

# Amplification des séquences répétées : mutations imprévues ou prédispositions ancestrales ?

Avec la découverte de la première mutation dynamique à l'origine de la maladie de l'X fragile, est apparu un nouveau mode de mutation, retrouvé dans plus d'une dizaine de maladies humaines, la dernière en date étant la maladie de Friedrich [1]. Et la liste n'est sûrement pas close. Pour chacune de ces maladies, on connaît la distribution du nombre des triplets dans la population normale, ainsi que le seuil au-delà duquel la séquence devient instable, avec des hyperextensions survenant à la méiose et qui sont retrouvées à la génération suivante chez les sujets atteints. Mais quels sont les mécanismes qui amènent la séquence à ce seuil ? Pourquoi, alors que la distribution des triplets reste stable dans la population normale, certaines familles en sont-elles parvenues à ce point de non retour ?

Deux approches semblent fournir quelques indices : l'analyse des haplotypes et la recherche des triplets variants dispersés dans la séquence répétée. Bien que les premiers résultats obtenus pour la maladie de l'X fragile (Xfra) soient encore difficiles à interpréter, nous allons voir qu'ils sont encourageants et que des démarches analogues dans les autres maladies à expansion de triplets permettront sans doute prochainement d'y voir plus clair.

Concernant la région *FRAXA*, quatre *loci* polymorphiques, *DXS548*, *FRAXAC1*, *FRAXAC2* [2] et *FRAXE* situés de part et d'autre de la séquence répétée permettent d'établir des haplotypes (figure 1A). En ne tenant compte que des trois premiers microsatellites, il apparaît clairement que certains haplotypes sont observés préférentiellement dans les familles avec

Xfra, et que d'autres n'y sont qu'exceptionnellement rencontrés. Le déséquilibre de liaison est d'autant plus net que la population est isolée, comme en Finlande par exemple où 75 % des porteurs d'Xfra ont le même haplotype [3]. De plus, dans

ce pays nordique où les registres de population sont conservés depuis de nombreuses générations, la généalogie de certaines familles montre qu'en fait, elles sont apparentées par de lointains ancêtres vivant au XVIII<sup>e</sup> ou au XVII<sup>e</sup> siècle. Force est donc

