



Correction des mésappariements et stabilité de l'information génétique

Jean-Pierre Claverys

Société Française de Génétique

Président

E. Moustacchi

Vice-présidents

P.-H. Gouyon
A. Nicolas
C. Stoll

Secrétaire général

R. Motta

Prière d'adresser toute correspondance au Secrétariat général de la SFG, Roland Motta, Institut de recherches scientifiques sur le cancer, laboratoire de recherches génétiques sur les modèles animaux, 7, rue Guy-Môquet, 94802 Villejuif Cedex, France.

Comité de rédaction

A. Bernheim
J.-C. Dreyfus
M. Fellous
J. Génormont
F. Minvielle
R. Motta
A. Nicolas
S. Sommer
D. de Vienne

Secrétaire

M.-L. Prunier

L'information génétique de toute cellule, contenue dans l'acide désoxyribonucléique (ADN), peut subir de nombreuses altérations (mutations) sous l'action d'agents mutagènes ou spontanément, en particulier lors de la réplication de l'ADN qui précède la division cellulaire. Ces mutations peuvent avoir des conséquences considérables pour le métabolisme cellulaire, pouvant aller jusqu'au déclenchement de processus de cancérisation chez les eucaryotes supérieurs. Ainsi, des oncogènes de type *ras* peuvent être activés par mutation ponctuelle. Les systèmes de réparation de l'ADN jouent donc un rôle essentiel en réduisant la probabilité de tels événements.

Parmi ces systèmes, ceux qui assurent la correction des mésappariements (bases non complémentaires) de l'ADN ont des effets multiples sur la stabilité de l'information génétique. Par leur action post-réplivative, certains systèmes de correction des mésappariements augmentent la fidélité de réplication. Par leur action sur les intermédiaires de recombinaison, les systèmes de correction peuvent aussi moduler la recombinaison. Selon les systèmes, il pourra s'ensuivre une augmentation de la stabilité de l'information génétique (dans le cas d'une inhibition de la recombinaison) ou un accroissement de

la diversité des recombinants formés. Ces différents aspects seront évoqués au travers de la description des systèmes de correction des mésappariements les plus étudiés.

Les mésappariements, leurs origines

Le terme de mésappariement désigne toute confrontation (au sens étymologique) de bases autres que les paires adénine-thymine (A-T) et guanine-cytosine (G-C) normalement présentes dans une molécule d'ADN bicaténaire. Par extension, ce terme est aussi utilisé pour désigner l'absence d'une ou de plusieurs bases complémentaires sur l'un des deux brins. Une molécule d'ADN comportant un ou plusieurs mésappariements est dite hétéroduplex et porte une information génétique différente sur chaque brin (*figure 1, centre*). En l'absence de correction, cette molécule donnera donc naissance, après réplication, à deux molécules homoduplex génétiquement différentes (*figure 1, droite*).

La réplication de l'ADN et la recombinaison génétique, deux processus universellement présents dans le monde vivant, peuvent être à l'origine de la formation de mésappariements. Dans le premier cas, les mésappariements

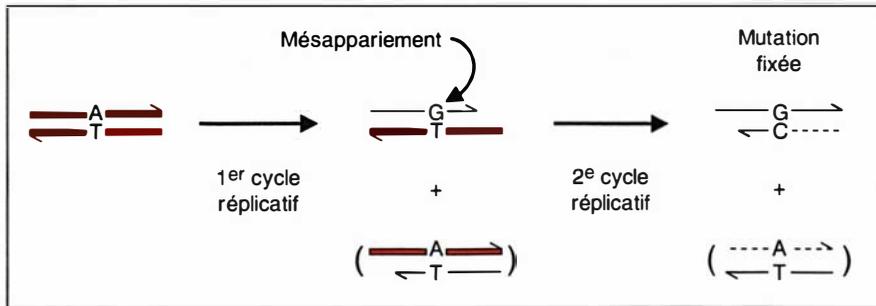


Figure 1. **Formation et ségrégation d'un mésappariement lors de la réplcation.** L'incorporation d'une guanine (G) face à la thymine (T), par suite d'une erreur de l'appareil réplcatif lors d'une première réplcation, aboutit à la formation d'un ADN hétéroduplex portant le mésappariement G/T (**centre**). Après un nouveau cycle de réplcation, deux molécules homoduplex génétiquement différentes sont obtenues, et la mutation est alors fixée (**droite**). Remarquons que la même mutation peut être provoquée par incorporation d'une cytosine (C) face à l'adénine (A) du brin complémentaire. Le mésappariement A/C est alors l'intermédiaire mutationnel.

résultent d'un dysfonctionnement de l'appareil réplcatif [1]. En l'absence de correction, ils ont pour conséquence une mutation qui sera définitivement fixée lors du cycle réplcatif suivant (figure 1). En revanche, la présence de mésappariements est une conséquence naturelle de la recombinaison homologue. Les intermédiaires de recombinaison sont constitués de deux brins complémentaires issus chacun d'un duplex d'ADN parental différent et associés par des liaisons hydrogène (ADN hybride) [2]. Ces intermédiaires sont des ADN hétéroduplex lorsqu'ils recouvrent une région où les partenaires de la recombinaison n'ont pas la même séquence. Une troisième source potentielle de mésappariements, commune à tous les organismes dont l'ADN contient des 5-méthylcytosine (5-meC), est la désamination spontanée des 5-meC en T qui produit spécifiquement le mésappariement G/T.

Les systèmes de correction des mésappariements

La correction des mésappariements est assurée par des systèmes variés qui peuvent être divisés en deux groupes, systèmes spécifiques et systèmes généralisés. Ces systèmes diffèrent non seulement par leur spécificité de substrat

mais aussi par l'orientation de la correction. Nous verrons que les systèmes spécifiques effectuent une correction directionnelle qui dépend uniquement du mésappariement. Ainsi, l'un de ces systèmes corrige le mésappariement G/T systématiquement en G-C (jamais en A-T). En revanche, l'orientation de la correction par les systèmes généralisés se fait à partir d'un signal distinct (et distant) du mésappariement. Cette dernière caractéristique a pour corollaire une autre différence entre systèmes spécifiques et systèmes généralisés : la correction est très localisée pour les premiers, alors qu'elle est beaucoup plus étendue pour les seconds (quelques centaines, voire quelques milliers de nucléotides). Les implications de ces différents modes de fonctionnement quant au rôle de ces systèmes dans la diversification des génomes et dans le maintien de leur homogénéité seront discutées plus loin.

Les systèmes de correction spécifique

Le premier système de correction spécifique d'un mésappariement mis en évidence est le système VSP (*very-short patch*) de la bactérie *Escherichia coli* [3]. Ce système corrige spécifiquement en G-C les mésappariements G/T provenant de la désamination spontanée des

5-meC à l'intérieur de la séquence CC(A ou T)GG. Les cytosines situées en position 2 dans cette séquence sont, chez *E. coli*, le substrat d'une cytosine méthylase, le produit du gène *dmc*. Immédiatement en aval de ce gène se trouve le gène *vsr* dont le produit est absolument requis pour la correction (pour revue, voir [4]). Cette organisation génétique et l'action directionnelle du système VSP suggèrent fortement que le rôle biologique de ce système est de protéger *E. coli* de l'effet mutateur potentiel de la désamination spontanée des 5-meC. Le mécanisme de correction est maintenant bien compris (figure 2, gauche). La protéine Vsr purifiée reconnaît le mésappariement G/T dans son contexte spécifique et fait une incision unique en 5' du résidu T [5]. L'ADN polymérase I (polymérase de réparation) élimine alors ce résidu par son activité exonucléasique 5' → 3' et assure la synthèse réparatrice sur une courte distance (une dizaine de nucléotides)[4].

