



Office de la Propriété
Intellectuelle
du Canada

Un organisme
d'Industrie Canada

Canadian
Intellectual Property
Office

An agency of
Industry Canada

CA 2072023 C 2005/03/29

(11)(21) **2 072 023**

(12) **BREVET CANADIEN
CANADIAN PATENT**

(13) **C**

(86) Date de dépôt PCT/PCT Filing Date: 1991/01/30
(87) Date publication PCT/PCT Publication Date: 1991/08/08
(45) Date de délivrance/Issue Date: 2005/03/29
(85) Entrée phase nationale/National Entry: 1992/06/19
(86) N° demande PCT/PCT Application No.: FR 1991/000054
(87) N° publication PCT/PCT Publication No.: 1991/011518
(30) Priorité/Priority: 1990/01/31 (90/01137) FR

(51) Cl.Int.⁵/Int.Cl.⁵ C12N 15/52, C12P 19/42, C12N 1/21,
C12N 9/00

(72) Inventeurs/Inventors:
BLANCHE, FRANCIS, FR;
CAMERON, BEATRICE, FR;
CROUZET, JOEL, FR;
DEBUSSCHE, LAURENT, FR;
LEVY-SCHIL, SOPHIE, FR;
THIBAUT, DENIS, FR

(73) Propriétaire/Owner:
RHONE-POULENC BIOCHIMIE, FR

(74) Agent: ROBIC

(54) Titre : POLYPEPTIDES IMPLIQUES DANS LA BIO-SYNTHESE DES COBALAMINES ET/OU DES COBAMIDES
(54) Title: POLYPEPTIDES INVOLVED IN THE BIOSYNTHESIS OF COBALAMINES AND/OR COBAMIDES, DNA
SEQUENCES CODING FOR THESE POLYPEPTIDES, AND THEIR PREPARATION AND USE

(57) **Abrégé/Abstract:**

La présente invention concerne de nouveaux polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides, et en particulier du coenzyme B₁₂. Elle concerne également le matériel génétique responsable de l'expression de ces polypeptides, ainsi qu'un procédé permettant leur préparation. Elle concerne enfin un procédé d'amplification de la production de cobalamines, et plus particulièrement du coenzyme B₁₂, par les techniques d'ADN recombinant.



PCT

ORGANISATION MONDIALE DE LA PROPRIÉTÉ INTELLECTUELLE
Bureau international



DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

(51) Classification internationale des brevets ⁵ : C12N 15/52, C07K 15/04 C12P 19/42, C12N 9/00, 1/21 C12P 19/42 // C12R 1/38 C12R 1/41C12R 1/01	A1	(11) Numéro de publication internationale: WO 91/11518 (43) Date de publication internationale: 8 août 1991 (08.08.91)
---	----	--

(21) Numéro de la demande internationale: PCT/FR91/00054

(22) Date de dépôt international: 30 janvier 1991 (30.01.91)

(30) Données relatives à la priorité:
90/01137 31 janvier 1990 (31.01.90) FR

(71) Déposant (pour tous les Etats désignés sauf US): RHONE-POULENC BIOCHIMIE [FR/FR]; 20, avenue Raymond-Aron, F-92160 Antony (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (US seulement): BLANCHE, Francis [FR/FR]; 41, rue des Solitaires, F-75019 Paris (FR). CAMERON, Béatrice [FR/FR]; 28, rue Henri-Barbusse, F-75005 Paris (FR). CROUZET, Joël [FR/FR]; 48-52, rue des Meuniers, F-75012 Paris (FR). DEBUSSCHE, Laurent [FR/FR]; 43, bd de l'Hôpital, F-75013 Paris (FR). LEVY-SCHIL, Sophie [FR/FR]; 2, rue Monttessuy, F-75007 Paris (FR). THIBAUT, Denis [FR/FR]; 28, rue Jean-Colly, F-75013 Paris (FR).

(74) Mandataire: BECKER, Philippe; Rhône-Poulenc Rorer S.A. - Direction Brevets, 20, avenue Raymond-Aron, F-92165 Antony Cédex (FR).

(81) Etats désignés: AT (brevet européen), BE (brevet européen), CA, CH (brevet européen), DE (brevet européen), DK (brevet européen), ES (brevet européen), FR (brevet européen), GB (brevet européen), GR (brevet européen), HU, IT (brevet européen), JP, LU (brevet européen), NL (brevet européen), SE (brevet européen), SU, US.

Publiée

*Avec rapport de recherche internationale.
Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si de telles modifications sont reçues.*

(54) Title: POLYPEPTIDES INVOLVED IN THE BIOSYNTHESIS OF COBALAMINES AND/OR COBAMIDES, DNA SEQUENCES CODING FOR THESE POLYPEPTIDES, AND THEIR PREPARATION AND USE

(54) Titre: POLYPEPTIDES IMPLIQUES DANS LA BIOSYNTHESE DES COBALAMINES ET/OU DES COBAMIDES, SEQUENCES D'ADN CODANT POUR CES POLYPEPTIDES, PROCEDE DE PREPARATION, ET LEUR UTILISATION

(57) Abstract

Novel polypeptides involved in the biosynthesis of cobalamines and/or cobamides, in particular coenzyme B₁₂, genetic material responsible for expressing these polypeptides, and a method for preparing them, are described. A method for amplifying the production of cobalamines, and particularly coenzyme B₁₂, using recombinant DNA techniques, are also described.

(57) Abrégé

La présente invention concerne de nouveaux polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides, et en particulier du coenzyme B₁₂. Elle concerne également le matériel génétique responsable de l'expression de ces polypeptides, ainsi qu'un procédé permettant leur préparation. Elle concerne enfin un procédé d'amplification de la production de cobalamines, et plus particulièrement du coenzyme B₁₂, par les techniques d'ADN recombinant.

2072023

La présente invention concerne de nouveaux polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides, et en particulier du coenzyme B₁₂. Elle concerne également le matériel génétique responsable de l'expression de ces polypeptides, ainsi qu'un
5 procédé permettant leur préparation. Elle concerne enfin un procédé d'amplification de la production de cobalamines, et plus particulièrement du coenzyme B₁₂, par les techniques d'ADN recombinant.

La vitamine B₁₂ fait partie des vitamines du groupe B. Il
10 s'agit d'une vitamine hydrosoluble qui a été identifiée comme étant le facteur permettant de traiter des malades souffrant d'anémie pernicieuse. Elle est généralement prescrite pour stimuler l'hématopoïèse chez les sujets fatigués, mais elle est aussi utilisée dans de nombreux autres cas qui comprennent les troubles hépatiques,
15 des déficiences nerveuses ou comme stimulant de l'appétit, principe tonifiant, ainsi qu'en dermatologie (Beck, 1982, Fraser et al., 1983). Dans les élevages industriels d'animaux non ruminants, l'alimentation étant essentiellement à base de protéines d'origines végétales, il est nécessaire d'incorporer aux rations alimentaires de la vitamine B₁₂ à
20 des quantités de 10 à 15 mg par tonne d'aliments (Barrère et al., 1981).

La vitamine B₁₂ fait partie d'une classe de molécules appelées cobalamines, dont la structure est présentée à la figure 1. Les cobamides diffèrent des cobalamines par la base du nucléotide
25 inférieur qui n'est plus le 5,6-diméthylbenzimidazole, mais une autre base, par exemple, le 5-hydroxybenzimidazole, pour la vitamine B₁₂-facteur III synthétisée entre autre par Clostridium thermoaceticum et Methanosarcina barkeri (Iron et al., 1984). Ces similitudes structurales expliquent que les voies métaboliques de biosynthèse des
30 cobalamines et des cobamides soient communes pour la majeure partie.

Les cobalamines sont synthétisées presque exclusivement par des bactéries, selon un processus complexe et encore très mal connu, qui peut se diviser en quatre étapes (figure 2) :

1) synthèse de l'uroporphyrinogène III (ou uro'gen III),

35 puis

ii) transformation de l'uro'gen III en acide cobyrique
suivie de

iii) transformation de celui-ci en cobinamide et

iv) construction de la boucle nucléotidique inférieure
5 avec incorporation de la base particulière (5,6-diméthylbenzimidazole
dans le cas des cobalamines).

Pour le coenzyme B₁₂, il est probable que l'addition du
groupement 5'-déoxy-5'-adénosyl se produit peu après que le noyau
corrine soit synthétisé (Huennekens et al., 1892).

10 Dans le cas des cobamides, seule l'étape de synthèse et
d'incorporation de la base inférieure est différente.

La première partie de la biosynthèse des cobalamines est
très bien connue puisqu'elle est commune à celle des hèmes ainsi qu'à
celle des chlorophylles (Battersby et al., 1980). Elle fait intervenir
15 successivement la δ -aminolévulinate synthase (EC 2.3.137), la
 δ -aminolévulinate déshydrase (EC 4.2.1.24), la porphobilinogène
déaminase (EC 4.3.1.8) et l'uro'gen III cosynthase (EC 4.2.1.75) qui
transforment le succinyl CoA et la glycine en uro'gen III. Toutefois
la première étape se fait chez certains organismes [par exemple *E.coli*
20 (Avissar et al., 1989) et chez les bactéries méthanogènes (Kannangara
et al., 1989), par exemple] par la conversion grâce à un complexe
multienzymatique de l'acide glutamique en acide δ -aminolévulinique.

Entre l'uro'gen III et l'acide cobyrique, seuls trois
dérivés d'intermédiaires ont été purifiés à ce jour ; il s'agit des
25 Facteurs FI, FII et FIII qui sont des produits d'oxydation
respectivement des trois intermédiaires précorrine-1, précorrine-2 et
précorrine-3 qui correspondent aux dérivés mono-, di- et triméthylés
de l'uro'gen III (figure 3) ; ces intermédiaires sont obtenus par des
transferts successifs de groupements méthyl à partir du SAM
30 (S-adénosyl-L-méthionine) sur l'uro'gen III aux positions C2, C7 et
C20 respectivement. Les autres réactions qui ont lieu pour donner
l'acide cobyrique sont, outre cinq autres transferts de groupements
méthyl à partir du SAM en C17, C12, C1, C15 et C5, l'élimination du
carbone en C20, la décarboxylation en C12 et l'insertion d'un atome de

cobalt (figure 4). Ces étapes de biosynthèse ont été déduites à partir d'expériences effectuées in vitro sur des extraits acellulaires de Propionibacterium shermanii ou de Clostridium tetanomorphum. Dans ces extraits, l'acide cobyrinique est obtenu par transformation de l'uro'gen III, après incubation dans des conditions appropriées en anaérobiose (Batterby et al., 1982). Aucun intermédiaire entre le précorrène-3 et l'acide cobyrinique, pouvant être transformé en corrinoïdes par incubations ultérieures avec des extraits de bactéries productrices de cobalamines, n'a été isolé à ce jour chez ces microorganismes. La difficulté d'isoler et d'identifier ces intermédiaires est liée à

- i) leur grande instabilité
- ii) leur sensibilité à l'oxygène et
- iii) leur faible niveau d'accumulation in vivo.

15 Dans cette partie de la voie, seule une enzyme de Pseudomonas denitrificans a été purifiée et étudiée ; il s'agit de la SAM:uro'gen III méthyltransférase (Blanche et al., 1989) appelée SUMT.

Entre l'acide cobyrinique et le cobinamide, les réactions suivantes sont effectuées :

- 20 i) addition du groupement 5'-désoxyadénosyl (s'il s'agit du coenzyme B₁₂ qui doit être synthétisé),
- ii) amidation de six des sept fonctions carboxyliques par addition de groupements amines,
- iii) amidation de la dernière fonction carboxylique
- 25 (chaîne d'acide propionique du noyau pyrrole D) par addition du R-1-amino-2-propanol (figure 2).

Il n'a pas été élucidé s'il existait réellement un ordre dans les amidations (Herbert et al., 1970). Enfin, aucun dosage d'activité dans cette partie de la voie n'a été décrit sauf en ce qui concerne l'addition du groupement 5'-désoxyadénosyl (Huennekens et al., 1982).

La dernière étape de la biosynthèse d'une cobalamine, par exemple le coenzyme B₁₂ comprend quatre phases successives décrites sur la figure 5 (Huennekens et al., 1982), à savoir :

- 35 1) la phosphorylation du groupement hydroxyl du résidu

2072023

aminopropanol du cobinamide pour donner le cobinamide phosphate, puis
 ii) l'addition d'une guanosine diphosphate par réaction
 avec la guanosine 5'-triphosphate ; le composé obtenu est le
 GDP-cobinamide (Friedmann, 1975) qui

- 5 iii) réagit avec le α -ribazole 5'-phosphate, lui-même
 synthétisé à partir de la riboflavine, pour donner l'adénosyl-
 cobalamine 5'-phosphate (Friedmann et al, 1968) qui
 iv) par déphosphorylation conduit au coenzyme B₁₂
 (Schneider et Friedmann, 1972).

- 10 Parmi les bactéries capables de produire des cobalamines,
 on peut citer notamment :

- Agrobacterium tumefaciens,
Agrobacterium radiobacter
Bacillus megaterium
 15 Clostridium sticklandii
Clostridium tetanomorphum
Clostridium thermoaceticum
Corynebacterium XG
Eubacterium limosum
 20 Methanobacterium arbophilicum
Methanobacterium ivanovii
Methanobacterium ruminantium
Methanobacterium thermoautotrophicum
Methanosarcina barkeri
 25 Propionobacterium shermanii
Protaminobacter ruber
Pseudomonas denitrificans
Pseudomonas putida
Rhizobium meliloti
 30 Rhodopseudomonas sphaeroides
Salmonella typhimurium
Spirulina platensis
Streptomyces antibioticus
Streptomyces aureofaciens

Streptomyces griseusStreptomyces olivaceus

Au niveau industriel, en raison de la grande complexité des mécanismes de biosynthèse, la production des cobalamines, et en particulier de la vitamine B₁₂, est exclusivement microbiologique. Elle est réalisée par des cultures en grands volumes des bactéries Pseudomonas denitrificans, Propionobacterium shermanii et Propionibacterium freudenreichii (Florent, 1986). Les souches utilisées pour la production industrielle sont issues de souches sauvages; elles peuvent avoir subi de nombreux cycles de mutation au hasard puis de sélection de clones améliorés pour la production de cobalamines (Florent, 1986). Les mutations sont obtenues par mutagénèse avec des agents mutagènes ou par traitements physiques tels que des traitements aux rayons ultra-violets (Barrère et al., 1981). Par cette méthode empirique, des mutations au hasard sont obtenues et améliorent la production de cobalamines. Par exemple, il est décrit qu'à partir de la souche originale de Pseudomonas denitrificans initialement isolée par Miller et Rosenblum (1960, brevet US 2 938 822) la production de ce microorganisme a été graduellement augmentée en dix ans, par les techniques citées ci-dessus de 0,6 mg/l à 60 mg/l (Florent, 1986). Pour les bactéries du genre Propionibacterium, [Propionobacterium shermanii (ATCC 13673) et freudenreichii (ATCC 6207)] les mêmes valeurs de production semblent être décrites dans la littérature ; par exemple une production de 65 mg/l a été décrite (brevet Européen 87920). Toutefois, aucun crible n'a encore été décrit permettant de sélectionner ou de repérer facilement soit des mutants surproducteurs de cobalamines soit des mutants nettement améliorés dans leur production de cobalamines.

Au niveau génétique, peu de travaux ont été effectués à ce jour. Le clonage de gènes cob (codant pour des enzymes impliquées dans le procédé de biosynthèse) a été décrit chez Bacillus megaterium (Brey et al., 1986). Onze groupes de complémentation ont été identifiés par complémentation de mutants cob de Bacillus megaterium avec des plasmides portant différents fragments d'ADN de Bacillus megaterium.

Ces gènes sont groupés sur le même locus, porté par un fragment de 12 kb.

Des études ont également été menées sur les gènes cob de Salmonella typhimurium. Sans que le clonage de ceux-ci ait été décrit, il a été montré que presque tous les gènes de biosynthèse des cobalamines sont regroupés entre les minutes 40 et 42 du chromosome (Jeter et Roth, 1987). Seul le locus cysG, qui doit permettre la transformation de l'uro'gen III en précorrène-2 ne fait pas partie de ce groupe de gènes. Toutefois, l'activité codée par ce locus, ainsi que ses propriétés biochimiques n'ont pas été décrites.

En outre, des phénotypes ont été associés aux mutations cob. Chez Salmonella typhimurium et chez Bacillus megaterium, les mutants cob ne montrent plus de croissance sur milieu minimum avec de l'éthanolamine comme source de carbone ou comme source d'azote (Roof et Roth, 1988). Ceci est dû au fait qu'une enzyme du catabolisme de l'éthanolamine, l'éthanolamine ammonia-lyase (EC 4.3.1.7), a pour cofacteur le coenzyme B₁₂; les mutants cob ne synthétisant plus de coenzyme B₁₂, ils ne peuvent plus croître avec l'éthanolamine comme source de carbone et/ou comme source d'azote. Des mutants metE de Salmonella typhimurium n'ont plus qu'une homocystéine-méthyltransférase (EC 2.1.1.13) méthylcobalamine dépendante. Les mutants cob de Salmonella typhimurium metE sont auxotrophes pour la méthionine (Jeter et al., 1984).

Chez Pseudomonas denitrificans et Agrobacterium tumefaciens des phénotypes associés à une déficience totale en synthèse de cobalamines n'ont pas été décrits à ce jour.

Enfin, des travaux sur Pseudomonas denitrificans (Cameron et al., 1989) ont abouti au clonage de fragments d'ADN portant des gènes cob de cette bactérie. Ceux-ci sont répartis en quatre groupes de complémentation portés par au moins 30 kb d'ADN. Au moins quatorze groupes de complémentation ont été identifiés par complémentation hétérologue de mutants cob d'Agrobacterium tumefaciens et de Pseudomonas putida avec des fragments d'ADN de Pseudomonas denitrificans portant des gènes cob.

Cependant, jusqu'à maintenant, aucun de ces gènes n'a été

purifié, et aucune séquence nucléotidique n'a été décrite. De même, aucune identification de protéine, ni aucune fonction catalytique attribuée au produit de ces gènes n'ont été décrites. De plus, aucune amélioration de production de cobalamines par les techniques d'ADN recombinant n'a pu être obtenue. L'amplification de gènes cob de Bacillus megaterium ne procure pas, chez la souche à partir de laquelle ils ont été clonés, une amélioration de production de cobalamines (Brey et al., 1986). Chez Salmonella typhimurium, des études physiologiques ont été menées afin de déterminer des conditions dans lesquelles une forte transcription des gènes cob étudiés était observée (Escalante et Roth, 1987). Dans ces conditions, il n'y a pas
10 d'amélioration de la production de cobalamines, bien que des gènes de la voie de biosynthèse soient plus exprimés que dans les conditions standard de culture.

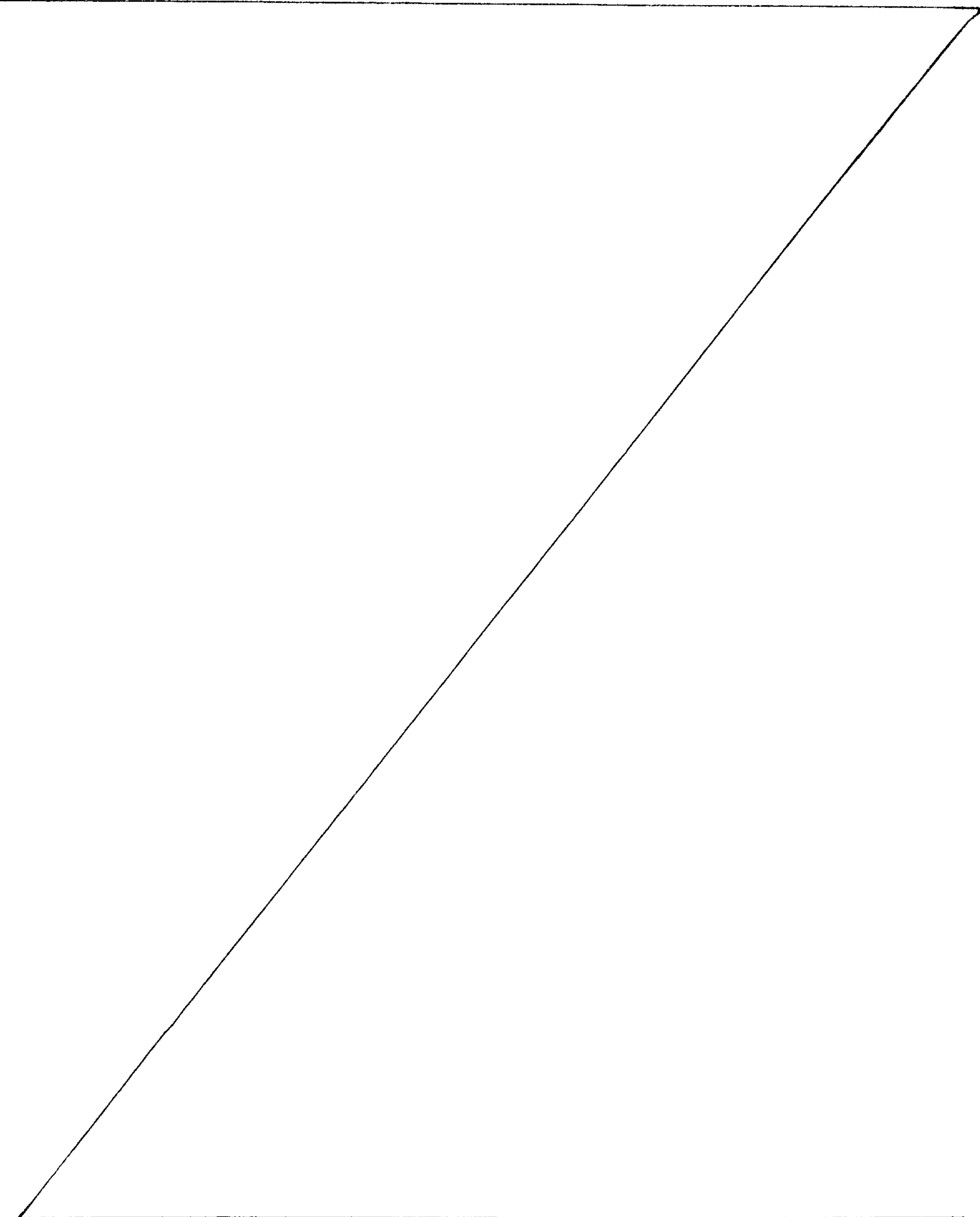
La présente invention résulte de l'identification précise de séquences d'ADN codant pour des polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides. Un objet de l'invention concerne donc une séquence d'ADN partiellement purifiée caractérisée en ce qu'elle code pour un polypeptide impliqué dans la biosynthèse des cobalamines et/ou des cobamides, ladite
20 séquence d'ADN étant caractérisée en ce qu'elle est choisie parmi:

- les gènes cobA, cobB, cobC, cobD, cobE, cobF, cobG, cobH, cobI, cobJ, cobK, cobL, cobM, cobN, cobO, cobP, cobQ, cobR, cobS, cobT, cobU, cobV, cobW, cobX, corA, présentés aux figures 15, 16, 40, 41, 47 et 52, ou
- les homologues de ces séquences résultant de la dégénérescence du code génétique.

L'invention a aussi pour objet les gènes contenant ces
30 séquences d'ADN.

7a

Les séquences d'ADN selon la présente invention ont été isolées à partir de différentes souches : une souche industrielle, Pseudomonas denitrificans SC510, dérivée de la souche MB580 (brevet US 3 018 225), par complémentation de mutants cob de A.tumefaciens et P.putida; et de Methanobacterium ivanovii. Les clones obtenus ont pu



être analysés précisément, notamment par cartographie à l'aide d'insertions d'un dérivé du transposon Tn5. Ces études génétiques ont permis de localiser les gènes cob ou cor sur la carte de restriction et de réaliser leur séquençage. Une analyse des phases ouvertes a
5 ensuite permis de mettre en évidence les régions codantes de ces fragments d'ADN.

