(19) RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE

PARIS

(11) N° de publication : **2 718 753**
(à n'utiliser que pour les commandes de reproduction)

(21) N° d'enregistrement national : **94 04554**

(51) Int Cl[6] : C 12 Q 1/68

(12) DEMANDE DE BREVET D'INVENTION A1

(22) **Date de dépôt** : 15.04.94.

(30) **Priorité** :

(43) **Date de la mise à disposition du public de la demande :** 20.10.95 Bulletin 95/42.

(56) **Liste des documents cités dans le rapport de recherche préliminaire :** *Se reporter à la fin du présent fascicule.*

(60) **Références à d'autres documents nationaux apparentés :**

(71) **Demandeur(s) :** *INSTITUT PASTEUR Fondation privée reconnue d'utilité publique* — FR et *CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE Etablissement public* — FR.

(72) **Inventeur(s) :** Canard Bruno et Sarfati Simon.

(73) **Titulaire(s) :**

(74) **Mandataire :** Gutmann Ernest - Plasseraud Yves S.A.

(54) Procédé de comptage de mono-, di- et trinucléotides répétés dans un génome eucaryote et trousse permettant la mise en œuvre de ce procédé.

(57) L'invention concerne un procédé de détermination du nombre de répétitions d'un motif possédant au plus trois des quatres bases nucléotidiques dans une séquence répétée, elle-même incluse dans une séquence d'acide nucléique, le cas échéant couplé à un support solide et caractérisé par les étapes suivantes:

a) hybridation d'une amorce nucléotidique d'élongation avec une séquence complémentaire d'une séquence située en 5' dudit microsatellite,

b) l'addition d'un mélange de 2'-désoxynucléotide 5'-triphosphates (A), de 2', 3'-didésoxynucléotide triphosphate (B) marqué ou non et de 2'-désoxynucléotide 5'-triphosphate substitué en 3' (C) et marqué ou non sur le groupe de substitution,

c) addition d'une acide nucléique polymérase dans des conditions de tampon et de température compatibles avec l'incorporation de n'importe lequel des trois désoxynucléotides (A), (B) et (C), et permettant la susdite incorporation,

d) action d'un composé chimique hydrolysant en 3' ou d'une enzyme de type estérase en 3' sur la chaîne en élongation,

e) séparation du surnageant de la séquence d'acide nucléique obtenue par élongation et mesure du marquage éventuel relargué dans ledit surnageant,

f) réitération des étapes b) à e),

g) comptage du nombre de répétitions après comparaison de la présence ou de l'absence de marquage entre les différentes réitérations.

FR 2 718 753 - A1

PROCEDE DE COMPTAGE DE MONO-, DI- ET TRINUCLEOTIDES REPETES DANS UN GENOME EUCARYOTE ET TROUSSE PERMETTANT LA MISE EN OEUVRE DE CE PROCEDE

La présente invention est relative à un procédé utilisant un mélange de 2'-désoxyribonucléotides 5'-triphosphates substitués ou non en 3' pour déterminer le nombre n de répétitions d'un, deux, trois nucléotides, ou d'un motif multinucléotidique ne contenant qualitativement que trois sur quatre des bases classiques de l'acide nucléique dans une séquence d'acide nucléique d'un génome eucaryote ou procaryote. De telles situations sont au moins au nombre de trois et extrêmement informatives : on pourra citer à titre d'exemple :

- certains motifs nucléotidiques répétés n fois dans le génome, dont un exemple évident est la séquence télomérique de type $(TTAGGG)_n$; le nombre n dans ce cas est lié au vieillissement cellulaire, et plus n est petit, plus la cellule est vieille,
- les minisatellites qui sont des régions du génome caractérisées par la répétition en tandem d'une même séquence d'ADN ; le nombre de répétitions varie d'un sujet à un autre, engendrant un multi-allélisme informatif en lui-même ; certains minisatellites sont localisés sur un seul locus, d'autres, au contraire, sont dispersés sur plusieurs chromosomes,
- les microsatellites qui sont des segments d'ADN contenant des répétitions en tandem de courts motifs "mono-", "di-" "tri-" ou "tétranucléotidiques" (dont le plus typique est le doublet CA/GT). Très nombreux et dispersés sur tout le génome, les microsatellites sont le siège de variations du nombre de répétitions, génératrices de polymorphismes multi-alléliques détectables,

contrairement aux minisatellites, uniquement après amplification par PCR.

Il a été découvert récemment que ces séquences microsatellites peuvent servir de marqueurs génétiques très utiles pour la cartographie des génomes des organismes supérieurs, ou encore dans l'analyse génétique de la transmission des maladies héréditaires (1).

Plus récemment encore, Wooster et al. (1994) (2) ont montré l'existence d'une instabilité des répétitions microsatellites dans des tumeurs ; initialement, cette instabilité avait été démontrée dans des tumeurs colorectales (3, 4 et 5). Cette instabilité des microsatellites a été maintenant détectée dans une grande variété de tumeurs incluant les cancers du sein, les sarcomes du cerveau et de l'ovaire et le procédé de l'invention peut également permettre la détection de telles cellules tumorales.

L'unité de base des séquences répétées dans le génome, notamment humain, s'étend d'une paire de base à plusieurs milliers de paires de base ; cependant, l'instabilité n'apparaît pas être équivalente dans chaque classe de répétition ; par exemple dans les minisatellites, il a été très rarement observé des allèles de taille augmentée ; au contraire, dans les microsatellites, plus la taille de la séquence de base répétée est petite, plus on observe des allèles de taille augmentée et cela a été notamment observé dans les répétitions de trinucléotides ou de tétranucléotides.

Une autre des caractéristiques essentielles de ces marqueurs génétiques que sont les microsatellites est leur polymorphisme, c'est-à-dire leur variabilité d'un individu à l'autre dans le nombre de répétitions et dans le motif nucléotidique qui comprend 1 (6), 2 (7), 3 (8) ou davantage de désoxynucléotides répétés dans l'ADN considéré.

Les répétitions de deux désoxynucléotides comme les $(dCdA)_n$ sont utilisées actuellement dans la construction de cartes génomiques dans le cadre de programmes mondiaux tel le projet "Génome Humain" (9) ou chez d'autres mammifères (10).

Les répétitions de trois désoxynucléotides sont liées à un nouveau groupe de maladies génétiques dans lequel la variation du nombre de répétitions du motif tridésoxynucléotidique a une signification diagnostique et pronostique très importante (8, 11, 12, 13).

Dans tous les cas, la détection du polymorphisme entre plusieurs échantillons, c'est-à-dire le nombre de répétitions du motif mono-, di- ou tridésoxynucléotidique fait appel aux techniques de biologie moléculaire classiques ; ces dernières sont l'amplification par exemple par PCR du fragment contenant le microsatellite, associée à l'analyse électrophorétique sur gel de polyacrylamide dénaturant, ou encore le séquençage nucléotidique du microsatellite par la technique de Sanger, améliorée grâce aux séquenceurs semi-automatiques (14).

Cependant, il a été remarqué que la détermination précise du nombre de répétitions est souvent problématique, voire très difficile, à l'aide de la technique PCR, car les patrons en échelle visibles (ladder pattern) sur gel de polyacrylamide après électrophorèse, ne peuvent pas être convenablement supprimés malgré les améliorations apportées dans les conditions dénaturantes (9). Dans le cas du séquençage des microsatellites (par exemple à l'aide du marqueur M13 de référence), la séquence répétée étant très courte, la détermination du nombre de répétitions est trop dépendante de la séquence considérée pour être évaluée sur les séquenceurs semi-automatiques (9). Enfin, quelle que soit la méthode utilisée parmi celles décrites ci-dessus, il est nécessaire de séparer les fragments et donc d'utiliser une électrophorèse sur gel

de polyacrylamide, ce qui oblige l'utilisateur à manipuler des produits dangereux (acrylamide, agent dénaturant, etc...), et surtout ce qui exclut une automatisation totale du processus.