Les organismes eucaryotes sont eux aussi confrontés au problème de l'instabilité des cytosines méthylées, présentes dans leur ADN à l'intérieur des séquences CpG (impliquées dans la régulation transcriptionnelle de plusieurs gènes). Une activité de correction spécifique du mésappariement G/T situé dans une séquence CpG a été mise en évidence à partir d'extraits nucléaires de cellules humaines Hela [4]. Cette activité a très probablement, comme le système VSP, un rôle antimutateur. Le processus de correction qui aboutit aussi à la resynthèse d'un court segment d'ADN diffère cependant du système VSP par le type d'activité enzymatique mis en jeu (figure 2, droite). Il fait intervenir séquentiellement une glycosylase spécifique qui élimine le résidu T mésapparié, une AP-endonucléase qui reconnaît le site abasique et incise le brin d'ADN à proximité immédiate de ce site, et enfin l'ADN polymérase β, une polymérase impliquée dans la réparation [4].

Un autre système de correction spécifique, mais du mésappariement G/A, a été mis en évidence chez *E. coli* (pour revue, voir [4]). Ce système assure la correction directionnelle des mésappa-

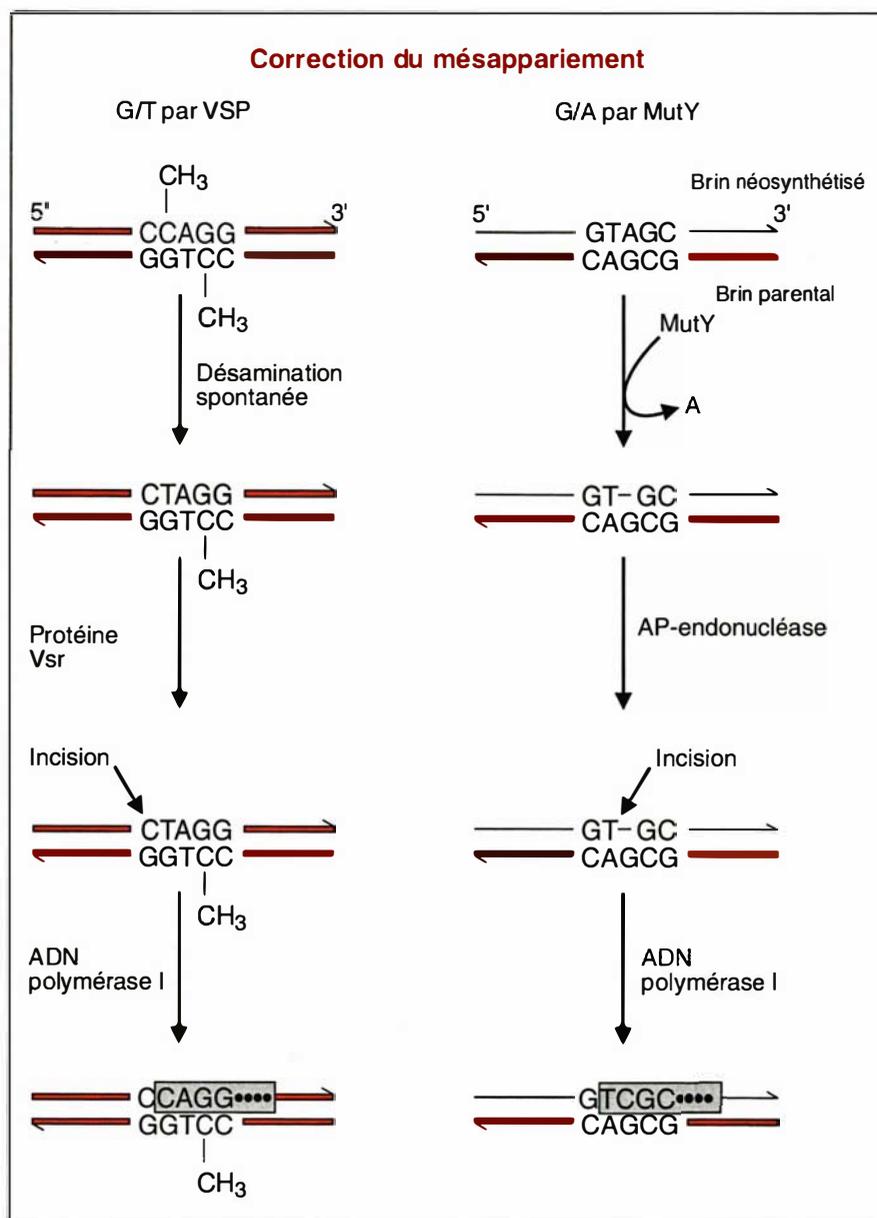


Figure 2. **Correction spécifique des mésappariements G/T (système VSP) et G/A (système MutY).** Ces systèmes se distinguent par la nature des intermédiaires conduisant à la correction d'un brin. **Gauche** : incision directe par la protéine Vsr de la liaison phosphodiester en 5' du résidu T mésapparié. **Droite** : incision par la protéine MutY de la liaison glycosidique du résidu A mésapparié (ou du résidu T mésapparié dans le cas de l'activité de correction des mésappariements G/T mise en évidence à partir d'extraits nucléaires de cellules humaines Hela). Le site abasique ainsi créé est reconnu par une AP-endonucléase qui incise la liaison phosphodiester adjacente. Dans les deux cas, la resynthèse d'un segment long d'une dizaine de nucléotides (zone encadrée) est prise en charge par l'ADN polymérase I.

m/s n° 6, vol. 8, juin-juillet 92

riements G/A en G/C. Le processus de correction est analogue à celui que nous venons de décrire (figure 2, droite). Il fait intervenir la protéine MutY, une glycosylase spécifique du résidu A mésapparié. Son rôle biologique est très probablement la correction des erreurs de réplication aboutissant à l'incorporation d'un résidu A face à un résidu G. Il semble, à première vue, surprenant qu'un tel système unidirectionnel puisse avoir une action antimutatrice. En effet, toute erreur de réplication conduisant à l'incorporation d'un résidu G face à un résidu A devrait être irrémédiablement fixée par ce système. Cette situation conflictuelle est évitée grâce à l'action de la protéine MutT, qui empêche précisément l'incorporation d'un résidu G face à un résidu A [6]. Prévention de la formation du mésappariement A/G et correction directionnelle du mésappariement G/A sont donc combinées pour réduire, chez *E. coli*, ce type d'événement mutationnel.

Cette combinaison d'activités antimutatrices spécifiques du couple de mésappariements A/G, G/A existe probablement chez d'autres organismes. Ainsi chez *Streptococcus pneumoniae*, un système de correction spécifique du mésappariement G/A en G-C a été mis en évidence [7] et la protéine MutX est l'équivalent fonctionnel de la protéine MutT (Méjean, Salles et Claverys, en préparation). L'existence d'une telle combinaison d'activités antimutatrices spécifiques est à rapprocher de la relative inefficacité de correction des mésappariements A/G, G/A par les systèmes de correction généralisée.

Les systèmes de correction généralisée

Le système Hex (*heteroduplex excision*) de *S. pneumoniae* et le système Mut de *E. coli* sont les deux systèmes de correction généralisée des mésappariements les mieux connus. Si l'étude du système Hex a permis l'émergence des concepts d'activité antirecombinogène et de rôle antimutateur par correction des erreurs de réplication, celle du système Mut a débouché sur l'enzymologie des systèmes de correction des mésappariements et a conduit à l'hypothèse d'une intervention de ces

systèmes dans la spéciation, par leur activité antirecombinogène. Nous passerons donc en revue les propriétés de ces deux systèmes, puis nous évoquerons les éléments qui indiquent une conservation de ces systèmes au cours de l'évolution, des eubactéries à l'homme, en passant par la levure, la drosophile et le xénope. Enfin, nous comparerons leurs effets sur l'homogénéisation des génomes et sur la recombinaison interspécifique.