La présente invention a aussi pour objet l'utilisation de ces séquences nucléotidiques pour cloner les gènes cob d'autres bactéries. En effet il est connu que pour des protéines catalysant les
10 mêmes activités, les séquences sont conservées avec, comme divergence, la divergence évolutive (Wein-Hsiung et al., 1985). Il est montré dans la présente invention qu'il y a une homologie significative entre les séquences nucléotidiques de gènes de différents microorganismes codant pour des polypeptides impliqués dans la biosynthèse des cobalamines
15 et/ou des cobamides. Les différences qui apparaissent résultent de la dégénérescence évolutive, et de la dégénérescence du code génétique qui est liée au pourcentage en GC du génome du microorganisme étudié (Wein-Hsiung et al., 1985).

Selon la présente invention, une sonde peut être faite
20 avec une ou plusieurs des séquences d'ADN de Pseudomonas denitrificans notamment, ou avec des fragments de celles-ci, ou avec des séquences analogues présentant un degré de dégénérescence spécifique, au niveau de l'usage des codons et du pourcentage en GC de l'ADN de la bactérie que l'on veut étudier. Dans ces conditions, il est possible de
25 détecter un signal d'hybridation spécifique entre la sonde et des fragments d'ADN génomique de la bactérie étudiée ; ce signal d'hybridation spécifique correspond à l'hybridation de la sonde avec les gènes cob isofonctionnels de la bactérie. Les gènes cob ainsi que leurs produits peuvent ensuite être isolés, purifiés et caractérisés.
30 l'invention fournit ainsi un moyen permettant, par hybridation, d'accéder aux séquences nucléotidiques et aux polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides de tout microorganisme.

La présente invention a également pour objet un ADN
35 recombinant contenant au moins une séquence d'ADN codant pour un

polypeptide impliqué dans la biosynthèse des cobalamines et/ou des cobamides, et notamment, un ADN recombinant dans lequel la ou lesdites séquences sont placées sous le contrôle de signaux d'expression.

A cet égard, on peut en particulier positionner en 5' de la séquence d'ADN des régions promotrices. De telles régions peuvent être homologues ou hétérologues de la séquence d'ADN. En particulier, des promoteurs bactériens forts, tels que le promoteur de l'opéron tryptophane P_{trp} ou de l'opéron lactose P_{lac} de E.coli, le promoteur gauche ou droit du bactériophage lambda, les promoteurs forts de phages de bactéries, telles que les corynebactéries, les promoteurs fonctionnels chez les bactéries gram-négatives, tel que le promoteur P_{tac} de E.coli, le promoteur P_{xylS} des gènes du catabolisme du xylène du plasmide TOL, le promoteur de l'amylase de Bacillus subtilis Pamy, pourront être utilisés. On peut citer également les promoteurs dérivés de gènes glycolytiques de levures, tels que les promoteurs des gènes codant pour le phosphoglycérate kinase, la glycéraldéhyde-3-phosphate déshydrogénase, la lactase ou l'énolase, qui pourront être utilisés lorsque l'ADN recombinant sera introduit dans un hôte eucaryote. Un site de fixation des ribosomes sera également positionné en 5' de la séquence d'ADN et il pourra être homologue ou hétérologue, tel le site de fixation des ribosomes du gène gII du bactériophage lambda.

Des signaux nécessaires à la terminaison de la transcription pourront être placés en 3' de la séquence d'ADN.

L'ADN recombinant selon la présente invention peut ensuite être introduit directement dans une cellule hôte compatible avec les signaux d'expression choisis ou être cloné sur un vecteur plasmidique pour permettre d'introduire de manière stable la séquence d'ADN en question dans la cellule hôte.

Un autre objet de l'invention concerne les plasmides ainsi obtenus contenant une séquence d'ADN codant pour un polypeptide impliqué dans la biosynthèse des cobalamines et/ou des cobamides. Plus précisément, ces plasmides contiennent aussi un système de réplication fonctionnel et un marqueur de sélection.

L'invention a aussi pour objet les cellules hôtes dans lesquelles une ou plusieurs séquences d'ADN telles que définies

précédemment, ou un plasmide tel que défini ci-avant, à été introduit.

Un autre objet de l'invention concerne un procédé de production de polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides. Selon ce procédé, ntrointroduit dans
5 une cellule hôte une séquence d'ADN telle que décrite précédemment, on cultive cette cellule recombinante dans des conditions d'expression de ladite séquence, puis on récupère les polypeptides produits.

Les cellules hôtes qui pourront être utilisées dans ce but sont aussi bien des procaryotes que des eucaryotes, des cellules
10 animales ou des cellules végétales. Préférentiellement, elles seront choisies parmi les bactéries et, en particulier, les bactéries du genre E.coli, P.denitrificans, A.tumefaciens ou R.melitoti.

Une autre utilisation des séquences d'ADN selon la présente invention réside dans un procédé d'amplification de la
15 production de cobalamines et/ou de cobamides ou de leurs précurseurs de biosynthèse, par les techniques d'ADN recombinant. En effet, si la limitation du flux métabolique de la biosynthèse des cobalamines et/ou des cobamides ou de leurs précurseurs est due à une limitation dans l'activité d'une enzyme dans la voie de biosynthèse, l'augmentation de
20 cette activité par augmentation de l'expression de cette même enzyme à l'aide des techniques d'ADN recombinant (amplification génique, substitution des signaux de transcription-traduction par des signaux plus efficaces ...) conduira à une augmentation de la biosynthèse des cobalamines et/ou des cobamides. Il est possible également que la
25 limitation de la production de cobalamines et/ou de cobamides résulte d'une régulation biochimique. Dans ce cas, le ou les gènes cob correspondant à l'enzyme régulée pourront être mutagénisés in vitro spécifiquement afin d'obtenir des gènes mutés dont les produits auront perdu les régulations allant à l'encontre d'une amélioration de la
30 production.

Le procédé selon la présente invention consiste à introduire dans un microorganisme producteur de cobalamines et/ou de cobamides, ou seulement potentiellement producteur de ces composés (c'est-à-dire déficient dans une ou plusieurs des étapes de la
35 biosynthèse), une séquence d'ADN telle que définie plus haut, puis à

cultiver ce microorganisme dans des conditions d'expression de ladite séquence et de synthèse des cobalamines et/ou des cobamides, et enfin à récupérer les cobalamines et/ou les cobamides produits. Un tel procédé est applicable en particulier à tous les microorganismes producteurs cités page 4, et plus spécifiquement aux microorganismes du genre P. denitrificans, Rhizobium meliloti, ou Agrobacterium tumefaciens. Dans un mode de réalisation préféré, le microorganisme est P. denitrificans, et en particulier, la souche SC510. Concernant les microorganismes potentiellement producteurs, les séquences d'ADN utilisées seront celles correspondant aux étapes de la biosynthèse que le microorganisme ne peut réaliser.

A l'aide de la présente invention et par les diverses stratégies exposées ci-dessus, une amélioration de la production de cobalamines et/ou de cobamides ou de leurs précurseurs pour tout microorganisme producteur, ou potentiellement producteur, de cobalamines et/ou de cobamides, pourra être obtenue. Il suffira de cultiver ce microorganisme recombiné dans des conditions appropriées pour la production de cobalamines et pour l'expression des séquences d'ADN introduites. Cette culture pourra se faire en batch ou bien en continu, et la purification des cobalamines pourra se faire par les méthodes déjà utilisées au niveau industriel (Florent, 1986). Ces méthodes comprennent entre autre :

i) la solubilisation des cobalamines et leur conversion en leur forme cyano (par exemple en traitant le moût de fermentation à la chaleur, avec du cyanure de potassium en présence de nitrite de sodium), puis

ii) la purification des cyanocobalamines en différentes étapes qui peuvent être par exemple

a) l'adsorption sur différents substrats tels que l'amberlite*IRC50, le Dowex*1X2 ou l'amberlite XAD 2 suivie d'une élution avec un mélange eau-alcool ou eau phénol, puis

b) l'extraction dans un solvant organique et enfin

c) la précipitation ou cristallisation à partir de la phase organique soit par addition de réactifs ou dilution dans un solvant approprié, soit par évaporation.

* (marques de commerce)

La présente invention montre de plus qu'il est possible, par les techniques d'ADN recombinant, d'améliorer la production de cobalamines d'une bactérie productrice de cobalamines en cumulant des améliorations. Ceci revient à obtenir une première amélioration comme
 5 cela est décrit ci-dessus puis à améliorer cette amélioration toujours à l'aide des techniques d'ADN recombinant, c'est-à-dire par exemple en amplifiant des gènes de biosynthèse des cobalamines.

Un autre objet de la présente invention concerne les polypeptides impliqués dans la biosynthèse des cobalamines et/ou des
 10 cobamides. En particulier, la présente invention a pour objet tous les polypeptides, ou des dérivés ou des fragments de ces polypeptides, qui sont codés par les séquences d'ADN décrites précédemment, et qui sont impliqués dans la voie de biosynthèse des cobalamines et/ou des cobamides. La séquence en acides aminés de ces polypeptides est
 15 décrite, ainsi que certaines de leurs caractéristiques physicochimiques. Une activité enzymatique ou des propriétés spécifiques ont également été associées à chacun d'entre-eux.

A cet égard, l'invention a pour objet les polypeptides intervenant dans la transformation du précorrène-3 en acide
 20 cobyrique a,c-diamide, et plus particulièrement dans le transfert d'un groupement méthyl du SAM aux positions C1, C5, C11, C15, et C17.

L'invention a aussi pour objet les polypeptide:

- . intervenant dans la transformation de l'acide cobyrique en cobinamide, ou
- 25 . possédant une activité S-adénosyl-L-méthionine:précorrène-2 méthyl transférase (SP2MT), ou
- . possédant une activité acide cobyrique et/ou hydrogénobyrique a,c-diamide synthase, ou
- . possédant une activité précorrène-8x mutase, ou
- 30 . possédant une activité nicotinate-nucléotide: diméthylbenzimidazole phosphoribosyltransférase, ou
- . possédant une activité cobalamine-(5'-phosphate) synthase, ou
- . possédant une activité cobyric acid synthase, ou
- 35 . possédant une activité cob(I)alamin adénosyltransférase,

ou

. possédant une activité precorrin-6x réductase, ou
. intervenant dans la transformation de l'acide hydrogénobyrrinique a,c-diamide en acide cobyrinique a,c-diamide.

5 Avantageusement, l'invention a pour objet un polypeptide choisi parmi les protéines COBA, COBB, COBC, COBD, COBE, COBF, COBG, COBH, COBI, COBJ, COBK, COBL, COBM, COBN, COBO, COBP, COBQ, COBS, COBT, COBU, COBV, COBW, COBX et CORA présentées aux figures 15, 16, 40, 41 et 47.

10 De plus, l'utilisation des sondes d'hybridation décrites précédemment permet, à partir de gènes isolés dans d'autres microorganismes, de caractériser et d'isoler les polypeptides isofonctionnels d'autres microorganismes. De cette manière, la présente invention montre que la séquence d'une protéine COB de
15 Pseudomonas denitrificans est significativement homologue aux séquences de protéines d'autres microorganismes présentant le même type d'activité. Entre ces protéines COB catalysant la même réaction chez des microorganismes différents, seules les distances évolutives ont introduit des variations (Wein-Hsiung et al., 1985). La présente
20 invention a également pour objet ces polypeptides isofonctionnels.

 L'attribution d'une activité enzymatique particulière est le résultat d'une analyse qui peut être effectuée selon diverses stratégies. En particulier, des études d'affinité in vitro, vis-à-vis du SAM (S-adénosyl-L-méthionine), permettent d'attribuer à une
25 protéine capable de fixer le SAM, une activité méthyltransférase, et donc son implication dans une des étapes de transfert des groupements méthyl qui se produisent entre l'uro'gen III et l'acide cobyrinique. Un autre moyen d'apprécier l'activité de ces polypeptides consiste à doser les intermédiaires de la voie de biosynthèse des cobalamines qui
30 sont accumulés chez des mutants incapables d'exprimer ces polypeptides (identifiés par des expériences de complémentation). Ces analyses permettent de déduire que le polypeptide en question a pour substrat l'intermédiaire accumulé, ce qui permet de situer et de définir son

activité dans la voie de biosynthèse. La présente invention décrit également un procédé de dosage des activités enzymatiques de la voie de biosynthèse, applicable à toute souche productrice de cobalamines et/ou de cobamides. Ces dosages permettent de purifier à partir de
5 toute souche productrice de ces composés, l'activité enzymatique dosée. A partir de cette activité purifiée, la séquence NH₂-terminale de la protéine COB en question, ou bien celle des sous-unités de cette protéine, peut être effectuée, ce qui permet d'identifier le ou les gènes de structure qui codent pour l'activité en question. Pour
10 Pseudomonas denitrificans, les gènes de structure qui codent pour des activités de la voie de biosynthèse sont identifiés en trouvant pour chaque séquence NH₂-terminale, la protéine COB ayant la même séquence NH₂-terminale.

La présente invention décrit aussi un procédé permettant
15 l'identification et le dosage, chez des souches productrices de cobalamines ou des mutants non producteurs, d'intermédiaires de la voie de biosynthèse des cobalamines ou d'autres corrinoïdes. Ces intermédiaires peuvent être dosés aussi bien dans des moûts de culture que dans les cellules elles-mêmes. Les intermédiaires qui peuvent être
20 dosés sont tous les corrinoïdes qui se trouvent dans la voie de biosynthèse après l'acide cobyrinique à savoir outre l'acide cobyrinique, l'acide cobyrinique monoamide, l'acide cobyrinique diamide, l'acide cobyrinique triamide, l'acide cobyrinique tétraamide, l'acide cobyrinique pentaamide, l'acide cobyrinique, le cobinamide, le
25 cobinamide phosphate, le GDP-cobinamide, le coenzyme B₁₂-phosphate et le coenzyme B₁₂. Les formes cyano et coenzyme de ces produits peuvent être dosées par cette technique.

D'autres objets et avantages de la présente invention apparaîtront à la lecture des exemples et des dessins suivants, qui
30 doivent être considérés comme illustratifs et non limitatifs.

DEFINITION DES TERMES EMPLOYES ET ABREVIATIONS

ACDAS: acide cobyrinique a,c-diamide synthase

ADN recombinant: ensemble de techniques qui permettent soit d'associer

- au sein du même microorganisme des séquences d'ADN qui ne le sont pas naturellement, soit de mutagéniser spécifiquement un fragment d'ADN.
- ATP: adénosine 5'-triphosphate
- BSA: sérum albumine bovine
- 5 CLHP: Chromatographie liquide à haute performance.
- cluster: groupe de gènes.
- Cob: correspond au phénotype de niveau réduit (au moins 10 fois moins que le contrôle) de production de cobalamines.
- Codon stop: codon de terminaison de traduction
- 10 Corrinoïdes : dérivés de l'acide cobyrique possédant le noyau corrine.
- dGTP: 2'-désoxyguanosine 5'-triphosphate
- DMBI: diméthylbenzimidazole
- dNTP: 2'-désoxyribonucléosides 5'-triphosphates
- 15 DTT: dithiothréitol
- gène cob: gène impliqué dans la biosynthèse du cobinamide à partir de l'uro'gen III.
- gène cor: gène impliqué dans la biosynthèse des corrinoïdes à partir de l'uro'gen III.
- 20 kb: kilobases
- NN:DMBI PRT :
- ORF: phase ouverte
- pb: paires de bases
- Protéine COB: protéine intervenant soit comme catalyseur dans la voie
- 25 de biosynthèse des cobalamines, soit comme protéine régulatrice dans le réseau de régulation des gènes cob, soit les deux.
- Protéine COR: protéine intervenant soit comme catalyseur dans la voie de biosynthèse des corrinoïdes, soit comme protéine régulatrice dans le réseau de régulation des gènes cor, soit les deux.
- 30 SAM: S-adénosyl-L-méthionine.
- SDS: sodium dodécyl sulfate
- SP2MT: SAM-L-méthionine:précorrine-2 méthyltransférase
- SUMT: SAM:uro'gen III méthyltransférase
- Uro'gen III: uroporphyrinogène III.

Légendes des figures:

Figure 1:

Structure du coenzyme B12; le groupement 5'-déoxyadénosyl est remplacé par un groupement CH₃ pour la méthylcobalamine, par un groupement cyano pour la cyanocobalamine, par un groupement hydroxyl pour l'hydroxocobalamine.

Figure 2:

Biosynthèse des cobalamines et différentes étapes de cette biosynthèse suivant la littérature. X: ligands axiaux du cobalt; le ligand en a peut être différent du ligand en b. R: ligand en a du cobalt qui définit le type de cobalamine (voir figure 1).

Figure 3:

Structures de l'uro'gen III, du précorrène-1, du précorrène-2 et du précorrène-3.

Figure 4:

Formules développées de l'uro'gen III et de l'acide cobyrinique. D'après la littérature, entre l'uro'gen III et l'acide cobyrinique il se produit 8 transferts de méthyles SAM-dépendants successivement en C2, C7, C20, C17, C12, C1, C15 et C5, une décarboxylation en C12, l'élimination du carbone en C20 et l'insertion de l'atome de cobalt. X: ligands axiaux du cobalt; le ligand en a peut être différent du ligand en b.

Figure 5:

Dernières étapes de la biosynthèse des cobalamines. Afin de clarifier le schéma des détails du noyau corrène ont été omis. Les cinq étapes enzymatiques sont représentées: 1, cobinamide kinase; 2, cobinamide phosphate guanylyltransférase; 3, cobalamine 5'-phosphate synthase; 4, cobalamine 5'-phosphate phosphohydrolase; 5, nicotinate nucléotide:DMBI phosphoribosyltransférase.

Figure 6: Cartes de restriction des fragments ClaI-HindIII-HindIII-HindIII de 5.4 kb; EcoRI de 8.7 kb; SalI-SalI-SalI-SalI-SalI-BglI de 4748 pb et SstI-SstI-BamHI de 3855 pb. Ne figurent que les 20 enzymes de restriction qui coupent le moins fréquemment l'ADN. Les sites de coupure de chaque

enzyme sont indiqués par un trait vertical.

Figure 7:

Séquence nucléotidique des deux brins du fragment ClaI-HindIII-HindIII-HindIII de 5378 pb de *Pseudomonas denitrificans*.

- 5 Le brin situé en haut est à lire 5' vers 3' dans le sens gauche droite qui correspond à l'orientation gauche droite du fragment de la carte de restriction présentée à la figure 6. Le site ClaI se trouve à la position 23 (début du site de coupure) car sur cette séquence se trouvent des sites de restriction PstI, SalI et XbaI qui sont apparus
- 10 lors des clonages dans des multisites en vue du séquençage. La séquence du fragment ClaI-HindIII-HindIII-HindIII commence donc à la position 23.

Figure 8:

- 15 Séquence nucléotidique des deux brins du fragment EcoRI de 8753 pb de *Pseudomonas denitrificans*. Le brin situé en haut est à lire 5' vers 3' dans le sens gauche droite qui correspond à l'orientation gauche-droite du fragment de la carte de restriction présentée à la figure 6.

Figure 9:

- 20 Analyse des probabilités des phases codantes d'après usage des codons en utilisant le programme de Staden et MacLachlan (1982) sur les 6 phases de lecture de la séquence du fragment ClaI-HindIII-HindIII-HindIII de 5378 pb. Pour les phases qui appartiennent au même brin codant, la phase la plus probable
- 25 correspond à celle où une ligne pointillée, non interrompue par des codons stop, est placée sous la ligne de probabilité de cette phase.

1. Séquence allant du nucléotide 1 au nucléotide 1200. Grâce à cette analyse, la phase ouverte 1 est identifiée. Elle commence à l'ATG en position 549 et se termine au TGA en position
- 30 1011.

2. Séquence allant du nucléotide 1000 au nucléotide 2200. Grâce à cette analyse, la phase ouverte 2 est identifiée. Elle commence à l'ATG en position 1141 et se termine au TGA en position 1981.

- 35 3. Séquence allant du nucléotide 1800 au nucléotide 3400.

Grâce à cette analyse, la phase ouverte 3 est identifiée. Elle commence à l'ATG en position 1980 et se termine au TGA en position 3282.

4. Séquence allant du nucléotide 3000 au nucléotide 4500.

5 Grâce à cette analyse, la phase ouverte ouverte 4 est identifiée. Elle commence à l'ATG en position 3281 et se termine au TGA en position 4280.

5. Séquence allant du nucléotide 3800 au nucléotide 5378.

Grâce à cette analyse, la phase ouverte 5 est identifiée. Elle commence au GTG en position 4284 et se termine au TGA en position 5253.

Figure 10:

Analyse des probabilités des phases codantes d'après usage des codons en utilisant le programme de Staden et MacLachlan (1982) sur les 6 phases de lecture du fragment EcoRI de 8753 pb. Pour les phases qui appartiennent au même brin codant, la phase la plus probable correspond à celle où une ligne pointillée, non interrompue par des codons stop, est placée sous la ligne de probabilité de cette phase.

1. Séquence allant du nucléotide 650 au nucléotide 1650.

20 Grâce à cette analyse, la phase ouverte 6 est identifiée. Elle commence à l'ATG en position 736 et se termine au TAG en position 1519.

2. Séquence allant du nucléotide 1400 au nucléotide 3100.

Grâce à cette analyse, la phase ouverte 7 est identifiée. Elle commence à l'ATG en position 1620 et se termine au TAG en position 2997.

3. Séquence allant du nucléotide 2700 au nucléotide 3700.

Grâce à cette analyse, la phase ouverte 8 est identifiée. Elle commence à l'ATG en position 3002 et se termine au TGA en position 3632.

4. Séquence allant du nucléotide 3500 au nucléotide 4100.

Grâce à cette analyse, la phase ouverte 9 est identifiée. Elle commence au GTG en position 3631 et se termine au TGA en position 4366.

5. Séquence allant du nucléotide 4150 au nucléotide 5150.

Grâce à cette analyse, la phase ouverte 10 est identifiée. Elle commence à l'ATG en position 4365 et se termine au TGA en position 5127.

6. Séquence allant du nucléotide 5000 au nucléotide 6000.

5 Grâce à cette analyse, la phase ouverte 11 est identifiée. Elle commence à l'ATG en position 5893 et se termine au TAG en position 5110.

7. Séquence allant du nucléotide 5700 au nucléotide 7200.

Grâce à cette analyse, la phase 12 est identifiée. Elle commence à l'ATG en position 5862 et se termine au TAA en position 7101.

8. Séquence allant du nucléotide 7000 au nucléotide 8000.

Grâce à cette analyse, la phase ouverte 13 est identifiée. Elle commence à l'ATG en position 7172 et se termine au TTG en position 7931.

15 Figure 11:

Construction des plasmides pXL556, pXL545 et pXL723. Un fragment ClaI-EcoRV de 2.4 kb contenant les gènes cobA et cobE est excisé du fragment de 5,4 kb puis purifié. Un "linker" EcoRI est ajouté au site EcoRV, puis le fragment est inséré dans le pXL59 entre les sites

20 ClaI-EcoRI. Le plasmide ainsi construit est nommé pXL556. La construction est comparable pour le pXL545 : un fragment ClaI-HindIII-HindIII de 1.9 kb est excisé du fragment de 5,4 kb puis purifié. Ce fragment contient uniquement le gène CobE. Un "linker" EcoRI est ajouté au site HindIII, puis le fragment est inséré dans le

25 pXL59 entre les sites ClaI-EcoRI.

Le pXL723 est construit comme suit: un fragment EcoRI-HindIII de 2.3 kb est excisé du fragment de 5,4 kb, purifié, puis les extrémités sont remplies par le grand fragment de l'ADN polymérase I de E. coli. Ce fragment est cloné dans le pRK290 (Ditta et al., 1981) digéré par

30 EcoRI puis traité avec le large fragment de l'ADN polymérase I de E. coli afin de remplir les extrémités. Les sites de restrictions qui sont figurés entre parenthèse correspondent à des sites qui ont disparu après traitement avec le large fragment de l'ADN polymérase I de E. coli.

