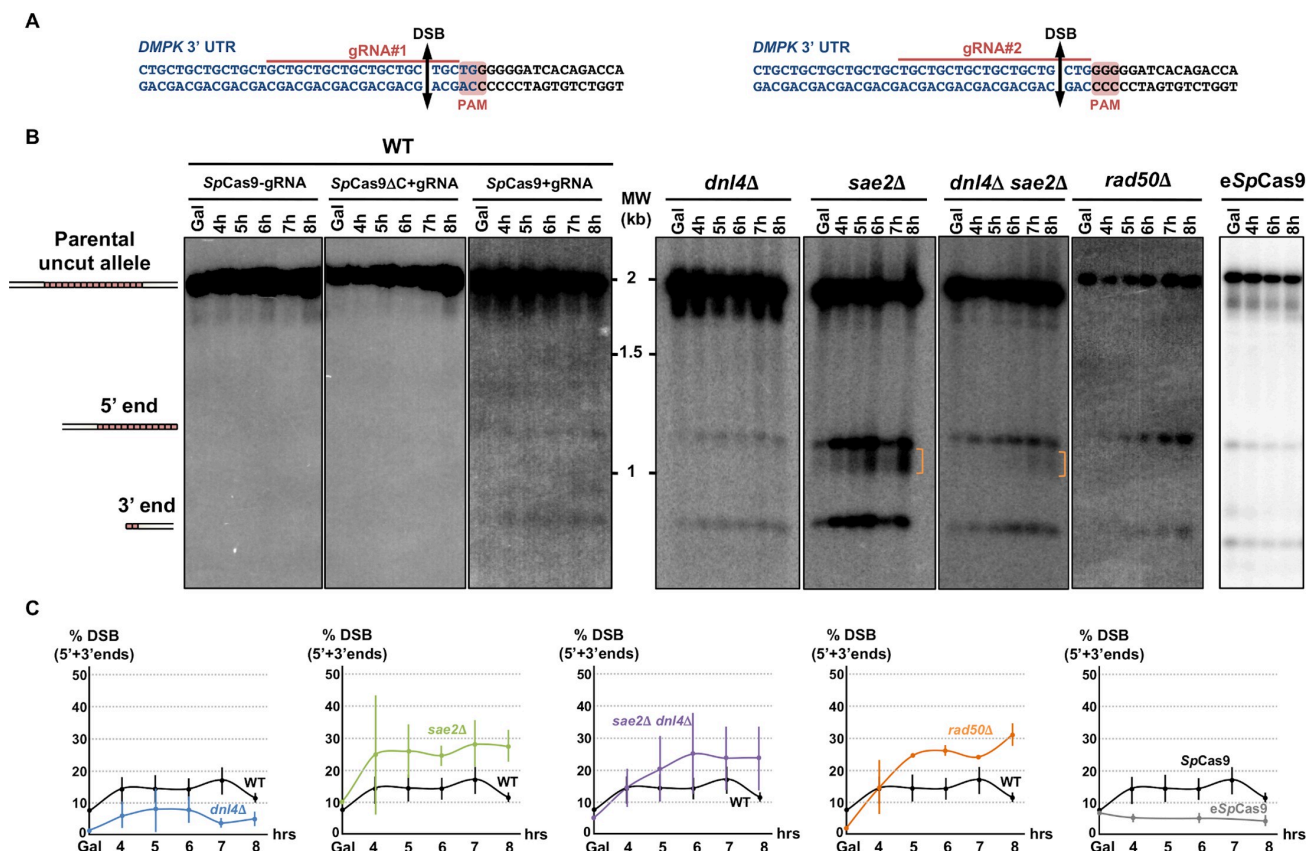


Fig 1. *SpCas9* and *eSpCas9* DSB induction in wild type and mutant strains. A: Sequence of the *SUP4::(CTG)_n* locus. The CTG trinucleotide repeat tract comes from a human DM1 patient and is shown in blue. The flanking non-repeated DNA is in black. For each guide RNA, the PAM, the gRNA sequence as well as the expected DSB site are indicated. B: Southern blots of yeast strains during time courses. Lanes labeled *SpCas9*-gRNA and *SpCas9ΔCter*+gRNA are control strains in which no DSB was visible. In the strain expressing both *SpCas9* and the gRNA, two bands are visible in addition to the parental allele (1966 bp). One band corresponds to the 3' end of the DSB containing a small number of triplets (821 bp), the other one corresponds to the 5' end of the DSB containing most of the repeat tract (1145 bp). C: Quantification of 5' and 3' DSB signals. For each time points, the total 5' + 3' signals were quantified and plotted as a ratio of the total signal in the lane. Three independent time courses were run in each strain background (except *rad50Δ* and *eSpCas9* for which two time courses were run) and plots show the average of three (or two) time courses. Error bars are 95% confidence intervals.

<https://doi.org/10.1371/journal.pgen.1008924.g001>

the repeat tract. Frequent rearrangements were found in surviving cells, with local deletions as well as more extensive ones involving recombination between retrotransposon LTRs. Survival and repair depended on *RAD50*, *RAD52*, *SAE2* and *DNL4* and double-strand break resection was abolished in *rad50Δ* and *sae2Δ* mutants. A more specific version of the nuclease, Enhanced *SpCas9*, generated the same rearrangements. In addition, we also discovered that gene conversion was less efficient when *SpCas9* was used to induce a DSB within a CTG repeat tract that could be repaired with a homologous template, suggesting that the trinucleotide repeat may interfere with gene conversion too.

Results

A Cas9-induced double-strand break within CTG repeats induces cell death and chromosomal deletions around the repeat tract

In previous work, we showed that a TALEN targeted at the CTG trinucleotide repeat from the human *DMPK* gene 3' UTR integrated in a yeast chromosome, was extremely efficient at

contracting the repeat tract below the pathological length [20,21]. In order to determine whether the CRISPR-Cas9 system could be used in the same manner, a plasmid-borne *Streptococcus pyogenes* Cas9 nuclease (*SpCas9*) was expressed in *Saccharomyces cerevisiae* from a *GAL1*-inducible promoter [22]. The same plasmid also carried a CTG guide RNA (hereafter named gRNA#1) under the control of the constitutive *SNR52* promoter. The PAM used in this experiment was the TGG sequence located right at the border between the CTG tract and non-repeated DNA (Fig 1A). As controls, we used the same *SpCas9*-containing plasmid without the gRNA or a frameshift mutant of the *SpCas9* gene resulting in a premature stop codon (*SpCas9ΔNdeI*) and the gRNA#1. The same genetic assay as previously described was used [20,21]. It is based on a modified suppressor tRNA gene (*SUP4*) in which a CTG trinucleotide repeat was integrated. The length of the CTG repeat at the start of the experiment was determined to be approximately 80 triplets. Four hours after transition from glucose to galactose medium, two faint bands were visible on a Southern blot, corresponding to the 5' and 3' ends of the *SpCas9* DSB. No signal was detected in control strains (Fig 1B). The DSB was quantified to be present in ca. 10%-15% of the cells at any given time point and remained the same for the duration of the time course. No evidence for repeat tract contraction was visible. Survival to the *SpCas9* break was low (17.9%±4%), as calculated from CFU on galactose plates over CFU on glucose plates (Fig 2, see Materials & Methods). Surviving colonies were picked, total

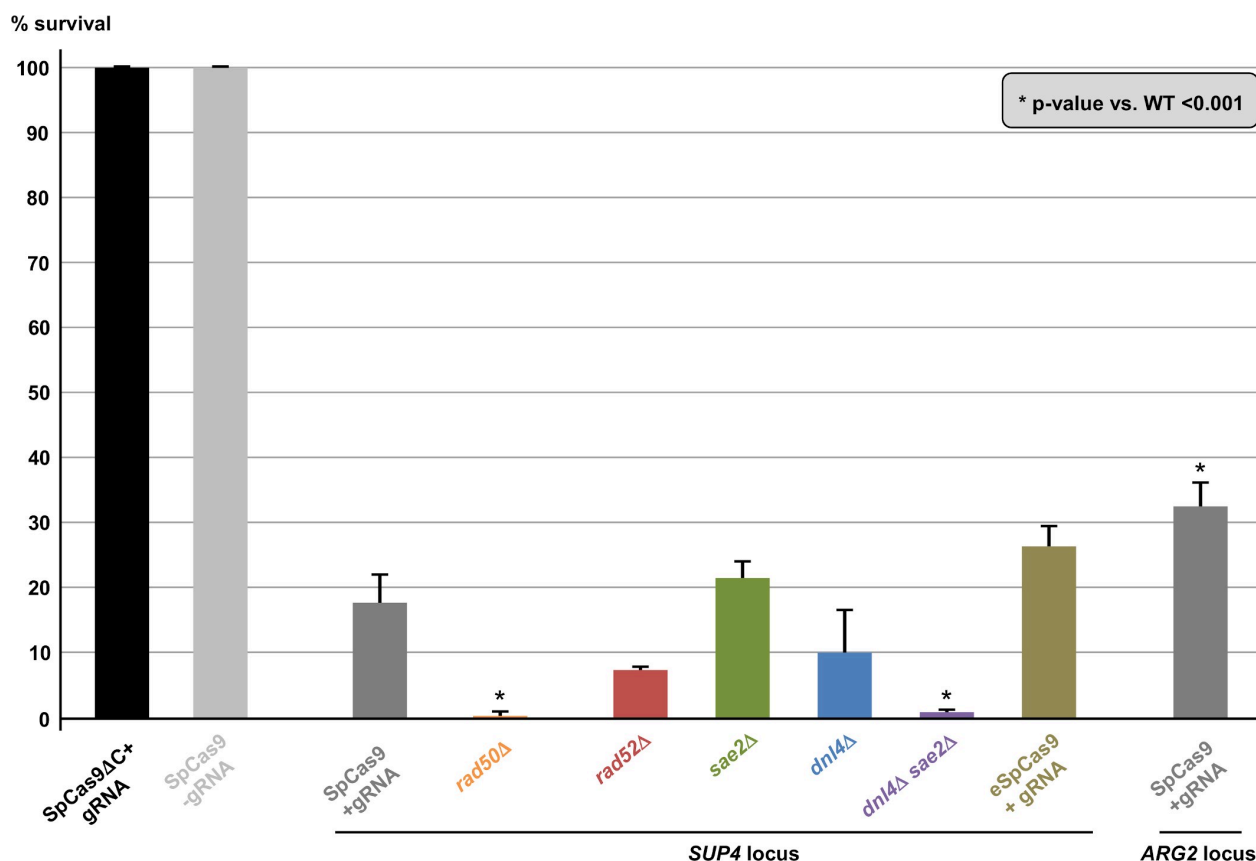


Fig 2. Yeast survival to Cas9 induction. For each strain, the same number of cells were plated on galactose and glucose plates and the survival was expressed in CFU number on galactose plates over CFU number on glucose plates. The mean and the 95% confidence interval are plotted for each strain. Significant t-test p-values when compared to wild-type *SpCas9* survival are indicated by asterisks. See Materials & Methods for statistics.

<https://doi.org/10.1371/journal.pgen.1008924.g002>

genomic DNA was extracted and the *SUP4::CTG* locus was analyzed by Southern blot. Patterns observed were remarkably different among clones, most of them showing bands of aberrant molecular weight, either much larger or much shorter than the repeat tract. In some clones, a total absence of signal suggested that the probe target was deleted and in other cases weakness of the signal was compatible with a partial deletion (Fig 3A). To understand these

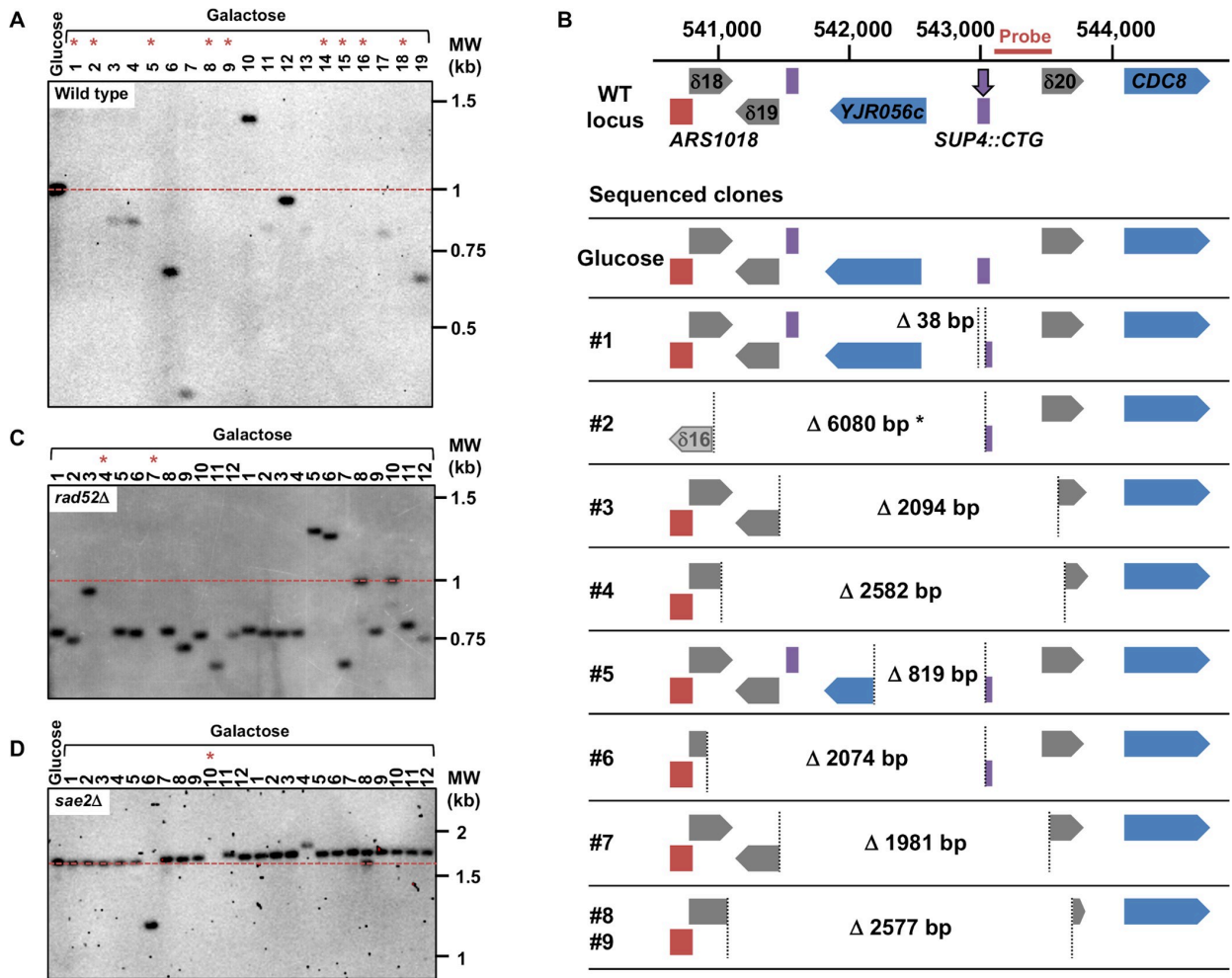


Fig 3. Chromosomal rearrangements following *SpCas9* induction. A: Southern blot of genomic DNA at the *SUP4* locus in the wild-type strain. The probe hybridizes ~300 bp downstream the repeat tract (see Fig 3B). The dotted red line shows the initial length of the CTG repeat tract. The lane labeled "Glucose" contains a clone in which Cas9 was not induced. Lanes numbered #1 through #19 contain independent clones in which Cas9 was induced. Asterisks point to lanes in which no signal was detected, meaning that the probe containing sequence was deleted. Note that signal intensities varies among lanes, showing that the probe did not fully bind to its target sequence, due to its partial deletion. B: Some examples of chromosome rearrangements following Cas9 induction in the wild-type strain. The genomic locus surrounding *SUP4* is shown on top, *ARS1018* is drawn in red, delta elements are in grey, protein-coding genes are colored in blue and tRNA genes in purple. The DSB (vertical purple arrow) is induced within *SUP4::(CTG)n*. Chromosome coordinates are indicated above and the probe used for hybridization is represented by an horizontal red bar. The locus sequence was retrieved from the Saccharomyces Genome Database (<http://yeastgenome.org/>, genome version R64-2-1, released 18th November 2014). Under the reference locus are cartooned the different chromosomal structures observed in some of the survivors. A yeast colony that was grown in glucose was also sequenced as a control. For each clone, vertical dotted lines represent junctions of rearrangements observed, with deletion sizes indicated in base pairs. Asterisk: clone #2 showed a complex rearrangement with a local inverted duplication involving the $\delta 16$ LTR and the 3' end of the *KCH1* gene 5 kb upstream *SUP4*. Two clones (#8 and #9) exhibit exactly the same chromosomal deletion at precisely the same nucleotides. Note that *CDC8* is an essential gene. C: Southern blot of genomic DNA at the *SUP4* locus in the *rad52 Δ strain. Legend as for Fig 3A. D: Southern blot of genomic DNA at the *SUP4* locus in the *sae2 Δ strain. Legend as for Fig 3A. Note that for this Southern blot genomic DNA was digested with *EcoRV* (instead of *Ssp I*, see [Methods](#)), therefore the expected CTG repeat length was around 1.8 kb, instead of 1 kb.**

<https://doi.org/10.1371/journal.pgen.1008924.g003>

abnormal patterns, the genome of nine surviving clones were totally sequenced by paired-end Illumina. As a control, one clone in which *SpCas9* had not been induced was also sequenced. In all nine cases, a deletion around the repeat tract was found, extending from a few nucleotides to several kilobases (Fig 3B). Some of the deletions involved flanking Ty1 retrotransposon LTRs, and in one case (clone #2), a complex event between a distant LTR ($\delta 16$) and the $\delta 20$ LTR close to the repeat tract was detected. Following this discovery, total genomic DNA was extracted from more surviving colonies, analyzed by Southern blot and deletion junctions were amplified by PCR and Sanger sequencing. Two different sets of primers were used to amplify the junction, su47/su48 allowing to amplify local deletions around the repeat tract, whereas su23/su42 were used to amplify larger deletions between Ty1 LTRs (Fig 4). Surviving colonies were classified into 12 different types, according to the *SUP4::CTG* locus after *SpCas9* induction: type I corresponded to a colony in which the repeat tract was unchanged (or was slightly expanded), type II to a colony in which a repeat contraction occurred, type III-V corresponded to local deletions around the repeat tract, types VI-XI to more extensive deletions, and type XII to a complex event between another δ element and $\delta 20$. A few examples of junction sequences are shown in S1 Fig.