Les techniques de séquençage elles-mêmes sont, à l'heure actuelle, en pleine évolution vers une automatisation complète. Une des conditions requises est la suppression de l'étape d'électrophorèse sur gel, comme en témoignent les travaux de Strezoska et al. (1991)(15) grâce à l'hybridation, de Driscoll et al. (1990)(16) et la microscopie à effet Tunnel de Gibbs et al. (1993)(demande WO93/05183), ou l'incorporation de nucléotides modifiés (demande de brevet FR9303538).

Dans le cas de cette dernière méthode, l'ADN à séquencer est mis à incuber avec un mélange de 2'-désoxynucléotides 5'-triphosphates substitués en 3' de manière réversible par un marqueur (3'-RT-NTP), une amorce et une polymérase. L'addition de la base correspondante, selon l'appariement Watson-Crick, permet, après déprotection en 3' et identification du marqueur, de déduire quelle base se trouvait sur l'ADN inconnu en face de la nucléobase incorporée.

Les 3'-RT-NTP peuvent être avantageusement de trois types dont aucun n'affecte l'incorporation du nucléotide qui est rapide et sélective :

- le premier type de dérivés représentés sur la figure 1 est constitué de 3'RT-NTP dans laquelle la substitution en 3' est une estérification par différents acides anthranyliques ; la libération du marqueur est réalisée dans une solution de soude 0,1 N ; cependant, ces conditions pouvant conduire à une instabilité du duplex amorce-brin d'ADN, une amorce en épingle à cheveu a été utilisée ;
- un deuxième type de dérivés 3'RT-NTP est constitué par des dérivés caproïlates présentant une substitution en 3' par un fluorochrome lié au

nucléotide par l'intermédiaire d'une chaîne carbonée de 2 à 12 atomes de carbone et de préférence de 4 à 8 atomes de carbone. L'avantage de ces dérivés est qu'ils sont non seulement hydrolysables par la soude mais également par une estérase, ce qui permet de préserver la stabilité du double brin amorce-brin d'ADN à séquencer, laquelle stabilité est compromise par l'utilisation de la soude ;

- un troisième type de dérivés 3'RT-NTP est constitué par les 2'-désoxynucléotides triphosphates acylés en 3' par l'acide 6-amino-2,3,6-tridésoxy-D-gluconique substitué sur la fonction amine par un fluorochrome, variable selon le cas.

Les avantages de ce dernier type de dérivés sont multiples : on peut citer le fait qu'ils soient déprotégeables en moins d'une minute à température ambiante, et à pH neutre, et la déprotection de l'hydroxyle en 3' pouvant se faire par oxydation par le périodate de manière quantitative.

Dans le présent texte, on appellera indifféremment microsatellite, séquence microsatellite ou nucléotide répété en tandem (NRT) toute série répétée de 1, 2 ou 3 nucléotides, le nombre de répétitions étant au moins égal à 2.

La présente invention est un procédé de détermination du nombre de répétitions d'un motif possédant au plus trois des quatre bases nucléotidiques dans une séquence répétée, elle-même incluse dans une séquence d'acide nucléique le cas échéant couplée à un support solide, et caractérisé par les étapes suivantes :

a) hybridation d'une amorce nucléotidique d'élongation avec une séquence complémentaire d'une séquence située en 5' dudit microsatellite,

ladite amorce permettant son élongation par incorporation enzymatique des désoxynucléotides,

b) addition d'un mélange de 2'-désoxynucléotide 5'-triphosphates (A), d'un ou de 2',3'-didésoxynucléotide triphosphate (B) marqué ou non, et d'un 2'-désoxynucléotide 5'-triphosphate substitué en 3' (C) et marqué ou non sur le groupe de substitution,

c) addition d'une acide nucléique polymérase dans des conditions de tampon et de température compatibles avec l'incorporation de n'importe lequel des trois désoxynucléotides (A), (B) ou (C), et permettant l'incorporation susdite,

d) action d'un composé chimique hydrolysant en 3' ou d'une enzyme de type estérase en 3' sur la chaîne en élongation,

e) séparation du surnageant de la séquence d'acide nucléique obtenue par élongation et mesure du marquage éventuel relargué dans ledit surnageant,

f) réitération des étapes b) à e),

g) comptage du nombre de répétitions après comparaison de la présence ou de l'absence de marquage entre les différentes réitérations.

Cette séquence d'étapes peut être mise en oeuvre tant pour les microsatellites, les minisatellites, que pour toutes les séquences répétées à partir du moment où le motif de base contient au plus trois des quatre bases nucléotidiques.

Le marquage en 3' peut être de toute nature : radioactivité, fluorescence, chimioluminescence, enzymatique ou autre ; à ce titre, on pourra parler indifféremment de "marqueur", ou l'anglicisme de type "group reporter", ou de "marquage sur l'ester en 3'", ou d'une combinaison de ces termes.

Le nombre de fois où ce processus est réitéré dépend de la détection ou, suivant le type de 2'-désoxynucléotide substitué en 3' utilisé, de l'absence de détection de marqueur dans le surnageant ; le nombre

de répétitions est directement corrélé au nombre de réitérations.

Chaque désoxynucléotide (A), (B) et (C) remplit une fonction particulière après avoir été incorporé dans la chaîne en élongation :

- les 2'-désoxynucléotides (A) assurent une polymérisation normale en présence de l'acide nucléique polymérase à partir du moment où la base complémentaire existe dans le motif de base du microsatellite ;
- les 2'-désoxynucléotides 3' substitués (C) assurent un arrêt de la polymérisation du fait de leur hydroxyle en 3' non disponible ; ce sont donc des "terminateurs de chaîne" ; mais, la substitution en 3' choisie est rendue réversible par les moyens chimiques ou enzymatiques expliqués plus haut ; donc, la polymérisation peut être reprise si l'expérimentateur déprotège la position 3' ; cette déprotection permet à la fois de détecter la présence d'une base particulière à cet endroit et permet également la reprise de la polymérisation par l'acide nucléique polymérase en présence des nucléotides (A), (B) ou (C) ;
- les 2',3'-didésoxynucléotides (B) assurent un arrêt définitif et irréversible de la polymérisation. Si la nature du ou des 2',3'-didésoxynucléotides est telle que leur incorporation n'a lieu qu'à l'extérieur du microsatellite, alors cette incorporation sera le signe de la fin de l'étendue du microsatellite au RNT ; cet ensemble de réactions est représenté sur la figure 2.

Il est intéressant de souligner plusieurs spécificités de ce procédé :

- le marqueur n'a pas besoin d'être spécifique de la base incorporée. Un seul marqueur peut être utilisé puisqu'un seul nucléotide est marqué, lequel nucléotide étant par exemple le deuxième d'un motif de deux ou le troisième d'un motif de trois, ce qui permet de mettre un choix très large de marqueurs réversiblement fixés en 3'. Les

critères dudit choix seront donc la facilité d'insertion par la polymérase, la facilité de déprotection, et la facilité de détection.