35 1 , fragment PstI-SstI de RSF1010 (De Graff et al., 1978); 2,

fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); 3, fragment BamHI-SstI contenant l'opéron lactose de E. coli sans son promoteur, l'opérateur, le site d'initiation de la traduction et les 8 premiers codons non essentiels de lacZ (Casadaban et al., 1983); 4, fragment Sau3AI de Pseudomonas putida KT2440 (Bagdasarian et al., 1981); ori, origine de répllication; nic, site de relaxation; mob, locus essentiel pour la mobilisation; Kmr gén-e de résistance à la kanamycine (Bagdasarian et al., 1981); B, BamHI; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI.

10 Figure 12:

Etudes des insertions des transposons Tn5Spr et Tn5 sur le fragment de 5378 bp. Les insertions du transposon Tn5 sur le plasmide pXL723 sont figurées comme à la figure 14; celles du transposon Tn5Spr, dans le chromosome de la souche SBL27 Rif^r, sont encadrées; les insertions dans le chromosome de SC510 Rif^r de cassettes portant le gène de résistance à la kanamycine (1630 et 1631) sont représentées avec une flèche, selon l'orientation de la transcription du gène de résistance à la kanamycine, sous le numéro de l'insertion. Les phases ouvertes déduites de la séquence sont portées sur cette figure (de cobA à cobE); des signes + ou - sont représentés sous chaque insertion de transposon ou de cassette de résistance pour indiquer que l'insertion est inactivationnelle (-) ou non (+) soit pour la complémentation de différents mutants (cas des insertions de transposons Tn5) ou que l'insertion abolit la production de cobalamines de la souche dans laquelle elle a lieu. Il y a une absence de complémentation lorsque le mutant recombiné synthétise moins de trois fois moins de cobalamines que le niveau de synthèse de la souche dont le mutant dérive. Les inserts des plasmides pXL545 Ω , pXL1500, pXL1397 et pXL302 sont représentés avec les sites de restriction se trouvant à leurs extrémités. Ces inserts sont clonés dans les plasmides à large spectre d'hôte pXL435 et pXL59 (Cameron et al., 1989).

Le plasmide pXL545 Ω correspond au plasmide pXL545 décrit à la figure 11 avec en plus le fragment BamHI de 2 kb du pHP45 Ω (Prentki et Krisch) contenant un gène de résistance à la spectinomycine cloné au site BamHI du pXL545.

Le plasmide pXL1500 correspond au fragment BglIII-SstI de 4.2 kb présenté sur cette figure cloné aux sites BamHI et SstI du pKT230 (Bagdasarian et al., 1981) présenté à la figure 30; le pXL1397 correspond au fragment HindIII-SstI de 2.4 kb désigné sur la figure
 5 inséré entre les sites HindIII et SstI du multisite du pXL435 (Cameron et al., 1989) décrit sur la figure 30; le plasmide pXL302 correspond au fragment EcoRI-HindIII de 2.3 kb tel qu'il est décrit sur la figure inséré entre les site EcoRI et HindIII du pXL59 (Cameron et al., 1989), décrit sur la figure 30, le site HindIII utilisé étant le site
 10 se trouvant dans le multisite de clonage du pXL59; le pXL723 est décrit à la figure 11 tout comme le pXL545.

Des signes + ou - sont représentés au dessus de chacun de ces inserts pour indiquer s'il y a complémentation par le plasmide en question des insertions chromosomiques représentées au-dessus. C, ClaI; E, EcoRI;
 15 H, HindIII; RV, EcoRV; Sau, Sau3AI; S, SstI.

Figure 13:

Construction des plasmides pXL253 et pXL367. Le fragment EcoRI de 8.7 kb est excisé puis purifié à partir du plasmide pXL151. Il est cloné au site EcoRI du pKT230 pour donner le pXL253. Ce même fragment est
 20 inséré au site EcoRI du pRK290 (Ditta et al., 1981) pour donner le pXL367. 1, fragment PstI-SstI de RSF1010 (De Graff et al., 1978); 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); ori, origine de répllication; nic, site de relaxation; mob, locus essentiel -pour la mobilisation (Bagdasarian et al., 1981); B, BamHI; C, ClaI;
 25 E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI; tetr gène de résistance à la tétracycline; Kmr, gène de résistance à la kanamycine.

Figure 14:

Etudes des insertions des transposons Tn3lacZ et Tn5 sur le fragment
 30 EcoRI de 8.7 kb cloné dans pRK290 (Ditta et al, 1980). Les insertions des transposons Tn3lacZ sont soulignées au contraire de celles des transposons Tn5. Les phases ouvertes déduites de la séquence (cobF à cobM) sont portées sur cette figure et les huit groupes d'insertions inactivationnelles (numérotés de 1 à 8 sont présentés; des signes + ou
 35 - sont représentés sous chaque insertion de transposon pour indiquer

que l'insertion est inactivationnelle (-) ou non (+) pour la complémentation de différents mutants). Il y a une absence de complémentation lorsque le mutant recombiné synthétise moins de trois fois moins de vitamine B12 que le niveau de synthèse de la souche dont le mutant dérive. Ces groupes d'insertions inactivationnelles correspondent aux mutants suivants: 1, G615; 2, G614 et G616; 3, G613 et G164; 4, G620; 5, G638; 6, G610 et G609; 7, G612; 8, G611. Ces mutants sont des mutants Cob d'*Agrobacterium tumefaciens* déjà décrits (Cameron et al, 1989). Une carte de restriction du fragment de 8.7 kb est portée au bas de la figure.

Figure 15:

Les séquences codantes de chacun des gènes du fragment de 5,4 kb, respectivement cobA à cobE, sont indiquées. La séquence des protéines COBA à COBE codées par ces séquences figurent sous leur séquence codante respective, cobA à cobE. la composition en acides aminés de chaque protéine, en nombre et en pourcentage, respectivement de COBA à COBE, sont présentées ainsi que le poids moléculaire, l'index de polarité, le point isoélectrique, la densité optique à 260 nm et à 280 nm d'une solution à 1 mg/ml de protéine purifiée. Le profil d'hydrophilicité de chaque protéine, respectivement de COBA à COBE, est représenté; il a été calculé d'après le programme de Hopp et Woods (1981). Les valeurs positives correspondent à des régions de la protéine qui sont hydrophiles. En abscisse la position des acides aminés est indiquée tandis qu'en ordonnée la valeur de l'index d'hydrophilicité est figurée; lorsque cette valeur est positive, cela indique que la région de la protéine est hydrophile.

Figure 16:

Les séquences codantes de chacun des gènes du fragment de 8,7 kb, respectivement cobF à cobM, sont indiquées. Les séquences des protéines COBF à COBM codées par ces séquences, figurent sous leur séquence. La légende est identique à celle de la figure 15. NB. Nous avons fait commencer la protéine COBF à l'ATG situé à la position 736; il est possible que l'ATG situé à la position 751 soit le véritable codon d'initiation de cette protéine.

Figure 17:

Réaction catalysée par l'acide cobyrinique a,c-diamide synthase. L'ACDAS catalyse l'amidation des fonctions acides carboxyliques des chaines acétate périphériques a et c de l'acide cobyrinique (hydrogénobyrique) pour donner l'acide cobyrinique diamide (hydrogénobyrique diamide) ; le donneur du groupement amine utilisé dans le test enzymatique est la L-glutamine; il donne par désamination l'acide L-glutamique. X correspond aux ligands axiaux du cobalt qui peuvent être différents l'un de l'autre.

Figure 18: Réaction catalysée par la SP2MT. La SP2MT catalyse le transfert d'un méthyle du SAM au dihydrosirohydrochlorine ou précorrène-2 pour donner le précorrène-3. Le groupement méthyle est transféré à la position C20 du noyau porphyrine.

Figure 19: Structure de l'acide hydrogénobyrique et de l'acide hydrogénobyrique a,c-diamide.

Figure 20:

Affinités des protéines COBA et COBF pour le SAM. Les courbes donnent en unités arbitraires la radioactivité à la sortie de la colonne de TSK-125 pour chaque protéine appliquée sur cette colonne. Les temps de rétention sont indiqués en minutes et le pic de radioactivité correspondant au SAM libre est observé au temps de 10 min 30 sec.

Figure 21:

Comparaison des séquences de COBA et de COBI. Seules les régions 1, 2 et 3, de forte homologie, sont présentées. Les signes = sont placés entre les résidus identiques et - entre les résidus homologues (H K R, L I V M, A G S T, Y F W, D E Q N B Z, P, C).

Figure 22:

Comparaison des séquences primaires des protéines COBA de *Pseudomonas denitrificans* et CYSG de *E. coli*. L'alignement a été fait suivant le programme de Kanehisa, 1984. Les signes = sont placés entre les résidus identiques et - entre les résidus homologues (H K R, L I V M, A G S T, Y F W, D E Q N B Z, P, C). Les régions 1, 2, 3 correspondent aux zones de forte homologie entre les protéines.

Figure 23:

Comparaison des séquences de CYSG de *E. coli* avec des protéines COB de *Pseudomonas denitrificans* (COBA, COBF, COBI, COBJ, COBL et COBM). Les

comparaisons portent sur les régions 1, 2 et 3, de fortes homologues, qui existent entre CYSG, COBA et COBI. Les positions sur les séquences protéiques des régions présentant des homologues sont présentées sur la figure. Nous avons considéré les groupes de résidus homologues
 5 suivants: H K R, L I V M, A G S T, Y F W, D E Q N B Z, P, C. Si, à une même position, il se trouve au moins 3 résidus homologues, nous avons encadré ces acides aminés.

Figure 24:

Construction des plasmides pXL1148 et pXL1149. Le pXL1148 est
 10 construit comme suit: le fragment BamHI-BamHI-SstI-SstI de 1.9 kb du fragment de 8,7 kb contenant les gènes cobH et cobI est purifié, et des "linkers" XbaI et EcoRI sont placés respectivement aux extrémités BamHI et SstI. Ce fragment est ensuite inséré entre les sites XbaI et EcoRI du plasmide à large spectre d'hôte pXL59 (Cameron et al., 1989)
 15 pour donner le plasmide pXL1148. Le pXL1149 est construit comme le pXL1148 si ce n'est que le fragment initialement purifié est le fragment BamHI-BamHI-SstI de 1.5 kb au lieu du fragment contenant en plus le petit fragment SstI de 400 pb utilisé pour le pXL1148. Le fragment subit ensuite les mêmes traitements enzymatiques et le même
 20 clonage dans le pXL59. 1, fragment PstI-SstI de RSF1010 (De Graff et al., 1978); 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); 3, fragment BamHI-SstI contenant l'opéron lactose de E. coli sans promoteur, opérateur, site d'initiation de la traduction et les 8 premiers codons non essentiels de lacZ (Casadaban et al., 1983); 4,
 25 fragment Sau3AI de Pseudomonas putida KT2440 (Bagdasarian et al., 1981); ori, origine de répllication; nic, site de relaxation; Kmr gène de résistance à la kanamycine; mob, locus essentiel pour la mobilisation (Bagdasarian et al., 1981); B, BamHI; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI.

30 Figure 25:

Protéines totales des souches SC510 Rif^r, SC510 Rif^r pKT230, SC510 Rif^r pXL1148, SC510 Rif^r pXL1149 analysées en PAGE-SDS à 10% comme cela est décrit. Les bactéries ont été cultivées pendant 4 jours en milieu PS4, puis des lysats des protéines totales ont été effectués.
 35 Piste 1, SC510 Rif^r; piste 2, SC510 Rif^r pXL1149; piste 3, SC510 Rif^r

pXL1148; piste 4, SC510 Rif^r pKT230. Les masses moléculaires des marqueurs de masse moléculaire sont indiquées. Les positions où les protéines COBI et COBH migrent sont indiquées.

Figure 26:

- 5 Construction des plasmides pXL1496 et pXL1546. Le plasmide pXL1496 permet de surexprimer la protéine COBF chez *E. coli* et le plasmide pXL1546 permet de surexprimer COBF chez *Pseudomonas denitrificans*. Le fragment EcoRI-XhoI de 2.2 kb est excisé et purifié à partir du fragment de 8.7 kb. Il est cloné au site EcoRI du phage M13mp19 pour
- 10 donner le plasmide pXL1405. Ensuite un site NdeI est introduit par mutagenèse dirigée, comme cela est décrit précédemment, à la position 733 de ce fragment; de cette manière un site NdeI se trouve juste sur le codon d'initiation présumé du gène *cobF*. Le nouveau plasmide ainsi obtenu est nommé pXL1406. Un fragment NdeI-SphI-SphI de 1.5 kb,
- 15 contenant le gène *cobF* à partir de son codon d'initiation présumé, est purifié après digestion partielle par les enzymes appropriées et ligaturé avec les fragments appropriés du plasmide pXL694 (fragment EcoRI-NdeI de 120 pb contenant des signaux d'expression de *E. coli* -voir texte- et fragment EcoRI-SphI de 3.1 kb contenant le gène de
- 20 résistance à l'ampicilline, les fonctions de répllication du plasmide ainsi que les terminateurs de l'opéron *rrnB* de *E. coli* comme cela est décrit dans le texte). Le plasmide ainsi construit est nommé pXL1496. Le pXL1546 est construit comme suit: le fragment EcoRI-BamHI-BamHI de 2 kb du pXL1496 est purifié par digestion partielle avec les enzymes
- 25 appropriées; ce fragment contient les signaux d'expression de *E. coli*, suivis du gène *cobF*, puis la partie 5' du gène *cobG*, elle-même suivie des terminateurs de l'opéron *rrnB* de *E. coli* comme cela est décrit dans le texte; ce fragment est cloné dans le plasmide multihôte pKT230 (Bagdasarian et al., 198-1) décrit sur la figure 30. B, BamHI; C,
- 30 ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI; Kmr, gène de résistance à la kanamycine; Amp, gène de résistance à l'ampicilline.

Figure 27:

- Protéines totales des souches SC510 Rif^r, SC510 Rif^r pKT230, SC510
- 35 Rif^r pXL1546 analysées en PAGE-SDS à 10% comme cela est décrit. Les

bactéries ont été cultivées pendant 4 jours en milieu PS4, puis des lysats des protéines totales ont été effectués. Piste 1, SC510 Rif^r; piste 2, SC510 Rif^r pKT230; piste 3, SC510 Rif^r pXL1546. Les masses moléculaires des marqueurs de masse moléculaire sont indiquées. La position où la protéine COBF migre est indiquée.

Figure 28:

Protéines totales des souches E. coli B et E. coli B pXL1496 analysées en PAGE-SDS à 10 % comme cela est décrit. Piste 1, E. coli pXL1496 cultivée en absence de tryptophane; piste 2, E. coli pXL1496 cultivée dans les mêmes conditions en présence de tryptophane; piste 3, E. coli cultivée en absence de tryptophane; piste 4, E. coli cultivée dans les mêmes conditions en présence de tryptophane. Les masses moléculaires des marqueurs sont indiquées. La position de migration de la protéine COBF est indiquée.

Figure 29:

Construction des plasmides pXL525 et pXL368. Le pXL368 est construit comme suit: le fragment EcoRV-ClaI de 2.4 kb (contenant les gènes *cobA* et *cobE*) est purifié à partir du plasmide pXL556 (B. Cameron et al., 1989) ce qui permet d'obtenir ce fragment avec un site BamHI et un site XbaI aux extrémités; ce fragment est cloné dans le pXL203 aux sites BamHI et XbaI. Pour la construction du pXL525 un "linker" XbaI est ajouté au site EcoRI situé à l'extrémité droite du fragment EcoRI de 8.7 kb; ce fragment EcoRI-XbaI de 8.7 kb est ensuite cocloné avec le fragment EcoRI-XbaI de 2.4 kb provenant du pXL556 et contenant *cobA* et *cobE*. Les sites de restrictions qui sont figurés entre parenthèse correspondent à des sites qui ont disparu après traitement avec le large fragment de l'ADN polymérase I de E. coli. 1, fragment PstI-SstI de RSF1010 (De Graff et al., 1978); 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); ori, origine de répllication; nic, site de relaxation; mob, locus essentiel pour la mobilisation; Kmr, gène de résistance à la kanamycine (Bagdasarian et al., 1981); B, BamHI; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI; tet, gène de résistance à la tétracycline; Ampr et Amp, gène de résistance à l'ampicilline.

Figure 30:

Plasmides, du groupe d'incompatibilité Q, à large spectre d'hôte chez les bactéries gram-négatives. Ces plasmides sont décrits dans une précédente publication (Cameron et al., 1989) et sont utilisés dans la présente invention. 1, fragment PstI-SstI de RSF1010 (De Graff et al., 1978); 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); 3, fragment BamHI-SstI contenant l'opéron lactose de E. coli sans promoteur, opérateur, site d'initiation de la traduction et les 8 premiers codons non essentiels de lacZ (Casadaban et al., 1983); 4, fragment Sau3AI de Pseudomonas putida KT2440 (Bagdasarian et al., 1981); ori, origine de répllication; nic, site de relaxation; Kmr gène de résistance à la kanamycine; Smr, gène de résistance à la streptomycine ; mob , locus essentiel pour la mobilisation (Bagdasarian et al., 1981); B, BamHI; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI.

15 Figure 31:

Temps de rétention de différents étalons corrinoïdes (1 mg/étalon) sur le système de séparation de décrit à l'exemple 9. La colonne utilisée est une colonne de Nucléosil C-18 (Macherey-Nagel). Au niveau de chaque pic d'absorbance un numéro est représenté correspondant au corrinoïde décrit ci-dessous. Le temps de rétention est porté en abscisse et l'absorbance à 371 nm se trouve en ordonnée. 1, acide cobyrique; 2, acide cobyrique α -amide; 3, acide cobyrique γ -amide; 4, acide cobyrique α,γ -diamide; 5, acide cobyrique ϵ -amide; 6, acide cobyrique ϵ,γ -diamide; 7, acide cobyrique α,ϵ -diamide; 8, acide cobyrique triamide; 9, acide cobyrique tetraamide; 10, acide cobyrique pentaamide; 11, acide cobyrique; 12, GDP-cobinamide; 13, cobinamide phosphate; 14, cobinamide; 15, cyanocobalamine 5'-phosphate; 16, cyanocobalamine.

Figure 32:

30 Séquence nucléotidique des deux brins du fragment SalI-SalI-SalI-SalI-BglI de 4748 pb de Pseudomonas denitrificans. Le brin situé en haut est à lire de 5' vers 3' dans le sens gauche droite qui correspond à l'orientation gauche droite du fragment de la carte de restriction présentée sur la figure 6.

35 Figure 33: Séquence nucléotidique des deux brins du fragment SstI-

SstI-BamHI de 3855 pb de *Pseudomonas denitrificans*. Le brin situé en haut est à lire de 5' vers 3' dans le sens gauche droite qui correspond à l'orientation gauche droite du fragment de la carte de restriction présentée sur la figure 6.

5 Figure 34:

Analyse des probabilités des phases codantes d'après l'usage des codons en utilisant le programme de Staden et MacLachlan (1982) sur les six phases de lecture du fragment SalI-SalI-SalI-SalI-SalI-BglI de 4748 pb. Pour les phases qui appartiennent au même brin codant, la phase la plus probable correspond à celle où une ligne pointillée, non interrompue par des codons stop, est placée sous la ligne de probabilité de cette phase. 4a. Analyse de la séquence correspondant aux nucléotides 200 à 800. Cette analyse permet d'identifier la phase ouverte 14. Elle commence à l'ATG en position 660 et se termine au TGA en position 379. 4b. Analyse de la séquence correspondant aux nucléotides 800 à 1500. Cette analyse permet d'identifier la phase ouverte 15. Elle commence au GTG en position 925 et se termine au TAA en position 1440. 4c. Analyse de la séquence correspondant aux nucléotides 1450 à 2600. Cette analyse permet d'identifier la phase ouverte 16. Elle commence à l'ATG en position 1512 et se termine au TGA en position 2510. 4d. Analyse de la séquence correspondant aux nucléotides 2500 à 4650. Cette analyse permet d'identifier la phase ouverte 17. Elle commence au GTG en position 2616 et se termine au TGA en position 4511.

25 Figure 35:

Analyse des probabilités des phases codantes d'après l'usage des codons en utilisant le programme de Staden et MacLachlan (1982) sur les six phases de lecture du fragment SstI-SstI-BamHI de 3855 pb. Pour les phases qui appartiennent au même brin codant, la phase la plus probable correspond à celle où une ligne pointillée, non interrompue par des codons stop, est placée sous la ligne de probabilité de cette phase. 5a. Analyse de la séquence correspondant aux nucléotides 1 à 905. Cette analyse permet d'identifier la phase ouverte 18. Elle commence à l'ATG en position 809 et se termine au TGA en position 108. 5b. Analyse de la séquence correspondant aux

nucléotides 955 à 2105. Cette analyse permet d'identifier la phase ouverte 19. Elle commence à l'ATG en position 1971 et se termine au TGA en position 1063. 5c. Analyse de la séquence correspondant aux nucléotides 2000 à 3300. Cette analyse permet d'identifier la phase ouverte 20. Elle commence à l'ATG en position 2099 et se termine au TAG en position 3115. 5d. Analyse de la séquence correspondant aux nucléotides 3250 à 3855. Cette analyse permet d'identifier la phase ouverte 21. Elle commence à l'ATG en position 3344 et se termine au TGA en position 3757.

10 Figure 36:

Construction des plasmides pXL233, pXL843 et pXL1558 à partir du pXL154. Les plasmides sont construits de la manière suivante. Le fragment EcoRI de 3,5 kb, contenant le gène cobS tronqué et la séquence en amont, est excisé du pXL154, puis purifié et cloné dans le site EcoRI du pKT230. Le plasmide ainsi construit est nommé pXL233. Le fragment EcoRI-XhoI-XhoI de 3,5 kb, contenant le gène cobT et la séquence en aval, est excisé et purifié à partir du pXL154 par digestions partielles. Le fragment EcoRI-EcoRI-EcoRI de 4,3 kb contenant le gène cobS et la séquence en amont est excisé et purifié à partir du pXL154, puis ligaturé au fragment de 3,5 kb précédent. Le fragment EcoRI-XhoI de 8 kb environ ainsi obtenu est cloné dans les sites EcoRI et SalI du pXL59 pour générer le plasmide pXL843. Le plasmide pXL1558 est construit de la façon suivante: le fragment HindIII-HindIII de 12 kb est excisé du pXL154 et purifié puis les extrémités sont remplies par le grand fragment de l'ADN polymérase I de E. coli. Cet insert est cloné dans le pRK290 (Ditta et al. 1981) digéré par EcoRI puis traité avec le large fragment de l'ADN polymérase I de E. coli pour rendre les extrémités franches. Les sites de restriction qui sont figurés entre parenthèse correspondent à des sites qui ont disparus au cours du clonage. 1, fragment PstI-SstI de RSF1010 (Degraff et al., 1978); 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); B, BamHI; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI; Tet gène de résistance à la tétracycline; Kmr gène de résistance à la kanamycine; Smr, gène de résistance à la streptomycine.

Figure 37:

Etude des insertions du transposon Tn5Sp sur l'insert HindIII-HindIII de 12 kb du pXL154. Les insertions du transposon sont cartographiées sur l'insert HindIII-HindIII de 12 kb, cloné sur le pXL1558. Les insertions chromosomiques dans la souche SC510 Rif^r sont encadrées, celle qui ne l'est pas, est introduite dans la souche SBL27 Rif^r. Un signe plus ou moins est représenté sous chaque insertion pour indiquer le phénotype Cob de la souche ayant cette insertion. L'absence de complémentation (ou la complémentation) de la souche G2035 par des plasmides pXL1558::Tn5Sp est indiquée par des signes moins (ou plus) au dessous de chaque insertion. Les inserts des plasmides, décrits dans la figure 36, sont représentés. Les signes plus (ou moins), sur ces plasmides et alignés avec les insertions au transposon, schématisent la complémentation (ou (l'absence) de la souche mutée au transposon par le plasmide. Les phases ouvertes déduites de la séquence sont aussi portées sur cette figure (ORF14 à 17 ainsi que les gènes cob correspondants (cobS et cobT). E : EcoRI; H : HindIII; X : XhoI.