Le système Hex

La mise en évidence du système Hex et sa caractérisation ont été menées à travers l'étude de ses effets sur la recombinaison homologe lors de la transformation (*pour revue, voir [8]*). La transformation est un processus inducible au cours duquel des fragments d'ADN exogène pénètrent sous forme monocaténaire dans les cellules. Ces fragments, lorsqu'il s'agit d'ADN homologue, participent à la formation d'un ADN hybride en établissant des liaisons hydrogène avec le brin complémentaire du génome récepteur (*figure 3*). Comme dans la recombinaison entre deux ADN duplex, l'ADN hybride comporte des mésappariements quand la séquence des partenaires de la recombinaison n'est pas totalement identique.

L'existence de variations importantes d'efficacité de transformation (de recombinaison) entre marqueurs génétiques d'une même région a conduit, dès 1966, à l'hypothèse de l'intervention d'un système de correction des mésappariements, responsable de l'élimination préférentielle du brin d'ADN exogène pour certains marqueurs génétiques, les marqueurs LE (*low efficiency* : faible efficacité de transformation) [9] (*figure 3B*). Dans cette hypothèse, les marqueurs HE (*highest efficiency*), qui donnent 20 fois plus de recombinants, correspondent à des mésappariements non reconnus (*figure 3A*). La validité de ce concept d'action antirecombinogène d'un système de correction des mésappariements a été démontrée par une série d'expériences génétiques et par l'isolement de mutants Hex⁻ intégrant marqueurs LE et HE avec la même haute efficacité [8].

L'ensemble des résultats génétiques obtenus suggèrent que la correction d'un mésappariement ponctuel s'accompagne de l'élimination de la majeure partie du fragment d'ADN exogène. Ainsi, l'efficacité de transformation d'un fragment d'ADN portant

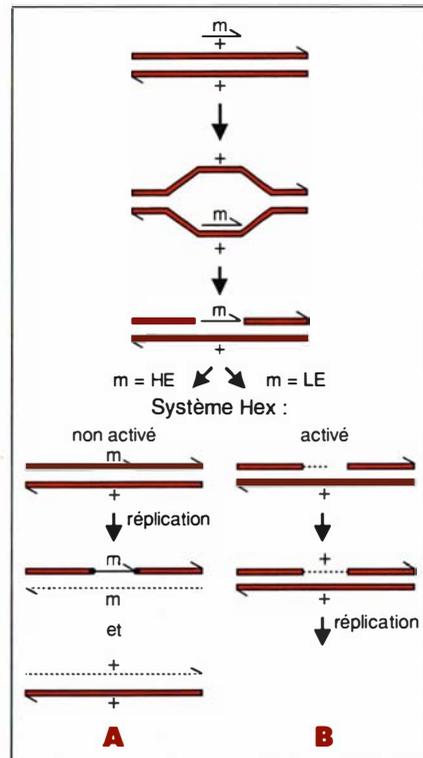


Figure 3. Correction des mésappariements par le système Hex au cours de la transformation. Un fragment exogène d'ADN monocaténaire (trait fin) déplace le brin récepteur (double trait) pour créer un hétéroduplex (la cinétique de disparition du brin déplacé n'est pas connue). **A.** Dans le cas des marqueurs HE (haute efficacité de transformation), l'établissement de liaisons covalentes (point noir) permet l'intégration définitive du fragment exogène. La structure hétéroduplex est résolue après passage de la fourche de réplication qui produit un duplex parental et un duplex transformé. **B.** Dans le cas des marqueurs LE (faible efficacité de transformation), l'action du système Hex aboutit à l'élimination de la totalité du fragment exogène. Une étape de resynthèse restaurerait l'information parentale (en l'absence d'information sur le devenir du brin déplacé, il est aussi possible d'envisager qu'il reprenne sa place pour reformer le duplex parental).

un marqueur HE à proximité d'un marqueur LE est celle du marqueur LE [8]. Le suivi du devenir d'un ADN exogène portant un marqueur LE en son centre et uniformément marqué au ³²P sur un seul brin démontre l'élimination des séquences exogènes situées aussi bien en 5' qu'en 3' du marqueur LE [10]. L'élimination de la totalité du fragment exogène, soit de 3 000 à 6 000 nucléotides, a lieu après formation de la structure hétéroduplex mais, très vraisemblablement, avant rétablissement de liaisons covalentes entre fragment exogène et brin d'ADN récepteur adjacent. Cette observation a conduit au modèle de correction fondé sur la reconnaissance du mésappariement, suivie de la recherche d'extrémités (celles du fragment exogène dans le cas de la transformation) à partir desquelles a lieu l'élimination du segment ainsi défini (*figure 4A*).

Le rôle biologique du système Hex dans la correction des erreurs de réplication a été déduit du phénotype mutateur des mutants Hex⁻ [11]. La correction des erreurs de réplication pose le problème de la distinction entre brins d'ADN parental et brins néosynthétisés. Ces derniers doivent être spécifiquement éliminés pour aboutir à la conservation de l'information génétique parentale. Le modèle de fonctionnement proposé pour le système Hex est calqué sur celui déduit de l'étude de son fonctionnement lors de la transformation et fondé sur la recherche d'interruptions sur l'ADN après reconnaissance d'un mésappariement provenant d'une erreur de réplication. L'ADN néosynthétisé possède au moins l'extrémité en cours d'élongation, voire plusieurs interruptions, en cas de synthèse discontinue, qui doivent permettre sa correction préférentielle (*figure 4B*). Deux gènes spécifiques essentiels au fonctionnement du système Hex, *hexA* et *hexB*, ont été identifiés [8].

Le système Mut

La mise en évidence et l'analyse génétique du système de correction généralisée des mésappariements de *E. coli*

ont été, pour une large part, fondées sur l'introduction dans les cellules de molécules hétéroduplex construites *in vitro* à partir de l'ADN de bactériophages (pour revue, voir [4, 8, 12]). L'analyse du devenir de marqueurs génétiques portés sur ces molécules hétéroduplex a permis de montrer que ce système, appelé Mut, était responsable de la correction de longs segments d'ADN (plusieurs centaines à quelques milliers de bases). Trois gènes spécifiques essentiels au fonctionnement de ce système, les gènes *mutH*, *mutL* et *mutS*, ont été identifiés. Les souches *mut⁻* présentent, comme les souches *hex⁻*, un phénotype mutateur, ce qui indique que le système Mut joue un rôle de correction des erreurs de réplication.

La distinction entre ADN parental et ADN néosynthétisé s'appuie, dans ce cas, sur l'absence transitoire de méthylation [13] de l'adénine dans les séquences GATC, sur les brins néosynthétisés (figure 5, p. VI). Cette méthylation est introduite par la désoxyadénosine méthylase Dam, de façon postréplivative [12]. Le phénotype mutateur associé à un défaut de méthylation (mutants *dam⁻*) ou à une hyperméthylation (souche surproductrice de Dam) [8] et l'influence directe de la présence et du nombre de sites GATC sur le taux de mutation d'un gène porté par un plasmide [14] sont en accord avec ce mode de reconnaissance du brin à corriger, de même que la correction préférentielle du brin non méthylé sur les ADN hétéroduplex [8]. L'activité des différentes protéines Mut et les autres activités requises pour une réaction complète de correction *in vitro* sont décrites plus loin.