Chromosomal deletions are under the control of *RAD50*, *RAD52*, *DNL4* and *SAE2*

We next decided to investigate the role of several genes known to be involved in DSB repair on chromosomal deletions generated by the *SpCas9* nuclease. In a *rad52* Δ strain, in which all homologous recombination was abolished, survival decreased but was not significantly different from wild type ($7.6\% \pm 0.7\%$, Fig 2). Molecular analysis of the survivors by Southern blot showed that deletions seemed to be less extensive than in the wild-type strain, fewer lanes showing a partial or total absence of signal (Fig 3C). Junction sequencing confirmed that deletions between Ty LTRs were lost (Type VI events), except for two cases in which the deletion occurred through annealing of eight or nine nucleotides and was therefore *RAD52* independent (Fig 4 and S1 Fig, clones #C4 and #C7). This result showed that about 50% of colonies growing on galactose plates survived the DSB by *RAD52*-dependent homologous recombination between two LTR elements flanking the trinucleotide repeat tract.

The possible role of non-homologous end-joining (NHEJ) in the observed deletions was also addressed by deleting the gene encoding yeast Ligase IV (*DNL4*). In the *dnl4* Δ strain, the level of detected DSBs was slightly lower than in wild type (Fig 1B and 1C). Survival was decreased, but not significantly different from wild type ($10.5\% \pm 6.3\%$, Fig 2). Molecular analysis of the survivors showed that local deletions were totally lost, whereas extensive deletions involving Ty LTR represented 84% of all events (Fig 4 and S1 Fig). Hence, we concluded that all local deletions were NHEJ dependent.

In a recent work, we showed that *SAE2* was essential to repair a DSB induced by a TALEN within a long trinucleotide repeat. In its absence, unrepaired breaks accumulated and DSB resection was lost on the trinucleotide repeat-containing end [20]. We therefore tested the effect of a *sae2* Δ mutation on a *SpCas9* DSB in the same experimental system. Southern blot analysis of repair intermediates showed that DSB ends accumulated twice as much in the *sae2* Δ mutant as compared to wild type (Fig 1B and 1C). In addition, a smear was detected below the 5' DSB end (Fig 1B, orange bracket), hallmark of an incomplete resection triggering a repair defect [23]. Survival was similar to wild type ($21.5\% \pm 2.9\%$, Fig 2). Southern blot analysis of surviving colonies displayed very little size changes as compared to uninduced controls (Fig 3D). However, sequencing showed that the most frequent event was an insertion (or sometimes a small deletion) of one to eight nucleotides between the PAM and the repeat tract

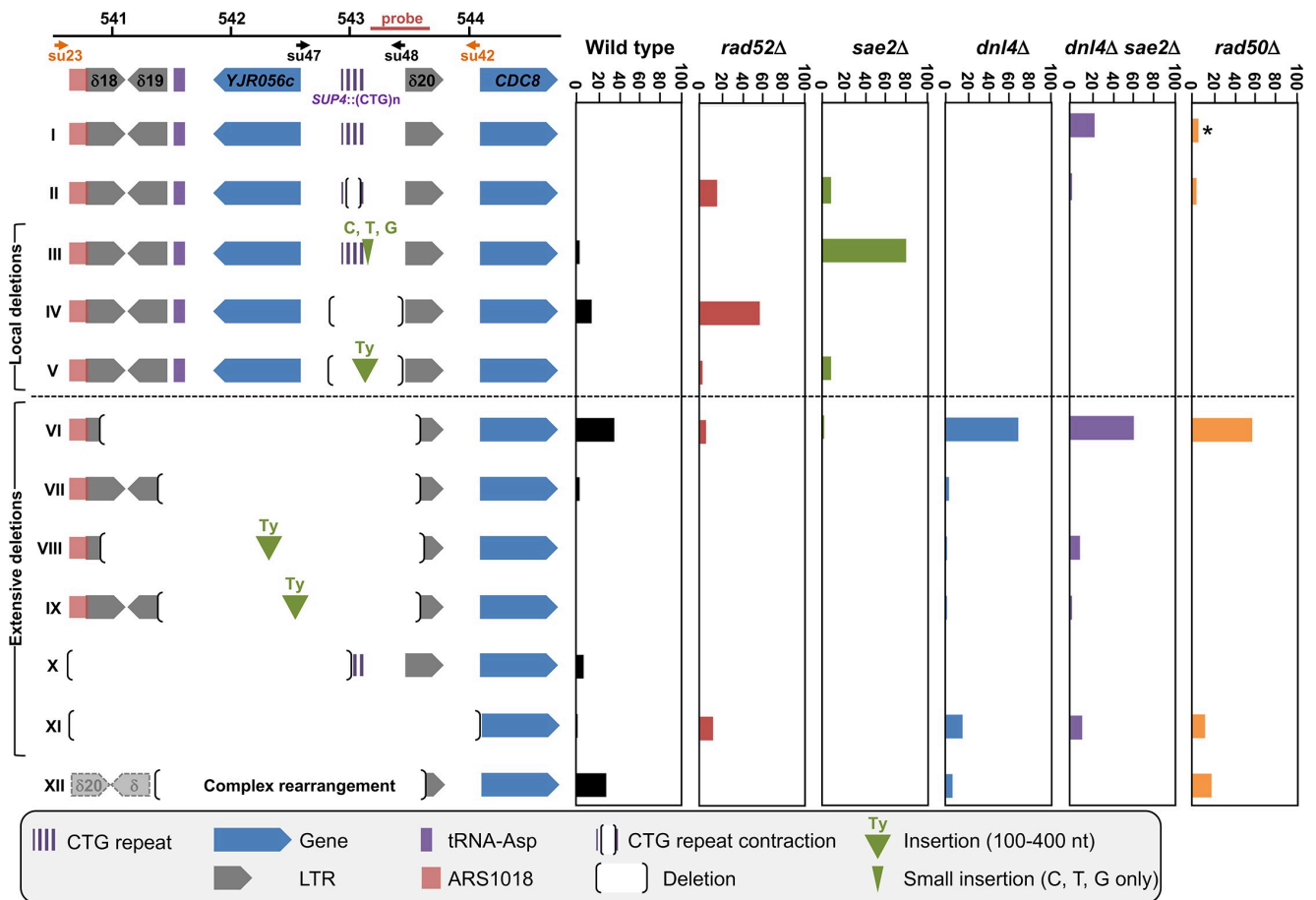


Fig 4. Summary of chromosomal rearrangements observed in wild-type and mutant strains, following *SpCas9* induction. **Left:** The twelve different possible outcomes following *SpCas9* induction are shown, subdivided in local and extensive deletions (see text for details). The *SUP4* locus is pictured and shows the position of each genetic element on yeast chromosome X. The probe used on Southern blots is shown, as well as both primer couples used to amplify the locus. In order to assess a given clone to a rearrangement type, the following rules were followed: i) when a band was detected by Southern blot, primers su47 and su48 were used to amplify the locus and sequence it. These events corresponded to types I-V. The absence of a PCR product indicated that primer su48 genomic sequence was probably deleted and therefore primers su23 and su42 were used to amplify and sequence the locus. These were classified as types IV-V events; ii) when no band was detected by Southern blot, primers su 23 and su42 were directly used to amplify and sequence the locus. These events were classified as types VI-X and XII. When no PCR product was obtained, it meant that at least one of the two primers genomic sequence was probably deleted and these events were classified as type XI. Note that this last category may also contain rare -but possible- chromosomal translocations that ended up in putting each primer in a separate chromosome, making unobtainable the PCR product. The extent of type XI deletions cannot go downstream the su42 primer, since the *CDC8* gene is essential. **Right:** The proportion of each type or event recovered is represented for wild type and mutants. The asterisk near Type I events in the *rad50Δ* strain indicates that some of them were small expansions (see text). Altogether, 262 surviving clones were sequenced, distributed as follows: WT: 51, *rad52Δ*: 29, *dnl4Δ*: 61, *sae2Δ*: 32, *dnl4Δ sae2Δ*: 47, *rad50Δ*: 42.

<https://doi.org/10.1371/journal.pgen.1008924.g004>

(Type III events, Fig 4). These local insertions represented 78% of all survivors, whereas only one Ty LTR recombination (Type VI) was detected (S1 Fig). This result showed that in the absence of *SAE2*, long range deletions were lost, probably due to the inability to resect the DSB into single-stranded DNA prone for homologous recombination.

The double mutant *sae2Δ dnl4Δ* was also built and showed an additive effect on survival, with a significant 30-fold reduction in CFU on galactose plates (0.6%±0.9%, Fig 2). This result proved that in the absence of one of the two genes repair could occur by the other pathway, but absence of both genes was almost lethal to yeast cells receiving a *SpCas9* DSB. Southern analysis showed that DSB levels were similar to *sae2Δ* levels (ca. 24% after 8 hrs versus 28% for

sae2Δ), showing that *SAE2* was epistatic to *DNL4*. The smear corresponding to resection defects was also visible (Fig 1B, orange bracket). Interestingly, 21% of survivors exhibited zero to two triplets lost, which could be attributed to natural microsatellite instability. These were classified as Type I events and were specific of the *sae2Δ dnl4Δ* double mutant (Fig 4 and S1 Fig). It is possible that given the low survival rate, cells in which *SpCas9* and/or the gRNA was mutated were positively selected during the time course in liquid culture and were therefore subsequently recovered on galactose plates. Remarkably, to the exception of the Type I events hereabove mentioned, all but one event corresponded to extensive deletions around the repeat tract, similarly to the single *dnl4Δ* mutant.

Finally, in a *rad50Δ* strain, the DSB accumulated over the duration of the time course at levels similar to *sae2Δ* mutants (Fig 1B and 1C). No smear was detected in this strain background, suggesting that the *sae2Δ* resection defect was specific of this mutant and did not involve the integrity of the MRX-Sae2 complex. Survival was very low in this strain background (0.3% ±0.4%), significantly different from wild type but not from the *sae2Δ dnl4Δ* double mutant (Fig 2). Survivor analysis showed a few repeat contractions (Type II), but most events were deletions between LTRs (Type VI), large deletions (Type XI) or complex rearrangements (Type XII), a pattern not significantly different from what was observed with the *sae2Δ dnl4Δ* double mutant (Fig 4). However, uniquely present in this strain background, were found three repeat expansions, two of them associated with the insertion of a 'C' in the first triplet preceding the PAM, most probably inhibiting Cas9 recognition and cutting (classified as Type I events in Fig 4, see S1 Fig). This suggests that repeat expansions are more frequent in a *rad50Δ* mutant in the presence of a DSB, as previously observed [24]. Such expansions were not recovered in the *sae2Δ dnl4Δ* double mutant, showing that this strain phenotype does not recapitulate exactly the *rad50Δ* phenotype, or that the rather limited number of survivors analyzed was not sufficient to detect a small number of expansions.

In conclusion, when a *SpCas9* DSB was induced into a long CTG trinucleotide repeat, cell survival was low and depended on *RAD50*, *RAD52*, *SAE2* and *DNL4*. Two classes of repair events were found: local deletions under the control of *RAD50* and *DNL4* and therefore the NHEJ pathway, and extensive deletions under the control of *SAE2* and *RAD52*. In addition, the deletion of *RAD50* almost completely recapitulated the *sae2Δ dnl4Δ* double mutation, except that the smear was not visible on Southern blots and a few expansions were recovered.

Enhanced *SpCas9* generates the same chromosomal deletions as *SpCas9*

Over the last four years, several mutants of the widely used *SpCas9* have been engineered or selected by genetic screens. *SpCas9*-HF1 and e*SpCas9* were built to exhibit less off-target DSBs [25,26], HypaCas9 was made to be even more accurate [27], Sniper-Cas9 also showed reduced off-target effects [28], while evoCas9 was selected in yeast for improved specificity [29]. We decided to explore the possibility that chromosomal deletions observed in our experimental system were partly due to the fact that *SpCas9* exhibited a high off-target activity on long CTG trinucleotide repeat tract, perhaps by generating more than one DSB within the repeat tract, or within the surrounding loci. In order to test this hypothesis, Enhanced *SpCas9* (e*SpCas9*) was expressed in yeast, along with the same guide RNA as previously (gRNA #1, Fig 1A). Survival was slightly higher than with *SpCas9* (26.3%±3.0%), but not significantly different (t test p-value = 0.06). DSB end accumulation was lower than *SpCas9* (Fig 1B and 1C). Molecular analysis of surviving yeast cells did not show any statistical difference between types of deletions observed with e*SpCas9* as compared to *SpCas9* (Chi2 p-value = 0.14) (Fig 5 and S1 Fig). A second guide RNA (gRNA#2) was designed, so that the DSB would be made two nucleotides