Figure 38:

Construction des plasmides pXL1286, pXL1303, pXL1324, pXL1490B et pXL1557 à partir du pXL519. La position du fragment séquencé est présentée dans la partie supérieure de la figure au dessus de la carte de restriction du cluster; il s'agit d'un SstI-SstI-BamHI de 3.9 kb. Les plasmides sont construits de la manière suivante. Le fragment BglII-EcoRI de 2 kb, contenant le gène cobU et la la séquence en aval, est excisé du pXL519 puis purifié et cloné aux sites BamHI et EcoRI du pKT230 pour générer le plasmide pXL1286. Le fragment SstI-EcoRI de 2,7 kb, contenant le gène cobV tronqué, le gène cobU et la séquence en aval, est excisé du pXL519 puis purifié et cloné aux sites SstI et EcoRI du pKT230 pour générer le plasmide pXL1324. Le fragment SstI-SstI de 1,6 kb, contenant le gène cobV tronqué et la séquence en amont, est excisé du pXL519 puis purifié et cloné au site SstI du pKT230 pour générer le plasmide pXL1303. Le fragment SstI-SstI-BamHI de 3.85 kb est purifié après digestion totale du pXL519 par BamHI et digestion partielle par SstI. Ce fragment est ensuite cloné aux sites

- BamHI et SstI du pKT230 pour générer le pXL1490B. Le plasmide pXL1557 est construit de la façon suivante: le fragment HindIII-BamHI de 9 kb est excisé du pXL519 et purifié puis les extrémités sont remplies par le grand fragment de l'ADN polymérase I de E. coli. Cet insert est
- 5 cloné dans le pRK290 (Ditta et al. 1981) digéré par EcoRI puis traité avec le large fragment de l'ADN polymérase I de E. coli pour rendre les extrémités franches. Les sites de restriction qui sont figurés entre parenthèse correspondent à des sites qui ont disparus au cours du clonage. 1, fragment PstI-SstI de RSF1010 (Degraff et al., 1978);
- 10 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); B, BamHI; Bg, BglII; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI; Tetr gène de résistance à la tétracycline; Kmr gène de résistance à la kanamycine; Smr, gène de résistance à la streptomycine.
- 15 Figure 39: Etude des insertions du transposon Tn5Sp sur l'insert HindIII-BamHI de 9 kb du pXL519. Les insertions du transposon sont cartographiées sur l'insert HindIII-BamHI de 9 kb, cloné sur le pXL1557. Les insertions chromosomiques dans la souche SC510 Rif^r sont encadrées, celles qui ne le sont pas, sont introduites dans la souche
- 20 SBL27 Rif^r. Un signe plus ou moins est représenté sous chaque insertion pour indiquer le phénotype Cob de la souche ayant cette insertion. L'absence de complémentation (ou la complémentation) de la souche G2040 par des plasmides pXL1557::Tn5Sp est indiquée par des signes moins (ou plus) au dessous de chaque insertion. Les inserts des
- 25 plasmides, décrits dans la figure 6, sont représentés. Les signes plus (ou moins), sur ces plasmides et alignés avec les insertions au transposon, schématisent la complémentation (ou l'absence) de la souche mutée au transposon par le plasmide. Les phases ouvertes déduites de la séquence sont aussi portées sur cette figure (ORF18 à
- 30 21) ainsi que les gènes cob correspondant (cobU et cobV).

Figure 40:

- Séquences codantes de chacun des gènes du fragment de 4,8 kb, respectivement cobX, cobS et cobT, sont indiquées. La séquence des protéines COBX, COBS et COBT codées par ces séquences figure sous les
- 35 séquences codantes respectives cobX, cobS et cobT. La légende est

identique à celle de la figure 15.

Figure 41:

Séquences codantes de chacun des gènes du fragment de 3,9 kb, respectivement cobU et cobV, sont indiquées. La séquence des protéines COBU et COBV codées par ces séquences figure sous les
5 séquences codantes respectives cobU et cobV. La légende est identique à celle de la figure 15.

Figure 42:

A. Protéines totales des souches E. coli BL21 pLysS pET3b, E. coli BL21
10 pLysS pXL1937 analysées sur PAGE-SDS 10%. Piste 1, BL21 pLyspET3b; piste 2, E. coli BL21 pLysS pXL1937.

B. Protéines totales des souches : E.coli BL21, E.coli BL21 pXL1874 et E.coli BL21 pXL1875 analysées sur PAGE-SDS 10%. Piste 1, E.coli BL21; piste 2, E.coli BL21 pXL1874; piste E.coli BL21 pXL1875.

15 Les masses moléculaires des marqueurs sont indiquées. La bande correspondant à la protéine surexprimée est indiquée par une flèche.

Figure 43: Séquence nucléotidique des deux brins du fragment SstI-SstI-SstI-SstI-BglII-BglII de 13144 pb de Pseudomonas denitrificans. Le brin situé en haut est à lire de 5' vers 3' dans le
20 sens gauche droite qui correspond à l'orientation gauche droite du fragment de la carte de restriction présentée sur la figure 46.

Figure 44: Carte de restriction du fragment SstI-SstI-SstI-SstI-BglII-SstI-BglII de 13144 pb de Pseudomonas denitrificans. La ou les positions de sites de restriction courant sont
25 indiquées par ordre croissant du nombre de coupure sur le fragment séquencé; les positions correspondent à la séquence présentée sur la figure 43.

Figure 45: Analyse des probabilités des phases codantes d'après l'usage des codons en utilisant le programme de Staden et MacLachlan (1982) sur
30 les six phases de lecture du fragment SstI-SstI-SstI-SstI-BglII-SstI-BglII de 13144 pb de Pseudomonas denitrificans. Pour les phases qui appartiennent au même brin codant,

la phase la plus probable correspond à celle où une ligne pointillée, non interrompue par des codons stop, est placée sous la ligne de probabilité de cette phase.

1. Séquence correspondant aux nucléotides 1 à 2266. Cette analyse permet d'identifier la phase ouverte 22. Elle commence à l'ATG en position 429 et se termine au TAG en position 1884.

2. Séquence correspondant aux nucléotides 2266 à 4000. Cette analyse permet d'identifier la phase ouverte 23. Elle commence à l'ATG en position 3364 et se termine au TGA en position 3886.

3. Séquence correspondant aux nucléotides 3800 à 5000. Cette analyse permet d'identifier la phase ouverte 24. Elle commence à l'ATG en position 3892 et se termine au TAG en position 4954.

4. Séquence correspondant aux nucléotides 5000 à 9000. Cette analyse permet d'identifier la phase ouverte 25. Elle commence à l'ATG en position 5060 et se termine au TAG en position 8885.

5. Séquence correspondant aux nucléotides 9000 à 9700. Cette analyse permet d'identifier la phase ouverte 26. Elle commence à l'ATG en position 9034 et se termine au TGA en position 9676.

6. Séquence correspondant aux nucléotides 9600 à 13144. Cette analyse permet d'identifier les phase ouvertes 27, 28, 29 et 30. Elles commencent respectivement aux ATG en position 9678, 10895, 11656 et 13059 et se terminent aux codons stop en position 10101, 10304, 12181 et 12366. Les phases ouvertes 28 et 30 se trouvent sur le brin complémentaire de celui codant correspondant à toutes les autres phases ouvertes.

Figure 46: Fragment EcoRI-BglII-EcoRI-BglII de 13.4 kb, positions des insertions des transposons Tn5Sp sur le fragment EcoRI de 9,1 kb, positions des insertions des transposons Tn5 sur l'insert du plasmide pXL189 ainsi que les inserts des divers plasmides utilisés lors des expériences de complémentation des souches SC510 Rif^r ::Tn5Sp. Les complémentations des mutants SC510 Rif^r ::Tn5Sp par les plasmides sont indiquées (+)-entre 5 % et 100 % du niveau de la souche parentale, SC510

- Rif^r-, (.)-complémentation partielle, entre 0.5 à 5 % du niveau de SC510 Rif^r-, ou (-)-absence de complémentation, c'est à dire, moins de mille fois moins que SC510 Rif^r- positionnés juste au dessus des traits schématisant l'insert des plasmides et alignés avec les sites d'insertion des mutant correspondants. Au dessous de la cartographie des insertions des transposons Tn5 sur l'insert du plasmide pXL189 est représentée la complémentation (+), ou l'absence de complémentation (-) de ces plasmides mutants pour les mutants d'Agrobacterium tumefaciens G632 et G633. Sur la partie droite de la figure se trouve un tableau représentant la complémentation des mutants G622, G623 et G630 (Cameron et al., 1989) par différents plasmides; (+)-complémentation totale, 100 % du niveau de la souche parentale, C58C9 Rif^r-, (.)-complémentation partielle, entre 10 et 50 % du niveau de C58C9 Rif^r-, ou (-)-absence de complémentation-.
- Les différents plasmides dont l'insert est représenté sont construits comme suit (les fragments sont excisés soit du pXL156 soit du pXL157):
- pXL618 correspond au fragment EcoRI-BamHI de 2.5 kb cloné aux mêmes sites de pKT230 (Bagdasarian et al., 1981);
- pXL593 correspond au fragment BamHI de 3.1 kb cloné au site BamHI de pKT230 (Bagdasarian et al., 1981);
- pXL623 correspond au fragment BamHI-XhoI de 1.9 kb cloné aux sites BamHI-SalI de pXL59 (Cameron et al., 1989);
- pXL1909 correspond au fragment BamHI-BamHI-BamHI de 8.4 kb cloné au site BamHI de pKT230 (Bagdasarian et al., 1981);
- le pXL221 correspond au fragment EcoRI-ClaI de 1.6 kb cloné aux mêmes site de pXL59 (le site ClaI dans lequel est cloné ce fragment est le site ClaI du multisite de pXL59 (Cameron et al., 1989);
- les pXL1908 et 1938 correspondent au même insert, fragment XhoI-BamHI-BamHI de 6.5 kb, auquel des linkers XbaI ont été ajoutés; cet insert est cloné dans les deux orientations au site XbaI de pXL435 (Cameron et al., 1989); une flèche positionnée sur la figure indique la position du gène de résistance à la kanamycine vis à vis des extrémités de l'insert des deux plasmides;
- pXL208 correspond au fragment BamHI de 5.2 kb cloné au site BamHI de pKT230 (Bagdasarian et al., 1981);

pXL297 correspond au fragment EcoRI de 9.1 kb cloné au site EcoRI de pKT230 (Bagdasarian et al., 1981).

Les phases ouvertes (PO) définies par le séquençage du fragment (PO 22 à 30) sont représentées ainsi que les gènes cob correspondants; une flèche indique la polarité de la transcription.

5 E, EcoRI; B, BamHI; Bg, BglII; Cl, ClaI; Sau, Sau3AI; X, XhoI;

Figure 47: Séquences codantes de chacun des gènes du fragment de 13.4 kb, respectivement cobQ, cobP et cobW, cobN et cobO sont indiquées. La séquence des protéines COBQ, COBP, COBW, COBN et COBO codées par ces séquences figurent sous leur séquence codante respective cobQ,
10 cobP, cobW, cobN et cobO. La légende est identique à celle de la figure 15.

figure 48 : A-Séquence NH2-terminale de la SUMT de M. ivanovii et séquence des oligonucléotides 923 946, 947; -, signifie qu'à cette position, le résidu n'a pas pu être déterminé; pour l'oligonucléotide
15 antisens les acides aminés indiqués au-dessus de la séquence correspondent aux anticodons présentés. B-présentation de l'amplification enzymatique d'un fragment interne au gène de structure de la SUMT de M. ivanovii avec les oligonuléotides 946 et 947.

Figure 49 : Constrution de la forme réplivative recombinante pG10. Le
20 fragment de 615 pb obtenu par amplification est digéré par HindIII et EcoRI puis purifié comme cela est décrit. Ce fragment est ensuite mis à ligaturer avec la forme réplivative du phage M13mp19 digérée par les mêmes enzymes. Le clone recombinant est trouvé comme cela est décrit dans le texte.

25 Figure 50 : Autoradiographie d'un blot d'ADN génomique de M. ivanovii digéré par diverses enzymes, séparé par électrophorèse en gel d'agarose puis transféré sur membrane de Nylon comme cela est décrit précédemment. La membrane est hybridée avec la sonde pG10 comme cela est décrit précédemment. 1, HindIII-BglII; 2, KpnI-BglII; 3, EcoRI-BglII; 4,
30 BglII-PstI. Les tailles des différents fragments qui hybrident avec la

sonde sont représentés en kb.

Figure 51 : Séquence nucléotidique des deux brins du fragment de 955 pb de *M. ivanovii*. Le brin situé en haut est à lire de 5' vers 3' dans le sens gauche droite.

- 5 Figure 52: Séquence codante du gène *corA* de *M. ivanovii* obtenue à partir de la séquence de 955 pb. La séquence primaire de la protéine CORA est aussi représentée. Les acides aminés sont représentés au-dessus de leur codon et le codon stop est figuré par une étoile. Les principales propriétés physiques de la protéine CORA de *M. ivanovii*, à savoir, la
- 10 composition en acides aminés, en nombre et en pourcentage, le poids moléculaire, l'index de polarité, le point isoélectrique, la densité optique à 280 nm d'une solution à 1 mg/l de protéine purifiée. Le profil d'hydrophobicité de la protéine CORA de *M. ivanovii*; ce profil a été réalisé d'après le programme de Hopp et Woods (1981). Les valeurs
- 15 positives correspondent à des régions de la protéine qui sont hydrophiles. Sont indiquées, la position des acides aminés en abscisse, et en ordonnée la valeur de l'index d'hydrophilicité; si cette valeur est positive, cela indique que dans cette région la protéine est hydrophile.
- 20 Figure 53 : Comparaison des séquences primaires des protéines COBA de *P. denitrificans* et CORA de *M. ivanovii*. Les protéines ont été alignées grâce au programme de Kanehisa (1984). =, acides aminés identiques; -, acides aminés homologues d'après les critères définis précédemment (voir figure 22 et 23).
- 25 Figure 54 : Construction des plasmides pXL1832 et pXL1841. Les légendes décrites placées sur la figure permettent de suivre les constructions.

Techniques générales de clonage, de biologie moléculaire et de biochimie.

Les méthodes classiques de biologie moléculaire telles que la centrifugation d'ADN plasmidique en gradient de chlorure de césium-bromure d'éthidium, les digestions par des enzymes de restriction, l'électrophorèse sur gel, l'électroélution des fragments d'ADN à partir de gels d'agarose, la transformation dans *E. coli*, etc, sont décrites dans la littérature (Maniatis et al., 1982, Ausubel et al., 1987).

Les enzymes de restriction ont été fournies par New-England Biolabs (Biolabs), Bethesda Research Laboratories (BRL) ou Amersham Ltd (Amersham). Les oligonucléotides "linkers" ont été fournis par Biolabs. Pour les ligatures, les fragments d'ADN sont séparés selon leur taille sur des gels d'agarose 0.7 % ou acrylamide 8 %, purifiés par électroélution, extraits au phénol, précipités à l'éthanol puis incubés dans un tampon Tris-HCl pH 7.4 50 mM, MgCl₂ 10 mM, DTT 10 mM, ATP 2 mM, en présence d'ADN ligase du phage T4 (Biolabs). Si nécessaire, les fragments d'ADN ayant des extrémités 5' proéminentes sont déphosphorylés par un traitement à la phosphatase alcaline d'intestin de veau (CIP, Pharmacia) à 37°C pendant 30 mn dans le tampon suivant: glycine 100 mM, MgCl₂ 1 mM, ZnCl₂ 1 mM, pH 10.5. La même technique est utilisée pour la déphosphorylation des extrémités 3' proéminentes ou franches, mais le traitement est de 15 mn à 37°C puis de 15 mn à 56 °C. L'enzyme est inactivée par chauffage du mélange réactionnel à 68°C pendant 15 mn -en présence de SDS 1% et de NaCl 100 mM suivi d'une extraction au phénol-chloroforme et d'une précipitation à l'éthanol. Le remplissage des extrémités 5' proéminentes est effectué par le fragment de Klenow de l'ADN polymérase I d'*E. coli* (Biolabs). La réaction est effectuée à température ambiante pendant 30 mn dans un tampon Tris-HCl pH 7.2 50 mM, dNTPs 0.4 mM, MgSO₄ 10 mM, DTT 0.1 mM, BSA 50 mg/ml. Le remplissage des extrémités 3' proéminentes est effectué en présence de l'ADN polymérase du phage T4 (Biolabs) selon les recommandations du fabricant. La digestion des extrémités proéminentes est effectuée par traitement limité à la nucléase S1 (BRL) selon les recommandations du fabricant. Des oligonucléotides "linkers" sont ajoutés sur des extrémités de fragments d'ADN comme cela est déjà décrit (Maniatis et al, 1982). La

mutagénèse in vitro par oligodéoxynucléotides est effectuée selon la méthode développée par Taylor et al., 1985, en utilisant le kit distribué par Amersham. Les ADN ligaturés sont utilisés pour transformer la souche rendue compétente: E. coli MC1060 [5 Δ (lacIOPZYA)X74, galU, galK, strAr, hsdR] pour les plasmides ou E. coli TG1 [Δ (lac proA,B), supE, thi, hsdD5/ F' traD36, proA+, B+, lacIq, lacZ Δ M15] pour les formes répliquatives de phages dérivés du bactériophage M13. Les ADN plasmidiques sont purifiés suivant la technique de Birnboim et Doly, 1979. Les minipréparations d'ADN plasmidique sont faites suivant le protocole de Klein et al., 1980. Les préparations d'ADN chromosomiques de bactéries gram-négatives sont réalisées comme cela a été déjà décrit (Cameron et al., 1989). Les sondes radioactives sont préparées par translation de coupure suivant la méthode déjà détaillée (Rigby et al., 1977). Les hybridations entre des séquences d'ADN ainsi que l'immobilisation des acides nucléiques sur membranes de nitrocellulose sont effectuées comme cela a été décrit (Cameron et al., 1989). Lors de clonages pour lesquels il y a une faible probabilité de trouver le clone recombinant recherché, ceux-ci sont trouvés après hybridation sur filtre comme cela est déjà décrit (Maniatis et al., 1982). La séquence nucléotidique de fragments d'ADN est réalisée par la méthode de terminaison de chaînes (Sanger et al., 1977). Dans le mélange réactionnel le dGTP est remplacé par le 7-déaza-dGTP, ceci afin d'éviter les compressions de bandes lors de l'électrophorèse sur gel d'acrylamide, provoquées par le pourcentage en GC élevé de l'ADN. Les milieux de culture utilisés pour la partie bactériologique ont déjà été présentés (Maniatis et al., 1982). Les cultures en milieu PS4 sont réalisées comme cela est déjà décrit (Cameron et al., 1989); les souches de Pseudomonas denitrificans SC510 Rifr et G2 Rifr sont cultivées en milieu PS4 comme suit: des Erlenmeyers de 250 ml contenant 25 ml de milieu PS4, avec si besoin l'antibiotique sélectif pour le plasmide porté par chaque souche, sont ensemencés avec une dilution au 1/100 de préculture saturée en milieu L (Miller 1972), avec, si besoin est, l'antibiotique sélectif pour le plasmide porté par chaque souche; ces cultures sont incubées 6 jours à 30°C, puis les moûts sont analysés pour leur titre en cobalamines ou

bien l'activité enzymatique de certaines enzymes de la voie. Les souches d'*Agrobacterium tumefaciens*, *Pseudomonas putida* et de *Rhizobium meliloti* sont cultivées à 30°C; sauf spécifications, elles sont cultivées en milieu L. Les conjugaisons bactériennes sont
5 réalisées comme cela est déjà décrit (Cameron et al., 1989). Les extraits de protéines totales sont réalisés comme cela est déjà décrit (Ausubel et al., 1987). Les électrophorèses (SDS-PAGE) d'analyse des protéines en gel d'acrylamide en conditions dénaturantes sont effectuées comme cela est déjà décrit (Ausubel et al., 1987).
10 L'appareil PhastSystem (Pharmacia) utilisant le système de tampons discontinus de Laemli (Laemli, 1970) est aussi utilisé; différents gels sont utilisés en fonction des poids moléculaires des protéines à analyser ainsi que leur pureté: PhastGel gradient 8-25 PhastGel Homogeneous 12.5 La coloration est effectuée soit au bleu de Coomassie
15 à l'aide du PhastGel Blue R (Pharmacia), soit au nitrate d'argent en utilisant le PhastGel silver Kit (Pharmacia) en se conformant aux instructions du fabricant. Les séquences NH₂-terminales des protéines sont déterminées par la technique de dégradation d'Edman en utilisant un séquenceur automatisé (Applied Biosystems modèle 407A) couplé à un
20 appareil de CLHP pour l'identification des dérivés phénylthiohydantoïnes.

EXEMPLE 1 - Isolement de fragments d'ADN de *P.denitrificans* contenant des gènes Cob

Cet exemple décrit l'isolement de fragments d'ADN de *Pseudomonas denitrificans* portant des gènes Cob. Ces fragments ont été mis en
25 évidence par des expériences de complémentation de mutants Cob d'*A.tumefaciens* et *P.putida* (Cameron et al., 1989).

Ces mutants Cob ont été obtenus par mutagénèse avec la N-méthyl-N'-nitro-N-nitrosoguanidine selon la technique de Miller
30 (Miller et al., 1972), ou par insertions du transposon Tn5. De cette manière, des souches incapables de synthétiser les cobalamines ont été mises en évidence, et en particulier, le mutant cob G572 de *P.putida*, et les mutants Cob G159, G161, G164, G169, G171, G258, G609, G610, G611, G612, G613, G614, G615, G616, G620, G622, G623, G630, G632,

G633, G634, G638, G642, G643, G2034, G2035, G2037, G2038, G2039, G2040, G2041, G2042 et G2043 de A.tumefaciens.

Parallèlement, une banque d'ADN génomique de P.denitrificans est réalisée dans un vecteur mobilisable à large spectre d'hôte, pXL59, par digestion, en présence d'enzymes de restriction, de 5 µg d'ADN (Cameron et al., 1989).

Par complémentation, plusieurs plasmides ont pu être isolés, permettant de compléter les mutants cob de P.putida et d'A.tumefaciens. Parmi ceux-ci on notera plus particulièrement les plasmides pXL151, pXL154, pXL156, pXL157 et pXL159.

Ces plasmides ont été isolés et des fragments d'ADN ont pu être excisés, purifiés et analysés par restriction. Ces fragments sont présentés sur les figures 6 et 44: Un fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb, un fragment EcoRI-EcoRI de 8,7 kb, un fragment SalI-SalI-SalI-SalI-SalI-BglI de 4,8 kb, un fragment SstI-SstI-BamHI de 3,9 kb et un fragment EcoRI-BglII-EcoRI-BglII de 13,4 kb.

EXEMPLE 2 - Séquençage des fragments d'ADN isolés

Cet exemple illustre le séquençage de fragments d'ADN portant des gènes cob de *Pseudomonas denitrificans* SC510.

2.1. Séquençage d'un fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb

Ce fragment est contenu dans le plasmide pXL157 décrit dans l'exemple 1. Après excision, des sous fragments du fragment de 5,4 kb ont été clonés dans les phages M13mp18, ou M13mp19 (Norlander et al., 1983), ou M13tg130, ou M13tg131 (Kieny et al., 1983) dans les deux orientations. Des délétions ont été ensuite réalisées in vitro par la méthode de Henikoff (1987). Ces délétions ont ensuite été séquencées avec le "primer universel" comme amorce de synthèse des réactions de terminaison de chaîne. Le recouvrement entre ces différentes délétions a permis d'établir la séquence totale, sur les deux brins, du fragment de 5,4 kb (figure 7). Ce fragment comprend 5378 pb. Sur la séquence décrite à la figure 7 figurent, avant le site

2072023

ClaI, trois sites de restriction (PstI, SalI et XbaI) qui sont apparus lors du clonage du fragment en question en vue du séquençage dans des multisites de clonage. Lorsque par la suite nous nous référerons, dans la présente invention, à la séquence de ce fragment

5 ClaI-HindIII-HindIII-HindIII, ce sera à la séquence présentée à la figure 7 où les 22 premières bases ne correspondent pas à de l'ADN de Pseudomonas denitrificans (ainsi toutes les positions de site de restriction ou de début de phase ouverte se réfèrent à la séquence présentée à la figure 7).