La spécificité de reconnaissance des mésappariements

L'analyse de la spécificité de correction du système Hex a montré que ce système reconnaît les mésappariements ponctuels avec des efficacités variables (pour revue, voir [8]). Les mésappariements A/C, G/T et G/G sont reconnus et corrigés avec une efficacité maximale et les autres mésappariements avec une efficacité décroissante :

m/s n° 6, vol. 8, juin-juillet 92

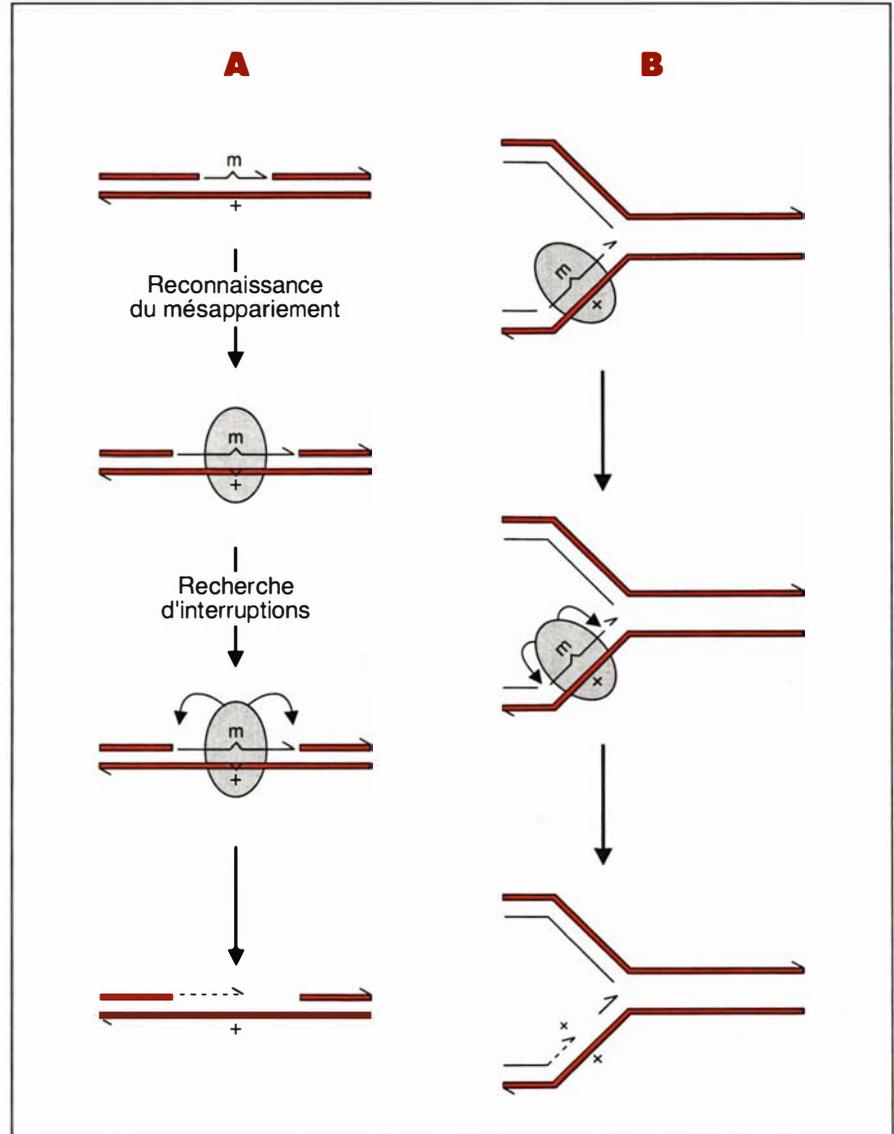


Figure 4. **Modèle d'action du système Hex sur les intermédiaires de recombinaison et sur les erreurs de réplication.** Après reconnaissance d'un mésappariement, le système Hex recherche des interruptions qui serviront de point d'entrée et de limites pour la correction. **A.** Dans le cas d'un hétéroduplex intermédiaire de recombinaison entre ADN exogène (trait fin) et ADN récepteur, les extrémités du fragment exogène sont des sites préférentiels d'initiation de la correction qui aboutira à l'élimination de ce fragment. **B.** Dans le cas d'une erreur de réplication, l'extrémité du brin en cours d'élongation (trait fin) et les interruptions éventuelles liées à une synthèse discontinue constituent autant de points d'entrée pour le système Hex (la correction pourrait être bornée, d'un côté, par le mésappariement). Dans les deux cas, la correction restaure l'information parentale.

C/T > A/A > T/T > A/G [15]. Le mésappariement C/C semble échapper à toute correction, voire inhiber la correction de mésappariements adjacents. Le système Hex corrige aussi efficacement les additions-délétions de 1 à

3 bases, mais pas celles de 5 bases et plus [8, 16]. La spécificité du système Mut s'est avérée très semblable à celle du système Hex (pour revue, voir [8, 12, 17]). Une influence des séquences environnantes sur la reconnaissance de cer-

tains mésappariements, en particulier A/G, a été observée dans les deux systèmes, sans qu'il soit possible d'établir de corrélation simple entre la nature de ces séquences et l'efficacité de reconnaissance. Ainsi, une relation directe avec la richesse en G-C des séquences environnantes, suggérée dans le cas du système Mut [18], n'a pas été confirmée [19] et n'a pas été constatée dans le cas du système Hex [15]. La plasticité conformationnelle du mésappariement A/G, beaucoup plus grande que celle des autres mésappariements (pour revue, voir [17]), pourrait expliquer l'influence des séquences environnantes sur sa reconnaissance ainsi que l'existence de plusieurs systèmes de correction spécifique de ce mésappariement, ou de prévention de sa formation.

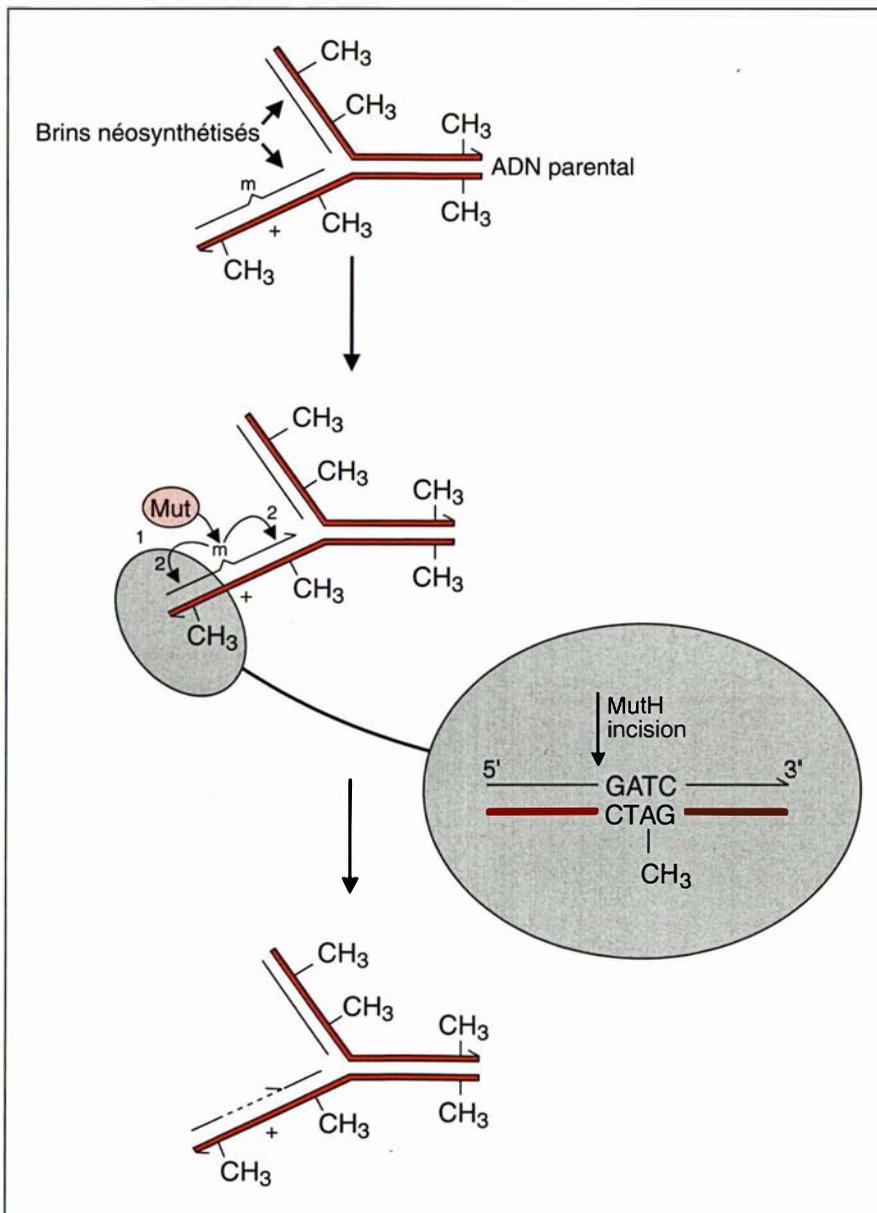


Figure 5. **Correction des erreurs de réplication par le système Mut.** La présence de groupes méthyl sur l'ADN parental est symbolisée par -CH₃. Après reconnaissance (1) d'un mésappariement (m/+), le système Mut effectue une recherche bidirectionnelle (2) de sites hémiméthylés. La protéine MutH incise alors le brin non méthylé comme indiqué pour initier la réaction de correction.