- le 2'-désoxynucléotide substitué en 3' (C) ("terminateur réversible de chaîne") n'est pas obligatoirement marqué en 3' ; il peut simplement être bloqué en 3' réversiblement par un groupement non détectable ; dans ce cas, le terminateur irréversible (B) 2',3'-didésoxynucléotide porte un marqueur en tout endroit de sa molécule qui n'altère pas ses capacités d'incorporation irréversibles. On pourra dans ce cas utiliser les didésoxynucléotides marqués par un fluorochrome sur la base disponible commercialement. Dans ce mode de réalisation, l'apparition de fluorescence dans la chaîne en élongation après rinçage et déprotection indique la fin du processus après n cycles tels que définis dans les étapes b) à e) citées plus haut.

- le mélange réactionnel doit comprendre des substrats désoxynucléotidiques (A), (B) et (C) d'affinité et de réactivité comparables pour éviter des incorporations non légitimes ou mésappariements par la polymérase. Toutefois, un certain biais peut être corrigé en utilisant des concentrations appropriées de chaque espèce de désoxynucléotide (A), (B) ou (C).

- la nature des désoxynucléotides utilisés est dictée par la connaissance préalable des NRT ou microsatellites analysés. Ainsi, le 2',3'-didésoxynucléotide sera le complémentaire de la base absente dans la séquence analysée, et son incorporation indiquera l'arrêt de la polymérisation hors de la séquence analysée. Le motif nucléotidique dont on cherche à connaître le nombre de répétitions n'est pas de taille limitée à 3 nucléotides. Le seul impératif est qu'il ne

contienne que trois des quatre bases possibles G, A, T ou C. Par exemple, les séquences télomériques de mammifères dont la séquence est généralement TTAGGG repété n fois se prêtent bien à la détermination du nombre n ; dans ce cas, le 2',3'-didésoxynucléotide utilisé sera le ddG, complémentaire à dC absent des répétitions. Dans le cas général, le 2',3'-didésoxynucléotide utilisé sera le complémentaire de la quatrième base absente dans la séquence analysée.

L'expérimentateur saura, au cas par cas, déterminer quel type de nucléotide doit être marqué réversiblement (nucléotide de type C) de façon à éviter tout problème d'interprétation du résultat du nombre n ainsi calculé ; autrement dit d'éviter le risque de compter comme motif de base de la répétition étudiée un autre motif qui aurait la caractéristique à la fois de contenir le nucléotide de type C et de ne pas contenir le ou les nucléotides absents du motif de base de la séquence répétée étudiée.

Afin d'assurer une meilleure fiabilité du procédé, il pourra être avantageux dans certains cas et quand la séquence d'acides nucléiques contenant le microsatellite est présente à une concentration non détectable, de multicopier ladite séquence par une méthode d'amplification quelconque, par exemple la méthode PCR (voir notamment les brevets européens EP201184 et EP200362).

Le procédé selon l'invention se caractérisé donc en ce que les étapes a) à g) citées ci-dessus peuvent être précédées par une étape d'amplification de la séquence d'acide nucléique.

En outre, cette amplification peut être réalisée avec une amorce ou un couple d'amorces modifiée(s) couplable(s) à un support solide et, par voie de conséquence, les produits d'amplification PCR peuvent également être couplés audit support solide. Le

couplage de l'amorce au support solide peut être avantageusement un couplage biotine-avidine ou biotine-streptavidine et le support peut être avantageusement une plaque de microtitrage ou un support magnétique.

Le procédé de l'invention est en outre caractérisé par le fait que le 2'-désoxynucléotide 5'-triphosphate substitué en 3' est substitué par l'un des groupements suivants :

- un groupement anthranylate,
- un ester de l'acide 6-amino-caproïque substitué par un groupement fluorescent ou radioactif,
- un ester de l'acide 6-amino-2,3,6-tridésoxy-D-gluconique marqué soit par un élément radioactif, soit substitué par un groupement fluorescent, soit par tout groupement détectable par un procédé chimique, enzymatique ou physique.

La séquence d'acide nucléique contenant le microsatellite peut être, le cas échéant, préalablement multicopiée et marquée à une de ses extrémités par la biotine, elle-même susceptible de se coupler à un support modifié par un groupement avidine ou streptavidine.

Les séquences multicopiées sont elles-mêmes couplables sur un support, puisqu'une amorce d'amplification est modifiée de telle façon qu'elle puisse être couplée sur ledit support, lequel avantageusement peut être magnétique.

L'invention décrit un procédé caractérisé en ce que l'hydrolyse de l'ester du nucléotide substitué en 3' est effectuée par voie enzymatique et notamment par une estérase.

L'invention concerne également une trousse d'identification du nombre de répétitions d'un motif d'une à trois bases nucléotidiques dans une séquence répétée, elle-même incluse dans une séquence d'acide

nucléique, le cas échéant couplée à un support solide, caractérisée en ce qu'elle contient au moins :

a) un ou deux désoxynucléotides 5'-triphosphates (A),

b) un 2',3'-didésoxynucléotide triphosphate (B) marqué ou non,

c) un 2'-désoxynucléotide 5'-triphosphate estérifié en 3', ledit ester étant hydrolysable par voie chimique ou enzymatique (C),

chacun des désoxynucléotides triphosphates en a), b) et c) étant porteur d'une base différente, (A) et (C) étant présents et (B) absent dans le motif répété, notamment une séquence microsatellite,

d) une polymérase ou tout moyen de polymérisation d'acide nucléique,

e) le cas échéant, une estérase susceptible d'hydrolyser l'ester en 3' de (C).

L'invention décrit également une trousse caractérisée en ce que le 2' désoxynucléotide triphosphate est estérifié par l'acide 6-amino-2,3,6-tridésoxy-D-gluconique et qu'il est hydrolysable par oxydation à l'aide de périodate.

L'invention a également pour objet une trousse caractérisée en ce que le 2'-désoxynucléotide 5'-triphosphate substitué en 3' est acylé par l'acide 6-amino-caproïque, lui-même marqué par un radio-isotope ou un fluorochrome.

L'invention est aussi relative à une trousse caractérisée en ce qu'elle contient en outre :

- une amorce ou un couple d'amorces,
- une ADN polymérase nécessaire à l'amplification de la séquence d'acide nucléique contenant ledit microsatellite,

l'amorce ou une des amorces étant susceptible d'être couplée chimiquement à un support solide, tous ces constituants permettant d'amplifier la séquence contenant le microsatellite, préalablement à la mesure du nombre de répétitions de la séquence de base.

L'utilisation d'une trousse de détermination selon l'invention à la construction de cartes génomiques chez l'homme ou l'animal ou encore à l'analyse génétique de la transmission de maladies héréditaires fait également partie de l'invention.

La détection du polymorphisme est souvent réalisée par la localisation de microsatellites de type $(CA)_n$; quant à la détection de maladies génétiques héréditaires, on a pu identifier plusieurs répétitions de trinucléotides, par exemple dans la Chorée de Huntington, la myopathie de Steinert, le retard mental dû au X-fragile, etc...

Ce type d'amplification de triplets est, à l'heure actuelle, reconnu comme étant lié à un nouveau groupe de maladies génétiques. Le nombre de répétitions est corrélé au pronostic de la pathologie, d'où l'importance du procédé de l'invention pour, par exemple, une prédiction in utero d'une telle pathologie.

La présente invention concerne également l'utilisation d'une trousse de détermination telle que décrite plus haut dans le diagnostic ou le pronostic de la présence d'un cancer ou de cellules cancéreuses chez l'homme ou l'animal.

Le procédé de l'invention est automatisable grâce au fait notamment qu'il n'utilise pas d'étape d'électrophorèse sur gel et que toutes les opérations nécessaires à sa réalisation (pipetage, rinçage, etc...) peuvent être effectuées par un robot.