10 2.2 Séquence nucléotidique d'un fragment EcoRI-EcoRI de 8,7 kb

Ce fragment est porté par le pXL151 décrit dans l'exemple 1. Le site EcoRI ainsi que les 70 pb adjacentes situées à la droite de ce fragment proviennent du pXL59 qui est le vecteur qui a servi à

15 construire le pXL151 par clonage d'un fragment Sau3AI de Pseudomonas denitrificans SC510. Après excision, des sous fragments du fragment de 8,7 kb ont été clonés dans les phages M13mp18, ou M13mp19 (Norrand et al., 1983), ou M13tg130, ou M13tg131 dans les deux orientations (Kieny et al., 1983). Des délétions ont été ensuite réalisées in vitro

20 par la méthode de Henikoff (1987). Ces délétions ont ensuite été séquencées avec le "primer universel" comme amorce de synthèse des réactions de terminaison de chaîne. Le recouvrement entre ces différentes délétions a permis d'établir la séquence totale, sur les deux brins, du fragment de 8,7 kb (figure 8). Ce fragment comprend

25 8753 pb.

2.3. Séquençage d'un fragment SalI-SalI-SalI-SalI-SalI-BglI de 4,8 kb.

Ce fragment est contenu dans le plasmide pXL154 décrit dans l'exemple 1. Le protocole est identique à celui utilisé dans

30 l'exemple 2.2. La séquence totale sur les deux brins du fragment de 4,8 kb est présentée sur la figure 32. Ce fragment contient 4749 pb.

2.4. Séquence nucléotidique d'un fragment

SstI-SstI-BamHI de 3,9 kb.

Ce fragment est inclus dans le plasmide pXL519 décrit dans l'exemple 1. Le protocole est identique à celui utilisé dans l'exemple 2.2. La séquence totale sur les deux brins du fragment de 3,9 kb est présentée sur la figure 33. Ce fragment contient 3855 pb.

2.5. Séquence nucléotidique d'un fragment

EcoRI-BglII-EcoRI-BglII de 13.4 kb.

Ce fragment est contenu dans les plasmides pXL156 et pXL157 décrits dans l'exemple 1. Le protocole utilisé est identique à celui de l'exemple 2.2. La séquence sur les deux brins du fragment de 13.15 kb est présentée sur la figure 43. Elle correspond à la séquence totale du fragment de 13.4 kb mis à part 250 pb, correspondant à un fragment EcoRI-SstI, se trouvant à l'extrémité gauche de celui-ci.

A partir de ces séquences nucléotidiques, des cartes de restriction ont été obtenues pour les enzymes qui coupent le moins fréquemment (figures 6 et 44). Le pourcentage en bases GC de l'ADN de Pseudomonas denitrificans SC150 est relativement élevé (65,5 %) et se traduit par des compressions sur les gels de séquence. Pour éviter ces problèmes deux solutions sont apportées :

- i) l'utilisation du 7-déaza-dGTP au lieu du dGTP dans les réactions de séquence pour diminuer les structures secondaires qui se forment pendant l'électrophorèse dans le gel de séquence et
- ii) le séquençage des deux brins.

25 EXEMPLE 3 - Analyse de ces séquences nucléotidiques : détermination des phases ouvertes

Les séquences nucléotidiques des fragments ClaI-HindIII-HindIII-HindIII de 5,4 kb (figure 7), EcoRI-EcoRI de 8,7 kb (figure 8), SalI-SalI-SalI-SalI-SalI-BglI de 4,8 kb (figure 32), SstI-SstI-BamHI de 3,9 kb (figure 33) et EcoRI-BglII-EcoRI-BglII de

13,4 kb (figure 43) permettent de définir des phases ouvertes. Comme il s'agit d'ADN à haut pourcentage en GC, les phases ouvertes sont nombreuses vu la faible fréquence de codons d'arrêt de traduction. Une étude de la probabilité des phases codantes d'après usage des codons en utilisant la méthode de Staden et MacLachlan (1982) est réalisée.

5 Elle caractérise les phases ouvertes qui ont la meilleure probabilité d'être codantes par rapport aux autres phases du même brin d'ADN, cette probabilité est fonction de l'usage des codons de gènes déjà séquencés provenant des bactéries du genre Pseudomonas. De cette manière :

10 3.1. Cinq phases ouvertes sont caractérisées pour le fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb. Elles sont nommées phases 1 à 5 et leurs positions sur la séquence du fragment de 5,4 kb sont les suivantes (sur la séquence 5'→3' du site ClaI vers les sites HindIII):

15 Tableau : Phases ouvertes probables du fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb. Les positions sur la séquence correspondent aux positions sur la séquence décrite à la figure 7 ; le brin codant est le brin 5'→3' correspondant au brin supérieur sur cette figure.

20

Numéro de la phase	Codon de démarrage de la traduction	Codon stop	Poids moléculaire en kD de la protéine codée
1	549	1011	15.5
2	1141	1981	29.2
25 3	1980	3282	45.7
4	3281	4280	35.0
5	4284	5253	34.1

Les représentations des probabilités que ces phases ouvertes soient des phases codantes, avec parallèlement celles

observées sur les autres phases (5 au total) sont portées sur la figure 9. Ces cinq phases sont codées par le même brin. Quatre d'entre elles (phases ouvertes 1 à 4) présentent les caractéristiques de phases codantes en couplage traductionnel (Normak et al., 1983), à savoir que le codon d'initiation de traduction de la phase x+1 chevauche le codon de terminaison traduction de la phase x ou bien que ceux-ci sont très proches.

3.2. Huit phases sont caractérisées pour le fragment EcoRI-EcoRI de 8,7 kb. Elles sont nommées phases 6 à 13 et leurs positions sur la séquence du fragment de 8,7 kb sont portées sur le tableau ci-dessous.

Tableau : Phases ouvertes probables du fragment EcoRI 8,7 kb. Les positions sur la séquence correspondent aux positions sur la séquence décrite sur la figure 8 ; sur cette figure, le brin codant est le brin supérieur sauf pour la phase 11.

Numéro de la phase de traduction	Codon de démarrage	Codon stop	Poids moléculaire en kD de la protéine codée
6	736	1519	28.9
7	1620	2997	46.7
8	3002	3632	22.0
9	3631	4366	25.8
10	4365	5127	27.1
11	5893	5110	28.0
12	5862	7101	42.9
13	7172	7931	26.8

Les représentations des probabilités de ces phases ouvertes, avec parallèlement celles observées sur les autres phases (6 phases au total) sont portées sur la figure 10. A l'exception de la

phase 11, ces phases sont codées par le même brin. Quatre d'entre elles (de 7 à 10) présentent les caractéristiques de phases codantes en couplage traductionnel (Normark et al., 1983), à savoir que le codon d'initiation de traduction de la phase $x+1$ chevauche le codon de terminaison de traduction de la phase x ou bien que ceux-ci sont très proches.

3.3. Quatre phases ouvertes sont caractérisées pour le fragment SalI-SalI-SalI-SalI-SalI-BglI de 4,8 kb. Elles sont nommées phases 14 à 17 et leurs positions sur la séquence du fragment de 4,8 kb sont les suivantes (sur la séquence 5'- 3' des sites SalI vers le site BglI):

Tableau: Phases ouvertes probables du fragment SalI-SalI-SalI-SalI-SalI-BglI de 4,8 kb. Les positions sur la séquence correspondent aux positions décrites sur la figure 32 où le brin supérieur est donné dans son orientation 5'- 3'. Les phases 15, 16 et 17 sont codées par le brin supérieur au contraire de la phase 14

Numéro de la phase	Codon de démarrage de traduction	Codon stop	Poids moléculaire en D de la protéine codée
14	660	379	10286
15	925	1440	18941
16	1512	2510	36983
17	2616	4511	70335

Les représentations des probabilités que ces phases ouvertes soient codantes, avec parallèlement celles observées sur les autres phases (au total 4) sont portées sur la figure 34. Les phases 15, 16 et 17 sont codées par le même brin, la phase 14 par le brin complémentaire.

3.4. Quatre phases sont caractérisées pour le fragment SstI-SstI-BamHI de 3,9 kb. Elles sont nommées 18 à 21 et leurs positions sur la séquence du fragment de 3,9 kb sont portées sur le tableau ci-dessous.

5 Tableau: Phases ouvertes probables du fragment SstI-SstI-BamHI de 3,9 kb. Les positions sur la séquence correspondent aux positions décrites sur la figure 33 où la polarité du brin supérieur est 5'-3'. Les phases 18 et 19 sont codées par le brin inférieur au contraire des phases 20 et 21.

10	Numéro de la phase	Codon de démarrage de traduction	Codon stop	Poids moléculaire en 'D de la protéine codée
	18	809	108	25148
15	19	1971	1063	30662
	20	2099	3115	34682
	21	3344	3757	14802

Les représentations des probabilités que ces phases ouvertes soient codantes, avec parallèlement celles observées sur les autres phases (au total 4) sont portées sur la figure 35. Les phases 19 et 20 sont transcrites de façon divergente.

3.5. Neuf phases ouvertes sont caractérisées pour le fragment EcoRI-BglII-EcoRI-BglII de 13.1 kb. Elles sont nommées phases 22 à 30 et leurs positions sur la séquence du fragment de 13,1 kb sont les suivantes (sur la séquence 5'-3' du site EcoRI vers le site BglII):

Tableau: Phases ouvertes probables du fragment EcoRI-BglII-EcoRI-BglII de 13.1 kb. Les positions sur la séquence correspondent aux positions

décrites sur la figure 43 où le brin supérieur est donné dans son orientation 5'-3'. Les phases 22, 23, 24, 25, 26, 27 et 29 sont codées par le brin supérieur au contraire des phases 28 et 30.

5		Numéro de la phase de traduction	Codon de démarrage	Codon stop	Poids moléculaire en D de la protéine codée
		22	429	1884	51 982
		23	3364	3886	19442
10		24	3892	4954	38121
		25	5060	8885	138 055
		26	9034	9676	24 027
		27	9678	10101	14 990
		28	10835	10306	21 057
15		29	11656	12181	19 183
		30	13059	12368	24 321

Les représentations des probabilités que les phases ouvertes 22, 23, 24, 25, et 26 soient codantes, avec parallèlement celles observées sur les autres phases (au total 5 phases) sont portées sur la figure 45.

20 Ces 5 phases sont codées par le même brin.

EXEMPLE 4 - Etudes génétiques sur les fragments d'ADN portant des gènes cob

Cet exemple montre la relation qui existe entre les différentes phases ouvertes identifiées précédemment et les gènes impliqués dans la biosynthèse des cobalamines et/ou des cobamides portés par ces mêmes fragments. Ces gènes sont identifiés par une étude génétique comme décrit ci-dessous.

25

4.1 - Etude génétique du fragment de 5,4 kb

Le plasmide pXL723 est le plasmide pRK290 (Ditta et al., 1980) contenant le fragment EcoRI-HindIII de 2264 pb, correspondant à la partie droite du fragment étudié, cloné au site EcoRI du pRK290 (figure 11). La construction des autres plasmides utilisés dans cette
5 étude (pXL302, pXL1397, pXL545, pXL545 Ω , pXL556 et pXL1500) est décrite dans la légende des figures 11 et 12.

Des insertions ont été obtenues sur le plasmide pXL723 en utilisant la technique de Bruijn et Lupski, 1984. Les insertions du transposon Tn5 sur le plasmide pXL723 ont été sélectionnées, puis
10 cartographiées dans le fragment de 5,4 kb (figure 12). Le pXL723 complémente le mutant Cob G572 de Pseudomonas putida et le mutant Cob G634 d'Agrobacterium tumefaciens. Ces insertions se classent en deux groupes d'insertions inactivationnelles: soit celles qui ne permettent plus de compléter le mutant Cob G572, soit celles qui abolissent la
15 complémentation du mutant Cob G634 (figure 12). Les insertions qui inactivent la complémentation du mutant G572 sont cartographiées dans la phase ouverte 4 (il s'agit des insertions 15, 27, 68, 81 et 97) ; la phase ouverte 4 correspond donc à un gène cob. Celui-ci est nommé cobC. Les insertions qui inactivent la complémentation du mutant G634
20 sont cartographiées dans la phase 5 (ce sont les insertions 66 et 107, figure 12); la phase ouverte 5 correspond donc à un gène cob. Celui-ci est nommé cobD. Par ailleurs des insertions avec un transposon Tn5Sp^r ont été réalisées. Le transposon Tn5Sp^r a été construit au laboratoire en clonant au site BamHI du transposon Tn5 (Jorgensen et al., 1979)
25 une cassette BamHI contenant le gène de résistance à la spectinomycine provenant du plasmide pHP45 Ω (Prentki et Krisch, 1984). Ces insertions ont été faites dans le chromosome de la souche de Pseudomonas denitrificans SBL27 Rif^r. La souche SBL27 est une souche de Pseudomonas denitrificans dont SC510 dérive par plusieurs mutagénèses.
30 SBL27 produit 10 fois moins de cobalamines que SC510 sur milieu PS4. Sur 10000 clones de la souche SBL27 Rif^r, portant chacun une insertion du transposon, plus de 30 d'entre eux avaient perdu la capacité de synthétiser des cobalamines. Certains de ces clones possédaient une insertion dans le fragment étudié dans cet exemple. Ces insertions ont

été cartographiées par analyses de restriction selon la méthode de Southern (Southern, 1975). Les sites d'insertions du transposon chez ces différents mutants sont portés sur la figure 12. Une de ces insertions numérotée 2639 se trouve dans le gène cobC; cette insertion
5 est complémentée par le plasmide pXL302 qui porte un fragment contenant le gène cobC (figure 12). Deux insertions, nommées 2636 et 2638, sont dans la phase ouverte 3. Ces mutants sont bloqués dans la biosynthèse des cobalamines, et ils sont complémentés par le plasmide pXL1397 qui ne contient que la phase ouverte 3, mais non complémentés
10 par le plasmide pXL302 qui contient les gènes cobC et cobD (figure 12). Ces deux insertions sont donc dans un autre gène. A la phase ouverte 3 nous associons le gène cobB. Une insertion 2933 est placée dans la phase ouverte 2 ; elle est complémentée par le plasmide pXL1500 qui contient la phase ouverte 2 ; cette insertion est non
15 complémentée par le plasmide pXL1397 qui contient le gène cobB et qui complémente les deux insertions dans cobB. Il s'agit donc ici d'une insertion dans un autre gène; à la phase ouverte 2 nous associons un gène nommé cobA.

Une cassette de résistance à la kanamycine provenant du
20 plasmide pUC4K (Barany et al., 1985) a été introduite au site NotI du fragment ClaI (position 0 sur la séquence)-RsaI (position 1686 sur la séquence) cloné dans un plasmide pUC8 (Viera et Messing, 1982) ; il s'agit du site NotI situé à la position 771 dans la phase 1 (voir la séquence sur la figure 7); deux insertions ont été retenues
25 correspondant chacune à une orientation différente de la cassette de résistance. Ces fragments portant chacun une insertion de la cassette de résistance ont été clonés dans le plasmide pRK404 (Ditta et al.) pour donner les plasmides pXL1630 et 1631. Ces plasmides ont été introduits par transfert conjugatif dans la souche de Pseudomonas
30 denitrificans SC510 Rif^r, puis par une série de cultures-dilutions en absence de l'antibiotique sélectif pour le plasmide (la tétracycline), des doubles recombinants ayant échangé le fragment plasmidique avec le fragment chromosomique et ayant perdu le plasmide ont été trouvés. Deux souches ont ainsi été caractérisées :

35 1) l'une est nommée SC510:1631 Rif^r, dans cette souche, la

cassette de résistance à la kanamycine se trouve insérée, dans le chromosome au site NotI (se trouvant dans la phase 1); les polarités des transcriptions du gène de résistance à la kanamycine et celle de la phase ouverte 1 sont opposées,

- 5 ii) l'autre insertion est nommée SC510:1630 Rif^r; la cassette de résistance est insérée au même site, mais la transcription du gène de résistance à la même polarité que celle de phase ouverte 1 entière.

10 Ces deux souches ont toutes les deux un taux de synthèse de cobalamines au moins 100 fois inférieur à celui de SC510.

15 Le plasmide pXL545Ω correspond au plasmide pXL545 dans lequel la cassette de résistance à la spectinomycine, du plasmide PHP45Ω a été insérée au niveau du site BamHI. Ce plasmide (figure 12) qui contient le fragment ClaI-HindIII de 814 pb (où seule la phase ouverte 1 est
20 entière) complémente seulement le mutant SC510:1630 Rif^r. Ceci suffit à définir un nouveau gène puisque ce mutant est complémenté par un plasmide qui contient uniquement la phase ouverte 1 entière. La phase ouverte 1 correspond à un gène de la voie de biosynthèse des cobalamines et/ou des cobamides. Ce gène est nommé cobE. L'absence de
25 complémentation du mutant SC510:1631 Rif^r par le plasmide pXL545Ω est peut être due au fait que les gènes cobA, cobB, cobC, cobD et cobE, ou une partie d'entre eux, appartiennent au même opéron et que l'insertion dans cobE qui conserve une transcription dans le sens de
transcription de l'opéron peut être complémentée seulement par
l'expression en trans du gène cobE. Au contraire le mutant SC510:1631 Rif^r lui ne peut être complémenté que par un plasmide qui permet l'expression en trans des gènes cobA à cobE.

 Le fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb contient donc cinq gènes cob nommés cobA, cobB, cobC, cobD et cobE.

30 4.2 - Etudes génétiques du fragment de 8,7 kb

 Le plasmide pXL367 est le pRK290 (Ditta et al., 1980) contenant le fragment EcoRI de 8,7 kb cloné au site EcoRI (figure 13).

 Des insertions du transposon Tn5 sur le plasmide pXL367

2072023

ont été sélectionnées en utilisant la technique déjà décrite (de Bruijn et Lupski, 1984). Les insertions dans le fragment de 8,7 kb ont été cartographiées. De la même manière des insertions du transposon Tn3lacZ ont été obtenues selon la méthode déjà décrite (Stachel et al., 1985) et cartographiées. 29 insertions du transposon Tn5 et 13 insertions du transposon Tn3lacZ ont ainsi été cartographiées. La position précise de ces insertions sur le fragment de 8,7 kb est portée à la figure 14. Des plasmides, portant chacun une seule insertion dans le fragment de 8,7 kb, ont été introduits par transferts conjugatifs dans les mutants Cob d'Agrobacterium tumefaciens G164, G609, G610, G611, G612, G613, G614, G615, G616, G620, G638. Ces mutants sont tous complémentés par le pXL367. Les insertions qui ne permettent plus la complémentation des différents mutants ont été recherchées. Elles correspondent à une insertion dans le gène responsable de la complémentation du mutant correspondant. Les résultats des complémentations des différents mutants pour leur caractère de production de cobalamines (phénotype Cob) sont portés sur la figure 14. Si le mutant recombiné produit moins de trois fois moins de cobalamines que ne produit le même mutant avec le plasmide pXL367, il est considéré comme étant non complémenté. Sur les mutants étudiés, G164, G609, G610, G611, G612, G613, G614, G615, G616, G620, G638, on observe huit classes différentes d'insertions inactivationnelles de transposons conduisant à un phénotype muté. Ces classes caractérisent des insertions par l'absence de complémentation d'un ou plusieurs mutants par les plasmides pXL367 portant ces mêmes insertions. Chaque classe correspond donc à un gène muté. On observe que les insertions appartenant à une même classe sont positionnées les unes à côté des autres. Huit classes d'insertions sont ainsi observées qui permettent de définir huit gènes. Chaque classe d'insertions définit un fragment minimum qui doit être contenu dans le gène correspondant. La figure 14 démontre un parfait recoupement entre les régions délimitées par chaque classe, au niveau de la carte de restriction, et les phases ouvertes décrites ci-dessus (exemple 3). On constate en effet que, pour chaque classe d'insertions, les transposons sont toujours insérés dans une partie du fragment de 8,7 kb qui est contenue dans une seule

phase ouverte. A chaque classe d'insertions, est donc associée une phase ouverte et une seule. Les phases ouvertes indiquées ci-dessus codent donc chacune pour une protéine impliquée dans la voie de biosynthèse des cobalamines et/ou des cobamides. Les phases ouvertes correspondent chacune à des gènes impliqués dans la biosynthèse des cobalamines et/ou des cobamides. Ces phases ouvertes sont appelées cobF, cobG, cobH, cobI, cobJ, cobK, cobL et cobM pour les phases 6 à 13 respectivement. La position de ces gènes par rapport à la carte de restriction est représentée sur la figure 14.

10 4.3 - Etude génétique du fragment de 4,8 kb.

Le plasmide pXL1558 est le plasmide pRK290 (Ditta et al., 1980) contenant le fragment HindIII-HindIII de 12 kb du pXL154 (Cameron et al., 1989) cloné au site EcoRI du pRK290 (figure 36). La construction des autres plasmides utilisés dans cette étude (pXL233 et 15 pXL843) est décrite dans la légende de la figure 36.

Des insertions Tn5Sp ont été obtenues sur le plasmide pXL1558. Tout d'abord une souche contenant un transposon Tn5Sp a été construite; ceci a été fait en transformant la souche C2110 (Stachel et al., 1985) à l'aide du plasmide pRK2013Tn5Sp (Blanche et al., 1989); le plasmide 20 pRK2013Tn5Sp ayant une origine de réplication ColE1 ne se réplique pas dans la souche C2110 qui est polA-. Les colonies obtenues après transformation qui sont résistantes à la spectinomycine ont donc le transposon Tn5Sp dans leur chromosome; une colonie est ensuite réisolée, puis l'insertion du transposon est ensuite transduite à 25 l'aide du phage P1 chez la souche MC1060 comme cela est décrit précédemment (Miller, 1972). La souche MC1060 Tn5Sp est transformée par le plasmide pXL1558; le plasmide pXL1558 est ensuite mobilisé par conjugaison à l'aide du pRK2013 chez C600 Rif^r. Des conjuguants résistants à la tétracycline (pour le plasmide pXL1558) et à la 30 spectinomycine (pour le transposon) sont ensuite sélectionnés; de tels conjuguants doivent contenir le plasmide pXL1558 dans lequel le transposon Tn5Sp s'est inséré. Des insertions portées sur le plasmide pXL1558, et plus précisément dans le fragment de 12 kb sont ensuite

cartographiées par digestion de restriction; 23 insertions sont ainsi obtenues et cartographiées sur le fragment de 12 kb; la position de ces différentes insertions sur le fragment est présentée sur la figure 37. Ces 23 insertions ont été introduites sur le chromosome de la souche SC510 Rif^r après transfert conjugatif du p-XL1558::Tn5Sp, puis introduction du plasmide pR751. Le plasmide pR751 est un plasmide, résistant à la triméthoprine, du même groupe d'incompatibilité que le pXL1558 (incP, Thomas et Smith, 1987). Par culture non sélective pour le pXL1558 (absence de tétracycline) mais sélective pour le pR751 et le transposon (présence de triméthoprine et de spectinomycine) l'échange de la mutation portée par le pXL1558::Tn5Sp avec le chromosome ainsi que la ségrégation du pXL1558 sont obtenus par la technique d'échange de marqueur par double recombinaison homologue comme cela est déjà décrit (Schell et al., 1988). Les souches ainsi sélectionnées portent le transposon dans leur chromosome. La double recombinaison homologue est vérifiée par la méthode de Southern (Southern, 1975). De cette manière 23 souches SC510 Rif^r::Tn5Sp dans le fragment de 12 kb ont été identifiées.

D'autre part une autre insertion Tn5Sp obtenue par mutagenèse au hasard du transposon Tn5Sp dans la souche SBL27 Rif^r (Blanche et al., 1989) a été cartographiée sur le fragment de 12 kb par analyse de restriction selon la méthode de Southern (Southern, 1975), voir figure 37; cette souche est nommée SBL27 Rif^r::Tn5Sp 1480.

Le taux de synthèse de cobalamines est déterminé pour ces 24 souches cultivées en milieu PS4 selon le protocole déjà décrit (Cameron et al., 1989) et le phénotype Cob⁻ est attribué aux souches produisant au moins 1000 fois moins (resp. 100) de vitamine B12 que la souche parente SC510 Rif^r (resp. SBL27 Rif^r), figure 37. Il est ainsi observé que 6 de ces insertions chromosomiques conduisent à un phénotype Cob⁻ chez *P. denitrificans*; il s'agit des insertions 31.1, 41.3, 45, 55, 22.1 et 1480.