La réaction de correction reconstituée *in vitro*

L'analyse *in vitro* du mécanisme de correction par le système Mut et la purification des différentes protéines ont pu être abordées par la mise en œuvre d'un test particulièrement élégant. Le substrat de la réaction est une molécule circulaire fermée contenant un mésappariement placé à la jonction de deux sites de restriction. La correction de l'un des deux brins restaure l'un des sites et donc la sensibilité à l'une des enzymes de restriction (figure 6). L'état de méthylation des sites GATC du substrat peut être contrôlé, à volonté, lors de sa préparation, pour orienter la réparation sur l'un ou l'autre brin.

Grâce à ce test, les protéines MutH, MutL et MutS ont été purifiées à quasi-homogénéité sous une forme biologiquement active (pour revue, voir [4]). La protéine MutS est capable de se fixer sur une molécule d'ADN présentant un mésappariement et recouvre alors une dizaine de bases de part et d'autre du mésappariement. MutL est la protéine dont le rôle est actuellement le moins bien compris. Cette protéine pourrait servir d'interface entre les protéines MutS et MutH [4]. Elle se fixe, en présence d'ATP, sur les complexes formés entre MutS et les molécules

d'ADN comportant un mésappariement. Cet ensemble est alors capable d'activer la protéine MutH pour l'incision des sites d(GATC) non méthylés, en 5' du G (figure 5). Les trois protéines MutH, MutL et MutS sont seules requises pour assurer les étapes de reconnaissance du mésappariement et de choix du brin à réparer. Outre ces protéines, l'ADN hélicase II (le produit du gène *uvrD*), la protéine Ssb, l'exonucléase I (activité 3' → 5'), l'exonucléase VII ou l'exonucléase RecJ (activités 5' → 3') et l'ADN polymérase III holoenzyme (polymérase répliquative) sont nécessaires, très vraisemblablement pour réaliser les étapes d'excision et de resynthèse (en présence de dNTP), et enfin l'ADN ligase pour recirculariser le substrat. L'ADN hélicase II pourrait intervenir dans l'étape d'excision du brin à corriger, en permettant l'action des différentes exonucléases qui sont toutes spécifiques de l'ADN monocaténaire.

Les systèmes Hex et Mut sont apparentés

L'hypothèse d'un apparentement des systèmes de correction généralisée des bactéries *S. pneumoniae* et *E. coli* [8], s'appuyant sur les similitudes de spécificité et de mode de correction, s'est avérée tout à fait fondée puisque des homologies significatives ont été découvertes entre les protéines HexA et MutS [20] et entre les protéines HexB et MutL [21]. La protéine HexA est, comme la protéine MutS, capable de se fixer sur une molécule d'ADN présentant un mésappariement (Prudhomme, Claverys et Modrich, en préparation). En dépit de leurs homologies, les protéines Hex ne peuvent remplacer les protéines Mut *in vivo* dans *E. coli* [22]. Cependant, la protéine HexA est capable d'interactions abortives avec une ou plusieurs protéines Mut [22].

La différence la plus marquante entre les systèmes Hex et Mut est l'existence de la protéine MutH et son action sur les sites GATC hémiméthylés pour orienter la correction. L'hypothèse d'une surimposition de ce mode d'orientation sur un système ancestral

orientant la correction à partir d'une interruption a été avancée [8]. Cette proposition s'est révélée exacte puisque la protéine MutH n'est pas nécessaire lorsque le substrat comporte une interruption, aussi bien *in vivo* qu'*in vitro* [4]. Il faut remarquer que le système MutH-hémiméthylation présente l'avantage de permettre une

reconnaissance spécifique des brins néosynthétisés sur une distance qui peut être estimée à 80 kb en aval de la fourche de réplication, alors que les interruptions naturelles n'existent probablement pas au-delà de 3 à 4 kb du point de réplication.

Les premières étapes de la correction

Il est maintenant admis que la reconnaissance d'un mésappariement par la protéine MutS (HexA) est l'événement initial de la réaction. L'orientation de la correction nécessitant la reconnaissance d'un signal (interruption d'un brin ou site GATC) distinct et distant du mésappariement pose le problème d'une interaction à distance. La capacité de la protéine MutS purifiée à former une figure en α (comportant une boucle) sur une molécule d'ADN contenant un mésappariement [4] est en faveur d'un modèle de migration [17]. La participation directe de la protéine MutL (HexB) à cette étape n'est pas établie. Sa fixation sur les complexes protéine MutS-ADN permet néanmoins de supposer son intervention. Les protéines HexB et MutL pourraient être chargées de la reconnaissance de l'interruption [21]. L'élimination de séquences situées en 5' et en 3' du mésappariement par le système Hex [10] suggère que la recherche d'interruptions est bidirectionnelle à partir du mésappariement. Dans le cas du système Mut, la correction peut se faire à partir d'un site GATC unique, mais situé aussi bien en aval qu'en amont du mésappariement [4]. Une recherche bidirectionnelle semble donc avoir lieu dans les deux systèmes. Cette recherche bidirectionnelle peut être intégrée dans un modèle qui suppose la séquestration du mésappariement à l'intérieur d'une boucle d'ADN (figure 7, p. VIII).

Dans le cas du système Hex, les résultats obtenus lors de la transformation suggèrent que la correction concerne un fragment défini par des interruptions situées de part et d'autre du mésappariement. En revanche, la région corrigée par le système Mut *in vitro* concernerait un intervalle compris entre un site GATC et quelques bases

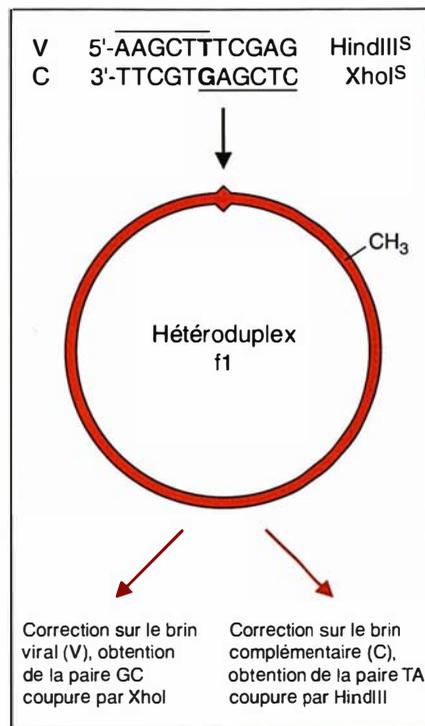
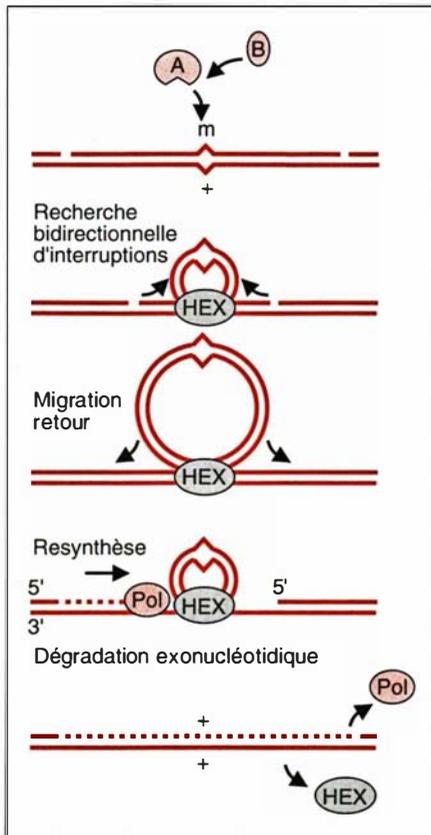