- Description détaillée de l'invention:

L'invention est décrite plus en détail dans la description ci-après accompagnée des figures dont la signification est la suivante.

La Figure 1 représente la structure des désoxynucléotides utilisés dans le mélange réactionnel. M peut être n'importe quel substituant n'empêchant pas

la régénération, par traitement physique, chimique ou enzymatique, de l'hydroxyle en 3'. En particulier, M peut être également porteur d'atomes radioactifs, ou être un radical fluorescent. De tels désoxynucléotides 3' substitués ont déjà été synthétisés et décrits (demandes de brevets WO93/05183 et FR9303538).

La Figure 2 représente le système opérationnel utilisé pour le comptage des (dCdA) répétés n fois dans une séquence d'ADN, c'est-à-dire la détermination du nombre n de dinucléotides inclus dans le microsatellite. Dans cet exemple, le marqueur utilisé est radioactif, et son clivage est réalisé par une enzyme possédant une activité estérase. L'amorce réactionnelle est hybridée sur la matrice d'ADN simple brin immobilisée sur un support solide.

Le nombre de dinucléotides (CA) du microsatellite est de 12. Le mélange réactionnel comprend un dGTP substitué en 3' par du 14C butyryle, du dTTP, du 2',3'-didésoxy CTP et du 2',3'-didésoxy ATP ; une ADN polymérase permet l'intégration dans une chaîne en croissance de ces différents éléments.

Dès que l'ADN polymérase incorpore le ddCTP ou le ddATP, l'arrêt de la croissance de la chaîne est immédiat et définitif et marque la terminaison de la répétition du microsatellite.

Au contraire, quand le 3'-14C-butyryle-dGTP incorporé en face d'une cytosine, il y a également arrêt de la chaîne, mais cet arrêt est réversible et la désestérification est effectuée par une estérase.

Après comptage de la radioactivité du surnageant, le cycle peut recommencer jusqu'à l'intégration d'un 2',3'-désoxynucléotide.

La Figure 3 décrit le système utilisé pour compter le nombre de fois où le triplet CAG est répété à l'intérieur de la séquence nucléotidique du gène codant pour le récepteur à l'androgène humain.

Dans cet exemple, le marqueur utilisé est fluorescent, son clivage hors du cycle ribose de la nucléobase insérée se fait en milieu alcalin, et l'amorce utilisée est liée de manière covalente au fragment d'ADN simple brin à étudier et ceci selon le protocole décrit dans la demande de brevet FR9303538 : il s'agit des amorces dites "en épingle à cheveu" dans lesquelles l'extrémité 5' de l'amorce est liée de façon covalente par l'intermédiaire 2 à 10 nucléobases de la partie complémentaire correspondante de la séquence d'acide nucléique porteuse de la séquence répétée à analyser.

Dans cette figure, le nombre de répétitions du triplet CAG (ou son complémentaire CTG) est de 43 et l'indication de fin du microsatellite est effectuée par incorporation de 2',3'-didésoxythymidine triphosphate.

Les 43 premiers cycles de polymérisation, hydrolyse enzymatique ou chimique, rinçage-comptage donnent une détection positive en signal de fluorescence, puis négative au 44ème cycle.

Dans le cas des figures 2 et 3, le cycle de réaction décrit est réalisé autant de fois que le marqueur est retrouvé dans le surnageant de déprotection. L'absence de marqueur dans ce dernier signifie la fin de la séquence répétée, et le nombre de cycles réalisés permet de déduire le nombre de répétitions. Il est intéressant de remarquer que le ou les didésoxynucléotides peuvent ne pas être incorporés juste après la séquence répétée si la séquence contigue à cette dernière ne contient pas les bases complémentaires au(x)dit(s) didésoxynucléotide(s) utilisé(s). Cela ne change pas l'exactitude ni l'utilité du procédé dans la mesure où les séquences encadrant les NRT sont connues à l'avance. Un tel cas est donc anticipé et ne gène pas la détermination du nombre de répétitions.

La Figure 4 décrit la synthèse de la 2'-désoxycytidine 3'-O-butyryle (^{14}C) 5'-O-triphosphate, nucléotide substitué en 3' de manière réversible :

Partant de dC, l'amine aromatique portée par la base et l'alcool en 5' du sucre sont respectivement protégés par des groupements benzyloxycarbonyles (Cbz) et diméthoxytrytyle (DMTr). L'hydroxyle libre en 3' est ensuite acylé par l'acide butyrique marqué au ^{14}C, en présence de dicyclocarbodiimide (DCCI). L'alcool en 5' est libéré en milieu acide puis phosphorylé par action du chlorure de dibenzyle phosphate. L'hydrogénolyse simultanée des groupements benzyloxycarbonyle et benzyle conduit à la 2'-désoxycytidine 5'-phosphate acylée en 3' par l'acide butyrique ^{14}C. L'activation de l'ester phosphorique, par action de la N,N'-carbonyldiimidazole et le déplacement de l'imidazolide résultant, par le sel de tri-n-butylammonium de l'acide pyrophosphorique, fournissent le dCTP substitué en 3' par l'acide butyrique ^{14}C.

L'exemple qui suit est donné à titre d'illustration et n'est en aucun cas limitatif ni de la séquence utilisée, ni de la polymérase utilisée, ni du marquage utilisé, ni de la méthode de liaison de la séquence d'acide nucléique contenant le microsatellite à un support solide, ni enfin de la méthode de révélation du marquage.

. **Détermination du nombre de CAG dans le gène du récepteur de l'androgène humain, dont le polymorphisme dans les CAG est impliqué, entre autre dans la maladie de Kennedy :**

Matériels et Méthodes utilisés :

- Réaction d'incorporation :

L'ADN polymérase Taq , les 2'-3'-didéoxynucléotides, les 2'-désoxynucléotide-5'-triphosphates, proviennent de chez Boehringer (FRA). Une unité d'ADN polymérase représente l'activité enzymatique qui incorpore 1 nmole de nucléotides totaux dans un produit insoluble en milieu acide en 1 minute à 37°C avec du poly(d(A-T)) comme matrice.

Les billes magnétiques revêtues de streptavidine (M-280 Dynabeads) proviennent de Dynal (Norvège).