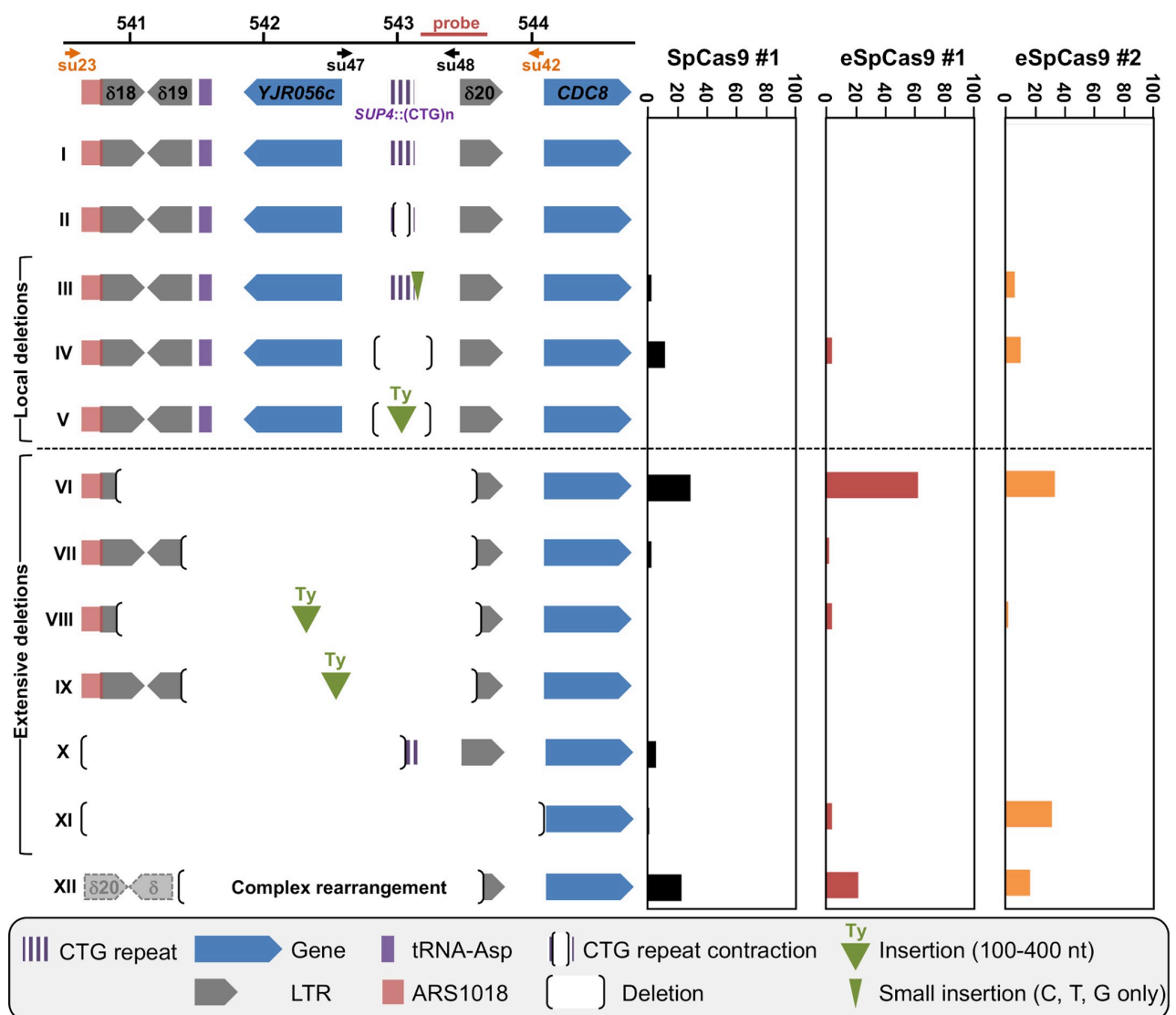


Fig 5. Summary of chromosomal rearrangements observed in following *SpCas9* and enhanced *SpCas9* (*eSpCas9*) inductions. See Fig 4 for legend. *eSpCas9* #1 and #2 refer respectively to guide RNAs #1 and #2 described in Fig 1A. The following number of surviving clones were analyzed: *SpCas9*: 51, *eSpCas9* #1: 49, *eSpCas9* #2: 48. See text for details.

<https://doi.org/10.1371/journal.pgen.1008924.g005>

closer to the repeat tract end (Fig 1A). Interestingly, the number of deletions involving a LTR (Types VI-IX) was lower than with gRNA#1 (35% with gRNA#2 vs 63% with gRNA#1) but the proportion of very large deletions (Type XI) significantly increased from 4% to 31% (Chi2 p-value = $1.6 \cdot 10^{-3}$). We concluded that moving the DSB cut site two nucleotides toward non-repeated DNA increased the outcome of very large deletions. Altogether, these results show that using a more specific version of *SpCas9* did not decrease chromosomal rearrangements, suggesting that deletions seen with *SpCas9* were probably not due to extra off-target DSBs within the CTG repeat tract or the surrounding loci.

Gene conversion efficacy is decreased when a Cas9 DSB is made within a long CTG trinucleotide repeat

Gene conversion is a very efficient DSB-repair mechanism in *S. cerevisiae*. We previously showed that a single DSB induced by the I-SceI meganuclease in a yeast chromosome was efficiently repaired using a CTG repeat-containing homologous template as a donor [24,30,31]. In order to determine whether a Cas9-induced DSB within a CTG repeat was properly repaired by the recombination machinery, we reused a similar experimental system in which two copies of the SUP4 allele were present on yeast chromosome X, one containing a (CTG)₆₀ repeat tract and the other copy containing an I-SceI recognition site (Fig 6A). In this ectopic gene conversion assay, 80.2%±2.3% of yeast cells survived after an I-SceI DSB and 100% of survivors were repaired by gene conversion using the ectopic SUP4::CTG copy as a donor [31]. When SpCas9 was induced in the same yeast strain along with gRNA#1, only 32.6%±3.8% of CFU formed on galactose plates (Fig 2). Molecular analysis of surviving cells showed that 89% (34 out of 38) repaired by ectopic gene conversion, as expected, and now contain two I-SceI recognition sites, one in each SUP4 copy (Fig 6B, GC events). However, one expansion event was also detected, as well as one local deletion (Type IV) and two events involving a deletion and a DNA insertion (Type V). Intriguingly, the DNA insertion was a 211 bp piece of DNA from the YAK1 gene, located 158 kilobases upstream the ARG2 locus, on chromosome X left arm. This gene contains a long and imperfect CAG/CTG repeat within its reading frame, like many yeast genes [32–35]. An unusual recombination event occurred between the YAK1 CAG/CTG repeat and the ARG2 repeat, leading to a chimeric repeat (Fig 6C). This rearrangement may be

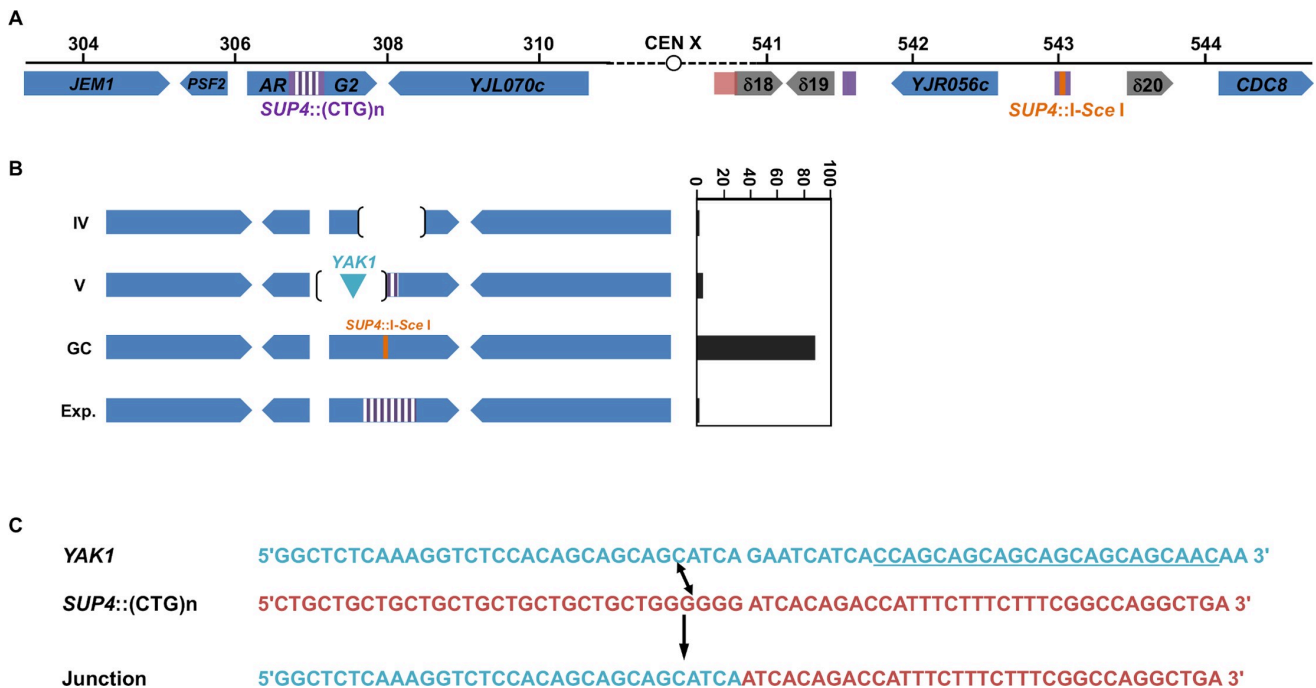


Fig 6. Chromosomal rearrangements observed at the ARG2 locus, following SpCas9 induction. A: ARG2 and SUP4 loci drawn to scale. A 2.6 kb piece of DNA containing 1.8 kb of the SUP4 locus in which a CTG repeat was integrated, as well as the TRP1 selection marker were integrated at ARG2 [31]. The TRP1 gene is not represented here but is centromere-proximal located. B: Types of rearrangements observed. Types IV and V deletions are explained in Fig 5. GC: gene conversion with SUP4::I-SceI. Exp.: CTG repeat expansion. C: Type V rearrangements involving the YAK1 gene. The off-target in YAK1 identified by CRISPOR is underlined. The blue sequence from YAK1 recombined with the red sequence from SUP4 to give a hybrid molecule called "Junction".

<https://doi.org/10.1371/journal.pgen.1008924.g006>

the result of an off-target DSB generated by *SpCas9* within the *YAK1* repeat, or an abnormal recombination event between the two CTG repeats following *SpCas9* induction. Using the CRISPOR *in silico* tools, off-target sites were examined for guide RNA #1 [36]. Two off-targets with zero mismatches were predicted, in the *NGR1* and *SGF73* genes. However, since they are using a non-canonical PAM (TGA), the Cutting Frequency Determination score (CFD score, ref. [37]) was very low at these two sites, suggesting that they would be poor substrate for *SpCas9* with this gRNA (S1 Table). When one mismatch was allowed, two hits were found in the *YAK1* gene, one of them with a canonical NGG PAM exhibiting the best CFD score of all predicted off-target sites (S1 Table). We therefore decided to check whether an off-target DSB could be induced within the *YAK1* gene by *SpCas9*. A time course was performed in conditions in which the nuclease was non-induced or induced, and the resulting Southern blot was hybridized with a *YAK1* specific probe. No evidence for a band that could correspond to a DSB at this locus could be seen. Faint signals were detected, both in non-induced and induced conditions, at molecular weights that did not fit the expected DSB size (S2 Fig). We concluded that, if an off-target DSB was made by *SpCas9* at the *YAK1* locus, it was too rare to be detected by Southern blot, and presumably could not influence cell survival, nor be sufficient to trigger frequent ectopic recombination events with the *SUP4* locus.

Resection of a Cas9-induced double-strand break

Quantitative PCR experiments were performed in order to determine the resection level in strains in which Cas9 was induced. The nuclease generates a DSB in the very last CTG triplets of the repeat tract (Fig 1A). Therefore, the 5' end of the break contains most of the 80 triplets whereas the 3' end contains only two triplets. This asymmetry allows to compare resection of a repeated and structured DNA end versus non-repeated DNA, concomitantly and in the same experimental setting. We took advantage of the convenient position of four *EcoRV* restriction sites, two on each side of the DSB, at different distances from the break (Fig 7A). Primers were designed in such a way that *EcoRV* digested DNA could not be PCR amplified. However, if DNA resection reached an *EcoRV* site, the resulting single-stranded DNA became resistant to digestion and therefore susceptible to amplification. In wild-type cells after eight hours, resection of the Cas9 DSB was always 100% at all *EcoRV* sites, except at the 3' distal site in which it was a little lower, around 70% (Fig 7A). In *dnl4Δ* cells, resection was not statistically different from wild type. In the *rad50Δ* mutant, DSB resection was totally abolished on the 5' end of the break that contains most of the repeat tract and severely impaired on the other end, showing that the MRX-Sae2 complex was essential on both DSB ends. Interestingly, the *sae2Δ* mutant exhibited a resection defect on the 5' end of the break but not on the other side. This was also true for the double mutant *sae2Δ dnl4Δ*. All these data prove that: i) Ligase IV plays no role in DSB resection; ii) Sae2 is essential to resect a long CTG trinucleotide repeat but is dispensable to resect a non-repeated DSB end.

Genome-wide mutation spectrum in cells expressing *SpCas9*

When carefully looking at deletion borders in haploid strains in which *SpCas9* was induced, they were found to be more extensive on the 5' side of the break than on the 3' side (Fig 4). This suggests that larger 3' deletions encompassing the essential gene *CDC8* or its promoter may not have been recovered because they would be lethal, probably counting for some of the lethality observed. In order to check this hypothesis, we expressed *SpCas9* in diploids containing *SUP4::(CTG)_n* repeat tracts on both homologues. In these cells, both chromosomes could be cut by the nuclease. We quantified by qPCR *CDC8* copy number in six independent diploid survivors. In all cases, it was reduced by half as compared to a control qPCR on another

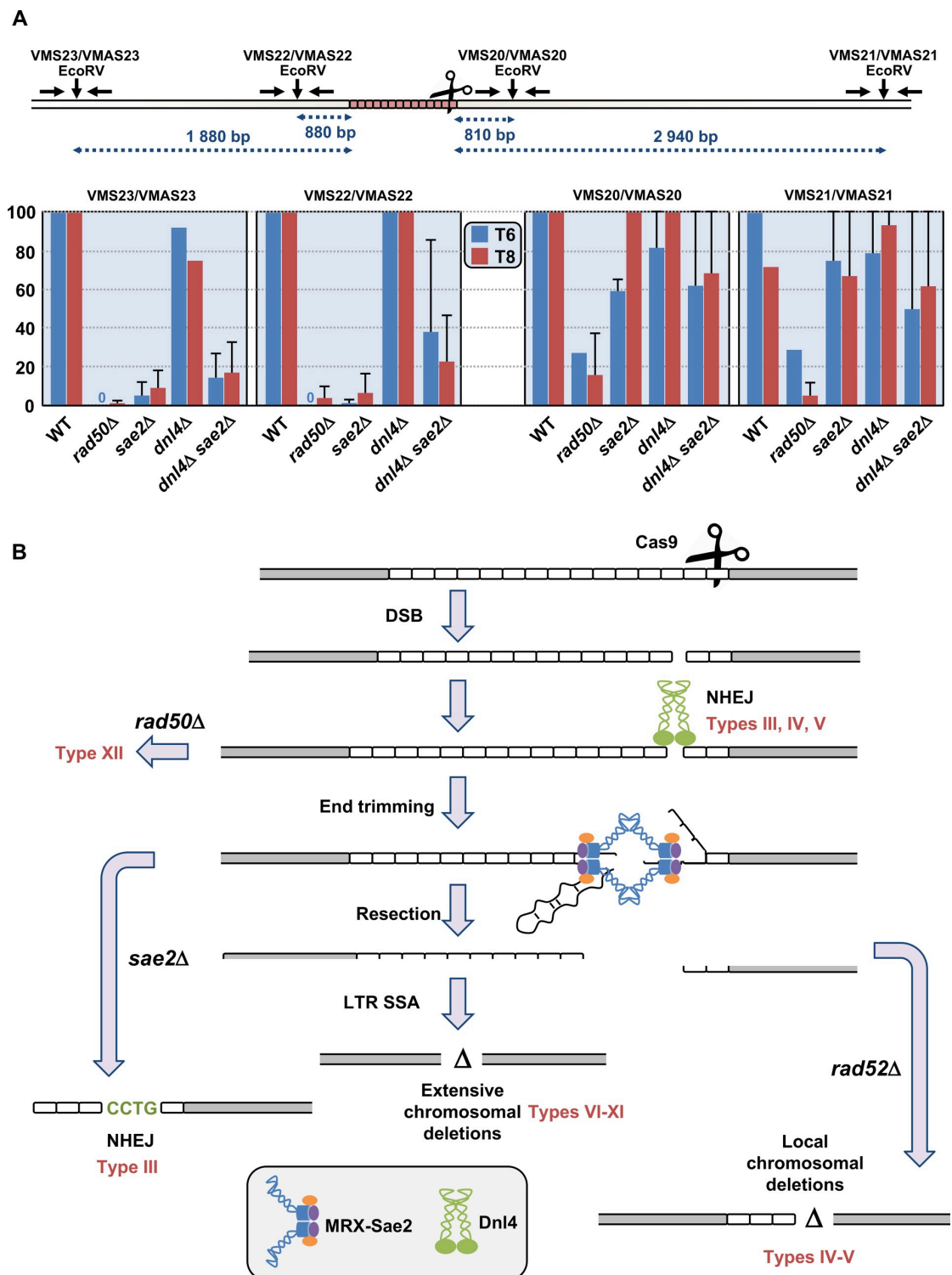


Fig 7. Quantification of double-strand break resection. A: Couples of primers used to amplify each EcoRV site are indicated above and can be found in S3 Table. EcoRV sites are shown by vertical arrows. Resection graphs are plotted for each primer pair. Average relative values of

resection as compared to the total DSB amount detected on Southern blots are shown at 6 hours (in blue) and 8 hours (in red), along with standard deviations. **B:** Mechanistic model for chromosomal deletions following a Cas9-induced DSB. See text for details. Resulting deletion types are indicated in red near each pathway.