Figure 6. **Test biochimique pour l'étude de la correction des mésappariements *in vitro*.** Le substrat est une molécule circulaire fermée (6440 pb) préparée à partir de dérivés du bactériophage f1 contenant un ou plusieurs sites GATC. Le test est fondé sur l'introduction d'un mésappariement (ici G/T) à l'intérieur de sites de reconnaissance pour des enzymes de restriction (ici HindIII, séquence surlignée, et XhoI, séquence soulignée). La présence de ce mésappariement rend ce site résistant à ces deux enzymes. La correction sur l'un ou l'autre brin restaure la sensibilité à l'une des deux enzymes, comme indiqué. L'état de méthylation des sites GATC peut être contrôlé lors de la préparation du substrat. La distance la plus courte entre le site GATC et le mésappariement est de 1024 pb.

au-delà du mésappariement [4]. Il est cependant possible que l'utilisation d'un substrat circulaire fermé et la situation dissymétrique des sites GATC introduisent des contraintes topologiques telles qu'elles empêchent l'utilisation normale de deux sites GATC. Il



est aussi possible que le système Mut ait évolué de façon à interrompre la réaction de migration dès la rencontre du premier site (GATC ou interruption) pour initier la correction. Une migration à rebours pourrait alors se produire qui permettrait une nouvelle fixation de la protéine MutS sur le mésappariement séquestré à l'intérieur de la boucle. Cette fixation « retour » pourrait être le signal de fin de correction (figure 7). Dans un modèle d'excision *in situ* du segment à réparer, il faut remarquer que l'utilisation d'un site GATC unique situé indifféremment en aval ou en amont du mésappariement suppose l'existence de deux activités exonucléasiques de polarité opposée. En outre, la resynthèse ne pourra accompagner l'excision que lorsque le site GATC est situé en 5' du mésappariement. Dans le cas contraire, une longue région d'ADN monocaténaire sera créée, qui pourrait contribuer à l'initiation d'événements de recombinaison.

L'universalité des systèmes de correction généralisée des mésappariements

Une série d'observations suggère que les systèmes de correction généralisée des mésappariements existent chez la plupart des êtres vivants. La première

de ces observations est la découverte que le système Pms de *Saccharomyces cerevisiae* a une spécificité de correction tout à fait similaire à celle du système Hex [8]. Les mutants Pms ont un phénotype mutateur, impliquant ce système dans la correction des erreurs de réplication. Plus récemment, des systèmes de correction généralisée des mésappariements ont été mis en évidence à partir d'œufs de xénope et d'extraits nucléaires de cellules de drosophile et de cellules humaines Hela [4]. Leur spécificité de correction est proche de celle des systèmes procaryotes et le segment réparé est long de plusieurs centaines de nucléotides. Comme pour le système Hex, l'orientation de la correction est dictée par la présence d'une interruption monocaténaire dans les systèmes humain et de drosophile. La polymérase intervenant dans ce cas est la polymérase α , c'est-à-dire une polymérase impliquée dans la réplication [4]. Ce résultat est à rapprocher de l'intervention de la polymérase répliquative dans le cas du système Mut. Il apparaît donc qu'un même mode de fonctionnement a été conservé au cours de l'évolution.

L'existence d'homologies significatives entre les protéines HexB-MutL et la protéine Pms1 [21] et l'identification d'un gène codant pour une protéine homologue des protéines HexA-MutS chez la souris et l'homme [4] suggèrent en outre que les différents systèmes de correction généralisée sont apparentés.

Correction des mésappariements et recombinaison génétique

Les conséquences de l'action des systèmes de correction sur les intermédiaires de recombinaison peuvent être multiples, voire antagonistes. Les systèmes de correction spécifique qui corrigent de très courts segments d'ADN sont potentiellement capables de produire de courtes conversions (figure 8A), c'est-à-dire des transferts unidirectionnels d'information génétique d'une molécule d'ADN à une autre [2]. Ces systèmes ont une action diversificatrice potentielle soulignée par de nombreux auteurs. La formation de séquences recombinantes du type + b+ (figu-

m/s n° 6, vol. 8, juin-juillet 92

re 8A) ou $a+c$ (figure 8B), événement rare entre sites proches au cours de la recombinaison, correspond à ce type d'action et a été démontrée, par exemple dans le cas du système VSP de *E. coli* [4]. Du fait de la longueur du segment corrigé, les systèmes de correction

généralisée vont, contrairement aux systèmes de correction spécifique, entraîner une homogénéisation de séquence. Suivant le choix du brin réparé, leur action aboutira soit à une conversion longue (figure 8C), soit à l'avortement de la recombinaison (figure 8D).

L'action antirecombinogène du système Hex est évidente au cours de la transformation ; c'est la base même de sa mise en évidence (voir plus haut). L'action antirecombinogène du système Mut est aussi apparente pendant la conjugaison [23]. Une action antirecombinogène du système Mut a aussi été observée lors de la recombinaison entre séquences partiellement divergentes (séquences homéologues) [24, 25]. (Les ADN hétéroduplex formés entre séquences homéologues comportent de nombreux mésappariements et constituent donc un substrat potentiel pour le système Mut.) A partir de ces observations, l'hypothèse d'une implication des systèmes de correction généralisée des mésappariements dans la stabilité des chromosomes et dans la spéciation a été proposée [12] : la recombinaison entre séquences répétées homéologues est susceptible de produire des réarrangements chromosomiques importants (inversions ou délétions de grands segments) ; son inhibition contribuerait à la stabilité chromosomique. De même, le maintien d'une barrière à la recombinaison entre séquences homéologues favoriserait la spéciation.

Correction généralisée des mésappariements, stabilité chromosomique et spéciation

En ce qui concerne la stabilité chromosomique, l'inactivation du système Mut a pour conséquence une augmentation d'un facteur 10 environ des duplications chromosomiques de grande taille [26]. Cet effet paraît suffisamment limité pour que l'on s'interroge sur le rôle biologique de ce système dans la stabilité chromosomique. Il est possible que d'autres mécanismes, par exemple la structuration du chromosome en domaines indépendants, assurent une protection plus efficace vis-à-vis de la recombinaison entre séquences dupliquées homéologues. D'une manière plus générale, d'autres hypothèses séduisantes ont été avancées qui peuvent expliquer la faible fréquence de recombinaison entre séquences répétées en méiose et donc la sta-