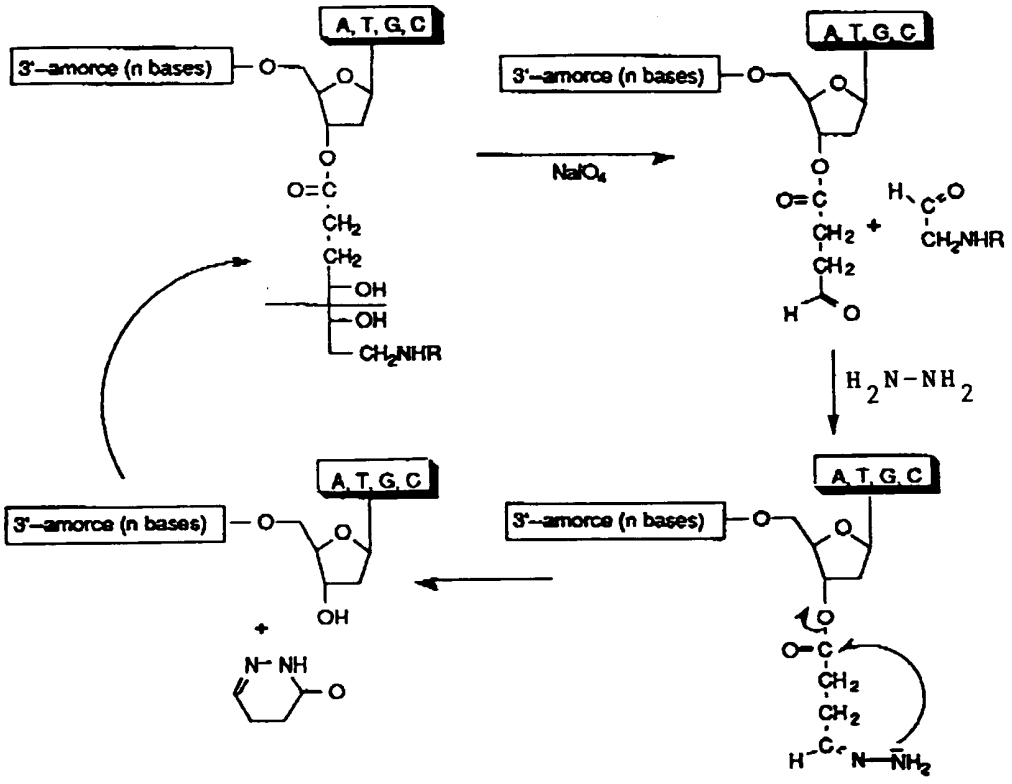
Les nucléotides triphosphates substitués en 3' peuvent être sinon de tous les types décrits, notamment dans le mémoire de Diplôme d'Etudes Approfondies (DEA) soutenu en septembre 1993 à l'Institut Pasteur dans l'Unité de Chimie Organique par Monsieur T. Berthod et intitulé "Synthèse d'analogues nucléotidiques triphosphates en vue de la mise au point d'une nouvelle méthode de 'miniséquençage' de l'ADN".

En synthèse oligonucléotidique, les sucres sont souvent temporairement protégés sous la forme d'esters de l'acide lévulinique, les lévulinates.

Entre autres intérêts de ces esters, la déprotection des alcools est quantitative en 1 minute à pH neutre. Dans ces conditions, les oligonucléotides sont stables.

Les nucléotides acylés en 3' par l'acide 6-amino-2,3,6-tridésoxy-D-gluconique oxydés par le périodate, libèrent quantitativement le fluorochrome et un ester qui ne diffère de l'ester lévulinique que par

le remplacement d'un méthyle par un hydrogène (une cétone par un aldéhyde) suivant le schéma suivant :



Les alcools sont régénérés par le mécanisme suivant: dans un mélange hydrazine-pyridine-acide acétique-eau, à pH neutre, la fonction cétone est transformée en hydrazone (18), le doublet libre sur l'azote attaque le carbonyle et l'alcool est libéré comme ceci est indiqué dans la figure ci-dessus.

Tout autre dérivé de dTNP estérifié en 3', par exemple ceux décrits dans la demande de brevet WO91/06678 ou la demande de brevet WO93/05183, est utilisable dans le procédé de l'invention.

- Préparation de la matrice d'ADN :

La préparation dont on étudie les répétitions en tandem, par exemple des CAG répétés dans le gène du récepteur de l'androgène humain, est effectuée de la façon décrite dans la figure 1 de la référence 17 : l'ADN génomique est multicopié par PCR à l'aide de deux amorces décrites dans cette légende. Toutes conditions égales par ailleurs, la seule modification apportée est le fait que l'amorce nucléotidique

5'-GCTGTGAAGGTTGCTGTTCCTC-3'

utilisée est biotinylée en 5' pour une immobilisation sur un support solide Dynabeads M-280.

L'amorce réactionnelle utilisée pour la détermination du nombre de CAG est :

5'-CGGCGCCAGTTTGCTGCTGCTG-3'.

La réaction de comptage des CAG est réalisée de la manière suivante : 5 picomoles d'ADN humain provenant de l'amplification PCR décrite ci-dessus (légende de la figure 1 de la référence 17) sont absorbés sur les billes magnétiques Dynabeads et la matrice simple brin préparée selon les instructions du fabricant : les billes magnétiques et citée plus haut.

L'amorce réactionnelle (10 picomoles) est hybridée 2 minutes à 70°C puis laissée 30 minutes au moins à refroidir jusqu'à une température inférieure à 30°C.

L'excès est enlevé par aimantation et rinçage des billes.

- Incorporation des désoxynucléotides triphosphates modifiés ou non :

La réaction est initiée par addition d'un mélange réactionnel comprenant les concentrations finales suivantes : 10 μM pour les dCTP et pour les dATP, 1 mM pour le dGTP-3'^{14}C-butyryle ester et 20 μM de ddTTP. Le surnageant est enlevé, les billes rincées abondamment jusqu'à l'obtention d'un surnageant exempt de radioactivité, et les billes sont mises à incuber en présence d'une estérase dans un tampon et à une température appropriée pendant 15 minutes. Après séparation magnétique, le surnageant est récupéré et sa radioactivité déterminée par scintillation liquide ou tout autre moyen équivalent. L'ensemble de ces étapes constitue un cycle.

Un score de 1+ est attribué à la détection significative de radioactivité, c'est-à-dire à la détection d'une activité spécifique corrélée avec la quantité de matrice utilisée, par exemple 5 picomoles de matrice simple brin d'ADN. Les billes sont lavées abondamment, et la réaction ré-initiée par l'ajout du mélange réactionnel décrit précédemment, les étapes suivantes ré-opérées jusqu'à l'attribution d'un nouveau score de 2+ lors de la détection significative de radioactivité. L'ensemble du procédé est réitéré 18 fois (soit 19 cycles), et à cette étape, le score est de 19+ car la radioactivité significative est toujours détectée dans le surnageant.

Au cycle suivant, la radioactivité n'est plus détectée, un score de 20- est déterminé et l'on peut aisément déduire de ces résultats que le nombre maximum de CAG répété est de 19.

Le diagnostic de la maladie de Kennedy peut ainsi être établi quand le nombre de cycles réalisés dépasse 20+.

Il est évident que l'ensemble de ces opérations peut être entièrement effectué de manière automatique, de la production de l'ADN à analyser par PCR, jusqu'à l'obtention informatisée des résultats, par exemple avec un robot de type Biomek (Beckman) et un compteur à scintillation liquide.

La technologie décrite permet de tester commodément, sur des plaques de microtitrage, beaucoup d'échantillons en parallèle à l'aide de manipulations robotisées.

Le principe décrit plus haut, ainsi que l'exemple l'illustrant, montrent clairement que, en jouant sur la composition du mélange de 2'-désoxynucléotides et sur la manière d'estérifier en 3' le désoxynucléotide destiné à "marquer" une des bases de l'unité répétée dans le microsatellite, tout type de mode de réalisation équivalent permettant par le principe décrit plus haut de déterminer le nombre de répétitions dans le microsatellite fait également partie de l'invention.