<https://doi.org/10.1371/journal.pgen.1008924.g007>

chromosome (S3A Fig). This showed that in these cells, only one of the two *CDC8* alleles was present, suggesting that the other was often deleted during DSB repair. In order to check if the whole chromosome could have been lost, we also amplified a region near the *JEM1* gene, on the other chromosomal arm, near the *ARG2* gene. Surviving clones showed a significantly higher signal, compatible with the presence of two chromosomes (Mann-Whitney-Wilcoxon rank test, p -value = 10^{-3}). Therefore, it was concluded that the mortality observed in haploid cells expressing *SpCas9* was at least partly due to the frequent deletion of the essential *CDC8* gene, but did not induce significant chromosome loss.

In order to detect possible off-target mutations, independent haploid and diploid colonies in which *SpCas9* had been induced were deep-sequenced, using Illumina paired-end technology. In diploid cells, two nucleotide substitutions were detected out of five independent clones when *SpCas9* was repressed (S3B Fig). When the nuclease was induced, six mutations were detected out of 18 sequenced clones, a similar proportion. One 36-bp deletion was found in the *FLO11* minisatellite and two deletions of one repeat unit were found in AT dinucleotide repeats, but no mutation was found in any other CAG/CTG repeat tract. Altogether, we concluded that *SpCas9* expression in diploid cells did not significantly increase genome-wide mutation frequency. The genome of 10 independent haploid cells in which *SpCas9* was induced was also completely sequenced. Eight mutations were detected among six of these survivors, all of them being nucleotide substitutions in non-repeated DNA (S3B Fig). This is statistically not different from what was observed in diploids (Fisher exact test p -value = 0.12). We concluded that besides chromosomal deletions around the *SUP4* locus observed in these haploids, *SpCas9* did not induce other mutations in yeast cells, to a level detectable with the present sequencing approach.

Discussion

SpCas9-induced DSB repair within CTG repeats generates chromosomal deletions

In previous work, in which we induced a DSB within a CTG repeat using a dedicated TALEN, 100% of surviving yeast colonies repaired the break by contracting the repeat tract (S4 Fig and ref. 21). These contractions occurred by single-strand annealing, and depended on *RAD52*, *RAD50* and *SAE2*, but was independent of *LIG4*, *POL32* and *RAD51* [20]. It was therefore striking and completely unexpected that a DSB made by the *SpCas9* nuclease (or by its more specific mutant version, *eSpCas9*) at exactly the same location within the very same CTG trinucleotide repeat induced frequent chromosomal deletions around the repeat tract and almost no repeat contraction.

With the TALEN, repeat contraction was proposed to be an iterative phenomenon, involving several rounds of cutting and contraction until the repeat tract was too short for the two TALEN arms to dimerize and induce a DSB [20,21]. A similar outcome was expected with *SpCas9*, iterative rounds of cutting and contraction could occur until the remaining CTG repeat tract would be too short for the gRNA to bind and induce a DSB. But this was not observed here, surprisingly proving that a *SpCas9*-induced DSB was differently repaired from a TALEN-induced DSB targeting the same exact repeated sequence. Previous terminal-transferase mediated PCR of four TALEN-induced DSB showed that the break left 2–11 bp of

homology, probably due to variable positioning of the left TALEN arm on the repeated sequence [20]. In comparison, Cas9 should not have any binding flexibility and should leave 5 bp of homology each time it cuts. We therefore think this is unlikely to explain the difference between the two nucleases, since even though the TALEN sometimes leave very little homology, no chromosomal rearrangement was observed. Reasons for this discrepancy may include different DSB ends (4 nucleotides 5' overhangs with the TALEN, blunt ends with Cas9), different substrate-enzyme kinetics of the TALEN as compared to Cas9, a role for the guide RNA in maintaining both ends together after cutting, or differences in checkpoint activation, these hypotheses being neither exhaustive nor mutually exclusive. These possibilities are now being investigated. Note that the use of PolQ-mediated micro homologies for end joining is very efficient in human cells and similar strategies might be more effective for contracting repeats in patient cells.

Spontaneous homologous recombination events between delta elements surrounding *SUP4* were already described by the past by Rothstein and colleagues [38]. Recombination between $\delta 18/ \delta 20$ or $\delta 19/ \delta 20$ was less frequent than between $\delta 16/ \delta 20$ or $\delta 17/ \delta 19$. This was the opposite in our experiments. This is probably due to the way recombination was triggered. In the Rothstein *et al.* article, recombination was spontaneous, therefore favoring LTRs with the highest sequence identity ($\delta 16/ \delta 20$ or $\delta 17/ \delta 19$). In our case, the initiating event was a DSB always at the same position, between $\delta 19$ and $\delta 20$. Resection occurred until the closest regions of homology were revealed, hence favoring $\delta 18/ \delta 20$ or $\delta 19/ \delta 20$ over recombination with the more distant $\delta 16$ or $\delta 17$. In addition, in the present case, Cas9-induced deletions also involved microhomology sequences or no homology at all, suggesting that repair depended on the initiating damage (replication-induced single-strand nicks vs. nuclease-induced double-strand breaks). Our data are also reminiscent of a previous work in which spontaneous deletions around the *URA2* gene were classified in seven different classes, six of them harboring microhomologies at their junctions and one showing no obvious homology [39].

In recent work, using a GFP reporter system in human cells, Cinesi and colleagues showed that *SpCas9* induced contractions as well as expansions of CTG trinucleotide repeats, whereas the nickase mutant *SpCas9*-D10A only induced contractions. By small pool-PCR analysis, a four-fold increase in the rate of CTG repeat tract contractions was observed when the nickase was expressed with a CTG-carrying gRNA, as compared to the expression of the nuclease alone ([40], Fig 2C). The sp-PCR experiment was not performed when the double-strand endonuclease *SpCas9* was expressed, so it is not possible to determine whether a similar increase in contractions would be observed. When individual clones were sequenced, local deletions of the triplet repeat tract were found in three clones out of 17 (17.6%) when a CAG-carrying gRNA was used, and two clones out of 11 (18.2%) when a CTG-carrying gRNA was expressed ([40], S3 Fig). These numbers were not statistically different from what was observed in the absence of the *SpCas9*-D10A nickase ([40], S1 Fig). There is no report of repeat tract sequencing when *SpCas9* was expressed, so it is not possible to know if the double-strand endonuclease would make more (or less) deletions in this experimental system. In our present experiments, eight clones out of 51 (15.7%) analyzed contained a local deletion (Types IV events, Fig 4) similar to those described by Cinesi and colleagues. However, it must be noted that in their work, only the top 1% brightest GFP-positive cells were sequenced, whereas no selection was used in our experiments; surviving colonies were randomly picked and analyzed at the molecular level. Therefore, the absolute frequency of local deletions cannot be reliably compared between the two experimental setups.

It must be noted that an approach using the *SpCas9*-D10A nickase was successfully implemented to delete the CAG trinucleotide repeat involved in Huntington disease, by making two single-strand nicks, upstream and downstream the repeat tract [41].

In another work looking at the effect of a Cas9-induced DSB at the *LYS2* locus in *S. cerevisiae*, the authors found frequent *POLA*-dependent small insertions (1–3 bp) in 42–68% of the survivors (depending on the PAM used) and local deletions (1–17 bp) in the remaining cases. However, given that there is no transposon or transposon remnant in the close proximity of the *LYS2* locus, the authors could not retrieve LTR deletions [42]. This strongly suggests that deletions observed heavily depend on the surrounding chromosomal location where the DSB is made.

CTG trinucleotide repeats interfere with SpCas9-triggered gene conversion

When an I-SceI DSB was induced within a *SUP4* allele, the break could be repaired by gene conversion with a CTG repeat-containing homologous donor at the *ARG2* locus. All yeast cells repaired by gene conversion with the donor, generating repeat contractions and expansions in the process [31]. Here, the exact reverse reaction was induced, the break was made within CTG repeats and repaired with a non-repeated sequence. DSB repair was much less efficient, since only 32.6% of the cells survived (Fig 2) and less specific since 10% of the repair events were unfaithful recombination (Fig 6B). This shows that when a Cas9 DSB was made into a CTG repeat, gene conversion was partially impaired, either by the repeat tract or by the Cas9 protein, or by both. Note that, in past experiments, when the DSB was made in the I-SceI recognition site, less than 10 nucleotides needed to be resected on each side of the break before homology with the other *SUP4* allele could be reached. In our present experiments, when the DSB was made within the CTG repeat tract, 64 nucleotides needed to be resected on one side and 304 on the other side of the break before homology with the other *SUP4* allele could be reached. This could also explain the better survival observed in the former case.

Ligase IV and Sae2 are respectively driving local and extensive chromosomal deletions

Yeast Ligase IV is encoded by the *DNL4* gene and is the enzyme used to ligate DSB ends during non homologous end-joining [43]. It was previously shown that *RAD50* and *SAE2* were essential to resect and process a TALEN-induced DSB but a *DNL4* deletion had no effect on break processing, cell survival or repair efficacy [20]. On the contrary, repair of a Cas9, an HO or an I-SceI DSB at the *MAT* locus, in the absence of any homologous donor cassette, was shown to be dependent on the product of the *DNL4* gene [42,44]. SpCas9 DSB repair has also been studied in human cells in the presence of a drug (NU7441), acting as a chemical inhibitor of non-homologous end-joining. In these conditions, the frequency of single-base insertions and small deletions decreased whereas larger deletions increased, suggesting that these repair events occurred by an alternative end-joining mechanism (alt-EJ/MMEJ) involving micro-homologies flanking the DSB [45,46]. Here, we showed that when *DNL4* was inactivated, local deletions were totally lost. However, survival was not significantly decreased because yeast cells could repair the DSB using LTR recombination, generating extensive deletions around the repeat tract (Fig 4). Supporting this model, the absence of any resection defect in the *dnl4Δ* mutant demonstrated that in the absence of end-joining, resection may take place very efficiently to repair the DSB by homologous recombination, using flanking homologies.

SAE2 is associated to the *MRE11-RAD50-XRS2* complex, whose roles are multiple during DSB repair [47] and it was proposed to encode an endonuclease activity essential to process DNA hairpins [48], as well as to resect I-SceI double-strand breaks [49]. We previously showed that it was essential to resect a TALEN-induced DSB end containing a long CTG trinucleotide repeat, but less important to resect the non-repeated end [20]. In the present experiments, extensive deletions involving LTR elements were lost in a *sae2Δ* mutant, and 97% of yeast cells

repaired the DSB by local deletions, most of them resulting in insertions or deletions between the PAM and the gRNA sequence (Fig 4), inactivating *SpCas9* capacity to induce another DSB. Small insertions of a few nucleotides were also frequently detected following *SpCas9* DSB induction at the VDJ locus in human B cells [50] or at the *MAT* locus in *S. cerevisiae* [42]. However, in our experiments, all nucleotides inserted were C, T or G, all three encoded by the gRNA. No insertion of an adenosine residue was found out of 28 insertions sequenced (S1 Fig). This intriguing observation suggests the possibility that the gRNA could be used as a template to repair the DSB, as it was demonstrated that a single-stranded RNA could be used to repair an HO-induced DSB into the *LEU2* gene [51].

Although *DNL4* and *SAE2* trigger different types of chromosomal deletions and none of the single mutants significantly decreased survival, the *dnl4Δ sae2Δ* double mutant abolished repair, like the *rad50Δ* mutant, since only 0.6% of the cells survived (Fig 2), showing the synthetic effect of both mutations. However, repair events in the double mutant were similar to those observed in *dnl4Δ* (Fig 4). It is possible that other nucleases, like *EXO1* or *DNA2*, could perform long range resection in the absence of *SAE2* [49,52], allowing the occurrence of extensive deletions (Types VI-XI).

In order to exclude the possibility that differences between wild type and mutants could be due to lower or higher expression of Cas9 in different backgrounds, the amount of Cas9 protein in each strain as compared to glucose-6-phosphate dehydrogenase (G6PDH, the product of the *ZWF1* gene in budding yeast) was determined by Western blots. G6PDH is one of the most abundant proteins in budding yeast, with ca. 14,000 molecules per cell (*Saccharomyces Genome Database*, <https://www.yeastgenome.org/locus/S000005185>). Cas9 level showed a small 2-fold decrease in *dnl4Δ*, *sae2Δ* and *dnl4Δ sae2Δ* strains, as compared to wild type, but no difference was observed in the *rad50Δ* mutant (S5 Fig). We concluded that the amount of *SpCas9* in all strains was comparable to G6PDH. Therefore, phenotypic differences observed in mutants could not be due to a much higher or much lower expression of *SpCas9* in these strains, as compared to wild type.

All these results are compatible with a model in which a *SpCas9* DSB was tentatively repaired by NHEJ first (Fig 7B). In the absence of *RAD50*, the MRX-Sae2 complex could not assemble, resection could not occur, NHEJ was impossible and DSB ends were lost, triggering high mortality. If *DNL4* was inactivated, resection proceeded normally and when reaching flanking LTRs, repair occurred by *RAD52*-mediated SSA. In the absence of this gene, the break was repaired by *RAD52*-independent local deletions. When *SAE2* was inactivated, resection was impeded on the 5' DSB end, leading to resection defects observed as smears on Southern blots. Mutagenic NHEJ was favored, leading to local insertions and deletions. It is unknown whether Sae2 would play the same essential role on other secondary structure-forming trinucleotide repeats, like GAA or CGG triplets, or if its activity is specific to CTG triplets, hence of a structure rather than a repeat, but this important question is currently under investigation.

Materials and methods

Yeast strains and plasmids

All mutant strains were built from strain GFY6162-3D by classical gene replacement method [53], using *KANMX4* or *HIS3* as marker (S2 Table). *KANMX4* cassettes were amplified from the EUROSCARF deletion library, using primers located 1kb upstream and downstream the cassette. VMS1/VMAS1 were used to amplify *rad52Δ::KANMX*, VMS2/VMAS2 were used to amplify *rad51Δ::KANMX*, VMS3/VMAS3 were used to amplify *pol32Δ::KANMX*, VMS4/VMAS4 were used to amplify *dnl4Δ::KANMX*, VMS6/VMAS6 were used to amplify *rad50Δ::KANMX* and *SAE2up/SAE2down* were used to amplify *sae2Δ::KANMX* (S3 Table). VMY350

and VMY352 strains were respectively used to construct VMY650 and VMY352 by mating-type switching, as follows: the pJH132 vector [54] carrying the HO endonuclease under the control of an inducible *GALI-10* promoter was transformed in the haploid *MAT α* strains. After 5h of growth in lactate medium, HO expression was induced by addition of 2% galactose (final concentration) and grown for 1.5 hour. Cells were then plated on YPD and mating type was checked three days later by crosses with both *MAT α* and *MAT α* tester strains.