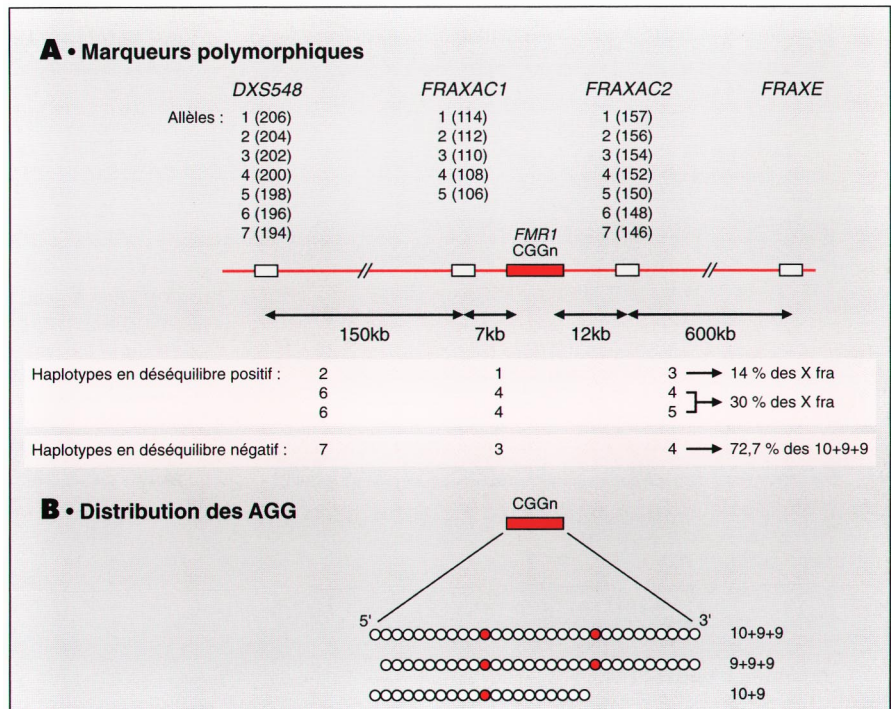


Figure 1. **Région FRAXA.** A. Position des marqueurs microsatellites autour de *FMR1*. *DXS548*, *FRAXAC1*, *FRAXAC2* et *FRAXE* flanquent la séquence *CGG* de *FMR1* et permettent de déterminer des haplotypes. Les allèles sont numérotés du plus grand au plus petit en fonction de la longueur du produit de PCR. *DXS548* a sept allèles (de 1 : 206 nucléotides à 7 : 194 nucléotides). *FRAXAC1* a cinq allèles et *FRAXAC2*, à 11 kb en amont de la séquence, sept allèles. B. Répartition des *AGG* dans la séquence répétée. Les *AGG* sont représentés par des cercles rouges et symbolisés dans le texte par un (+) après le nombre des *CGG* qui les précèdent et qui les suivent (d'après [8] et [10]).

d'admettre que certains allèles, prédisposés à subir un jour une hyper-extension, furent transmis silencieusement pendant des dizaines de générations [4]. Puisqu'on sait, par ailleurs, qu'il y a une absence de continuité entre l'allèle *FMRI* normal (contenant de 5 à 50 triplets CGG) et l'allèle prémuté (contenant de 59 à 200 CGG) [5], deux auteurs américains ont proposé un modèle multialélique en plusieurs étapes [6]: de l'allèle stable N, on passerait, par une petite expansion, à l'allèle S qui resterait encore stable pendant de nombreuses générations avant de devenir l'allèle Z prémuté. Le passage à la mutation complète L se ferait alors brutalement à travers les méioses féminines. Puisque l'allèle S reste stable longtemps, un événement particulier doit survenir qui le prédispose à l'instabilité, et cet événement devrait déjà être inclus potentiellement dans les haplotypes en déséquilibre de liaison. Il fallait donc examiner soigneusement les haplotypes prédisposés, « protoprémutes » en quelque sorte, pour voir en quoi ils différaient des autres.

On avait remarqué que la présence de triplets AGG disséminés parmi les triplets CGG conféraient à la séquence répétée une certaine stabilité; au contraire, dès qu'une séquence CGG pure dépasse 34 à 37 triplets CGG, elle devient instable, peut-être par la formation de boucles en épingle à cheveu, comme on en observe *in vitro* [7]. Il fallait donc recenser la distribution des AGG dans les allèles prédisposés.

Les AGG sont distribués de façon non aléatoire dans les différents haplotypes de la population générale [8] (figure 1B). Les distributions 10+9+9, 9+9+9, et 10+9 correspondent à 51,5% des chromosomes X étudiés, quels que soient les haplotypes. Le premier AGG en 5', souvent en position 10 ou 11, est assez stable. En revanche, le troisième AGG, en 3', est souvent porté disparu.

Chez les Européens (sur 214 sujets témoins et 30 sujets prémutés de famille d'Xfra), en se fondant sur les trois microsatellites *DXS548*, *FRAXA1* et *FRAXA2*, on observe que sur 28 haplotypes, 18 n'ont jamais été trouvés chez des prémutés et que deux