BIBLIOGRAPHIE

1. Weissenbach J. et al. (1992) Nature, 359:794-801).
2. Wooster R. et al. (février 1994) Nature Genetics, 6:152-156.
3. Aaltonen L.A. et al. (1993) "Clues to the Pathogenesis of Familial Colorectal Cancer" Science, 260:812-816.
4. Thibodeau S.N. et al. (1993) "Microsatellite Instability in Cancer of the Proximal Colon" Science, 260:816-819.
5. Peltomaki P. et al. (1993) "Genetic Mapping of a Locus Predisposing to Human Colorectal Cancer" Science, 260:810-812.
6. Economou E.P. et al. (1990) P.N.A.S. 87:2951-2954.
7. Weber J.L. (1990) Genomics, 7:524-530.
8. Kuhl D.P.A. et Caskey C.T. (1993) Current Opinion in Genetics and Development, 3:404-407.
9. Weissenbach J. (1993) Current Opinion in Genetics and Development, 3:414-417.
10. Dietrich W. et al. (1992) Genetics, 131:423-447.
11. Richards R.I. et Sutherland G.R. (1992) Cell, 70:709-712.
12. Sutherland G.R. et al. (1991) Lancet, 338:289-292.
13. Duyao M. et al. (1993) Nature Genetics, 4:387-392.
14. Sanger F., Nicklen S. and Coucson A.R. (1977) Proc., Natl. Acad. Sci. USA, 74, 5463-5467.
15. Strezoska Z. et al. (1991) P.N.A.S., 88:10089-10093.
16. Driscoll R.J. et al. (1990) Nature, 346:294-296.
17. La spada A.R. et al. (1991) Nature, 352:77-79.
18. Van Boom et al. (1976) Tetrahedron Lett. 48:75.

REVENDICATIONS

1. Procédé de détermination du nombre de répétitions d'un motif possédant au plus trois des quatres bases nucléotidiques dans une séquence répétée, elle-même incluse dans une séquence d'acide nucléique, le cas échéant couplé à un support solide et caractérisé par les étapes suivantes :

a) hybridation d'une amorce nucléotidique d'élongation avec une séquence complémentaire d'une séquence située en 5' dudit microsatellite,

b) l'addition d'un mélange de 2'-désoxynucléotide 5'-triphosphates (A), de 2',3'-didésoxynucléotide triphosphate (B) marqué ou non et de 2'-désoxynucléotide 5'-triphosphate substitué en 3' (C) et marqué ou non sur le groupe de substitution,

(A) autorisant une élongation normale de la chaîne,

(B) bloquant une chaîne en élongation irréversiblement et

(C) bloquant réversiblement une chaîne en élongation,

c) addition d'une acide nucléique polymérase dans des conditions de tampon et de température compatibles avec l'incorporation de n'importe lequel des trois désoxynucléotides (A), (B) et (C), et permettant la susdite incorporation,

d) action d'un composé chimique hydrolysant en 3' ou d'une enzyme de type estérase en 3' sur la chaîne en élongation,

e) séparation du surnageant de la séquence d'acide nucléique obtenue par élongation et mesure du marquage éventuel relargué dans ledit surnageant,

f) réitération des étapes b) à e),

g) comptage du nombre de répétitions après comparaison de la présence ou de l'absence de marquage entre les différentes réitérations.

2. Procédé selon la revendication 1, caractérisé en ce que la séquence répétée est une séquence microsatellite.

3. Procédé selon la revendication 1 ou 2, caractérisé en ce que les étapes a) à g) sont précédées par une étape d'amplification de la séquence d'acides nucléiques, les amorces d'amplification étant couplées chimiquement ou susceptibles d'être couplées chimiquement à un support solide.

4. Procédé selon l'une des revendications 1 à 3, caractérisé en ce que l'amorce nucléotidique d'élongation est constituée d'une séquence de 5 à 100 nucléotides de long et est contigue au microsatellite et pouvant, le cas échéant, comprendre de 1 à 4 motifs de base dudit microsatellite.

5. Procédé selon l'une des revendications 1 ou 3, caractérisé en ce que le 2'-désoxynucléotide 5'-triphosphate substitué en 3' est substitué par l'un des groupements suivants :

- un groupement anthranylate,
- un ester de l'acide 6-amino-caproïque substitué par un groupement fluorescent ou radioactif,
- un ester de l'acide 6-amino-2,3,6-tridésoxy-D-gluconique marqué soit par un élément radioactif, soit substitué par un groupement fluorescent, soit par tout groupement détectable par un procédé chimique, enzymatique ou physique.

6. Procédé selon l'une des revendication 1 à 5, caractérisé en ce que la séquence d'acide nucléique contenant le microsatellite est, le cas échéant, préalablement multicopiée et marquée à une de ses extrémités par la biotine, elle-même étant susceptible de se coupler à un support modifié par un groupement avidine ou streptavidine.

7. Procédé selon l'une des revendications 1 à 6, caractérisé en ce que l'hydrolyse de l'ester du

nucléotide 3' substitué est effectuée par voie enzymatique et notamment par une estérase.

8. Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce que la séquence d'acide nucléique est couplée par l'intermédiaire d'une substance chimique à un support de type magnétique ou autre.

9. Trousse de détermination du nombre de répétitions d'un motif possédant au plus trois des quatre bases nucléotidiques dans une séquence répétée, notamment dans une séquence microsatellite ou dans une séquence minisatellite, elle-même incluse dans une séquence d'acide nucléique le cas échéant couplée à un support solide et caractérisée en ce qu'elle contient au moins :

a) un ou deux désoxynucléotides 5'-triphosphates (A),

b) un 2',3'-didésoxynucléotide triphosphate (B) marqué ou non,

c) un 2'-désoxynucléotide 5'-triphosphate estérifié en 3', ledit ester étant hydrolysable par voie chimique ou enzymatique (C),

chacun des désoxynucléotides triphosphates en a), b) et c) étant porteur d'une base différente, (A) et (C) étant présents et (B) absent dans le motif répété du microsatellite,

d) une polymérase ou tout autre moyen de polymérisation,

e) le cas échéant, une estérase susceptible d'hydrolyser l'ester en 3' de (C).

10. Trousse selon la revendication 9, caractérisée en ce que le 2',3'-didésoxynucléotide triphosphate est estérifié par l'acide 6-amino-2,3,6-tridésoxy-D-gluconique et qu'il est hydrolysable par oxydation à l'aide de périodate.

11. Trousse selon la revendication 9, caractérisée en ce que le 2'-désoxynucléotide 5'-triphosphate substitué en 3' est substitué par l'acide

6-amino-caproïque, lui-même marqué par un radio-isotope ou un fluorochrome.

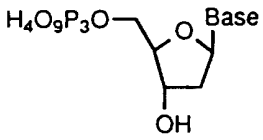
12. Trousse selon l'une des revendications 9 à 11, caractérisée en ce qu'elle contient en outre :
- une amorce ou un couple d'amorces,
- une ADN polymérase nécessaire à la multicopie de la séquence d'acide nucléique contenant ledit microsatellite,

l'amorce ou une des amorces étant susceptible d'être couplée chimiquement à un support solide.

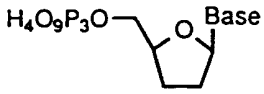
13. Utilisation d'une trousse de détermination du nombre de répétitions d'un motif d'une à trois bases nucléotidiques dans une séquence répétée selon l'une des revendications 9 à 12, à la construction de cartes génomiques chez l'homme ou l'animal.

14. Utilisation d'une trousse de détermination du nombre de répétitions d'un motif d'une à trois bases nucléotidiques dans une séquence répétée selon l'une des revendications 9 à 12, dans le diagnostic ou le pronostic d'un cancer ou à la détection de cellules cancéreuses chez l'homme ou l'animal.

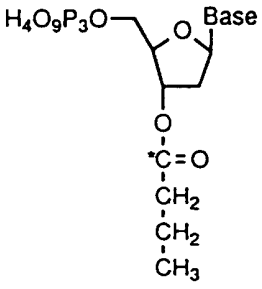
15. Utilisation d'une trousse de détermination du nombre de répétitions d'un motif d'une à trois bases nucléotidiques dans une séquence répétée selon l'une des revendications 9 à 12, dans l'analyse génétique de la transmission de maladies héréditaires.