For *SpCas9* inductions, addgene plasmid #43804 containing the nuclease under the control of the GalL promoter and the *LEU2* selection marker was digested with HpaI and cloned into yeast by homology-driven recombination [55] with a single PCR amplified fragment containing the *SNR52* promoter, the gRNA#1 and the *SUP4* terminator, using primers SNR52Left and SNR52Right (S3 Table) to give plasmid pTRi203. A frameshift was then introduced in this plasmid by NdeI digestion followed by T4 DNA polymerase treatment and religation of the plasmid on itself, to give plasmid pTRi206. In this plasmid, the *SpCas9* gene is interrupted by a stop codon after amino acid Ile₁₆₁. The haploid GFY6162-3D strain (or its mutant derivatives), was subsequently transformed with pTRi203 or pTRi206 and transformants were selected on SC-Leu. The plasmid containing Enhanced *SpCas9* (version 1.1, Addgene #71814, Slaymaker et al., 2016) was a generous gift of Carine Giovannangeli from the *Museum National d'Histoire Naturelle*. The e*SpCas9* gene was amplified using primers LP400 and LP401 (S3 Table) and cloned into yeast cells in the Addgene#43804 plasmid digested with BamHI, by homology-driven recombination, with 34-bp homology on one side and 40-bp homology on the other side [55], to give plasmid pLPX11. For the gRNA#1, plasmid pLPX11 was digested with HpaI and cloned into yeast by homology-driven recombination [55] with a single PCR amplified fragment containing the *SNR52* promoter, the gRNA#1 and the *SUP4* terminator, as above to give plasmid pTRi207. For the gRNA#2, a guide RNA cassette was ordered from ThermoFisher (GeneArt), flanked by *EcoRI* sites and was cloned in pRS416 [56] using standard procedures to give plasmid pLPX210.

In silico simulations of off-target sites

To assess the number of off-target sites for *SpCas9* in *Saccharomyces cerevisiae*, online tools were used. CRISPOR is a software that evaluates the specificity of a guide RNA through an alignment algorithm that maps sequences to a reference genome to identify putative on- and off-target sites [57]. To predict off-target sites, the online tool sequentially introduces changes in the sequence of the gRNA and checks for homologies in the specified genome [36]. The Cutting Frequency Determination score (CFD score) relies on several criteria to assess off-target probabilities: nucleotide deletions, insertions, mismatches, as well as the position and identity of the mismatch(es) [37].

Cas9 inductions

Before nuclease induction, Southern blot analyses were conducted on several independent subclones to select one containing ca. 80 CTG triplets. For Cas9 inductions, yeast cells were grown overnight at 30°C in liquid SC-Leu medium, then washed with sterile water to remove any trace of glucose. Cells were split in two cultures, half of the cells were grown in synthetic -Leu medium supplemented with 2% galactose (final concentration) and the other half were grown in synthetic -Leu medium supplemented with 2% glucose (final concentration). Around 4×10^8 cells were collected at different time points ($T = 0, 4, 5, 6, 7$ and 8 hours) and killed by addition of sodium azide (0.01% final). Cells were washed with water, and frozen in dry ice before DNA extraction. To determine survival to Cas9 induction, 24 hours after the T0 time point, cells were diluted to an appropriate concentration, then plated on SC-Leu plates

containing either 20 g/l glucose or galactose. After 3–5 days of growth at 30°C, ratio of CFU on galactose plates over CFU on glucose plates was considered to be the survival rate.

Double-strand break analysis and quantification

Total genomic DNA (4 µg) of cells collected at each time point was digested for 6h with EcoRV (40 U) (NEB) loaded on a 1% agarose gel (15x20 cm) and run overnight at 1 V/cm. The gel was vacuum transferred in alkaline conditions to a Hybond-XL nylon membrane (GE Healthcare) and hybridized with two randomly-labeled probes specific of each side of the repeat tract, upstream and downstream the *SUP4* gene [58]. After washing, the membrane was overnight exposed to a phosphor screen and signals were read and quantified on a FujiFilm FLA-9000. For the *YAK1* Southern blot, DNA was digested with BamHI and the probe was a 719 bp probe amplified with YAK1f and YAK1r primers (S3 Table).

SUP4 locus analysis after Cas9 induction

Several colonies from each induced or repressed plates were picked, total genomic DNA (4 µg) was extracted with Zymolyase, digested for 6h by SspI (20 U) (NEB), loaded on a 1% agarose gel (15x20 cm) and run overnight at 1V/cm. The gel was vacuum transferred in alkaline conditions to a Hybond-XL nylon membrane (GE Healthcare) and hybridized with a randomly-labeled PCR fragment specific of a region downstream the *SUP4* gene, amplified from the su8-su9 primer couple (S3 Table). After washing, the membrane was overnight exposed on a phosphor screen and signals were revealed on a FujiFilm FLA-9000. Genomic DNA of each clone for which a signal was detected by Southern blot was subsequently amplified with su47-su48 primers and sequenced using su47 (S3 Table). Genomic DNA of clones for which no signal was detected by Southern blot of no PCR product was obtained with su47-su48 were subsequently amplified with su23-su42 primers and sequenced using su42 (S3 Table). Sanger sequencing was performed by GATC biotech.

Western blots

Liquid cultures of each strain were grown to exponential phase in synthetic glucose medium or synthetic galactose medium without leucine, on order to maintain the Cas9 plasmid. Proteins were extracted on 2×10^8 cells in 200 µl Laemmli solution with 100 µl glass beads. Proteins were separated on a 10% acrylamide gel in standard conditions and blotted to a nitrocellulose membrane (Optitran BA-S 83 reinforced NC, Schleicher & Schuell). For Cas9 detection, a monoclonal HRP-conjugated mouse antibody was used (Abcam [7A9-3A3], ab202580, 1/1000 dilution). Note that blocking was achieved in 10 mM Tris-HCl pH8.0, NaCl 150 mM, 0.05% Tween 20, 3% dry milk (instead of the regular 5%). For G6PDH detection, the primary antibody was a polyclonal rabbit antibody (Sigma-Aldrich (A9521), 1/100 000 dilution). A secondary goat anti-rabbit antibody conjugated to horseradish peroxidase was used for detection of G6PDH (Thermo Scientific, 0.16 µg/ml final concentration). Quantification was performed using a ChemiDoc MP Imager (Bio-Rad) with the dedicated Image Lab software. The molecular weight marker used was the Precision Plus Protein Kaleidoscope marker (Bio-Rad).

Analysis of Cas9-induced DSB end resection by qPCR

A real-time PCR assay, using primer pairs flanking EcoRV sites 0.81 kb and 2.94 kb away from the 3' end of the CTG repeat tract (VMS20/VMAS20 and VMS21/VMAS21 respectively) and 0.88 kb and 1.88 kb away from the 5' end of the CTG repeat tract (VMS22/VMAS22 and VMS23/VMAS23 respectively), was used to quantify end resection. Another pair of primers

was used to amplify a region of chromosome X near the *ARG2* gene [11], to serve as an internal control of DNA amount (JEM1f-JEM1r). Genomic DNA of cells collected at T = 0h, T = 6h and T = 8h was split in two fractions, incubated at 80°C for 10 minutes in order to inactivate any remaining active DNA nuclease, then one fraction was used for EcoRV digestion and the other one for a mock digestion in a final volume of 15 µl. Samples were incubated for 5h at 37°C, then the enzyme was inactivated for 20 min at 80°C. DNA was subsequently diluted by adding 55 µl of ice-cold water, and 4 µl was used for each real-time PCR reaction in a final volume of 25 µl. PCRs were performed with the Absolute SYBR Green Fluorescein mix (Thermo Scientific) in the Mastercycler S realplex (Eppendorf), using the following program: 95°C 15min, 95°C 15sec, 55°C 30 sec, 72°C 30 sec repeated 40 times, followed by a 20 min melting curve. Reactions were performed in triplicates and the mean value was used to determine the amount of resected DNA, using the following formula: raw resection = $2/(1+2^{\Delta Ct})$ with $\Delta Ct = C_{t,EcoRV} - C_{t,mock}$. Relative resection values were calculated by dividing raw resection values by the percentage of DSB quantified at the corresponding time point.

The same protocol was used to determine the relative amount of *CDC8* and chromosome X in surviving clones after Cas9 induction, except that total genomic DNA was not digested prior to real-time PCR. Primer couples VMS23-VMAS23 were used to amplify *CDC8* and JEM1f-JEM1r for chromosome X left arm. Primers Chromo4_f and Chromo4_r were used to amplify a region of chromosome IV as an internal control for total DNA amount. See S3 Table for all primer sequences.

Library preparation for deep-sequencing

Approximately 10 µg of total genomic DNA was extracted and sonicated to an average size of 500 bp, on a Covaris S220 (LGC Genomics) in microtubes AFA (6x16 mm) using the following setup: Peak Incident Power: 105 Watts, Duty Factor: 5%, 200 cycles, 80 seconds. DNA ends were subsequently repaired with T4 DNA polymerase (15 units, NEBiolabs) and Klenow DNA polymerase (5 units, NEBiolabs) and phosphorylated with T4 DNA kinase (50 units, NEBiolabs). Repaired DNA was purified on two MinElute columns (Qiagen) and eluted in 16 µl (32 µl final for each library). Addition of a 3' dATP was performed with Klenow DNA polymerase (exo-) (15 units, NEBiolabs). Home-made adapters containing a 4-bp unique tag used for multiplexing, were ligated with 2 µl T4 DNA ligase (NEBiolabs, 400,000 units/ml). DNA was size fractionated on 1% agarose gels and 500–750 bp DNA fragments were gel extracted with the Qiaquick gel extraction kit (Qiagen). DNA was PCR amplified for 12 cycles with Illumina primers PE1.0 and PE2.0 and Phusion DNA polymerase (1 unit, Thermo Scientific). Six PCR reactions were pooled for each library, and purified on a Qiagen purification column. Elution was performed in 30 µl and DNA was quantified on a spectrophotometer and on agarose gel.

Analysis of paired-end Illumina reads

Multiplexed libraries were loaded on a HiSeq2500 (Illumina), 110 bp paired-end reads for haploids and 260 bp paired-end reads for diploids were generated. Reads quality was evaluated by FastQC v.0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were mapped along S288C chromosome reference sequence (*Saccharomyces Genome Database*, release R64-2-1, November 2014), using the paired-end mapping mode of BWA v0.7.4-r385 with default parameters [57]. The output SAM files were converted and sorted to BAM files using SAMtools v0.1.19-44428cd [59]. The command *IndelRealigner* from GATK v2.4–9 [60] was used to realign the reads. Duplicated reads were removed using the option “*MarkDuplicates*” implemented in Picard v1.94 (<http://picard.sourceforge.net/>). Reads uniquely mapped

to the reference sequence with a minimum mapping quality of 20 (Phred-scaled) were kept. Mpileup files were generated by SAMtools without BAQ adjustments. SNPs and INDELS were called by the options “*mpileup2snp*” and “*mpileup2indel*” of *Varscan2* v2.3.6 [61] with a minimum depth of 10 reads for haploids and 20 reads for diploids. Average read coverage was 255X for diploid cells ($\sigma = 187X$) and 190X for haploids ($\sigma = 43X$). Diploid strains are homozygous except for selection markers and some specific loci like *MAT* and *SUP4*. Therefore, *de novo* heterozygous mutations should represent 50% of reads, on the average. Taking that into account, lower and upper thresholds for variant allele frequency were respectively set between 30% and 70% in diploids. For haploids, the threshold for minimum variant allele frequency was set at 70%. Mutations less than 10 bp away from each other were discarded to avoid mapping problems due to paralogous genes or repeated sequences. To assess microsatellite mutations, we only retained reads uniquely anchored at least 20 bp on each side of the microsatellite [62]. All detected mutations were manually examined using the IGV software (version 2.3.77), and compared between all sequenced libraries for interpretation. All the scripts used in order to process data are available on github (<https://github.com/sdeclere/nuclease>). All Illumina sequences were uploaded in the European Nucleotide Archive (ENA), accession number PRJEB16068.

Statistical analyses

All analyses were performed using the R package (version 3.6.3) [63]. Survival rates after DSB induction were compared using the t-test (Fig 2). When comparing rates after DSB induction at *SUP4* in wild type vs mutant strains, the Bonferroni correction for multiple testing was applied. Deletions and rearrangement types between different strains were compared using the Chi2 test.

Supporting information

S1 Fig. Sequences at the left and right of junctions in rearranged haploid clones. Junctions were deduced from Illumina read mapping (when available) and confirmed by subsequent PCR and Sanger sequencing. Nucleotides in red are those used to anneal each DSB end, and are therefore present in only one copy in the genomic sequence. The extent of calculated deletions (Δ) is indicated in parentheses. Nucleotides in red in parentheses correspond to small deletions. Nucleotides in green correspond to insertions. The length of Ty insertions is indicated along with the LTR it comes from. Nucleotides in purple (*SpCas9* at the *ARG2* locus) correspond to the *I-SceI* site (see text). Nucleotides in light blue correspond to homeologies between the left and right junction sequences that were lost after deletion (the junction sequence shows the nucleotide in blue, not the one in red). Nucleotides in light blue in parentheses correspond to homeologies that were removed during the deletion. Note that extended homologies between LTRs does not always allow to determine the exact breakpoint with a high precision.

(PDF)

S2 Fig. Southern blot at the *YAK1* locus. The time course was run in non-induced (glucose) and induced (galactose) conditions, as previously. The uncut locus is clearly visible as a 2612 bp band, but no signal can be seen at the expected size for a DSB (2294 bp). Two fuzzy bands present in both conditions and corresponding to faint cross-hybridizations are indicated by asterisks. Note that even when the blot was overexposed no signal could be detected at the expected DSB size.

(PDF)

S3 Fig. Genome-wide mutation spectrum observed in haploid and diploid cells following Cas9 induction. **A:** Real-time PCR quantification of *CDC8* and *JEM1* amounts relative to an internal control on chromosome IV, in diploid cells in which Cas9 was induced. Half the amount of *CDC8* product was detected in each clone analyzed. This was significantly different from the amount of product amplified from the *JEM1* gene located on the other chromosome X arm. **B:** Illumina results for diploid and haploid cells. For each clone, the number of mutations detected is shown. Substit.: nucleotide substitution; Indel: insertion or deletion; Indel micro.: insertion or deletion of one repeat unit in a microsatellite. The asterisk corresponds to a 36 bp deletion in the *FLO11* minisatellite (36 bp repeat).
(PDF)

S4 Fig. Fate of a DSB made within a trinucleotide repeat tract using different endonucleases. The DSB induced by I-SceI (**A**), a TALEN (**B**) or SpCas9 (**C**). In each case, survivors were separated in three different categories: deletions (local or large) around the repeat tract, removing partially or totally the repeat (left), repeat contraction without other mutation (right) or all other kinds of rearrangements (middle). Note that in A, five triplets flank the repeat tract on each side, whereas in B and C the break is made at the end of a long (80 triplets) repeat tract, leaving 1–4 triplets downstream and the remaining upstream the DSB.
(PDF)

S5 Fig. Western blots. **A:** Western blots. Proteins were extracted in non-induced (Glu) and induced (Gal) conditions for wild type and each mutant strain. Glucose-6-phosphate deshydrogenase (G6PDH) was used as a loading control. **B:** Ratios of Cas9 over G6PDH signals, for each strain in each condition.
(PDF)

S1 Table. Off-target sites in the yeast genome, ranked by decreasing CFD score, using the CRISPOR tool.
(XLS)

S2 Table. List of strains used in the present study.
(DOCX)

S3 Table. List and sequence of primers used in this study.
(DOCX)

Acknowledgments

We thank Carine Giovannangeli for the generous gift of the Enhanced SpCas9 plasmid.

Author Contributions

Conceptualization: Guy-Franck Richard.

Formal analysis: Valentine Mosbach, David Viterbo, Lucie Poggi, Wilhelm Vaysse-Zinkhöfer.

Funding acquisition: Guy-Franck Richard.

Investigation: Valentine Mosbach, David Viterbo, Lucie Poggi, Wilhelm Vaysse-Zinkhöfer.

Methodology: Valentine Mosbach, Guy-Franck Richard.

Project administration: Guy-Franck Richard.

Software: Stéphane Descorps-Declère.

Supervision: Guy-Franck Richard.

Writing – original draft: Guy-Franck Richard.

Writing – review & editing: Valentine Mosbach, Guy-Franck Richard.