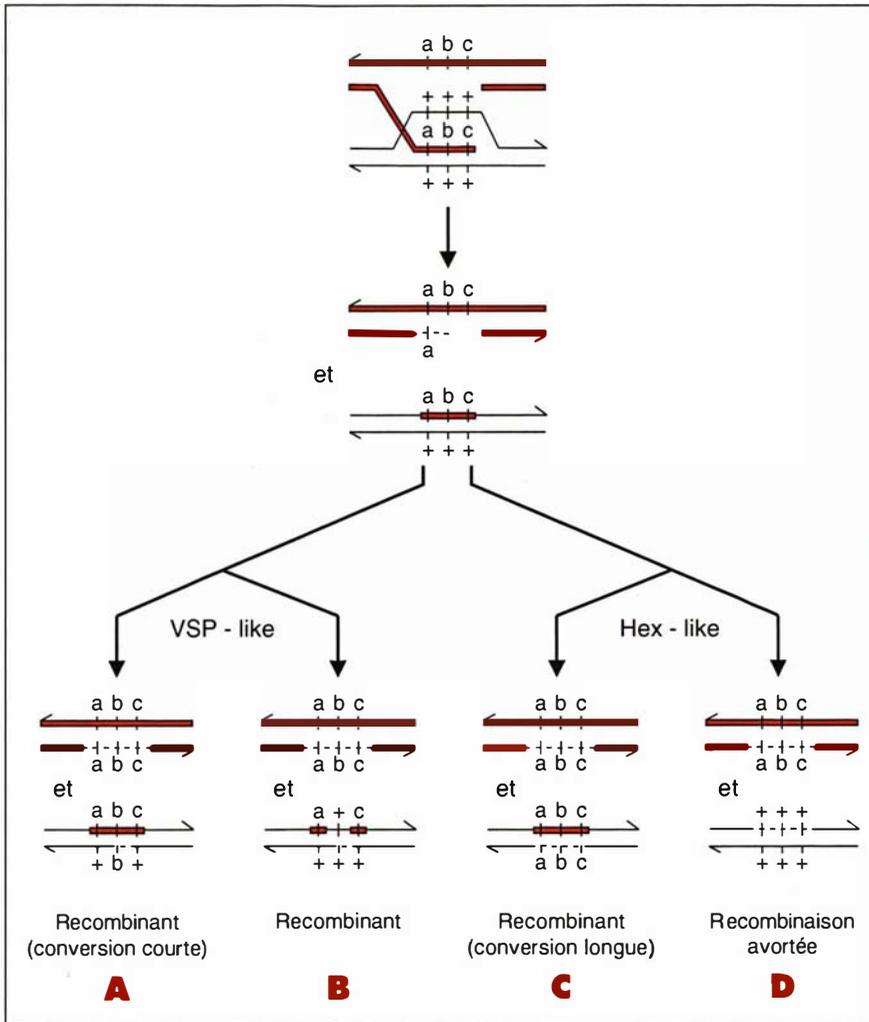


Figure 8. **Correction des mésappariements et recombinaison génétique.** Illustration des conséquences possibles de l'action des différents systèmes de correction des mésappariements sur les intermédiaires de recombinaison entre deux ADN duplex. La recombinaison, initiée (arbitrairement) par l'invasion d'un brin, conduit à la formation d'un hétéroduplex comportant trois mésappariements $a/+$, $b/+$ et $c/+$. La correction spécifique du mésappariement $b/+$ par un système analogue au système VSP produira, par correction en b/b , une conversion courte et un excès de recombinants $+b+$ (A) ou, par correction en $+/+$, un excès de recombinants $a+c$ (B). (Il faut noter qu'en l'absence de marqueurs flanquants a et c , cette correction de $b/+$ en $+/+$ aurait un effet antirecombinogène apparent, en réduisant le nombre de recombinants.) Un système de correction généralisée des mésappariements analogue au système Hex pourra produire une conversion longue (C) ou faire avorter la recombinaison en détruisant le brin envahissant (D).

bilité des chromosomes [2]. L'importance du système Mut dans la spéciation est par contre attestée par la démonstration que ce système est responsable des très faibles fréquences de recombinaison entre *E. coli* et *Salmonella typhimurium*, deux espèces dont la divergence est estimée à 15-20 % [25]. La recombinaison interspécifique est près de 1 000 fois plus efficace dans une souche *mut⁻*.

L'implication des systèmes de correction généralisée des mésappariements dans la spéciation au travers de l'inhibition de la recombinaison homéologue est-elle générale ? Pour tenter de répondre à cette question, nous avons récemment examiné l'action du système Hex au cours de la transformation homéologue chez *S. pneumoniae*. Nous avons constaté que, au-delà d'un certain seuil de divergence, le système Hex s'avérait incapable d'inhiber la recombinaison : la fréquence absolue de recombinants obtenue avec un fragment d'ADN présentant une divergence de 5 % est, dans une souche *hex⁺*, près de 10 fois supérieure à celle d'un fragment dont la divergence est de 2,2 % (Humbert et Claverys, résultats non publiés). Il y a blocage du système Hex, attribué à la saturation de sa capacité de correction par un hétéroduplex comportant un trop grand nombre de mésappariements. L'analyse de ce phénomène est actuellement en cours pour tenter de déterminer quels sont les éléments limitants du système Hex et le mécanisme de ce blocage. Il n'est en effet pas impossible qu'une correction partielle ait lieu, en particulier aux extrémités du fragment homéologue. Dans cette hypothèse, le système Hex aurait alors une action diversificatrice potentielle lors de la transformation homéologue. Une modulation de l'action de ce système lors de la recombinaison homéologue (par modulation de la capacité de correction des mésappariements ou de la capacité de formation des hétéroduplex) pourrait d'ailleurs contribuer à l'adaptabilité de cette bactérie dans des conditions changeantes d'environnement. Quelles que soient les conclusions à venir, il paraît d'ores et déjà évident que le système Hex ne joue pas le rôle de barrière à la recom-

* **GLOSSAIRE** *

Mésappariement : toute confrontation de bases autre que les paires adénine-thymine (A-T) et guanine-cytosine (G-C) dans une molécule d'ADN bicaténaire. Il existe huit mésappariements ponctuels : G/T, A/C, A/A, T/T, G/G, C/C, A/G et C/T.

ADN hétéroduplex : molécule d'ADN comportant au moins un mésappariement (voir ci-dessus).

Recombinaison génétique : processus conduisant à un changement dans l'association physique entre deux segments d'ADN.

ADN homéologues : molécules ayant une homologie imparfaite (degré de divergence de quelques % à 20-25 %).

Conversion : transfert unidirectionnel d'information d'un duplex d'ADN parental vers l'autre, ou d'un site chromosomique vers un autre dans le cas de séquences répétées

binaison interspécifique. Tout indique que les marqueurs génétiques que nous avons utilisés, issus d'isolats cliniques de *S. pneumoniae*, proviennent d'un transfert interspécifique [27]. Nos résultats suggèrent fortement que ce transfert a eu lieu par transformation interspécifique naturelle de souches *hex⁺*. Les deux systèmes de correction généralisée des mésappariements les mieux connus, les systèmes Hex et Mut, n'ont donc pas le même effet inhibiteur de la recombinaison homéologue. L'implication directe des systèmes de ce type dans la spéciation par inhibition de la recombinaison homéologue pourrait néanmoins être générale et le système Hex ne représenter qu'un cas particulier lié au statut d'organisme transformable de *S. pneumoniae*. La transformation, processus hautement spécialisé et très efficace de transfert génétique, pourrait avoir évolué pour permettre une grande adaptabilité de

l'espèce, y compris par acquisition de séquences hétérosécifiques. L'étude d'autres systèmes de correction généralisée des mésappariements permettra d'évaluer le caractère général de la contribution de ces systèmes à la spéciation. Actuellement, seules des données fragmentaires concernant le système Pms de *S. cerevisiae* sont disponibles. L'augmentation de la recombinaison homéologue mitotique observée dans un mutant *pms1* est limitée à un facteur 4,5 [28]. Il serait intéressant de connaître l'effet de ce système sur la recombinaison homéologue méiotique.