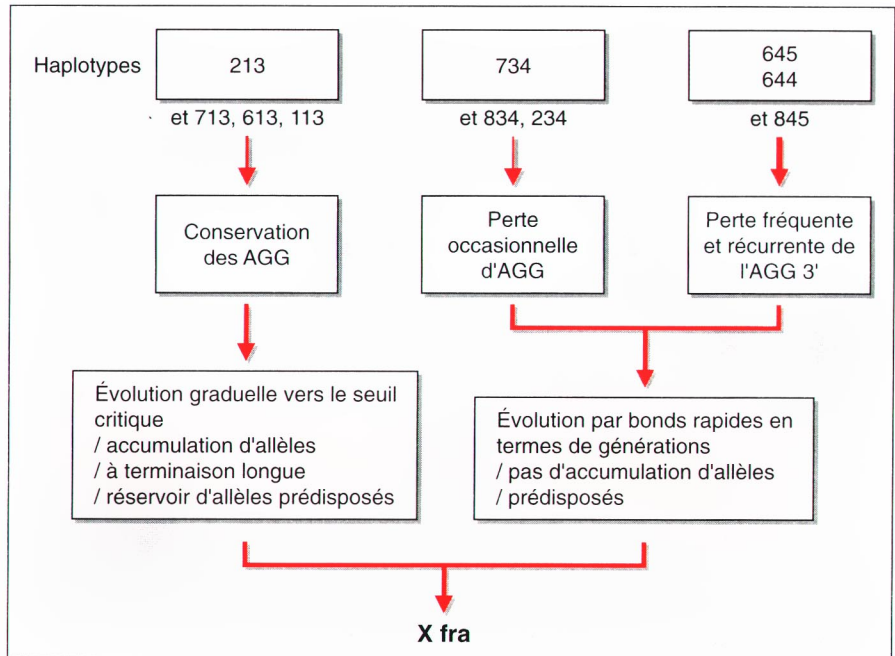


Figure 2. **L'hypothèse de la double évolution vers le syndrome Xfra.** En se fondant sur la distribution des AGG dans chacun des haplotypes et en utilisant la méthode de parcimonie, on distingue deux modes d'évolution vers le seuil critique, graduelle, ou par bonds avec perte du triplet AGG en 3' (d'après [8]).

groupes d'haplotypes sont en déséquilibre de liaison: les haplotypes 213 d'une part, et 644 ou 645 d'autre part. La fréquence et la position des AGG furent donc étudiées pour ces deux groupes ainsi que le nombre de CGG et la longueur de la séquence: (1) pour l'haplotype 213, les AGG restent très conservés, et les CGG ont une distribution régulière progressive; (2) les haplotypes 645 et 644 perdent au contraire leurs AGG et deviennent rapidement instables.

En analysant aussi la distribution des AGG dans d'autres haplotypes, rarement associés à l'Xfra, l'équipe de Patricia Jacobs [8] aboutit à la conclusion qu'il existe deux mécanismes conduisant à une séquence prémutée: soit un allongement progressif avec conservation des triplets AGG, soit une perte des AGG avec progression par bonds vers la prémutation (figure 2).

Pour parvenir à ces conclusions et tenter de retrouver les haplotypes ancestraux, ce même groupe a eu recours à l'analyse de parcimonie [9]. Cette méthode consiste à analyser toutes les substitutions observées

pour chaque site, puis de rassembler par deux, puis par groupes, ceux qui possèdent des singularités communes. L'analyse des résultats suggère que l'allèle 213 ancestral est probablement celui dont la distribution est 9+9+9. L'amplification se ferait lentement dans la région 3' en conservant au moins deux AGG et toute une série de transitions précéderait la prémutation. Au cours de l'évolution, la séquence de l'haplotype 213 se serait allongée progressivement jusqu'à atteindre un nombre de CGG assez important pour passer au stade prémuté. Par conséquent, cet haplotype a pu constituer, au cours du temps, un réservoir d'allèles à risque.

Les haplotypes 645 et 644, quant à eux, auraient eu tendance à perdre l'AGG situé en 3'. Cette suppression allongée d'un seul coup la séquence pure CGG. La survenue de la prémutation se serait donc faite par bonds, la séquence arrivant très vite à 70 et devenant instable.

Pour d'autres haplotypes de la population générale (734 par exemple), la perte d'un AGG peut survenir,

mais seulement occasionnellement, ce qui expliquerait que ces haplotypes soient rarement observés dans les familles d'Xfra (figure 2).

Des déséquilibres de liaison ont aussi été recherchés dans d'autres groupes humains, en étudiant à la fois les haplotypes et la position des AGG (aborigènes de Bornéo, Tibétains, groupes ethniques du Sénégal, et d'Indiens du Mexique et d'Argentine) [10]. Chez les Asiatiques et les Africains, on ne retrouve pas de déséquilibre de liaison mais la distribution des haplotypes est différente de celle des Européens. Pour les Indiens d'Amérique, on retrouve une plus grande homogénéité, probablement en raison du goulot d'étranglement au moment des migrations américaines. Leur séquence CGG est ponctuée de trois AGG, ce qui donne à penser que ces triplets se sont maintenus depuis des millénaires.