References

1. Richard G-F, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev.* 2008; 72: 686–727. <https://doi.org/10.1128/MMBR.00011-08> PMID: 19052325
2. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature.* 2004; 431: 931–945. <https://doi.org/10.1038/nature03001> PMID: 15496913
3. Orr HT, Zoghbi HY. Trinucleotide repeat disorders. *Annu Rev Neurosci.* 2007; 30: 575–621. <https://doi.org/10.1146/annurev.neuro.29.051605.113042> PMID: 17417937
4. Gacy AM, Goellner G, Juranic N, Macura S, McMurray CT. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell.* 1995; 81: 533–540. [https://doi.org/10.1016/0092-8674\(95\)90074-8](https://doi.org/10.1016/0092-8674(95)90074-8) PMID: 7758107
5. Liu G, Chen X, Bissler JJ, Sinden RR, Leffak M. Replication-dependent instability at (CTG)_x (CAG)_n repeat hairpins in human cells. *Nat Chem Biol.* 2010; 6: 652–9. <https://doi.org/10.1038/nchembio.416> PMID: 20676085
6. Axford MM, Wang YH, Nakamori M, Zannis-Hadjopoulos M, Thornton CA, Pearson CE. Detection of Slipped-DNAs at the Trinucleotide Repeats of the Myotonic Dystrophy Type I Disease Locus in Patient Tissues. *PLoS Genet.* 2013; 9: 1–13. <https://doi.org/10.1371/journal.pgen.1003866> PMID: 24367268
7. Anand RP, Shah KA, Niu H, Sung P, Mirkin SM, Freudenreich CH. Overcoming natural replication barriers: differential helicase requirements. *Nucleic Acids Res.* 2012; 40: 1091–105. <https://doi.org/10.1093/nar/gkr836> PMID: 21984413
8. Nguyen JHG, Viterbo D, Anand RP, Verra L, Sloan L, Richard G-F, et al. Differential requirement of Srs2 helicase and Rad51 displacement activities in replication of hairpin-forming CAG/CTG repeats. *Nucleic Acids Res.* 45: 4519–4531. <https://doi.org/10.1093/nar/gkx088> PMID: 28175398
9. Pelletier R, Krasilnikova MM, Samadashwily GM, Lahue R, Mirkin SM. Replication and expansion of trinucleotide repeats in yeast. *Mol Cell Biol.* 2003; 23: 1349–57. <https://doi.org/10.1128/mcb.23.4.1349-1357.2003> PMID: 12556494
10. Samadashwily G, Raca G, Mirkin SM. Trinucleotide repeats affect DNA replication in vivo. *Nat Genet.* 1997; 17: 298–304. <https://doi.org/10.1038/ng1197-298> PMID: 9354793
11. Viterbo D, Michoud G, Mosbach V, Dujon B, Richard G-F. Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. *DNA Repair.* 2016; 42: 94–106. <https://doi.org/10.1016/j.dnarep.2016.03.002> PMID: 27045900
12. Sutherland GR, Baker E, Richards RI. Fragile sites still breaking. *Trends Genet.* 1998; 14: 501–506. [https://doi.org/10.1016/s0168-9525\(98\)01628-x](https://doi.org/10.1016/s0168-9525(98)01628-x) PMID: 9865156
13. Balakumaran BS, Freudenreich CH, Zakian VA. CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in *Saccharomyces cerevisiae*. *Hum Mol Genet.* 2000; 9: 93–100. <https://doi.org/10.1093/hmg/9.1.93> PMID: 10587583
14. Freudenreich CH, Kantrow SM, Zakian VA. Expansion and length-dependent fragility of CTG repeats in yeast. *Science.* 1998; 279: 853–856. <https://doi.org/10.1126/science.279.5352.853> PMID: 9452383
15. Haber JE. *Genome stability.* Summer Scholl. New York: Garland Science; 2014.
16. Fairhead C, Dujon B. Consequences of unique double-stranded breaks in yeast chromosomes: death or homozygosis. *Mol Gen Genet.* 1993; 240: 170–180. <https://doi.org/10.1007/BF00277054> PMID: 8355651
17. Haber JE. In vivo biochemistry: physical monitoring of recombination induced by site-specific endonucleases. *BioEssays.* 1995; 17: 609–620. <https://doi.org/10.1002/bies.950170707> PMID: 7646483
18. Plessis A, Perrin A, Haber JE, Dujon B. Site-specific recombination determined by I-Sce I, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics.* 1992; 130: 451–460. PMID: 1551570
19. Nelms BE, Maser RS, MacKay JF, Lagally MG, Petrini JHJ. In Situ Visualization of DNA Double-Strand Break Repair in Human Fibroblasts. *Science.* 1998; 280: 590–592. <https://doi.org/10.1126/science.280.5363.590> PMID: 9554850

20. Mosbach V, Poggi L, Viterbo D, Charpentier M, Richard G-F. TALEN-induced double-strand break repair of CTG trinucleotide repeats. *Cell Rep*. 2018; 22: 2146–2159. <https://doi.org/10.1016/j.celrep.2018.01.083> PMID: 29466740
21. Richard G-F, Viterbo D, Khanna V, Mosbach V, Castelain L, Dujon B. Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. *PLoS ONE*. 2014; 9: e95611. <https://doi.org/10.1371/journal.pone.0095611> PMID: 24748175
22. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res*. 2013; 41: 4336–43. <https://doi.org/10.1093/nar/gkt135> PMID: 23460208
23. Chen H, Lisby M, Symington LS. RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Mol Cell*. 2013; 50: 589–600. <https://doi.org/10.1016/j.molcel.2013.04.032> PMID: 23706822
24. Richard G-F, Goellner GM, McMurray CT, Haber JE. Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11/RAD50/XRS2 complex. *EMBO J*. 2000; 19: 2381–2390. <https://doi.org/10.1093/emboj/19.10.2381> PMID: 10811629
25. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. 2016; 529: 490–495. <https://doi.org/10.1038/nature16526> PMID: 26735016
26. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2016; 351: 84–88. <https://doi.org/10.1126/science.aad5227> PMID: 26628643
27. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature*. 2017; 550: 407–410. <https://doi.org/10.1038/nature24268> PMID: 28931002
28. Lee JK, Jeong E, Lee J, Jung M, Shin E, Kim Y, et al. Directed evolution of CRISPR–Cas9 to increase its specificity. *Nat Commun*. 2018; 9: 3048. <https://doi.org/10.1038/s41467-018-05477-x> PMID: 30082838
29. Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G, et al. A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. *Nat Biotechnol*. 2018; 36: 265–271. <https://doi.org/10.1038/nbt.4066> PMID: 29431739
30. Richard G-F, Dujon B, Haber JE. Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats. *Mol Gen Genet*. 1999; 261: 871–882. <https://doi.org/10.1007/s004380050031> PMID: 10394925
31. Richard G-F, Cyncynatus C, Dujon B. Contractions and expansions of CAG/CTG trinucleotide repeats occur during ectopic gene conversion in yeast, by a MUS81-independent mechanism. *J Mol Biol*. 2003; 326: 769–782. [https://doi.org/10.1016/s0022-2836\(02\)01405-5](https://doi.org/10.1016/s0022-2836(02)01405-5) PMID: 12581639
32. Field D, Wills C. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proc Natl Acad Sci USA*. 1998; 95: 1647–1652. <https://doi.org/10.1073/pnas.95.4.1647> PMID: 9465070
33. Malpertuy A, Dujon B, Richard G-F. Analysis of microsatellites in 13 hemiascomycetous yeast species: mechanisms involved in genome dynamics. *J Mol Evol*. 2003; 56: 730–741. <https://doi.org/10.1007/s00239-002-2447-5> PMID: 12911036
34. Richard G-F, Dujon B. Distribution and variability of trinucleotide repeats in the genome of the yeast *Saccharomyces cerevisiae*. *Gene*. 1996; 174: 165–174. [https://doi.org/10.1016/0378-1119\(96\)00514-8](https://doi.org/10.1016/0378-1119(96)00514-8) PMID: 8863744
35. Richard G-F, Hennequin C, Thierry A, Dujon B. Trinucleotide repeats and other microsatellites in yeasts. *Res Microbiol*. 1999; 150: 589–602. [https://doi.org/10.1016/s0923-2508\(99\)00131-x](https://doi.org/10.1016/s0923-2508(99)00131-x) PMID: 10672999
36. Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud J-B, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol*. 2016; 17. <https://doi.org/10.1186/s13059-016-1012-2> PMID: 27380939
37. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat Biotechnol*. 2016; 34: 184–191. <https://doi.org/10.1038/nbt.3437> PMID: 26780180
38. Rothstein R, Helms C, Rosenberg N. Concerted deletions and inversions are caused by mitotic recombination between delta sequences in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 1987; 7: 1198–1207. <https://doi.org/10.1128/mcb.7.3.1198> PMID: 3550432

39. Welcker AJ, de Montigny J, Potier S, Souciet JL. Involvement of very short DNA tandem repeats and the influence of the RAD52 gene on the occurrence of deletions in *Saccharomyces cerevisiae*. *Genetics*. 2000; 156: 549–557. PMID: [11014805](#)
40. Cinesi C, Aeschbach L, Yang B, Dion V. Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nat Commun*. 2016; 7: 13272. <https://doi.org/10.1038/ncomms13272> PMID: [27827362](#)
41. Dabrowska M, Juzwa W, Krzyzosiak WJ, Olejniczak M. Precise Excision of the CAG Tract from the Huntingtin Gene by Cas9 Nickases. *Front Neurosci*. 2018; 12. <https://doi.org/10.3389/fnins.2018.00075> PMID: [29535594](#)
42. Lemos BR, Kaplan AC, Bae JE, Ferrazzoli AE, Kuo J, Anand RP, et al. CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc Natl Acad Sci*. 2018; 115: E2040–E2047. <https://doi.org/10.1073/pnas.1716855115> PMID: [29440496](#)
43. Wilson TE, Grawunder U, Lieber MR. Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature*. 1997; 388: 495–498. <https://doi.org/10.1038/41365> PMID: [9242411](#)
44. Frank-Vaillant M, Marcand S. NHEJ regulation by mating type is exercised through a novel protein, Lif2p, essential to the Ligase IV pathway. *Genes&Development*. 2001; 15: 3005–3012.
45. Charpentier M, Khedher AHY, Menoret S, Brion A, Lamribet K, Dardillac E, et al. CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat Commun*. 2018; 9: 1133. <https://doi.org/10.1038/s41467-018-03475-7> PMID: [29556040](#)
46. van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, et al. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol Cell*. 2016; 63: 633–646. <https://doi.org/10.1016/j.molcel.2016.06.037> PMID: [27499295](#)
47. Haber JE. The many interfaces of Mre11. *Cell*. 1998; 95: 583–586. [https://doi.org/10.1016/s0092-8674\(00\)81626-8](https://doi.org/10.1016/s0092-8674(00)81626-8) PMID: [9845359](#)
48. Lengsfeld BM, Rattray AJ, Bhaskara V, Ghirlando R, Paull TT. Sae2 Is an Endonuclease that Processes Hairpin DNA Cooperatively with the Mre11/Rad50/Xrs2 Complex. *Mol Cell*. 2007; 28: 638–651. <https://doi.org/10.1016/j.molcel.2007.11.001> PMID: [18042458](#)
49. Mimitou EP, Symington LS. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature*. 2008; 455: 770–4. <https://doi.org/10.1038/nature07312> PMID: [18806779](#)
50. So CC, Martin A. DSB structure impacts DNA recombination leading to class switching and chromosomal translocations in human B cells. *PLOS Genet*. 2019; 15: e1008101. <https://doi.org/10.1371/journal.pgen.1008101> PMID: [30946744](#)
51. Storici F, Bebenek K, Kunkel TA, Gordenin DA, Resnick MA. RNA-templated DNA repair. *Nature*. 2007; 447: 338–341. <https://doi.org/10.1038/nature05720> PMID: [17429354](#)
52. Zhu Z, Chung WH, Shim EY, Lee SE, Ira G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell*. 2008; 134: 981–94. <https://doi.org/10.1016/j.cell.2008.08.037> PMID: [18805091](#)
53. Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci U S A*. 1981; 78: 6354–6358. <https://doi.org/10.1073/pnas.78.10.6354> PMID: [6273866](#)
54. Holmes AM, Haber JE. Double-strand break repair in yeast requires both leading and lagging strand DNA polymerases. *Cell*. 1999; 96: 415–424. [https://doi.org/10.1016/s0092-8674\(00\)80554-1](https://doi.org/10.1016/s0092-8674(00)80554-1) PMID: [10025407](#)
55. Muller H, Annaluru N, Schwerzmann JW, Richardson SM, Dymond JS, Cooper EM, et al. Assembling large DNA segments in yeast. *Methods Mol Biol Clifton NJ*. 2012; 852: 133–150. https://doi.org/10.1007/978-1-61779-564-0_11 PMID: [22328431](#)
56. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*. 1989; 122: 19–27. PMID: [2659436](#)
57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl*. 2009; 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: [19451168](#)
58. Viterbo D, Marchal A, Mosbach V, Poggi L, Vaysse-Zinkhöfer W, Richard G-F. A fast, sensitive and cost-effective method for nucleic acid detection using non-radioactive probes. *Biol Methods Protoc*. 2018; 3. <https://doi.org/10.1093/biomethods/bpy006> PMID: [32161800](#)
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: [19505943](#)
60. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011; 43: 491–8. <https://doi.org/10.1038/ng.806> PMID: [21478889](#)

61. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012; 22: 568–76. <https://doi.org/10.1101/gr.129684.111> PMID: 22300766
62. Fungtammasan A, Ananda G, Hile SE, Su MS-W, Sun C, Harris R, et al. Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res.* 2015 [cited 13 Oct 2016]. <https://doi.org/10.1101/gr.185892.114> PMID: 25823460
63. Millot G. Comprendre et réaliser les tests statistiques à l'aide de R. 2nd ed. Brussels: de boeck; 2011.

Annexe 2: A fast, sensitive and cost-effective method for nucleic acid detection using non-radioactive probes

Protocole de détection qualitative par sondes froides d'ADN par Southern blot et comparaison des signaux obtenues avec le signal de sondes radioactives. J'ai réalisé le gel de la figure 1 en appliquant le protocole de l'article.

METHODS MANUSCRIPT

A fast, sensitive and cost-effective method for nucleic acid detection using non-radioactive probes

David Viterbo^{1,2}, Astrid Marchal^{1,2}, Valentine Mosbach^{1,2}, Lucie Poggi^{1,2,3,4}, Wilhelm Vaysse-Zinkhöfer^{1,2,3} and Guy-Franck Richard^{1,2,*}

¹Department Genomes & Genetics, Institut Pasteur, 25 rue du Dr Roux, F-75015 Paris, France, ²CNRS, UMR3525, F-75015 Paris, France, ³Collège Doctoral, Sorbonne Université, 4 Place Jussieu, F-75005 Paris, France and ⁴Biologics Research, Sanofi R&D, 13 Quai Jules Guesde, 94403 Vitry sur Seine, France

*Correspondence address: Department Genomes & Genetics, Institut Pasteur, 25 rue du Dr Roux, F-75015 Paris, France. Tel: +33-1-45-68-84-36, E-mail: gfrichar@pasteur.fr

Abstract

Nucleic acid detection and quantification using a labeled DNA probe is a very common molecular biology procedure. Here, we describe a new method, based on commonly used laboratory solutions, for nucleic acid hybridization and detection with digoxigenin-labeled DNA probes. The protocol described is faster, more sensitive and much cheaper than a standard protocol using commercial solutions. Comparison with a classical radioactive detection method shows that the latter exhibits less background and shows a greater linear response. Hence, the proposed protocol may be routinely performed for qualitative detection of nucleic acid, but when precise signal quantitation needs to be obtained, radioactive probe hybridization associated to phosphorimaging technology is more reliable.