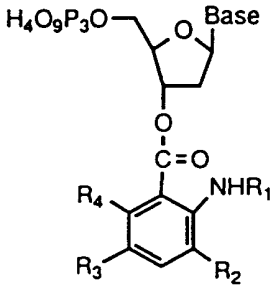
2'- désoxynucléotide 5'- O- triphosphate



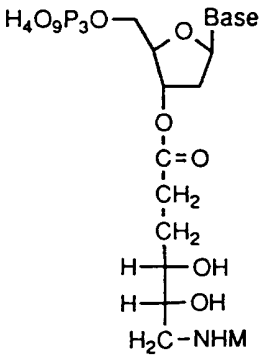
2',3'-di- désoxynucléotide 5'- O- triphosphate



2'- désoxynucléotide 3'-O- butyryle (^{14}C) 5'- O- triphosphate

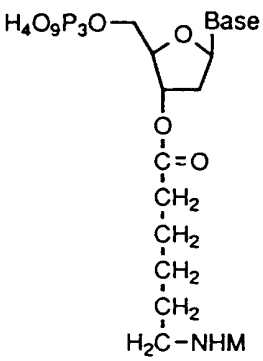


2'- désoxynucléotide 3'-O- Ant 5'- O- triphosphate



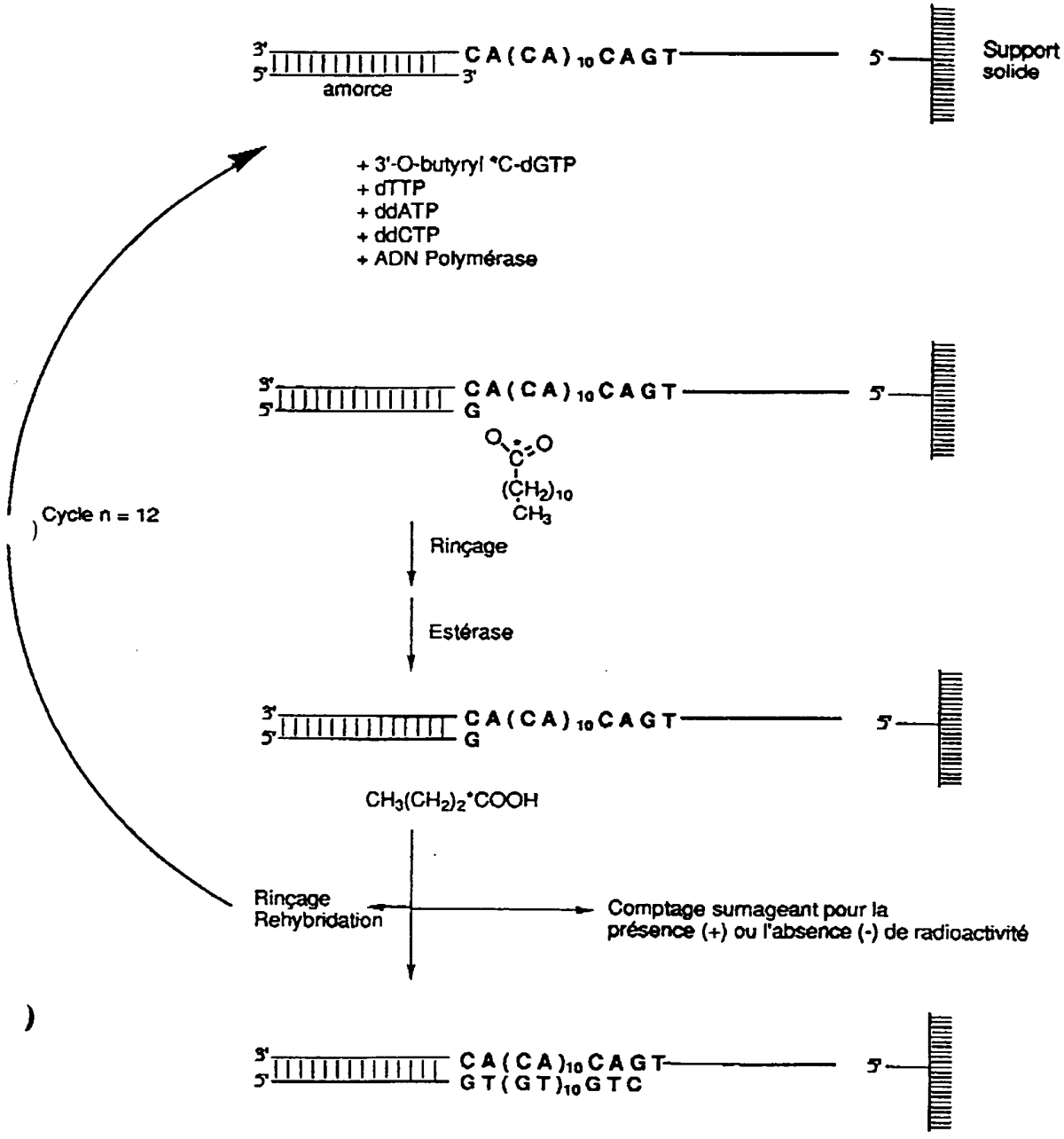
2'- désoxynucléotide 5'- O- triphosphate
estérifié en 3' par l'acide 6-amino 2-3-6- tridésoxy
- D-gluconique substitué par différents marqueurs

M = Différents marqueurs



2'- désoxynucléotide 5'- O- triphosphate
estérifié en 3' par l'acide 6-amino caproïque
substitué par différents marqueurs