Mais ces interprétations, pour intéressantes qu'elles soient, sont encore bien fragiles. Car les haplotypes liés à l'Xfra sont souvent ceux où les microsatellites sont les plus longs. Or, il est possible qu'au cours de l'évolution (les études chez les primates tendraient à le prouver [11]), il existe une tendance générale unidirectionnelle à l'allongement des microsatellites. On aurait alors une fausse

impression d'un effet fondateur: la séquence répétitive et les microsatellites proches évolueraient ensemble vers l'allongement, par exemple sous l'influence d'une mutation d'un gène de réparation des mésappariements [12]. Il faut donc se garder de conclure pour l'instant, tout en suivant attentivement les études analogues portant sur d'autres maladies à expansion de triplets [13, 14].

## S.G.

1. Koenig M, Campuzano V, Cossée M, Mandel JL. Ataxie de Friedreich: les expansions de triplets frappent encore. *médecine/sciences* 1996; 12: 431-5.
2. Zhong N, Dobkin C, Brown WT. A complex mutable polymorphism located with the fragile X gene. *Nature Genet* 1993; 5: 248-53.
3. Haataja R, Vaisanen ML, Li M, Ryyanen M, Leisti J. The fragile X syndrome in Finland: demonstration of a founder effect by analysis of microsatellite haplotype. *Hum Genet* 1994; 94: 479-83.
4. Oudet C, von Koskull H, Nordstrom AM, Oeippo M, Mandel JL. Striking founder effect for the fragile X syndrome in Finland. *Eur J Hum Genet* 1993; 1: 181-9.
5. Dreyfus JC. Les mutations dynamiques, résultat de l'expansion de séquences répétitives. *médecine/sciences* 1992; 9: 983-4.
6. Morton NE, MacPherson JN. Population genetics of the fragile X syndrome: a multiallelic model for the FMR1 locus. *Proc Natl Acad Sci USA* 1992; 89: 4215-7.
7. Gacy MA, Goellner G, Juranic N, Macura S, McMurray CT. Trinucleotide repeats that expand

in human disease form hairpin structures *in vitro*. *Cell* 1995; 81: 533-40.

8. Eichler EE, MacPherson JN, Murray A, Jacobs PA, Chakravarti A, Nelson DL. Haplotype and interspersed analysis of the FMR1 CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome. *Hum Mol Genet* 1996; 5: 319-30.

9. Philippe H, Germot A, le Guyader H, Adoutte A. Que savons-nous de l'histoire évolutive des eucaryotes? 1. L'arbre universel du vivant et les difficultés de la reconstruction phylogénétique. *médecine/sciences* 1995; 11: I-XIII.

10. Kunst CB, Zervinick C, Karickhoff L, Eichler E, Bullard J, Chalifoux M, Holden JA, Torroni A, Nelson DL. FMR1 in global populations. *Am J Hum Genet* 1996; 58: 513-22.

11. Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li S-H, Margolis RL, et al. Microsatellite evolution - evidence for directionality and variation in rate between species. *Nature Genet* 1995; 10: 337-43.

12. Zhong N, Yang W, Dobkin C, Brown WT. Fragile X gene instability: anchoring AGGs and linked microsatellites. *Am J Hum Genet* 1995; 57: 351-61.

13. Brown XT, Zhong N, Dobkin C. Positive fragile X microsatellite associations point to a common mechanism of dynamic mutation evolution. *Am J Hum Genet* 1996; 58: 641-3.

14. Limprasert P, Nouri N, Heyman RA, Nopparatana C, Kamonsilp M, Deininger PL, Keats BJB. Analysis of CAG repeat of the Machado-Joseph gene in human, chimpanzee and monkey populations: a variant nucleotide is associated with the number of CAG repeats. *Hum Mol Genet* 1996; 5: 207-13.

15. Yanagisawa H, Fujii K, Nagafuchi S, Nakahori Y, Nagakome Y, Akane A, Nakamura M, Sano A, Komure O, Kondo I, et al. A unique and multi-step process for the generation of expanded DR-PLA triplet repeats. *Hum Mol Genet* 1996; 5: 373-9.

## ERRATUM

*médecine/sciences* n° 5, vol. 12, Mai 1996, pages 610 et 611

Dans l'article de Sylvie G. Bernier et Gaétan Guillemette, un incident technique a provoqué l'inversion des documents des figures 1 et 2.

La Rédaction