Trois plasmides pXL233, pXL837 (Cameron et al.) et pXL843 sont introduits par transferts conjugatifs dans trois souches présentant le phénotype Cob⁻ soient SC510 Rif^r::Tn5Sp 31.1, SC510 Rif^r::Tn5Sp 45, SBL27 Rif^r::Tn5Sp 1480. Ces trois mutants ont chacun un profil de

complémentation différent pour la synthèse des cobalamines. En effet SBL27 Rif^r::Tn5Sp 1480 est complémenté par le pXL837 et le pXL843 mais pas par le pXL233; le mutant SC510 Rif^r::Tn5Sp 45 n'est complémenté que par le pXL843; le mutant SC510 Rif^r::Tn5Sp 31.1 est complémenté par le plasmide pXL843 ainsi que par le plasmide pXL233 (voir figure 37). Les données présentées permettent donc, d'après les résultats sur les complémentations des trois mutants, de conclure que les trois mutants sont différents et que pour chacun d'entre eux, le transposon Tn5Sp s'est inséré dans un gène cob différent.

D'autre part les plasmides pXL1558::Tn5Sp 41.3, pXL1558::Tn5Sp 45, et pXL1558::Tn5Sp 22.1 sont introduits par transferts conjugatifs dans la souche G2035 (Cameron et al., 1989) et ne la complémentent pas. Le plasmide pXL1558 complémente ce mutant au contraire du plasmide pXL1558::Tn5Sp 31.1.

Les données de phénotype et complémentation nous permettent de définir 3 classes d'insertions; chacune de ces classes est représentée par les insertions suivantes : 31.1, classe 1; 45, 41.3, 55 et 22.1, classe 2; 1480 classe 3.

Pour chaque classe d'insertions, les transposons sont toujours insérés dans une partie du fragment de 4,8 kb qui est contenue dans une seule phase ouverte (ORF14, ORF16 et ORF17, comme définies à l'exemple 3). A chaque classe d'insertions est associée une seule phase ouverte. Les phases ouvertes indiquées ci-dessus codent donc pour une protéine impliquée dans la voie de biosynthèse des cobalamines et/ou des cobinamides. Ces phases ouvertes sont appelées cobX, cobS et cobT pour les phases 14, 16 et 17. La position de ces gènes par rapport à la carte de restriction est représentée sur la figure 37. La phase ouverte 15 n'est pas un gène impliqué dans la biosynthèse du coenzyme B12.

30

4.4 - Etudes génétiques du fragment de 3,9 kb

Le plasmide pXL1557 est le plasmide pRK290 (Ditta et al., 1980) contenant le fragment HindIII-BamHI de 9 kb du pXL519 cloné au site EcoRI du pRK290 (figure 38). La construction des autres

plasmides utilisés dans cette étude (pXL1286, pXL1303, pXL1324) est décrite dans la légende de la figure 38. Par ailleurs, le fragment BglII-XhoI de 2 kb (positions sur la séquence présentée figure 33 251 et 2234) du plasmide pXL519 est cloné aux sites BamHI-SalI du plasmide pXL435 (Cameron et al) pour générer le plasmide pXL699.

Des insertions Tn5Sp ont été obtenues sur le plasmide pXL1557 selon la technique décrite dans l'exemple 4.3. Les insertions du transposon Tn5Sp sur le plasmide pXL1557, alors nommées pXL1557::Tn5Sp, ont été sélectionnées. Celles qui sont cartographiées dans le fragment de 9 kb (figure 39) ont été introduites sur le chromosome de la souche SC510 Rif^r après transfert conjugatif du pXL1557::Tn5Sp et échange de marqueur par double recombinaison homologue comme cela est décrit en 4.3.

La double recombinaison homologue est vérifiée par la méthode de Southern (Southern, 1975). De cette manière 20 souches SC510 Rif^r::Tn5Sp ont été identifiées.

D'autre part deux autres insertions Tn5Sp obtenues par mutagenèse au hasard du transposon Tn5Sp dans la souche SBL27 Rif^r (Blanche et al., 1989) ont été cartographiées sur le fragment de 9 kb par analyse de restriction selon la méthode de Southern (Southern, 1975), voir sur la figure 39 les insertions 1003 et 1147.

Le taux de synthèse de cobalamines est déterminé pour ces 22 souches cultivées en milieu PS4 selon le protocole déjà décrit (Cameron et al., 1989) et le phénotype Cob⁻ est attribué aux souches produisant 1000 fois moins (resp. 100) de vitamine B12 que la souche parente SC510 Rif^r (resp. SBL27 Rif^r), figure 39. Seules les 4 insertions 1, 1003, 23 et 1147 se traduisent par un phénotype Cob⁻ chez P. denitrificans.

Quatre plasmides pXL699, pXL1286, pXL1303 et pXL1324 sont introduits par transferts conjugatifs dans les quatre souches présentant le phénotype Cob⁻ soient SC510 Rif^r::Tn5Sp 1, SBL27 Rif^r::Tn5Sp 1003, SC510 Rif^r::Tn5Sp 23 et SBL27 Rif^r::Tn5Sp 1147. Le plasmide pXL699 complémente les deux premiers mutants (SC510 Rif^r::Tn5Sp 1, SBL27 Rif^r::Tn5Sp 1003) mais le plasmide pXL1303 ne les complémente pas; le plasmide pXL1324 complémente les deux autres mutants (SC510

Rifr::Tn5Sp 23 et SBL27 Rifr::Tn5Sp 1147) mais le plasmide pXL1286 ne les complémente pas.

D'autre part le plasmide pXL1557::Tn5Sp 1, est introduit par transfert conjuguatif dans la souche G2040 et ne la complémente pas alors que
 5 les plasmides pXL1557, pXL1557::Tn5Sp 6A, pXL1557::Tn5Sp 54, pXL1557::Tn5Sp 48, pXL1557::Tn5Sp 21, pXL1557::Tn5Sp 8, pXL1557::Tn5Sp 23 aussi introduits par transferts conjugatifs, la complémentent (voir figure 39).

Les données de phénotype et complémentation permettent de
 10 définir 2 classes d'insertions. Pour chaque classe d'insertions, les transposons sont toujours insérés dans une partie du fragment de 3,9 kb qui est contenue dans une seule phase ouverte (ORF 19 et ORF20 comme définie à l'exemple 3).

A chaque classe d'insertions est associée une seule phase ouverte. Les
 15 phases ouvertes indiquées ci-dessus codent pour une protéine impliquée dans la voie de biosynthèse des cobalamines et/ou des cobinamides. Ces phases ouvertes sont appelées cobV et cobU pour les phases 19 et 20. Les phases 18 et 21 ne sont pas des gènes impliqués dans la voie de biosynthèse du coenzyme B12. La position de ces gènes par rapport à la
 20 carte de restriction est représentée sur la figure 39. Les insertions 48, 21 et 8 sont cartographiées entre les gènes cobU et cobV

4.5 - Etudes génétiques du fragment de 13,4 kb.

4.5.1. Etudes sur le fragment EcoRI-BglII de 4327 pb.

Le plasmide pXL189 (Cameron et al., 1989), qui contient au moins un
 25 gène cob porte un insert de 3.1 kb qui excepté 300 pb correspond à un fragment EcoRI-ClaI de 4,26 kb (voir figure 45). Le pXL189 a été soumis à une mutagénèse au transposon Tn5, comme cela est décrit précédemment (De Bruijn et Lupski (1984). 13 insertions ont ainsi été cartographiées dans l'insert du pXL189 comme cela est présenté sur la figure 46. Ces
 30 13 plasmides mutants ainsi que pXL189 ont été conjugués chez deux mutants d'A. tumefaciens, G632 et G633, qui sont des mutants complémentés par pXL189 (Cameron et al., 1989). Seule l'insertion 58 s'est avérée être une insertion inactivationnelle. Ce résultat montre

que les deux mutants G632 et G633 correspondent à une mutation dans un même gène et que d'autre part le seul gène de *P. denitrificans* qui puisse être responsable de leur complémentation correspond à la phase ouverte 26 (voir figure 46) puisque l'insertion 58 est cartographiée dans cette phase ouverte; de plus il s'agit de la seule insertion parmi les 13, qui est cartographiée dans cette phase ouverte. Un gène cob, nommé cob0 est donc associé à la phase ouverte 26.

Pour savoir si les quatre autres phases ouvertes (phases ouverte 27 à 30) identifiées sur ce fragment correspondent à des gènes cob, une cassette de résistance à la spectinomycine, du plasmide pHP45 Ω (Prentki et Krisch, 1984) a été spécifiquement insérée dans chacun de ces gènes puis introduite dans le chromosome de *P. denitrificans* SC510 Rif^r par recombinaison homologue de manière à obtenir des mutants d'insertions dans chacune de ces phases ouvertes. Pour ce faire, le fragment EcoRI-ClaI (positions respectives 8818 et 13082 sur la séquence présentée sur la figure 43) a été utilisé. Ce fragment, qui porte les phases ouvertes 27 à 30, a été purifié à partir du pXL157 (Cameron et al., 1989); un linker EcoRI a été ajouté sur l'extrémité ClaI après que celle-ci ait été remplie par le fragment Klenow de l'ADN polymérase d'*E. coli*. Ce fragment a ensuite été cloné dans le plasmide pUC13 (Viera et al., 1982) au site EcoRI. Le plasmide ainsi construit a été appelé pXL332. Des insertions de la cassette de résistance à la spectinomycine du plasmide pHP45 Ω (Prentki et Krisch, 1984) ont été réalisées sur le pXL332. Ces insertions ont été faites séparément aux sites SmaI (position 9868, phase ouverte 27), BamHI (position 10664, phase ouverte 28), ClaI (position 11687, phase ouverte 29) et NcoI (position 12474, phase ouverte 30) par digestions totales ou partielles du pXL332 avec les enzymes correspondantes puis, si cela est nécessaire, remplissage de ces extrémités avec le fragment de Klenow de l'ADN polymérase d'*E. coli*, puis ligature avec le fragment de SmaI 2 kb du pHP45 Ω (Prentki et Krisch, 1984) contenant un gène de résistance à la spectinomycine; ces insertions sont nommées respectivement n2, n1, n3 et n4 comme cela est présenté sur la figure 46. Les fragments EcoRI portant ces différentes insertions ont ensuite été clonés sur le pRK404 (Ditta et al., 1985) à l'un des deux sites EcoRI. Les 4 plasmides

portant ces différentes insertions ont ensuite été introduits par conjugaison chez SC510 Rif^r, comme cela est décrit précédemment. Le plasmide pR751 (Thomas et Smith, 1987) a ensuite été introduit dans les transconjuguants. L'échange des mutations portées par les 4 différents
5 dérivés de pRK404 et le chromosome de SC510 Rif^r a pu être sélectionné comme cela est décrit (voir exemple 4.3). 4 souches ont ainsi été obtenues. Ces souches portent chacune une insertions de la cassette de résistance dans une des quatre phases ouvertes 27 à 30. Ces insertions ont été vérifiées par analyse de l'ADN génomique par Southern blot
10 (Southern, 1975). La production de cobalamines de ces différentes souches a été étudiée. Elles ont toutes montré un phénotype Cob⁺ par culture en milieu PS4. Ce résultat indique que ces phases ouvertes n'interviennent pas dans la biosynthèse du coenzyme B12. Toutefois il est possible qu'une ou plusieurs de ces phases codent pour des
15 protéines qui interviennent par exemple dans la transformation du coenzyme B12 en méthylcobalamine par exemple, c'est à dire la synthèse d'une autre cobalamine voir d'un autre corrinoïde.

4.5.2 Etude du fragment EcoRI-EcoRI de 9.1 kb.

Différents plasmides sont utilisés dans cette étude;
20 le plasmide pXL1560 est le plasmide pRK290 (Ditta et al., 1980) contenant le fragment EcoRI-EcoRI de 9,1 kb kb du pXL156 (exemple 1) cloné au site EcoRI du pRK290 (voir figure 46). La construction des autres plasmides utilisés dans cette étude (pXL618, pXL593, pXL623, pXL1909, pXL1938, pXL1908, pXL221, pXL208, pXL297) est décrite dans la
25 légende de la figure 45.

Des insertions Tn5Sp ont été obtenues sur le plasmide pXL1560. La souche MC1060 Tn5Sp transformée par le plasmide pXL1560 a servi à obtenir des insertions du transposon Tn5Sp dans le fragment pXL1560; 27 insertions ont ainsi été obtenues et cartographiées sur le fragment de
30 9,1 kb; la position de ces différentes insertions sur le fragment est présentée sur la figure 4. Ces 27 insertions ont été introduites sur le chromosome de la souche SC510 Rif^r après transfert conjugatif des

pXL1560::Tn5Sp, puis introduction du plasmide pR751. Le plasmide pR751 est un plasmide, résistant à la triméthoprine, du même groupe d'incompatibilité que le pXL1560 (incP, Thomas et Smith, 1987). Par culture non sélective pour le pXL1560 (absence de tétracycline) mais
5 sélective pour le pR751 et le transposon (présence de triméthoprine et de spectinomycine) l'échange de la mutation portée par le pXL1560::Tn5Sp avec le chromosome ainsi que la ségrégation du pXL1560 sont obtenus; cette technique d'échange de marqueur par double recombinaison homologue est équivalente à celle déjà décrite par Schell
10 et al., 1988. Les souches ainsi sélectionnées portent le transposon dans leur chromosome.

La double recombinaison homologue est vérifiée par la méthode de Southern (Southern, 1975). De cette manière 27 souches SC510 Rif^r::Tn5Sp possédant chacune une insertion différente du transposon
15 Tn5Sp dans le fragment de 9.1 kb ont été identifiées.

Le taux de synthèse de cobalamines est déterminé pour ces 27 souches cultivées en milieu PS4, et le phénotype Cob⁻ est attribué aux souches produisant au moins 1000 fois moins de vitamine B12 que la souche parente SC510 Rif^r, figure 46. Il est ainsi observé que 18 de
20 ces insertions chromosomiques, sur les 27, conduisent à un phénotype Cob⁻ chez *P. denitrificans* comme cela est montré sur la figure 46.

Les insertions 19, 32, 24, 27, 37, 39, 26, 11 et 14 sont cartographiées dans la phase ouverte 22 (voir figure 46). Toutes ces insertions sont complémentées par le plasmide pXL618 qui ne contient que la phase
25 ouverte 22. Nous en déduisons que la phase ouverte 22 correspond à un gène cob que nous appelons cobQ. Aucune insertion n'a été obtenue dans la phase ouverte 23; cependant le plasmide pXL623 qui ne contient que cette phase ouverte (voir figure 46), complémente deux mutants cob d'*Agrobacterium tumefaciens*, G642 et G2043 (Cameron et al., 1989). La
30 phase ouverte 23 correspond donc à un gène cob nommé cobP. Les insertions 23, 13, 12, 30, 22, 40, 35, 10 et 17 qui sont cartographiées dans les phases ouvertes 24 et 25 entraînent un phénotype Cob⁻ chez SC510 Rif^r. Il semble donc qu'il s'agisse de deux phases ouvertes dont le produit est impliqué dans la biosynthèse des cobalamines. Toutefois
35 on ne peut exclure que ces insertions aient des effets polaires sur les

gènes positionnés en 3' tels que cobO. Il convient donc d'étudier la complémentation de ces mutants afin de voir si leur phénotype Cob- ne résulte pas d'un effet polaire.

- Les mutants Cob d'Agrobacterium tumefaciens G622, G623 et G630
 5 complémentés par le pXL156 ont été étudiés. Ces mutants ne sont pas complémentés par le plasmide pXL189 (Cameron et al., 1989) qui contient cobO comme seul gène cob. Au contraire ils sont complémentés par le plasmide pXL1908 qui contient cobO et la phase ouverte 25 en plus des phases ouvertes 27 à 30 (voir figure 45). Ces dernières ne peuvent être
 10 responsables de la complémentation de ces mutants puisque les protéines pour lesquelles elles codent n'interviennent pas dans la voie du coenzyme B₁₂. Les complémentations observées ne peuvent donc être que le seul fait de la phase ouverte 25. De plus les mutants SC510 Rif^r Tn5Sp cartographiés dans cette même phase ouverte (il s'agit des
 15 mutants 22, 40, 35, 10 et 17) sont complémentés par le plasmide pXL1908, voir figure 46, (portant cobO et la phase 25) alors qu'au moins deux d'entre eux ne sont pas complémentés par le pXL189 qui ne contient que cobO comme gène cob. Ces résultats montrent clairement que la phase ouverte 25 est un gène cob; ce gène cob est nommé cobN.
- 20 Les mutants SC510 Rif^r Tn5Sp 23, 13 et 12, qui ont le phénotype Cob-, sont cartographiés dans la phase ouverte 24. Ces mutants ne sont pas complémentés par le plasmide pXL623 qui ne contient que le gène cobP. Au contraire ces mutants sont complémentés par le plasmide pXL593 qui contient cobP et la phase ouverte 24 ce qui indique que la phase
 25 ouverte 24 est responsable de leur complémentation. La phase ouverte 24 est donc un gène cob qui est nommé cobW.

EXEMPLE 5 - Gènes et protéines.

5.1 - Fragment de 5,4 kb

- Cinq gènes (cobA, cobB, cobC, cobD et cobE) sont donc
 30 définis sur le fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb. Ils codent respectivement pour les protéines COB suivantes: COBA, COBB, COBC, COBD et COBE. Les parties codantes des gènes (cobA à cobE) sont

décrites à la figure 15, ainsi que les séquences des protéines COBA à COBE. Des propriétés de chacune de ces protéines sont aussi présentées (composition en acides aminés, point isoélectrique, index de polarité et profil d'hydrophilicité).

5

5.2 - Fragment de 8,7 kb

Huit gènes sont donc définis sur le fragment de 8,7 kb. Ces gènes cobF à cobM, codent respectivement pour les protéines COB suivantes : COBF, COBG, COBH, COBI, COBJ, COBK, COBL, COBM. Les parties codantes des gènes (cobF à cobM) sont décrites sur la figure 10 16, ainsi que les séquences des protéines COBF à COBM. Des propriétés de chacune de ces protéines sont aussi présentées (composition en acides aminés, poids moléculaire, point isoélectrique, index de polarité et profil d'hydrophilicité).

5.3 - Fragment de 4,8 kb

15

Trois gènes (cobX , cobS, cobT) sont définis sur le fragment SalI-SalI-SalI-SalI-SalI-BglI de 4,8 kb. Ils codent respectivement pour les protéines suivantes: COBX, COBS et COBT. Les parties codantes de ces gènes sont décrites sur la figure 40, ainsi que les séquences des protéines COBX, COBS et COBT. Arbitrairement, 20 l'ATG a la position 1512 de cobS a été choisi comme codon d'initiation plutôt que celui situé à la position 1485 (voir figure 32). Des propriétés de chacune de ces protéines sont aussi représentées (composition en acides aminés, point isoélectrique, index de polarité et profil d'hydrophobicité). COBT présente une poche hydrophile 25 correspondant aux acides aminés 214 à 305.

5.4 - Fragment de 3,9 kb

Deux gènes (cobU et cobV) sont définis sur le fragment SstI-SstI-BamHI de 3,9 kb. Ils codent respectivement pour les protéines suivantes: COBU et COBV. Les parties codantes de ces gènes

sont décrites sur la figure 41, ainsi que les séquences des protéines COBU à COBV. Des propriétés de chacune de ces protéines sont aussi représentées (composition en acides aminés, point isoélectrique, index de polarité et profil d'hydrophobicité).

5 5.5 - Fragment de 13,4 kb

Cinq gènes cob sont définis sur le fragment de 13.4 kb (cobQ, cobP, cobW, cobN et cobO et cobV). Ils codent respectivement pour les protéines suivantes: COBQ, COBP, COBW, COBN et COBO. Les parties codantes de ces gènes (cobQ, cobP, cobW, cobN et cobO) sont décrites
10 sur la figure 46, ainsi que les séquences des protéines COBQ, COBP, COBW, COBN et COBO. Des propriétés de chacune de ces protéines sont aussi représentées (composition en acides aminés, point isoélectrique, index de polarité et profil d'hydrophobicité).

D'après les profils d'hydrophilicité, qui ont été réalisés
15 suivant les programmes de Hopp et Woods (1981), toutes les protéines COB, à l'exception de COBV, sont vraisemblablement des protéines solubles, par opposition à des protéines membranaires, puisque l'on constate l'absence de grands domaines hydrophobes. COBV est soit une protéine membranaire, puisque l'on constate 4 longs domaines
20 hydrophobes (voir figure 41) soit une protéine cytoplasmique ayant d'importants domaines hydrophobes.

Pour toutes les séquences en acides aminés des protéines COB, il est indiqué comme premier acide aminé en position NH₂-terminale une méthionine. Il est entendu que celle-ci peut être
25 excisée in vivo (Ben Bassat et Bauer, 1984). On sait que des règles concernant l'excision in vivo de la méthionine NH₂-terminale par la méthionine aminopeptidase ont été proposées (Hirel et al., 1989).

Par ailleurs, ces séquences protéiques ont été comparées aux protéines de Genpro*, qui est une extraction protéique de Genbank
30 (version 59) augmentée des parties codantes putatives supérieures à

* (marque de commerce)

200 acides aminés, suivant le programme de Kanehisa (1984). Aucune homologie significative n'a pu être mise en évidence avec les paramètres utilisés sur la version 59 de Genbank, sauf pour COBT. En effet, la protéine COBT présente un "core d'acide aminés acides" entre les positions (en acides aminés) 224 et 293 (voir figure 40); dans cette portion de la protéine, plus d'un acide aminé sur 2 est un résidu acide glutamique ou aspartique; ce noyau d'acides aminés acides rend la protéine homologue sur cette région, suivant le programme de Kanehisa (1984), à d'autres protéines ayant elles aussi un tel noyau acide. Les protéines les plus homologues sont : la protéine GARP de *Plasmodium falciparum* (Triglia et al., 1988), la troponine T cardiaque du rat (Jin et Lin, 1989), la prothymosine humaine et de rat (Eschenfeld et Berger, 1986), une protéine du rat androgène-dépendante se liant à la spermine (Chang et al., 1987), les protéines "mid-size neurofilament subunit" humaine, de rat et de poulet (Myers et al., 1987, Levy et al., 1987, Zopf et al., 1987). La fonctionnalité de ces noyaux riches en résidus acides n'est pas connue; cependant ce noyau acide devrait soit permettre la fixation de cation métalliques tels que le Co^{++} , ce qui rendrait à la protéine COBT un rôle de métallothionéine à cobalt, ou bien permettre des interactions avec d'autres protéines.

EXEMPLE 6 - Etudes enzymatiques

6.1 - Identification de protéines COB et de leurs gènes à partir d'activités enzymatiques

Cet exemple décrit comment, à partir d'une protéine purifiée, après avoir établi sa séquence NH_2 -terminale, il est possible de trouver le gène de structure correspondant parmi des gènes cob séquencés.

6.1.1. Identification de la protéine COBA codée par le gène coba

La purification de la SUMT de *Pseudomonas denitrificans* a

été décrite (F. Blanche et al., 1989). La séquence NH₂-terminale de la protéine ainsi purifiée a pu être réalisée selon la technique décrite ci-dessus. Les dix premiers acides aminés ont été identifiés :

	1	2	3	4	5	6	7	8	9	10
5	Met	Ile	Asp	Asp	Leu	Phe	Ala	Gly	Leu	Pro

La séquence NH₂-terminale de la protéine COBA (figure 15) correspond exactement à cette séquence. Le poids moléculaire de la SUMT purifiée estimé par électrophorèse PAGE-SDS à 12,5 % est de 30 000. La protéine COBA a un poids moléculaire déduit de sa séquence de 29 234 (figure 15). Les correspondances entre les séquences NH₂-terminales et les poids moléculaires indiquent clairement que la protéine COBA correspond à la SUMT. Le gène coba est le gène de structure de la SUMT.