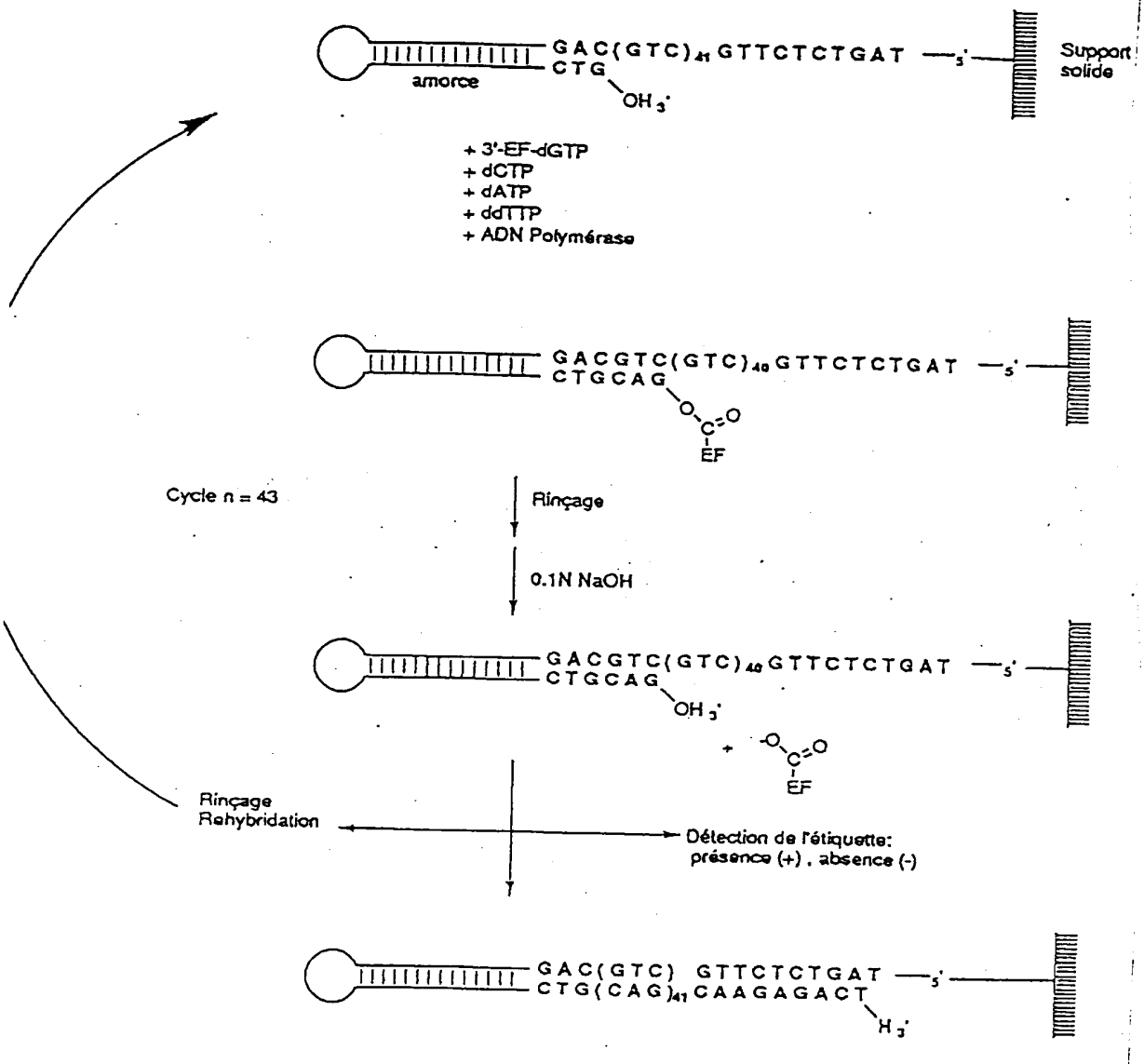
FIGURE 1



Score =

n	°C
1	+
.	.
.	.
12	+
13	-

FIGURE 2



EF = Etiquette fluorescente

FIGURE 3

Score =

n	°C
1	+
.	.
.	.
.	.
43	+
44	—

SYNTHESE DE LA 2'-DESOXY 3'-O-BUTYRYLE (^{14}C) 5'-0-TRIPHOSPHATE CYTIDINE

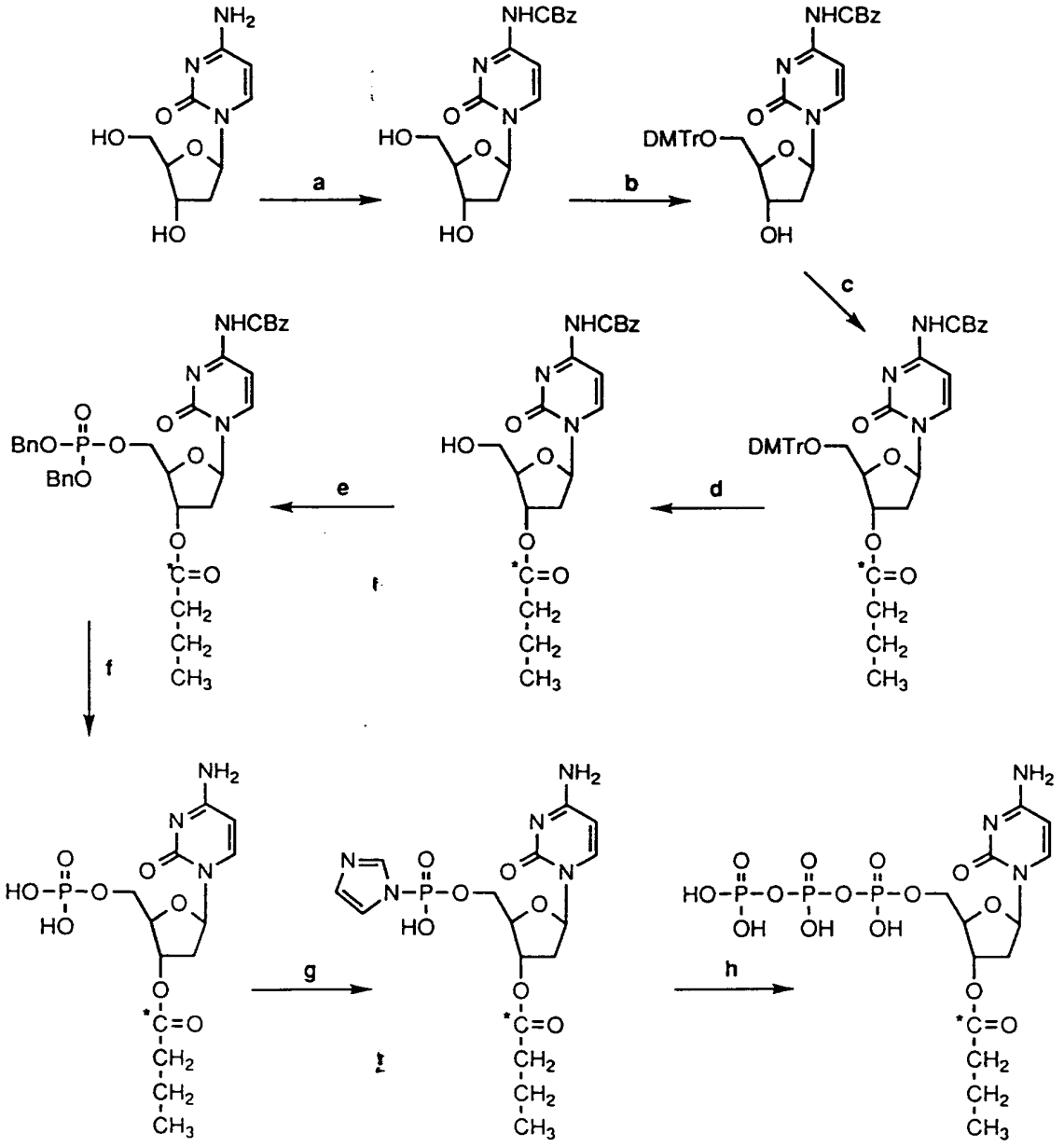


FIGURE 4

REPUBLIQUE FRANÇAISE

INSTITUT NATIONAL
de la
PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE PRELIMINAIRE

établi sur la base des dernières revendications déposées avant le commencement de la recherche

2718753
N° d'enregistrement national

FA 499975
FR 9404554

DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendications concernées de la demande examinée
A	WO-A-93 05183 (BAYLOR COLLEGE OF MEDICINE) * le document en entier *	1-8
A	WO-A-91 06678 (SRI INTERNATIONAL) * le document en entier *	1-8
A	WO-A-93 21340 (MEDICAL RESEARCH COUNCIL) * le document en entier *	1-8
A	WO-A-93 17104 (MASSACHUSETTS INSTITUTE OF TECHNOLOGY)	
A	WO-A-93 16197 (UNIV. WALES COLLEGE OF MEDICINE)	

DOMAINES TECHNIQUES RECHERCHES (Int.Cl.5)

C12Q

Date d'achèvement de la recherche: 21 Décembre 1994

Examinateur: Osborne, H

CATEGORIE DES DOCUMENTS CITES

X : particulièrement pertinent à lui seul
Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie
A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général
O : divulgation non-écrite
P : document intercalaire

T : théorie ou principe à la base de l'invention
E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure.
D : cité dans la demande
L : cité pour d'autres raisons

& : membre de la même famille, document correspondant

EPO FORM 1503 03.82 (P04C13)