Keywords: nucleic acid hybridization, nucleic acid quantification, digoxigenin-labeled probe

Introduction

Detection of nucleic acids (DNA or RNA) bound to an hybridization membrane, using a specific probe, has been a widely used method in molecular biology laboratories for more than four decades [1]. Over the years, several radioactive or non-radioactive methods were designed to label probes. Random priming of a DNA molecule using random oligonucleotides as short primers is a popular technique aiming at obtaining single-stranded radioactive DNA probes exhibiting a high specific activity [2]. Such probes can be hybridized on denatured DNA or RNA molecules bound to a nylon membrane, in a sodium-phosphate buffer (also called “Church buffer”) [3]. The use of phosphor screen technology to quantify the amount of radioactivity at specific locations on the hybridization

membrane has allowed for almost three decades precise and reliable quantification of DNA or RNA amounts, with high sensitivity on a linear range. However, manipulation of radioactive phosphorus is tedious, restricted to authorized areas and must be performed behind a protective screen by properly trained scientists. In addition, radioactive products handling and shipping are more and more regulated and their price has been dramatically increasing for the last 10 years, to such a point that regular ordering of such products weighs significantly on small to medium size laboratories budget. Non-radioactive hybridization methods have been developed for many years as alternatives to radioactive approaches. Enzymatic incorporation of one of the four dNTP covalently linked to biotin or digoxigenin into single-stranded DNA probes were commonly used to that end. Comparisons of both

Received: 13 March 2018; Revised: 1 May 2018; Editorial decision: 3 May 2018; Accepted: 23 May 2018

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Table 1: Comparison of non-radioactive hybridization methods

	Commercial protocol	Time (°C)	Homemade protocol	Time (°C)
Probe hybridization	Tampon DIG Easy Hyb	Overnight (42°C)	250 mM sodium phosphate buffer, 7% SDS, 1 mM EDTA	Overnight (65°C)
Hybridization washes	SSC 2×, 0.1% SDS	2×5' (RT)	20 mM sodium phosphate buffer, 1% SDS, 1 mM EDTA	2×10' (65°C)
	SSC 1×, 0.1% SDS	2×15' (60°C)		
	B1 washing buffer	1' (RT)		
Membrane blocking	10× blocking solution diluted in B2 solution	60' (RT)	75 mM maleic acid pH 7.5, 200 mM NaCl, 5% non-fat dry milk powder	60' (RT)
Antibody binding	B3 10× blocking solution diluted in B2 solution	60' (RT)	75 mM maleic acid pH 7.5, 200 mM NaCl, 5% non-fat dry milk powder	30'–60' (RT)
Membrane washes	B1 washing buffer	2×15' (RT)	75 mM maleic acid, 200 mM NaCl, 0.3% Tween 20 ^a	2×15' (RT)
	B4 detection buffer	5' (RT)	100 mM Tris pH 9.5, 100 mM NaCl	5' (RT)

RT, room temperature.

^aThe tween is optional, it slightly reduces the background with some probes.

methods reached the conclusion that they were equally sensitive [4] but that digoxigenin probes were slightly more specific [5]. Several kits are commercially available to label DNA with digoxigenin-dUTP but this step may be easily implemented with homemade laboratory solutions [6]. However, subsequent steps of DNA hybridization, washes, antibody incubation, and washes are usually tedious, time consuming and expensive when using commercially available solutions. Here, we describe a simple and cheap protocol for non-radioactive probes in nucleic acid detection and quantification. All reagents can be simply prepared from stock solutions available in any molecular biology laboratory. Sensitivity was found to be generally better than with a commonly performed commercial protocol and comparable to radioactive probes coupled to phosphorimaging technology.

Materials and methods

Dot blots

Total genomic DNA was extracted from *Saccharomyces cerevisiae* (strain BY4741) and quantified on gel. Two-fold serial dilutions of this genomic DNA were made in 0.4M NaOH, incubated at room temperature for 10 min, then spotted on Hybond-XL nylon membranes (Amersham RPN 203S), before proceeding to hybridization.

Radioactive probe synthesis

The probe was labeled by random priming as follows. A 750-bp DNA probe corresponding to the 5' end of the ARG2 gene was amplified using oligonucleotides AR1 (AGGAGAATATTCGCGC ATGAA) and AR2 (AAGATATCTCATCTTTTAAACGT). After gel purification, 50–60 ng of this PCR product were mixed with 150 pmoles poly-deoxynucleotides hexamers (pd(N)₆, Takara) in 15 μ l of 1× random priming buffer (500 mM Tris-HCl (pH 7), 100 mM MgSO₄, 1 mM DTT). This mix was denatured at 95°C for 5 min before being put on ice. To this denatured DNA, 3 μ l of α -³²P dATP (6000 Ci/mmol), 1 μ l of the three remaining dNTP (10 mM each) and 10 units of Klenow polymerase fragment were added. The probe was incubated at 37°C for 1 h, then denatured 5 min at 95°C before being added to the hybridization buffer. Alternatively, it could be purified on ProbeQuant G50 micro sepharose columns (GE Healthcare) to allow probe quantification. Specific activities ranged from 6, 3×10⁷ to 1, 4×10⁹ cpm/ μ g DNA, depending on the probe (mean \pm 99% confidence interval = 5, 6×10⁸ \pm 1, 9×10⁸). The whole labeled probe was used in each hybridization (3 000 000 cpm on the average).

Non-radioactive probe synthesis

Non-radioactive probes were synthesized from the same purified PCR fragment covering 750 bp of the ARG2 gene 5' end. Labeling was performed in 100 μ l volume reactions containing 50–60 ng ARG2 PCR template, 1 M of each AR1 and AR2 primers, 10 μ l PCR DIG labeling mix (Roche 11 585 550 910), 10 μ l × PCR buffer and 5 units of Taq polymerase (Thermo Fischer). The PCR program used started by an initial denaturation step at 98°C for 2 min, followed by 35 cycles of: 98°C denaturation for 30 s, 50°C annealing for 30 s, 72°C extension for 1 min. A final 10 min extension step at 72°C was added at the end. A 5 μ l aliquot was loaded on gel for quantification and approximately 250 ng of the probe was used for hybridization, after a 5 min denaturation step at 95°C.

Probe hybridization

Hybridization of radioactive or non-radioactive probes was performed overnight, at 65°C in a rotating tube (Hybaid hybridization oven, Thermo Scientific) in 250 mM sodium phosphate buffer (pH 7.4), 7% SDS, 1 mM EDTA. Two washes were performed in 20 mM sodium phosphate buffer (pH 7.4), 1% SDS, 1 mM EDTA. When using the commercial protocol with DIG-labeled probes, the membrane was hybridized and washed as recommended by the manufacturer (Roche DIG Easy Hyb 11 796 895 001, see Table 1). Membranes hybridized with radioactive probes were exposed 4 or 20 h on a phosphor screen before quantification. Membranes hybridized with non-radioactive probes were subsequently treated with an anti-DIG antibody (see below).

Revealing the probe with the anti-digoxigenin antibody

When using the commercial protocol, the membrane was treated as recommended by the manufacturer (Roche Wash and Block Buffer Set 11 585 762 001, see Table 1). When using the homemade protocol, the membrane was incubated in 100 ml of blocking buffer (75 mM maleic acid, 200 mM NaCl, pH adjusted to 7.5 with 32% NaOH solution), supplemented with 5% non fat dry milk powder, for 1 h at room temperature. The anti-digoxigenin antibody (Fab fragment, Roche 11 093 274 910) was diluted (1/10 000) in 50 ml blocking buffer supplemented with milk powder and incubated with the membrane for 30 min to 1 h at room temperature. Subsequently, two washes were performed in blocking buffer supplemented with 0.3% Tween 20 for 15 min each at room temperature. One quick final wash in 100 mM Tris-HCl (pH 9.5), 100 mM NaCl was performed for 5 min at room temperature before the chemiluminescent substrate (CSPD, Roche 11 655 884 001) was diluted (1:100 to 1:10 in 100 mM

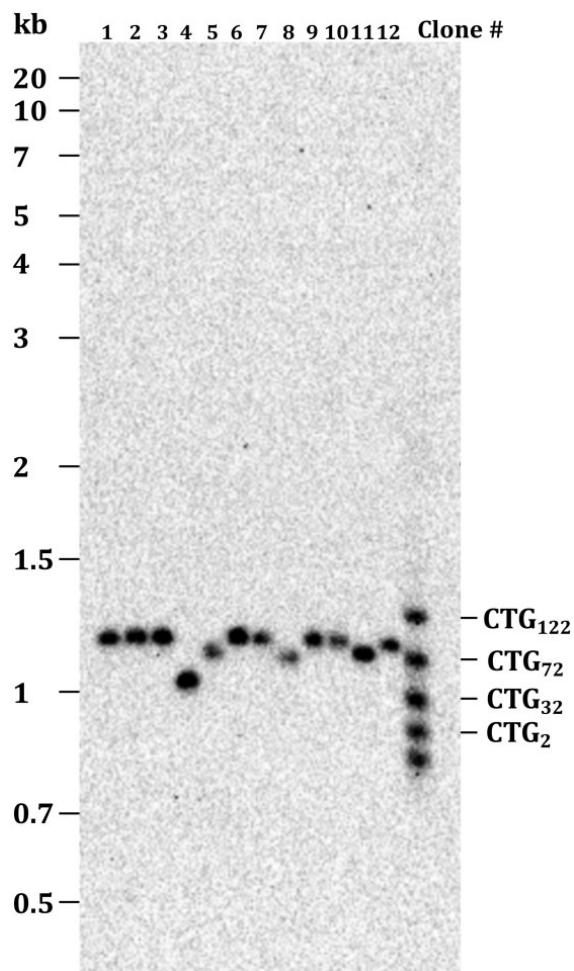


Figure 1: Southern blot of *S. cerevisiae* BY4741 genomic DNA. Total genomic DNA was prepared from yeast independent colonies grown overnight in YPD medium as previously described [8] and 2–3 μ g were digested with 20 units of SspI (NEBiolabs) during 4 h at 37°C. Digestions were loaded on a 1% agarose gel run overnight at 1 V/cm. It was alkaline transferred on a Hybond-XL membrane (Amersham) and hybridized with the DIG-labeled SUP4 probe. The membrane was treated as described in the “Materials and methods” section, using our homemade protocol. The CSPD was used undiluted. The molecular weight marker on the left is the GeneRuler 1 kb Plus (Thermo Scientific #SM1331). The molecular weight ladder on the right is a homemade PCR product hybridizing with the probe and corresponding to different CTG trinucleotide repeat lengths, as described previously [9].

Tris-HCl pH 9.5, 100 mM NaCl for dot blots, or used undiluted for the Southern blot in Fig. 1) and poured on the membrane (1 ml/10 cm²). It was subsequently incubated during 5 min at room temperature, the CSPD was removed and a further 10 min incubation at 37°C was performed to enhance the chemoluminescent reaction, as recommended by the manufacturer.

Signal quantification

Radioactive signals were read on a FujiFilm FLA-9000 phosphor-imager with the Image Reader software and quantified with Multi Gauge (v 3.0). Light signals were detected on a ChemiDoc Imaging System (Bio-Rad) and quantified with the dedicated

Image Lab software. Alternatively, an open source software ImageJ may also be used, for very similar results.

Results and discussion

The homemade non-radioactive protocol is faster, cheaper and more sensitive

Three different probes were assayed, targeting three different loci: ARG2, CAN1, and SUP4. The yeast genome being AT-rich, the three probes show similar low GC content, respectively, 38, 33 and 34%. Probe lengths were 305 bp for CAN1, 488 bp for SUP4, and 751 bp for ARG2. Efficacy of both hybridization and wash protocols were compared on dot blots containing serial dilutions of *S. cerevisiae* total genomic DNA. Commercial and homemade protocols are extensively described in Table 1 and in the “Materials and methods” section. With the commercial protocol, the minimal amount of target molecules detected by the DIG-labeled probe was 370 000 genome equivalent (For *S. cerevisiae* (12.5-Mb genome), one genome equivalent corresponds to 0.0137 pg of genomic DNA, the amount needed to contain one genome copy, on average.) for ARG2 (5 ng DNA, Fig. 2A), 2 690 000 genome equivalent for the CAN1 probe (38 ng DNA, Fig. 3A) and 80 000 genome equivalent for the SUP4 probe (1 ng DNA, Fig. 4A). With our homemade protocol, sensitivity increased to 80 000 genome equivalent for ARG2 and CAN1 (1 ng DNA, Figs. 2B and 3B) and was unchanged for the SUP4 probe (Fig. 4B). Hence, our homemade protocol was found to be at least as sensitive as the commercial method (SUP4 probe) and at best 5–30 times more sensitive (ARG2 and CAN1 probes). Linearity of the signal response as compared to DNA amount was found to be in the same range. The ratio signal/background was very comparable with both methods and quite low in all three experiments, as shown by the high background visible. The total time spent on washes with our homemade protocol was 55 min altogether, for a 20-min time saving as compared to the commercial protocol. So, the use of our buffers and method was faster and generally more sensitive than the commercial solution used here.

Radioactive hybridization is more specific, cleaner and more linear

In order to compare non-radioactive methods to a well-established protocol of hybridization with a radioactive probe, we hybridized similar dot blots with alpha-³²P dATP labeled probes (“Materials and methods” section). Membranes were exposed 4 or 20 h on a phosphor imaging screen. The ARG2 and CAN1 radioactive probes were able to detect 80 000 genome equivalent (1 ng DNA), similar to what was achieved with our homemade DIG probe protocol. Sensitivity was slightly better with the SUP4 probe that was able to detect 20 000 genome equivalent (300 pg DNA) after 20 h of exposure (Fig. 4D). The ratio signal/background was extremely high, in all three experiments, as compared to what was obtained with non-radioactive protocols. The outcome was a cleaner and nicer image of each dot blot, with low background. Linearity of the signal response as compared to DNA amount was excellent, especially at the highest exposure time, for which it was almost perfectly linear, the CAN1 probe giving a slightly poorer linear response. This allowed for very precise quantification of the amount of DNA present on each blot. Therefore, we concluded that although the DIG probe was as sensitive as the ³²P probe, this last one was a much better choice when precise signal quantification was required.

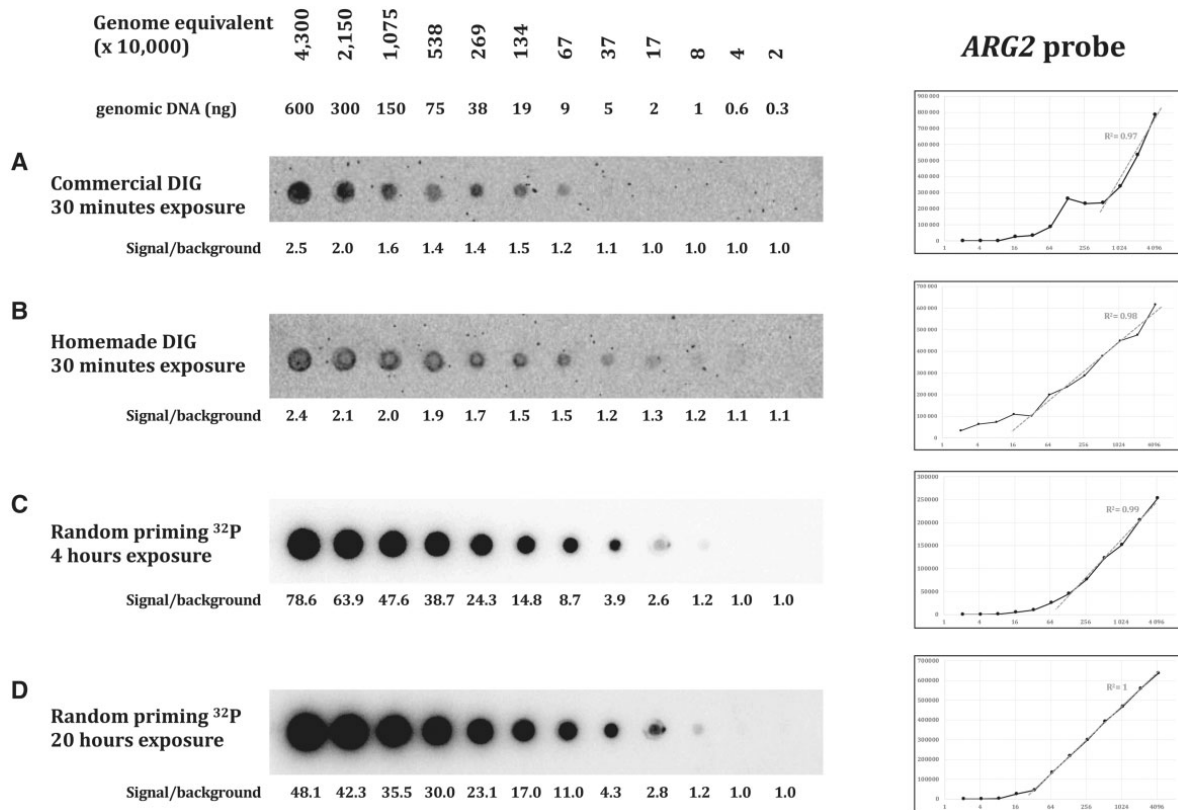


Figure 2: *S. cerevisiae* BY4741 total genomic DNA was 2-fold diluted and spotted on nylon membranes. From 600 to 0.3 ng genomic DNA, corresponding to 4.3×10^7 to 20 000 genome equivalents were spotted. **A** Non-radioactive ARG2 DIG probe with commercial protocol, CSPD diluted 1:100, 30 min exposure on ChemiDoc. **B** Non-radioactive ARG2 DIG probe with homemade protocol, CSPD diluted 1:100, 30 min exposure on ChemiDoc. **C** Radioactive ARG2 probe, 4 h exposure on phosphor screen. **D** Radioactive ARG2 probe, 20 h exposure on phosphor screen. Signal quantification graphs are shown to the right, for each experiment. x-axis: genome equivalent; y-axis: signal quantification. Linearity is estimated by the correlation coefficient value of the linear part of the curve.