6.1.2. Identification de la protéine COBB codée par le gène CobB

a) Dosage de l'activité acide cobyrinique a,c-diamide synthase

Cet exemple illustre le dosage d'une activité de la voie de biosynthèse des corrinoïdes qui n'a encore jamais été décrite. Il s'agit de l'acide cobyrinique a,c-diamide synthase (ACDAS) qui catalyse l'amidation de deux fonctions acides carboxyliques du noyau corrine ou descobaltocorrine aux positions a et c (figure 17). Le donneur de groupement NH₂ est la L-glutamine et la réaction consomme 1 molécule d'ATP par amidation de chaque fonction acide carboxylique. Le dosage qui est décrit ci-dessous s'applique à la réaction de diamidation de l'acide cobyrinique; avec quelques modifications (détection en CLHP à 330 nm en particulier) il s'applique à la réaction de diamidation de l'acide hydrogénobyrrinique.

Le mélange d'incubation (250 µl de tris-HCl 0,1 M pH 7,6) contenant ATP (1 mM), MgCl₂ (2,5 mM), glutamine (1 mM), acide cobyrinique (25 µM) ou acide hydrogénobyrrinique (5 µM), cobyrinique a,c-diamide synthase (environ 1 unité d'activité) est incubé durant 1 heure à 30°C. A la fin de l'incubation, 125 µl d'une solution aqueuse

de KCN (2,6 g/l) et 125 µl d'HCl 0,2 M sont ajoutés au mélange qui est ensuite chauffé à 80°C pendant 10 minutes puis centrifugé 5 minutes à 5 000 g. 50 µl du surnageant de centrifugation sont analysés en CLHP. Ils sont injectés sur une colonne Nucleosil 5-C₁₈ de 25 cm et élués avec un gradient de 0 à 100 % de tampon B dans A en 30 minutes ;
5 tampon A : phosphate de potassium 0,1 M pH 6,5, 10 mM KCN ; tampon B : phosphate de potassium 0,1 M pH 8, 10 mM KCN/acétonitrile (1/1). Les corrinoïdes sont détectés grâce à leur absorption UV à 371 nm. L'unité d'activité enzymatique est définie comme la quantité d'enzyme
10 nécessaire pour synthétiser 1 nmole de groupements amides par heure dans les conditions décrites.

b) Purification de l'activité acide cobyrinique a,c-diamide synthase de Pseudomonas denitrificans

Cette expérience illustre comment une protéine de
15 Pseudomonas denitrificans intervenant dans la voie de biosynthèse des cobalamines peut être purifiée.

A partir du dosage décrit à l'exemple 6.1.2 a), la purification de l'acide cobyrinique a,c-diamide synthase de Pseudomonas denitrificans est réalisée comme décrit ci-dessous.

20 Dans une expérience typique de purification, 7 g de cellules humides de la souche SC 510 Rif^r dans laquelle on a introduit le plasmide pXL1500 (voir exemple 4.1. pour la description du pXL1500 ainsi que la figure 12) sont suspendues dans 30 ml de Tris-HCl 0,1 M pH 7,7 et soniquées durant 15 minutes à 4°C. L'extrait brut est
25 ensuite récupéré par centrifugation 1 heure à 50 000 g puis 10 ml de cet extrait sont injectés sur une colonne de Mono Q HR 10/10 équilibrée avec le même tampon. Les protéines sont éluées avec un gradient linéaire de KCl (0 à 0,5 M). Les fractions contenant l'activité enzymatique sont regroupées et concentrées à 2,5 ml. Après
30 dilution avec 1 ml de Tris-HCl 25 mM pH 7,7, les protéines sont fractionnées sur une Mono Q HR 5/5 en utilisant le gradient de KCl (0 à 0,5 M) précédent. Les fractions actives sont regroupées, 1 ml de Tris-HCl 0,1 M pH 7,7 contenant 1,7 M de sulfate d'ammonium est ajouté à l'échantillon qui est ensuite chromatographié sur une colonne de

Phényl-Superose (Pharmacia) avec un gradient décroissant de sulfate d'ammonium (1,0 M à 0 M). Les fractions contenant l'activité recherchée sont rassemblées et chromatographiées sur une colonne Bio-Gel HPHT (Bio-Rad) avec un gradient de phosphate de potassium (0 à 0,35 M).

Après cette étape, l'enzyme est pure à plus de 95 %. Elle ne présente aucune protéine contaminante en PAGE-SDS. La pureté de la protéine est confirmée par l'unicité de la séquence NH₂-terminale. Son poids moléculaire dans cette technique est de 45 000. Les différentes étapes de purification de l'ACDAS, avec leur facteur de purification et leur rendement sont portées sur le tableau ci-dessous.

Tableau : Purification de l'ACDAS

16	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. (u/mg de protéines)	Rendement	Facteur de purification ¹
20	Extrait brut	10	200	8.5	-	-
	MonoQ 10/10	12	15.1	108	96	12.7
	MonoQ 5/5	3	3.75	272	60	32
	Phényl-Superose	1	0.865	850	43	100
	Bio-Gel HPHT	2	0.451	1320	35	155

¹ce facteur est calculé d'après l'augmentation de l'activité spécifique des fractions au cours de la purification.

c) Séquence NH₂-terminale de l'acide cobyrinique a,c-diamide synthase de Pseudomonas denitrificans et identification du gène de structure de Pseudomonas denitrificans codant pour cette activité

Cet exemple illustre comment la séquence NH₂-terminale d'une protéine qui intervient dans la voie de biosynthèse des cobalamines permet d'identifier le gène de structure qui code pour

cette protéine.

La séquence NH₂-terminale de l'acide cobyrique a,c-diamide synthase de Pseudomonas denitrificans purifiée comme décrit à l'exemple 6.1.2 b) a été réalisée comme décrit précédemment.

5 15 résidus ont été identifiés :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ser	Gly	Leu	Leu	Ile	Ala	Ala	Pro	Ala	Ser	Gly	Ser	Gly	Lys	Thr

10 La séquence NH₂-terminale de la protéine COBB (figure 15) correspond exactement à cette séquence, si ce n'est que dans la séquence présentée sur la figure 15, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la méthionine aminotermale est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids moléculaire de l'ACDAS purifiée estimé par électrophorèse PAGE-SDS à 15 12,5 % est de 45 000. La protéine COBB a un poids moléculaire déduit de sa séquence de 45 676 (figure 15). Les correspondances entre les séquences NH₂-terminales et les poids moléculaires indiquent clairement que la protéine COBB correspond à l'ACDAS. Le gène cobB est le gène de structure de l'ACDAS.

20 6.1.3. Identification de la protéine COBI codée par le gène CobI

a) Dosage d'une activité S-adénosyl-L-méthionine:précorrine-2 méthyltransférase

25 Cet exemple illustre le dosage d'une activité enzymatique de la voie de biosynthèse des corronoïdes qui n'a encore jamais été décrite. Il s'agit de la S-adénosyl-L-méthionine:-2 méthyltransférase (SP₂MT) qui catalyse le transfert d'un groupement méthyl de la S-adénosyl-L-méthionine (SAM) au précorrine-2 pour donner le précorrine-3 (figure 18). Les facteurs II et III, produits d'oxydation 30 respectivement du précorrine-2 et du précorrine-3 ont déjà été purifiés à partir d'extraits cellulaires de Propionibacterium shermanii (Battersby et MacDonald, 1982, Scott et al., 1984); le

précorrine-2 et le précorrine-3 sont reconnus comme des intermédiaires présumés de biosynthèse du coenzyme B₁₂ mais ils n'ont jamais été purifiés. Pour cette raison, l'activité correspondante n'a jamais été ni dosée, ni purifiée auparavant. Le substrat de la réaction enzymatique, précorrine-2, est une molécule très labile qu'il n'est pas possible de conserver car elle s'oxyde spontanément en présence de traces même infimes d'oxygène (Battersby et MacDonald, 1982). Le principe de ce test enzymatique repose donc sur la possibilité de générer extemporanément, à l'aide d'un extrait enzymatique de la souche SC510 Rif^r dans laquelle on a introduit le plasmide pXL1500, le précorrine-2 à partir du SAM et d'acide δ-aminolévulinique. L'incubation doit être effectuée en conditions de stricte anaérobie.

Les fractions contenant la SP₂MT sont incubées dans 1 ml de Tris-HCl 0,1 M pH 7,7 en présence de 5 mM DTT, 1 mM EDTA, 100 µM [methyl-³H]-SAM (1 µCi), 0,8 mM acide δ-aminolévulinique et 6 mg d'extrait enzymatique brut de la souche de Pseudomonas denitrificans SC510 Rif^r pXL1500 pendant 3 heures à 30°C. La souche SC510 Rif^r pXL1500 contient une forte activité SUMT (F. Blanche et al., 1989). Les composés tétrapyrroliques produits durant l'incubation sont fixés sur une colonne d'échangeur d'anions DEAE-Sephadex et estérifiés dans le méthanol à 5 % d'acide sulfurique en l'absence d'oxygène. Les dérivés diméthylés et triméthylés de l'uro'gen III sont ensuite séparés par chromatographie sur couche mince de silice en utilisant le dichlorométhane/méthanol (98,3/1,7) comme système éluant (F. Blanche et al., 1989). L'activité de la SP₂MT est exprimée par le rapport de la quantité de dérivés triméthylés obtenus sur l'ensemble des dérivés (di- et tri-) méthylés produits ramenée, à la quantité de protéine. L'extrait de SC510 Rif^r pXL1500 introduit dans le test ne présente pas d'activité SP₂MT détectable dans les conditions de dosage (le ratio précorrine-3 produits sur précorrine-2 produits durant le test est inférieur à 0,05).

b) Purification de la S-adénosyl-L-méthionine:précorrine-2 méthyltransférase de Pseudomonas denitrificans

Cette expérience illustre comment une protéine de

Pseudomonas denitrificans intervenant dans la voie de biosynthèse de cobalamines peut être purifiée lorsqu'un dosage de l'activité en question existe.

La protéine est purifiée à partir de cellules de SC510
5 Rif^r contenant le plasmide pXL253. Il s'agit du plasmide pKT230 sur lequel a été inséré le fragment EcoRI de 8,7 kb (figure 13). Dans une expérience typique de purification, 50 g de cellules humides de la souche SC150 Rif^r dans laquelle a été introduit le plasmide pXL253 sont suspendues dans 250 ml de phosphate de potassium 0,1M pH 7,7, 5 mM DTT
10 et soniquées durant 15 minutes à 4°C. Après centrifugation à 50 000 g durant 1 heure, le surnageant est passé à travers une colonne de DEAE-Sephadex (10 ml de gel) pour éliminer les composés tétrapyrroliques. Le pH de l'extrait brut ainsi obtenu est ajusté à pH 7,7 avec KOH 0,1 M. Les protéines précipitant entre 33 % et 45 % de
15 saturation en sulfate d'ammonium sont collectées et dissoutes dans 40 ml de Tris-HCl 0,1M pH 7,7, 5 mM DTT. Cette solution est passée à travers une colonne de Sephadex G-25 éluee avec du Tris-HCl 10 mM pH 7,7, 5 mM DTT et les protéines collectées sont injectées sur une colonne de DEAE-Trisacryl-M. Les protéines sont éluées avec un
20 gradient linéaire de 0 à 0,25 M KCl et les fractions contenant l'activité SP₂MT sont regroupées et passées une seconde fois à travers une colonne de Sephadex G-25 comme ci-dessus. La fraction protéique est injectée sur une colonne d'Ultrogel HA (IBF) équilibrée dans le Tris-HCl 10 mM pH 7,7, 5 mM DTT. Les protéines sont éluées avec un
25 gradient linéaire de 0 à 50 mM de phosphate de potassium pH 7,8 contenant 5 mM DTT. Les fractions contenant l'activité recherchée sont regroupées et injectées sur une MonoQ HR 5/5 (Pharmacia) équilibrée avec du Tris-HCl 50 mM pH 7,7, 5 mM DTT. La SP₂MT est éluee avec un gradient linéaire (0 à 0,25 M) de KCl. A la sortie de l'étape MonoQ,
30 l'électrophorèse en PAGE-SDS (12,5 %) avec coloration aux sels d'argent révèle que l'enzyme est pure à plus de 99 %. Ceci est confirmé par l'unicité de la séquence NH₂-terminale de la protéine. Le poids moléculaire calculé à partir de l'électrophorèse en conditions dénaturantes (PAGE-SDS à 12,5 %) est de 26 500. Les étapes de
35 purification de la SP₂MT avec leurs rendements, sont décrites dans le

tableau ci-dessous.

Tableau : Purification de la SP₂MT

	Etape de purification	Vol (ml)	Protéines (mg)	Facteur de purification ¹
5	Extrait brut	300	6000	-
	Précipitation (33-45 %)	40	1530	3.9
	DEAE-Trisacryl-M	57	355	16.9
	Ultrogel HA	30	71	85
10	MonoQ HR 5/5	12	33.5	179

¹ce facteur est calculé d'après le rendement en protéines.

c) Séquence NH₂-terminale de la SP₂MT et identification du gène de structure codant pour cette activité

Cet exemple illustre comment la séquence NH₂-terminale d'une protéine qui intervient dans la voie de biosynthèse permet d'identifier le gène de structure qui code pour cette protéine. Dans l'exemple présent, il s'agit du gène de structure de la SP₂MT.

La séquence NH₂-terminale de la protéine purifiée a été réalisée comme décrit précédemment. Les 15 premiers acides aminés ont été identifiés :

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Ser Gly Val Gly Val Gly Arg Leu Ile Gly Val Gly Thr Gly Pro

La séquence NH₂-terminale de la protéine COBI (figure 16) correspond exactement à cette séquence, si ce n'est que dans la séquence présentée sur la figure 16 une méthionine précède la séquence peptidique déduite de la séquence nucléotidique. Il en résulte que la méthionine aminoterminal est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids

moléculaire de la SP₂MT purifiée estimé par électrophorèse PAGE-SDS à 12,5 % est de 26 500. La protéine COBI a un poids moléculaire déduit de sa séquence en acides aminés de 25 878 (figure 16). Les correspondances entre les séquences NH₂-terminales et les poids moléculaires indiquent clairement que la protéine COBI correspond à la SP₂MT. Le gène cobI est le gène de structure de la SP₂MT.

6.1.4. Identification de la protéine COBH codée par le gène cobH

a) Dosage de l'activité precorrin-8x mutase.

10 Cet exemple illustre le dosage d'une activité enzymatique de la voie de biosynthèse des cobalamines qui n'a jusqu'à ce jour jamais été décrite. Il s'agit de la precorrin-8x mutase. Cette enzyme catalyse le transfert du groupement methyl de la position C-11 à la position C-12 lors de la transformation du precorrin-8x en acide hydrogénobyrique (voir la nomenclature des carbones FIG. 19; PL. 68). Plus généralement, c'est l'enzyme catalysant le transfert du groupement méthyl de C-11 vers C-12, conduisant ainsi au noyau corrine. L'enzyme est ici appelée mutase, bien qu'il ne soit pas formellement démontré que le transfert du groupement méthyl soit intramoléculaire, même si cela est très vraisemblable.

L'activité enzymatique est mise en évidence par la transformation du precorrin-8x (5 µM) en acide hydrogénobyrique au cours d'incubations en présence de fractions enzymatiques dans du Tris-HCl 0.1 M pH 7.7, 1 mM EDTA, à 30 °C durant 1 h. En fin d'incubation, la réaction est stoppée par chauffage à 80 °C durant 10 min et après centrifugation à 3000 x g durant 10 min, l'acide hydrogénobyrique formé présent le surnageant est analysé par CLHP (Cf exemple 6.1.2.a).

b) Purification de la precorrin-8x mutase.

30 La purification de la precorrin-8x mutase de *Pseudomonas denitrificans*

est réalisée comme décrit ci-dessous.

Durant cette purification, toutes les solutions tampons sont ajustées à pH 7.7.

Dans une expérience typique de purification, 50 g de cellules de la souche SC510 Rif^r portant le plasmide pXL253 (plasmide pKT230 sur lequel a été cloné, au site EcoRI, le fragment de 8.7 kb, figure 13), obtenues après culture en milieu PS4, sont resuspendues dans 200 ml de tampon phosphate de potassium 0.1 M, et soniquées durant 12 min. Après centrifugation à 50 000 x g durant 1 heure, le surnageant est passé à travers une colonne de DEAE-Sephadex* (10 ml de gel) pour éliminer les composés tetrapyrroliques. Le pH de la solution est immédiatement ajusté à 7.7 avec une solution 1 M KOH. La fraction protéique précipitant entre 40 et 60% de saturation en sulfate d'ammonium est collectée par centrifugation et dissoute dans 50 ml de Tris-HCl 0.1 M. Cet échantillon est ensuite injecté sur une colonne d'Ultrogel* ACA 54 (IBF, France) (volume de gel 1000 ml), et les protéines sont éluées à un débit de 60 ml/h avec du Tris-HCl 50 mM. Les fractions contenant l'activité sont regroupées et injectées sur une colonne de DEAE-Trisacryl M (IBF, France) équilibrée avec du Tris-HCl 50 mM, et les protéines sont éluées avec un gradient de 0 à 0.2 M KCl. Les fractions contenant la protéine à purifier sont regroupées, passées à travers une colonne de Sephadex* G25 équilibrée en Tris-HCl 10 mM. La fraction protéique est injectée sur une colonne d'Ultrogel* HA (IBF, France) équilibrée avec du Tris-HCl 10 mM, et les protéines sont éluées avec un gradient de 0 à 0.1 M de phosphate de potassium, puis la fraction active est chromatographiée sur une colonne de Phenyl-Sepharose CL 4B (Pharmacia) dans du phosphate de potassium 10 mM, éluee avec un gradient de 0.65 à 0 M de sulfate d'ammonium. Les fractions actives sont regroupées. La protéine ainsi obtenue est pure à plus de 95% (d'après les résultats d'électrophorèse PAGE-SDS à 12.5% et coloration aux sels d'argent). La pureté de la protéine est confirmée par l'unicité de la séquence N-terminale. Son poids moléculaire calculé à l'aide de cette technique est de 22 000. Les étapes de purification de la precorrin-8x mutase avec leurs

* (marques de commerce)

rendements de purification sont décrites sur le tableau ci-dessous.

tableau: Purification de la precorrin-8x mutase.

	Etape de . purification	Vol (ml)	Protéines (mg)	Facteur de purification ¹
5	Extrait brut	250	6000	-
	Précipitation (40-60 %)	50	2350	2.6
	Ultrogel ACA 54	70	655	9.2
	DEAE-Trisacryl M	30	271	22
10	Ultrogel HA	22	93	65
	Phényl Sépharose	12	31	194

¹ce facteur est calculé d'après le rendement en protéines

c) Séquence NH₂-terminale de la précorrin-8x mutase et identification de son gène de structure.

15 Cet exemple illustre comment la séquence NH₂-terminale d'une protéine qui intervient dans la voie de biosynthèse permet d'identifier le gène de structure qui code pour cette protéine.

La séquence NH₂-terminale de cette protéine a été réalisée comme décrit précédemment. 15 résidus ont été identifiés :

20 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
 Pro Glu Tyr Asp Tyr Ile Arg Asp Gly Asn Ala Ile Tyr Glu Arg

25 La séquence NH₂-terminale de la protéine COBH (figure 16) correspond exactement à cette séquence, si ce n'est que dans la séquence présentée sur la figure 16 une méthionine précède la séquence peptidique déterminée par le séquençage décrit ci-dessus. Il en résulte que la méthionine aminoternale est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le

deuxième résidu étant une proline, cette excision est conforme aux règles déjà énoncées (Hirel et al., 1989). Le poids moléculaire de la precorrin-8x mutase purifiée estimé par électrophorèse PAGE-SDS à 12,5 % est de 22 000. La protéine COBH a un poids moléculaire déduit de sa séquence de 22 050 (figure 16). Les correspondances entre les séquences NH₂-terminales et les poids moléculaires, de ces protéines, indiquent clairement que la protéine COBH correspond à la precorrin-8x mutase. cobH est le gène de structure de la precorrin-8x mutase.

d) Préparation, isolement et identification du precorrin-8x.

Dans une expérience typique de préparation du precorrin-8x, un extrait enzymatique brut de la souche SC510 Rif^r pXL253 (1000 mg de protéines) est incubé en anaérobiose durant 20 h à 30°C dans 100 ml de tampon Tris-HCl 0.1 M pH 7.7 avec 1000 nmol de trimethylisobacteriochlorin préparé comme décrit précédemment (Battersby et al., 1982), EDTA (1 mM), ATP (100 µmol), MgCl₂ (250 µmol), NADH (50 µmol), NADPH (50 µmol), SAM (50 µmol) et acide hydrogénobyrrinique (20 µmol). En fin d'incubation, le precorrin-8x est le produit tétrapyrrolique formé majoritaire. Il est isolé et purifié par CLHP sur une colonne µBondapak^{*}C18 (Waters) en utilisant un gradient d'élution linéaire de 0 à 50% d'acétonitrile dans un tampon phosphate de potassium, pH 5,8. La masse du precorrin-8x (m/z = 880) et la masse de son dérivé méthyl ester (m/z = 978) indiquent qu'il s'agit d'un composé ayant la même formule brute que l'acide hydrogénobyrrinique. Les caractéristiques UV/Visible et de fluorescence sont très différentes de celles de l'acide hydrogénobyrrinique et indiquent que la molécule possède deux chromophores séparés. La seule réaction enzymatique d'isomérisation entre precorrin-6x (Thibaut et al, 1990) et l'acide hydrogénobyrrinique étant la migration du méthyl de C-11 vers C-12, le precorrin-8x est le dernier intermédiaire avant l'acide hydrogénobyrrinique et la réaction correspondante est la migration du méthyl de C-11 vers C-12, catalysée par la precorrin-8x mutase.

6.1.5. Identification de la protéine COBU codée par le

* (marque de commerce)

gène cobU

a) Dosage de l'activité

nicotinate-nucleotide:dimethylbenzimidazole phosphoribosyltransferase (figure 5, réaction 5). Cet exemple illustre le dosage d'une activité

5 enzymatique directement liée à la voie de biosynthèse des cobalamines.

Il s'agit de la nicotinate-nucleotide:dimethylbenzimidazole phosphoribosyltransferase (NN:DMBI PRT) (EC 2.4.2.21). Les fractions

contenant l'activité NN:DMBI PRT (environ 5 unités) sont incubées à 30°C durant 8 mn dans 500 µl de tampon glycine:NaOH 0.1 M pH 9.7 en

10 présence de 1 mM NaMN (acide nicotinique mononucéotide) et 10 µM DMBI.

La réaction est ensuite stoppée par chauffage à 80°C durant 10 mn, le mélange réactionnel est dilué par 4 volumes d'eau et 100 µl de cette

solution sont injectés sur une colonne CLHP Nucleosil 5-C8 de 15 cm éluee avec un mélange 0.1 M phosphate de potassium pH 2.9:acétonitrile

15 (93:7) à un débit de 1 ml/min. Le α-ribazole 5'-phosphate est

détecté et quantifié par fluorimétrie (excitation : 260 nm; émission >370 nm). L'unité d'activité enzymatique est définie comme la quantité

d'enzyme nécessaire pour générer 1 nmol de α-ribazole 5'-phosphate par heure dans ces conditions.