Conclusion et perspectives

L'implication des systèmes de correction généralisée des mésappariements dans la stabilité de l'information génétique par la correction des erreurs de réplication est bien établie. La coïncidence de spécificité de correction de ces systèmes avec certaines des erreurs les plus fréquentes de l'appareil répliatif (pour revue, voir [29]) peut être considérée comme un indice supplémentaire de ce rôle antimutateur. En revanche, la généralité de l'implication de ces systèmes dans la spéciation (au travers de l'inhibition de la recombinaison homéologue) reste à établir puisque, contrairement au système Mut, le système Hex ne semble pas capable de jouer un tel rôle. Néanmoins, la conservation des systèmes de correction généralisée des mésappariements tout au long de l'évolution atteste de leur importance biologique. Leur inactivation a probablement des conséquences très dommageables chez les eucaryotes supérieurs. Il n'est pas impossible que, dans un proche avenir, une relation directe soit établie entre certaines maladies génétiques telles que le syndrome de Bloom [30] et l'inactivation de l'un de ces systèmes.

J.-P. Claverys : directeur de recherche au Cnrs. Microbiologie et génétique moléculaires, Cnrs UPR 9007, université Paul-Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex, France.

Références

- Radman M, Wagner R. La réplication de l'ADN : un mécanisme « haute fidélité ». *Pour la Science* 1988 ; 132 : 28-34.
- Rossignol JL. La recombinaison homologue : mécanismes et conséquences. *médecine/sciences* 1990 ; 7 : IV-XI.
- Lieb M. Spontaneous mutation at a 5-methylcytosine hotspot is prevented by very short patch (VSP) mismatch repair. *Genetics* 1991 ; 128 : 23-7.
- Modrich P. Mechanisms and biological effects of mismatch repair. *Annu Rev Genet* 1991 ; 25 : 229-53.
- Hennecke F, Kolmar H, Bründl K, Fritz HJ. The *usr* gene product of *E. coli* K-12 is a strand — and sequence — specific DNA mismatch endonuclease. *Nature* 1991 ; 353 : 776-8.
- Bhatnagar SK, Bullions LC, Bessman MJ. Characterization of the *mutT* nucleoside triphosphatase of *Escherichia coli*. *J Biol Chem* 1991 ; 266 : 9050-4.
- Mostachfi P, Sicard AM. Polarity of localized conversion in *Streptococcus pneumoniae* transformation. *Mol Gen Genet* 1987 ; 208 : 361-3.
- Claverys JP, Lacks SA. Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiol Rev* 1986 ; 50 : 133-65.
- Ephrussi-Taylor H, Gray TC. Genetic studies of recombining DNA in pneumococcal transformation. *J Gen Physiol* 1966 ; 49 : 211-31.
- Prudhomme M, Méjean V, Martin B, Humbert O, Claverys JP. Generalized mismatch repair in *Streptococcus pneumoniae*. In : Dunny GM, Cleary PP, McKay LL, eds, *Genetics and Molecular Biology of Streptococci, Lactococci and Enterococci*. Washington DC : American Society for Microbiology, 1991 : 67-70.
- Tiraby G, Fox MS. Marker discrimination in transformation and mutation of pneumococcus. *Proc Natl Acad Sci USA* 1973 ; 70 : 3541-5.
- Radman M. Mismatch repair and genetic recombination. In : Kucherlapati R, Smith GR, eds, *Genetic Recombination*. Washington DC : American Society for Microbiology, 1988 ; 169-92.
- Wagner R, Meselson M. Repair tracts in mismatched DNA heteroduplexes. *Proc Natl Acad Sci USA* 1976 ; 73 : 4135-9.
- Claverys JP, Méjean V. Strand targeting signal(s) for *in vivo* mutation avoidance by post-replication mismatch repair in *Escherichia coli*. *Mol Gen Genet* 1988 ; 214 : 574-8.
- Gasc AM, Sicard AM, Claverys JP. Repair of single — and multiple — substitution mismatches during recombination in *Streptococcus pneumoniae*. *Genetics* 1989 ; 121 : 29-36.
- Gasc AM, Garcia P, Baty D, Sicard AM. Mismatch repair during pneumococcal transformation of small deletions produced by site-directed mutagenesis. *Mol Gen Genet* 1987 ; 210 : 369-72.
- Modrich P. DNA mismatch correction. *Annu Rev Biochem* 1987 ; 56 : 435-66.
- Jones M, Wagner R, Radman M. Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genetics* 1987 ; 115 : 605-10.
- Lu AL, Chang DY. Repair of single base-pair transversion mismatches of *Escherichia coli in vitro* : correction of certain A/G mismatches is independent of *dam* methylation and host *mutHLS* gene functions. *Genetics* 1988 ; 118 : 593-600.
- Priebe S, Hadi S, Greenberg B, Lacks SA. Nucleotide sequence of the *hexA* gene for DNA mismatch repair in *Streptococcus pneumoniae* and homology of HexA to MutS of *Escherichia coli* and *Salmonella typhimurium*. *J Bacteriol* 1988 ; 170 : 190-6.
- Prudhomme M, Martin B, Méjean V, Claverys JP. Nucleotide sequence of the *Streptococcus pneumoniae hexB* mismatch repair gene : homology of HexB to MutL of *Salmonella typhimurium* and to PMS1 of *Saccharomyces cerevisiae*. *J Bacteriol* 1989 ; 171 : 5332-8.
- Prudhomme M, Méjean V, Martin B, Claverys JP. Mismatch repair genes of *Streptococcus pneumoniae* : HexA confers a mutator phenotype in *Escherichia coli* by negative complementation. *J Bacteriol* 1991 ; 173 : 7196-203.
- Feinstein SI, Low KB. Hyper-recombining recipient strains in bacterial conjugation. *Genetics* 1986 ; 113 : 13-33.
- Shen P, Huang HV. Effect of base pair mismatches on recombination *via* the RecBCD pathway. *Mol Gen Genet* 1989 ; 218 : 358-60.
- Rayssiguier C, Thaler DS, Radman M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* 1989 ; 342 : 396-401.
- Petit MA, Dimpfl J, Radman M, Echols H. Control of large chromosomal duplications in *Escherichia coli* by the mismatch repair system. *Genetics* 1991 ; 129 : 327-32.
- Dowson CG, Hutchison A, Brannigan JA, et al. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proc Natl Acad Sci USA* 1989 ; 86 : 8842-6.
- Bailis AM, Rothstein R. A defect in mismatch repair in *Saccharomyces cerevisiae* stimulates ectopic recombination between homeologous genes by an excision repair dependent process. *Genetics* 1990 ; 126 : 535-47.
- Echols H, Goodman MF. Fidelity mechanisms in DNA replication. *Annu Rev Biochem* 1991 ; 60 : 477-511.
- Groden J, Nakamura Y, German J. Molecular evidence that homologous recombination occurs in proliferating human somatic cells. *Proc Natl Acad Sci USA* 1990 ; 87 : 4315-9.

Summary

Mismatch repair and genetic information stability

Mismatch repair systems are widespread in nature. Depending on the kinds of mismatch they process, these systems can be divided into two groups, generalized and specific. The former, found in cells of several organisms from bacteria to humans, are likely to have evolved from a common ancestor, and appear to serve as means of correcting DNA replication errors. Recognition of newly synthesized DNA strands, a prerequisite for antimutator action, is achieved through recognition of a strand targeting signal that is distinct from and distal to the mismatch. Thus generalized mismatch repair systems usually correct long DNA stretches. Specific mismatch systems on the other hand frequently correct mismatches present in resting DNA. This is the case for the VSP system of *Escherichia coli* which corrects G/T mismat-

ches originating through deamination of 5-methyl-C in specific sequences. Antimutator action is then achieved through directional repair (to G-C) and correction is restricted to a very limited number of bases. Due to their different mechanisms of correction, mismatch repair systems diversely affect genetic stability through action on (heteroduplex) DNA recombination intermediates. VSP-like systems are responsible for apparent multiple exchange events, potentially resulting in sequence diversification. Generalized mismatch repair systems frequently abort recombination. It has been proposed that this antirecombination effect reflects their functional role in controlling the fidelity of genetic exchange. These systems would contribute to chromosomal stability (by limiting exchanges between members of a gene family) and represent the major barrier to recombination between bacterial species. The generality of this hypothesis is discussed in light of results obtained with the two best characterized bacterial systems.