The homemade non-radioactive protocol exhibits a good signal specificity

In order to check probe specificity, a Southern blot of total yeast genomic DNA was hybridized with a DIG-labeled *SUP4* probe, in our homemade conditions. The yeast strain analyzed contained a CTG trinucleotide repeat integrated at the *SUP4* locus, whose length varies among different subclones [7], as shown in Fig. 1. Length polymorphism corresponding to shorter CTG repeat tracts was clearly visible in four clones (4, 5, 8, and 11). The signal was unique and no other band was detected under or above the expected band, showing the good specificity of our homemade protocol.

Hybridization of nucleic acids to a specific labeled probe is very commonly used in molecular biology experiments such as dot blots, Southern or Northern blots. With the recent increase in price and regulation of radioactive compounds, non-radioactive probes such as those labeled with digoxigenin-dUTP are more and more attractive for molecular biology laboratories. Here, we show that using homemade buffers and solutions that are common and cheap, we achieved a more sensitive and faster detection than what was obtained with dedicated commercial solutions. The protocol was also greatly simplified, since the probe hybridization buffer was the same as the buffer used for radioactive probes ("Church buffer"), and only two

different solutions were used for antibody blocking, binding, and washes, instead of four when using commercial solutions. In addition, our method was also cost-effective, since homemade buffers led to a 12-fold decrease in experimental cost as compared to commercial solutions. That last result will be a definitive asset for laboratories using hybridization methods on a regular basis.

Comparison of non-radioactive methods with ³²P probe hybridization showed a clear advantage to the latter for low background and high linearity of the response. It is particularly visible after 20 h of exposure, although sensitivity did not increase with longer exposure times. Despite the high signal level, the phosphor screen was remarkably not saturated, even for larger DNA amounts.

In conclusion, we recommend to use our non-radioactive detection protocol to obtain a fast, sensitive, and cost-effective response to nucleic acid hybridization, when qualitative rather than quantitative results are expected. However, as long as signal quantitation needs to be accurate, for example to compare relative levels of low intensity bands on a Southern or Northern blot, or to quantify pausing signal level with 2D gel electrophoresis [8], radioactive hybridization and phosphor screen technology must be chosen for reliable and reproducible results.

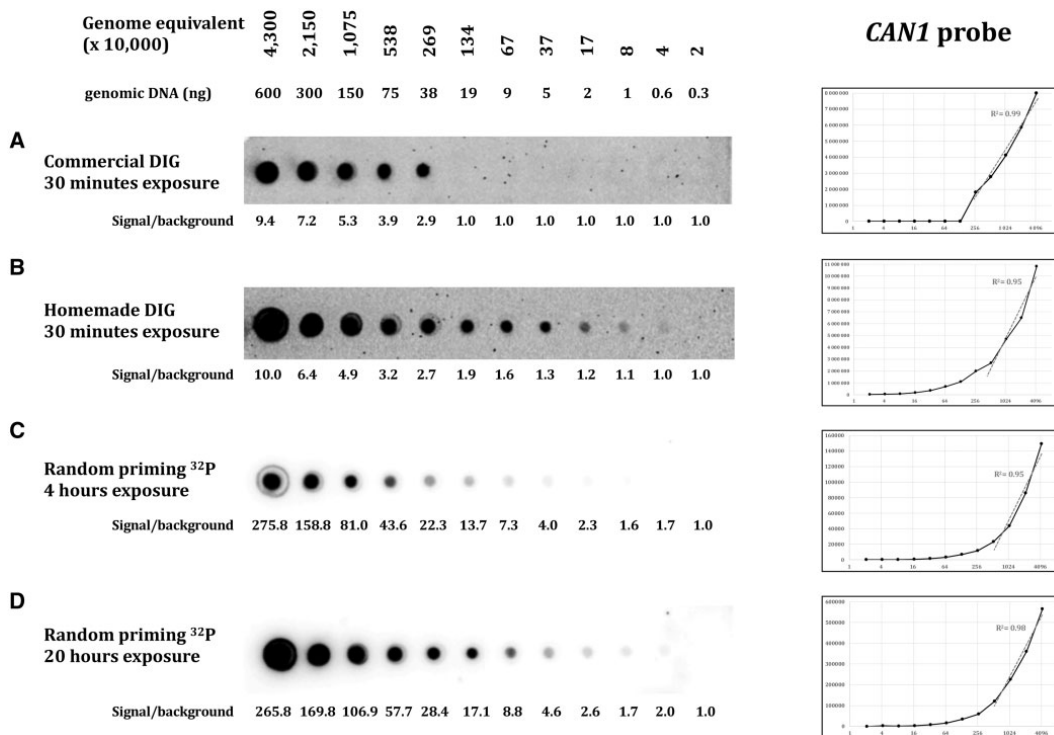


Figure 3: Same as Fig. 2, with the CAN1 probe. CSPD diluted 1:10 in both DIG protocols.

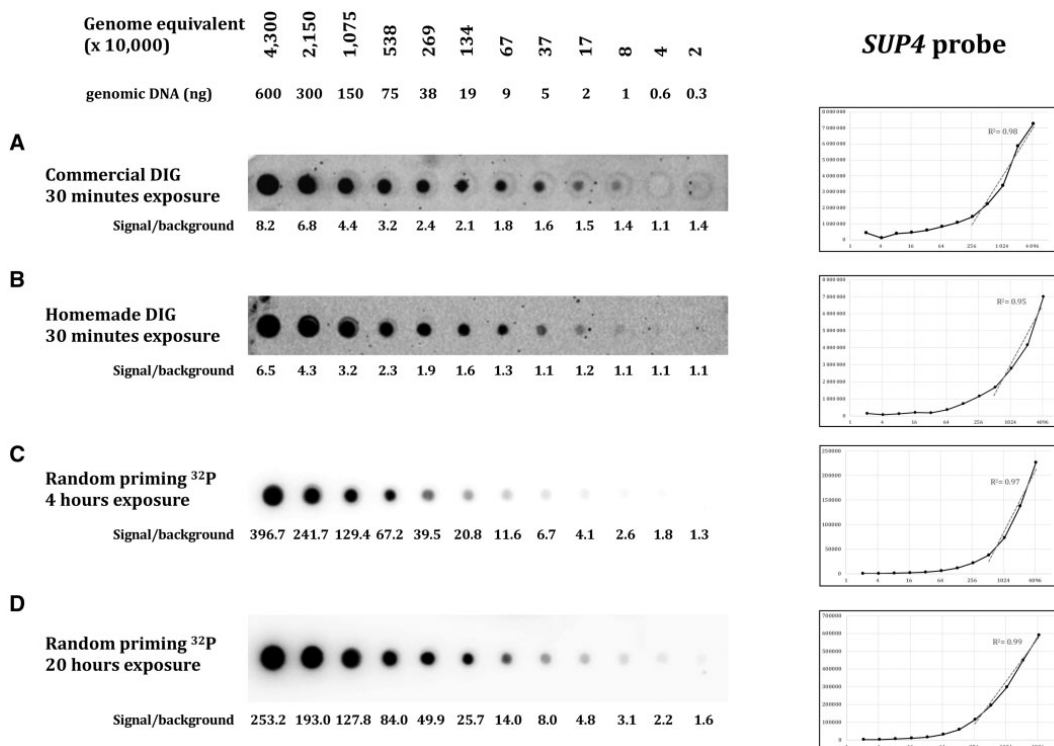


Figure 4: Same as Fig. 2, with the SUP4 probe. CSPD diluted 1:10 in both DIG protocols.

Author contributions

D.V., A.M., V.M., L.P., and W.V.-Z. ran hybridizations with radioactive or non-radioactive probes in different experimental conditions and quantified results. G.-F.R. wrote the manuscript.

Conflict of interest statement. None declared.

Funding

V.M. was supported by Fondation Guy Nicolas and Fondation Hardy. L.P. is the recipient of a PhD student CIFRE fellowship from SANOFI. W.V.-Z. is the recipient of a PhD student fellowship from the Ligue Nationale contre le Cancer.

References

1. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975;**98**:503–17.
2. Feinberg AP, Vogelstein B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 1983;**132**:6–13.
3. Church GM, Gilbert W. Genomic sequencing. *Proc Natl Acad Sci USA* 1984;**81**:1991–5.
4. McQuaid S, McMahon J, Allan GM. A comparison of digoxigenin and biotin labelled DNA and RNA probes for in situ hybridization. *Biotech Histochem* 1995;**70**:147–54.
5. Buti M, Jardi R, Rodriguez-Frias F et al. Digoxigenin-labelled DNA-probe: a rapid non-radioactive method for hepatitis B virus DNA detection in serum. *Eur J Clin Chem Clin Biochem J Forum Eur Clin Chem Soc* 1991;**29**:731735.
6. Emanuel JR. Simple and efficient system for synthesis of non-radioactive nucleic acid hybridization probes using PCR. *Nucleic Acids Res* 1991;**19**:2790.
7. Richard GF, Viterbo D, Khanna V et al. Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. *PLoS One* 2014;**9**:e95611.
8. Viterbo D, Michoud G, Mosbach V et al. Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. *DNA Repair (Amst)* 2016;**42**:94–106.
9. Mosbach V, Poggi L, Viterbo D et al. TALEN-induced double-strand break repair of CTG trinucleotide repeats. *Cell Rep* 2018;**22**:2146–59.

*Viens, mon beau chat, sur mon coeur amoureux ;
Retiens les griffes de ta patte,
Et laisse-moi plonger dans tes beaux yeux,
Mêlés de métal et d'agate.*

*Lorsque mes doigts caressent à loisir
Ta tête et ton dos élastique,
Et que ma main s'enivre du plaisir
De palper ton corps électrique,*

*Je vois ma femme en esprit. Son regard,
Comme le tien, aimable bête
Profond et froid, coupe et fend comme un dard,*

*Et, des pieds jusques à la tête,
Un air subtil, un dangereux parfum
Nagent autour de son corps brun.*

Charles BAUDELAIRE – Le chat

Abstract :

Les microsatellites sont des répétitions en tandem d'un motif compris entre une et neuf paires de bases. Ces répétitions sont retrouvées dans tous les organismes de façon ubiquitaire mais sont particulièrement abondantes dans les organismes eucaryotes, couvrant 3% du génome humain. Certains microsatellites sont enclins à une expansion, parfois massive, conduisant à de nombreuses maladies neurodégénératives chez l'homme telle que la maladie de Huntington, la dystrophie myotonique de type 1 (DM1), le syndrome du X fragile ou encore l'ataxie de Friedreich. Toutes ces répétitions sont capables de former des structures secondaires *in vitro* et possiblement *in vivo*. Les mécanismes conduisant à l'expansion massive des microsatellites ne sont pas encore connus. La maladie neurodégénérative la plus fréquemment transmise est la DM1. La DM1 est une maladie caractérisée par une faiblesse musculaire progressive et une myotonie. L'apparition et la sévérité des symptômes sont positivement corrélée avec le nombre de répétitions, localisées dans le 3'UTR du gène *DMPK* (Dystrophia Myotonica Protein Kinase). Il n'existe à ce jour aucun traitement connu pour aucune maladie neurodégénérative. Dans des travaux précédents du laboratoire, une TALE nucléase (TALEN), élaborée dans le but d'introduire une cassure double-brin au sein d'un microsatellite (CTG)_n provenant d'un patient DM1, s'est montrée très efficace à contracter ce microsatellite (>99% des cellules montrent une contraction) et très spécifique. La compréhension des mécanismes conduisant à la contraction des répétitions chez la levure est nécessaire si l'on souhaite en comprendre les mécanismes chez l'homme. Ainsi, plus récemment des expériences ont été menées dans des cellules dont les systèmes de réparation des CDB ont été altérés, montrant que *RAD51* (conversion génique), *POL32* (Break Induced Replication) et *DNL4* (non-homologous end joining) n'étaient pas nécessaires à la réparation des CDB au sein des microsatellites. Seul *RAD50* (complexe MRX) et *RAD52* (recombinaison homologue) semblent nécessaires, indiquant que la cellule répare les CDB dans les régions répétées par single-strand annealing (SSA) entre les motifs répétés. L'objectif de cette thèse a été d'étudier le rôle de plusieurs gènes (*MRE11*, *EXO1*, *SGS1*, *DNA2*, *SAE2*), impliqués dans la résection et la réparation d'une unique CDB au sein d'une région répétée CTG, chez la levure, ainsi que d'étudier l'interaction de la machinerie de maintenance télomérique avec la réparation d'une cassure double-brin au sein d'une région répétée, au travers de l'étude des gènes *RIF1* et *RIF2*. Des cinétiques visant à quantifier le pourcentage de cassure (par Southern blot), de résection (par QPCR), de survie et d'évènements de recombinaison (par Southern blot) chez les survivants ont été réalisées.

Mots clés : [maladies à triplets ; TALENs ; Réparation cassure double-brin ; Instabilité microsatellites ; Résection]

Microsatellites are tandem repetitions of a pattern between one and nine base pairs. These repeats are found ubiquitously in all organisms but are particularly abundant in eukaryotic organisms, covering 3% of the human genome. Some microsatellites are prone to expansion, sometimes massive, leading to many neurodegenerative diseases in humans such as Huntington's disease, myotonic dystrophy type 1 (DM1), fragile X syndrome or Friedreich's ataxia. All these repeats are capable of forming secondary structures *in vitro* and possibly *in vivo*. The mechanisms leading to the massive expansion of microsatellites are not yet known. The most frequently transmitted neurodegenerative disease is DM1. DM1 is a disease characterized by progressive muscle weakness and myotonia. The onset and severity of symptoms are positively correlated with the number of repetitions, located in the 3'UTR of the *DMPK* (Dystrophia Myotonica Protein Kinase) gene. To date, there is no known treatment for any neurodegenerative disease. In previous work in the laboratory, a TALE nuclease (TALEN), developed for the purpose of introducing a double-stranded break in a microsatellite (CTG)_n from a DM1 patient, has been shown to be very effective in contracting this microsatellite (>99% of the cells show a contraction) and very specific. Understanding the mechanisms leading to the contraction of repeats in yeast is necessary to understand the mechanisms in humans. Thus, more recently experiments have been conducted in cells with altered CBD repair systems, showing that *RAD51* (gene conversion), *POL32* (Break Induced Replication) and *DNL4* (non-homologous end joining) were not necessary for CBD repair within microsatellites. Only *RAD50* (MRX complex) and *RAD52* (homologous recombination) appeared to be necessary, indicating that the cell repairs CBDs in the repeated regions by single-strand annealing (SSA) between the repeated patterns. The objective of this thesis was to study the role of several genes (*MRE11*, *EXO1*, *SGS1*, *DNA2*, *SAE2*), involved in the resection and repair of a single CBD within a CTG repeated region, in yeast, as well as to study the interaction of telomeric maintenance machinery with the repair of a double-strand break within a repeated region, through the study of the *RIF1* and *RIF2* genes. Kinetics to quantify the percentage of breakage (by Southern blot), resection (by QPCR), survival and recombination events (by Southern blot) in survivors were performed.

Keywords : [microsatellite disorders ; TALENs ; Double-strand break repair ; Microsatellites instability ; Resection]