20 b) Purification de l'activité NN:DMBI PRT de *Pseudomonas denitrificans*

Cette expérience illustre comment une protéine de *P. denitrificans* intervenant dans la voie de biosynthèse des cobalamines peut être

purifiée. A partir du dosage décrit dans l'exemple 6.1.5.a), la purification de la NM:DMBI PRT de *Pseudomonas denitrificans* est

25 réalisée comme décrit ci-dessous. Dans une expérience typique de

purification on utilise 10 g de cellules humides de la souche SC510 Rifr dans laquelle le plasmide pXL1490B a été introduit comme cela est

décrit précédemment. Le plasmide pXL1490B est décrit sur la figure 38; ce plasmide a été obtenu par clonage du fragment BamHI-SstI-SstI de

30 3.85 kb du pXL519 (voir figure 38). Ce plasmide porte donc les gènes

cobU et cobV de *P. denitrificans*. Les cellules cultivées en milieu PS4, supplémenté en lividomycine, comme cela est décrit précédemment,

sont récoltées après 96 heures de culture en milieu PS4. Elles sont resuspendues dans 25 ml de tampon 0.1M Tris-HCl pH 7.2 et soniquées

durant 15 mn à 4°C. L'extrait brut est ensuite récupéré par centrifugation durant 1 h à 50 000 g puis passé à travers une colonne de DEAE-Trisacryl M (IBF, France) équilibrée avec le même tampon. 10 % de l'éluat (120 mg de protéines) est fractionné sur une colonne de
 5 Mon-o*Q HR 10/10 en utilisant un gradient de KCl (de 0 à 0.6 M). Les fractions actives sont regroupées et concentrées à 2 ml par ultrafiltration puis après mélange avec un volume de tampon Tris-HCl 30 mM pH 7.2, l'échantillon est fractionné une seconde fois sur une Mono Q HR 5/5 comme précédemment. Les fractions actives sont
 10 regroupées puis l'échantillon est amené à une molarité de 1 M à l'aide de sulfate d'ammonium et chromatographié sur une colonne de Phényl-Superose* HR 5/5 élué avec un gradient décroissant de sulfate d'ammonium (de 1 M à 0 M). Les fractions contenant l'activité recherchée sont rassemblées, concentrées par ultrafiltration, et
 15 chromatographiées sur une colonne de perméation de gel Bio-Sil 250, éluée avec du phosphate de sodium 20 mM-sulfate de sodium 50 mM pH 6.8.

Après cette étape, l'enzyme est pure à plus de 95 %. Elle ne présente aucune protéine contaminante en PAGE-SDS. Cette pureté est confirmée
 20 par l'unicité de la séquence NH₂-terminale. Son poids moléculaire dans cette technique est de 35 000. Les différentes étapes de purification de la NN:DMBI PRT sont portées sur le tableau ci-dessous.

Tableau : Purification de la NN:DMBI PRT de P. denitrificans

* (marques de commerce)

	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. (u/mg de protéines)	Rendement	Facteur de purification ¹
	Extrait brut	6,0	120	2650	-	-
5	MonoQ 10/10	6,0	12,07	13515	51,3	5,1
	MonoQ 5/5	3,0	6,19	20140	39,2	7,6
	Phényl-Superose	1,5	2,60	35510	29,0	13,4
	Bio-Sil 250	1,2	1,92	39750	24,0	15,0

c) Séquence NH2-terminale de la NN:DMBI PRT de *P. denitrificans* et
10 identification du gène de structure de *Pseudomonas denitrificans*
codant pour cette activité. La séquence NH2-terminale de la NN:DMBI
PRT de *Pseudomonas denitrificans* purifiée comme décrit dans l'exemple
6.1.5b) a été réalisée selon la technique décrite ci-dessus. Les 15
premiers résidus ont été identifiés :

15 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
 Ser Ala Ser Gly Leu Pro Phe Asp Asp Phe Arg Glu Leu LeuArg

La séquence NH2-terminale de la protéine COBU (figure 41) correspond à
cette séquence, si ce n'est que dans la séquence présentée sur la
figure 41, une méthionine précède le premier acide aminé de la
séquence peptidique déterminée par séquençage direct. Il en résulte
20 que la méthionine aminoterminal est certainement excisée in vivo par
la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids
moléculaire de la N-transglycosidase purifiée estimé par
électrophorèse PAGE-SDS à 12,5 % est de 35 000. La protéine COBU a un
25 poids moléculaire déduit de sa séquence de 34 642 (figure 41). Les
correspondances entre les séquences NH2-terminales et les poids
moléculaires indiquent clairement que la protéine COBU correspond à la
NN:DMBI PRT. Le gène cobU est le gène de structure de la NN:DMBI PRT.

d) Spécificité de la NN:DBI PRT pour de DBI. Cet exemple illustre comment l'étude de la spécificité de la NN:DMBI PRT de *P. denitrificans* permet de faire biosynthétiser à *P. denitrificans* diverses cobamides en utilisant les propriétés catalytiques de la NN:DMBI PRT de *P. denitrificans* pour effectuer la synthèse de la base nucléotidique en question.

Le substrat de l'enzyme pour synthétiser des cobalamines est le 5,6-diméthylbenzimidazole. Le benzimidazole et le 5-méthylbenzimidazole, respectivement, sont des substrats de la réaction avec des vitesses de réaction de 157 % et de 92 %, respectivement, comparé au substrat naturel (5,6-diméthylbenzimidazole), la concentration en NaMN étant fixée à 2 mM. La spécificité de la NN:DMBI PRT de *P. denitrificans* est donc faible pour les substrats à noyau benzimidazole. On peut donc utiliser la souche de *P. denitrificans* SC510 Rifr (Cameron et al., 1989), la cultiver en milieu PS4 où le 5,6-diméthylbenzimidazole est remplacé par du benzimidazole ou du 5-diméthylbenzimidazole, respectivement, afin de faire synthétiser à la bactérie de la Co α -(benzimidazolyl)-Co β -cyanocobamide, Co α -(5-méthylbenzimidazolyl)-Co β -cyanocobamide respectivement. Il est certain que d'autres cobamides pourraient être synthétisées de cette manière.

6.1.6. Identification de la protéine COBV codée par le gène CobV.

Cet exemple illustre comment le dosage d'une activité de la voie de biosynthèse du coenzyme B12 chez *P. denitrificans*, puis la purification partielle de cette activité peut permettre d'identifier le gène de structure de cette enzyme chez *P. denitrificans*.

a) dosage de l'activité GDP-cobinamide: α -ribazole (5'-phosphate) cobinamidephosphotransférase (ou cobalamine (5'-phosphate) synthase) Cet exemple illustre le dosage d'une activité directement liée à la voie de biosynthèse des cobalamines. Il s'agit de la cobalamine (5'-phosphate) synthase. Les fractions contenant l'activité (environ 5

à 10 unités) sont incubées à l'obscurité à 30°C dans 500 µl de tampon Tris-HCl 0.3 M pH 9.0 en présence de 1 mM EDTA, 12.5 mM MgCl₂, 50 µM α-ribazole 5'-phosphate et 20 µM GDP-cobinamide [sous forme 5'-deoxy-5'-adénosyl (Ado) ou coenzyme]. Après 15 mn d'incubation, 500 µl de cyanure de potassium 20 mM sont ajoutés et la solution est chauffée à 80°C durant 10 mn. Après centrifugation pour éliminer le matériel précipité, la vitamine B12 5'-phosphate présente dans le surnageant est dosée comme cela est décrit à l'exemple 9. Une unité de cobalamine (5'-phosphate) synthase est définie comme la quantité d'enzyme nécessaire pour générer 1 nmol de cobalamine 5'-phosphate par h dans les conditions décrites ci-dessus.

L'Ado-GDP-cobinamide est obtenu par incubation de l'Ado-cobinamide phosphate (Blanche et al., 1989) avec un extrait de SC510 Rif^r pXL623 dans les conditions du dosage de la cobinamide phosphate guanylyltransferase (voir 6.1.11.b). Le α-ribazole et le α-ribazole-5'-phosphate sont isolés à partir de cultures de SC510 Rif^r et purifiés par CLHP dans les conditions du dosage décrit dans l'exemple 6.1.5a).

b) Purification partielle de la cobalamine (5'-phosphate) synthase. Cette expérience illustre comment une activité enzymatique de *P. denitrificans* intervenant dans la voie de biosynthèse des cobalamines de *P. denitrificans* peut être partiellement purifiée. A partir du dosage décrit ci-dessus, la purification de la cobalamine (5'-phosphate) synthase est réalisée. Pour ce faire, dans une expérience typique de purification, 10 g de cellules humides de la souche SC510 Rif^r dans laquelle le plasmide pXL1490B a été introduit comme cela est décrit précédemment. Le plasmide pXL1490B est décrit sur la figure 38; ce plasmide correspond au fragment SstI-SstI-BamHI de 3.85 kb cloné dans le pKT230. Ce plasmide porte les gènes cobU et cobV de *P. denitrificans*. La présence de ce plasmide chez *P. denitrificans* SC510 Rif^r conduit à une amplification de l'activité cobalamine (5'-phosphate) synthase d'un facteur 100 environ; il est donc probable que l'insert porté par le plasmide pXL1490B contient le gène de structure de cette enzyme; ce gène ne peut donc être que cobU ou cobV. Les cellules SC510 Rif^r pXL1490B sont obtenues par culture en

milieu PS4 supplémenté avec de la lividomycine comme cela est décrit ci-dessus. Les cellules sont centrifugées puis resuspendues dans 25 ml de tampon Tris-HCl 0.1 M (pH 8.3)-1 mM EDTA (tampon A) et soniqués pendant 15 mn à 4°C. L'extrait brut est ensuite récupéré par centrifugation 1 h à 50 000 g et passé à travers une colonne de Sephadex*G-25 équilibrée avec du tampon A. La fraction protéique est récupérée et injectée par fraction de 300 µl (7.5 mg de protéines) sur une colonne de Superose 12 HR 10/30 éluée dans du tampon A. La fraction exclue est récupérée, mélangée à un volume égal de tampon A-1.0 M sulfate d'ammonium et chromatographiée sur une colonne de Phényl-Supérose* HR 5/5. Les protéines sont éluées avec un gradient décroissant de sulfate d'ammonium (de 0.5 M à 0 M) dans du tampon A, suivi d'un plateau à 0 M de sulfate d'ammonium dans le but d'éluer l'activité cobalamine (5'-phosphate) synthase. La purification partielle de cette enzyme est décrite dans le tableau ci-dessous, sur la base de 75 mg de protéines introduits au départ dans le processus de purification.

Tableau : Purification partielle de la cobalamine (5'-phosphate) synthase de *P. denitrificans*.

20	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. (u/mg de protéines)	Rendement	Facteur de purification ¹
	Extrait brut	3.0	75	325	-	-
	Superose*12HR	50.0	2.9	6 810	81	21
25	Phenyl-Superose*	4.5	0.35	17 850	26	55

c) Spécificité de la cobalamine (5'-phosphate) synthase. Le Km pour le (Ado)GDP-cobinamide est de 0.9 µM. Toutefois, l'enzyme présente la même affinité et une vitesse de réaction pratiquement identiques pour la forme (CN, aq) du substrat. Le Km de l'enzyme pour le α-ribazole

* (marques de commerce)

5'-phosphate est de 2.7 μM environ. De plus, les préparations les plus pures de cobalamine (5'-phosphate) synthase catalysent la réaction du Ado-GDP-cobinamide avec le α -ribazole pour donner le coenzyme B12 et dans ces conditions aucune accumulation de cobalamine 5'-phosphate n'est observée. Le K_m de l'enzyme pour le α -ribazole est de 7.8 μM . Des concentrations intracellulaires en α -ribazole 5'-phosphate et en α -ribazole de 30 et de 700 μM respectivement, ont été mesurées par CLHP au cours de la production de cobalamines de SC510 Rifr en milieu PS4 dans les conditions de cultures décrites dans l'exemple 6.1.5a). Ceci montre que le coenzyme B12 peut être généré directement à partir du Ado-GDP-cobinamide par la cobalamine (5'-phosphate) synthase sans l'intervention d'une cobalamine 5'-phosphatase. L'absence d'accumulation ou la présence de traces de cobalamine 5'-phosphate dans les cultures de *P. denitrificans* SC510 Rifr confirme que le coenzyme B12 est produit par la réaction directe du Ado-GDP-cobinamide avec le α -ribazole in vivo. Cette réaction directe a déjà été observée et décrite in vitro chez *Propionibacterium shermanii* (Ronzio et al., 1967; Renz, 1968). Comme le gène de structure de la cobalamine (5'-phosphate) synthase ne peut être que cobU ou cobV puisque l'amplification chez *P. denitrificans* d'un fragment portant ces deux gènes cob de *P. denitrificans* conduit à une augmentation par un facteur 100 de l'activité cobalamine (5'-phosphate) synthase et que le gène cobU est le gène de structure de la NN:DMBI PRT, cobV est donc le gène de structure de la cobalamine (5'-phosphate) synthase.

6.1.7. Identification de la protéine COBK codée par le gène cobK.

a) dosage de l'activité precorrin-6x réductase.

Cet exemple illustre le dosage d'une activité enzymatique nouvelle directement liée à la voie de biosynthèse des cobalamines. Il s'agit de la precorrin-6x réductase.

- Les fractions contenant l'activité precorrin-6x réductase (environ 0.05 unités, U) sont incubées à 30°C durant 60 min dans 250 µl de tampon Tris-HCl 0.1 M pH 7.7 en présence de 1 mM EDTA, 500 µM NADPH, 25 µM [methyl-3H]SAM (80 µCi/µmol), 4 µM precorrin-6x (Thibaut et al., 1990), et 0.5 U de dihydroprecorrin-6x méthylase partiellement purifiée (voir préparation ci-dessous). La réaction est ensuite stoppée par chauffage à 80°C durant 5 min et, après centrifugation à 5000 x g durant 5 min, le surnageant est injecté sur une colonne de DEAE-Sephadex (contenant 200 µl de gel). La colonne est ensuite lavée extensivement avec le tampon Tris-HCl, et les composés fixés sont élués avec 4 ml de 1 M HCl. La radioactivité dans cet éluat est comptée en scintillation liquide. L'unité d'activité enzymatique est définie comme la quantité d'enzyme nécessaire pour réduire 1 nmol de precorrin-6x par h dans ces conditions.
- La dihydroprecorrin-6x méthylase est partiellement purifiée à partir d'un extrait brut de SC510 Rifr pXL253, sur une colonne d'échange d'anions Mono Q HR 5/5 (Pharmacia). La colonne est éluee avec un gradient linéaire de 0 à 0.4 M de KCl dans du tampon 0.1 M Tris-HCl pH 7.7. L'activité enzymatique est éluee à 0.35 M KCl. Cette activité est détectée et quantifiée grâce au test d'activité de la precorrin-6x reductase défini ci-dessus (en présence de 0.5 U de precorrin-6x reductase dans le milieu d'incubation). Après l'étape de Mono Q, les fractions contenant de l'activité dihydroprecorrin-6x méthylase sont totalement dépourvues d'activité precorrin-6x réductase. L'unité d'activité méthylase est définie comme la quantité d'enzyme nécessaire pour transférer 1 nmol de groupements méthyls sur le dihydroprecorrin-6x par h dans les conditions décrites ci-dessus.

b) purification de l'activité precorrin-6x réductase.

- A partir du dosage décrit ci-dessus, la purification de la precorrin-6x réductase de *Pseudomonas denitrificans* est réalisée comme décrit ci-dessous.

Dans une expérience typique de purification, 100 g de cellules humides

de la souche SC510 Rifr dans laquelle on a introduit le plasmide pXL253 (plasmide pKT230 sur lequel a été cloné, au site EcoRI, le fragment de 8.7 kb, figure 13) sont suspendues dans 200 ml de tampon Tris-HCl 0.1 M pH 7.7-1 mM EDTA (tampon A), et soniquées 5 durant 15 min à 4°C. L'extrait brut est ensuite récupéré par centrifugation durant 1 h à 50 000 x g et passé en trois fois à travers une colonne de Sephadex G 25 équilibrée avec le tampon A. Les trois fractions exclues du gel sont poolées et ajustées à 1 l avec le tampon A. Les protéines précipitant entre 25 et 40% de saturation 10 en sulfate d'ammonium sont collectées par centrifugation et resuspendues dans 50 ml de tampon A et cette solution est désalée à travers une colonne de Sephadex G 25 équilibrée avec du tampon B (25 mM Tris-HCl-500 µM DTT-15% glycérol). La solution de protéines est ensuite injectée à 2.5 ml/min sur une colonne de Q 15 Sepharose Fast Flow (Pharmacia) équilibrée avec du tampon B, et les protéines sont éluées avec un mélange tampon B-0.2 M KCL. Cette fraction est désalée sur une colonne de Sephadex G 25 équilibré avec du tampon C (50 mM Tris-HCl-500 µM DTT-15% glycérol). La solution protéique est ensuite fractionnée (100 mg de protéines à chaque 20 chromatographie) sur une colonne Mono Q HR 10/10 (Pharmacia) à l'aide d'un gradient de 0 à 0.4 M KCL dans le tampon C, puis la fraction contenant l'activité est chromatographiée sur une colonne Phenyl-Superose HR 10/10 (Pharmacia) dans un gradient linéaire décroissant de sulfate d'ammonium (de 1 à 0 M). La fraction active est 25 désalée et la precorrin-6x réductase est repurifiée sur une colonne Mono Q HR 5/5. Elle est éluée dans le tampon 50 mM Tris-HCl pH 8.1-500 µM DTT-15% glycérol, avec un gradient de 0 à 0.2 M de KCL. Pour parfaire la purification, la protéine est finalement chromatographiée sur une colonne Bio-Sil 250 (Bio-Rad) éluée avec 30 20 mM potassium phosphate-50 mM sodium sulfate, pH 6.8-500 µM DTT-15 % glycérol. Après cette étape, l'enzyme est pure à plus de 95%. Elle ne présente aucune protéine contaminante en PAGE-SDS, les protéines étant révélées au nitrate d'argent. Ce degré de pureté est confirmé par l'unicité de la séquence NH2-terminale. Son poids 35 moléculaire dans cette technique est de 31 000. Les différentes étapes

de purification de la precorrin-6x réductase, avec leur facteur de purification et leur rendement sont portées sur le tableau ci-dessous.

Tableau: Purification de la precorrin-6x reductase.

5	Etape de purification	vol (ml)	Protéines (mg)	Activité.spé. (u/mg de protéines)	Rendement	Facteur de purification ¹
	Extrait brut	270	9600	0.535	-	-
10	S.A. 25 40%	100	4160	1.14	92	2.1
	Q Sepharose	150	1044	3.64	74	6.8
	Mono Q 10/10	55	67	24.5	32	46
	Phényl-superose	10	2.2	325	14	607
	Mono Q 5/5	2.5	0.082	5750	9.2	10750
15	Bio-sil 250	1.0	0.055	7650	8.2	14300

c) Séquence NH2-terminale et séquences partielles internes de la precorrin-6x réductase de Pseudomonas denitrificans et identification du gène de structure de Pseudomonas denitrificans codant pour cette activité.

20 La séquence NH2-terminale de la precorrin-6x réductase de Pseudomonas denitrificans purifiée comme décrit ci dessus a été déterminée comme décrit précédemment. Six résidus ont été identifiés:

Ala-Gly-Ser-Leu-Phe-Asp

25 De même, après digestion tryptique et séparation des fragment CLHP sur une colonne de phase inverse C-18, trois séquences internes ont été obtenues:

Ile-Gly-Gly-Phe-Gly-Gly-Ala-Asp-Gly-Leu

Arg-Pro-Glu-Trp-Val-Pro-Leu-Pro-Gly-Asp-Arg

Val-Phe-Leu-Ala-Ile-Gly

La séquence NH₂-terminale de la protéine COBK (figure 16) correspond exactement à la séquence NH₂-terminale de la precorrin-6x réductase, si ce n'est que dans la séquence présentée sur la figure 16, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la méthionine aminotermale est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). De même, les trois séquences internes correspondent aux troisséquences 60 à 69, 112 à 122, et 143 à 148 de la protéine COBK. Le poids moléculaire de la precorrin-6x réductase purifiée est estimé par électrophorèse PAGE-SDS à 31 000. La protéine COBK a un poids moléculaire déduit de sa séquence de 28 000 (figure 16). Les correspondances entre les séquences NH₂-terminales, internes, et les poids moléculaires indiquent clairement que la protéine COBK correspond à la precorrin-6x réductase. Le gène cobK est le gène de structure de la precorrin-6x réductase.

d) Réaction catalysée par la precorrin-6x réductase.

La réaction enzymatique de réduction du precorrin-6x est NADPH-dépendante de façon stricte chez *P. denitrificans*. Le NADPH ne peut pas être remplacé par le NADH. Lorsque l'enzyme purifiée (ou une fraction active en cours de purification, ou même un extrait enzymatique brut) est incubée dans les conditions du dosage de l'activité, mais en l'absence de SAM et de dihydroprecorrin-6x méthylase, le produit de la réaction peut alors être purifié par CLHP dans le système décrit pour la purification du precorrin-6x (Cf exemple 6.1.4.d). Après désalage et estérification (méthanol sulfurique à 4%, 20°C, 24h, atmosphère argon), l'ester correspondant a une masse $m/z = 1008$. Le produit de la réaction catalysée par la precorrin-6x reductase est donc le dihydroprecorrin-6x, appelé aussi

precorrin-6y.

6.1.8. Identification de la protéine COBQ codée par le gène CobQ.

a) Dosage de l'activité cobyric acid synthase.

5 Cet exemple illustre le dosage d'une activité enzymatique de la voie de biosynthèse des cobalamines qui n'a jusqu'à ce jour jamais été décrite. Il s'agit de la cobyric acid synthase. Cette enzyme catalyse l'amidation des fonctions acides carboxyliques périphériques en positions b, d, e, et g sur le noyau corrine (voir FIG. 19; PL. 68). Le
10 donneur de groupements NH₂ est la L-glutamine et chaque réaction d'amidation s'accompagne de la consommation d'une molécule d'ATP.

La fraction à doser est incubée à l'obscurité à 30°C pendant 60 min dans 250 µl de tampon 0.1 M Tris hydrochloride pH 7.5 contenant 1 mM DTT, 1 mM EDTA, 1 mM ATP, 2.5 mM MgCl₂, 1 mM glutamine, 10 µM
15 acide Ado-cobyrique di- ou pentaamide. La réaction est stoppée en ajoutant 25 µl d'une solution aqueuse 0.1 M de cyanure de potassium. Après chauffage à 80°C durant 10 min et centrifugation à 3000 x g durant 10 min, les composés formés présents dans le surnageant sont analysés par CLHP. L'unité d'activité est définie comme la quantité
20 d'enzyme nécessaire pour générer 1 nmol de fonctions amides par h dans ces conditions.

Les acides 5'-deoxy-5'-adenosyl(Ado)-cobyriques diamide et pentaamide sont isolés de cultures de la souche SC510 en milieu PS4, en utilisant la méthode dont le principe est décrit dans l'exemple 9.

25 b) Purification de la cobyric acid synthase.

A partir du dosage décrit à l'exemple 6.1.8 a), la purification de la cobyric acid synthase de *Pseudomonas denitrificans* est réalisée comme décrit ci-dessous.

Dans une expérience typique de purification, 6 g de cellules humides
30 de SC510 Rif^r dans laquelle on a introduit le plasmide pXL618 (voir

exemple 4.5.2) sont soniqués dans 15 ml de tampon Tris-HCl 0.1 M pH 7.7, DTT 1 mM, EDTA 1mM. Après centrifugation (50 000 x g durant 1 h), l'extrait est amené à 20% de glycerol (vol/vol). A 8.5 ml de l'extrait brut (203.5 mg de protéines) sont ajoutés 24 ml de 5 tampon Tris-HCl 10 mM, DTT 1 mM, glycerol 20%. La solution est injectée sur Mono Q HR 10/10 (Pharmacia) à 2 ml/min équilibrée avec du tampon Tris-HCl 50 mM pH 7.7, DTT 1 mM, glycérol 20%. Les protéines sont éluées avec un gradient linéaire de 0.5 M NaCl et les fractions actives regroupées et amenées à 1 mM EDTA. La solution est amenée 10 à 0.85 M en sulfate d'ammonium et injectée sur une colonne Phenyl-Superose HR 5/5 (Pharmacia), équilibrée dans le tampon Tris-HCl pH 7.7, DTT 1 mM, sulfate d'ammonium 0.85 M et les protéines sont éluées avec un gradient linéaire décroissant de 0.85 M à 0 M de sulfate d'ammonium. Les fractions sont immédiatement amenées à 20% de 15 glycérol. La fraction active est concentré à 2.5 ml par ultrafiltration et chromatographiée sur une colonne PD 10 (Pharmacia) équilibrée et éluee avec du tampon Tris-HCl 50 mM pH 8.3, DTT 1 mM, glycérol 20% (vol/vol). La fraction protéique est recueillie et injectée sur une Mono Q HR 5/5 équilibrée avec le même tampon et les protéines 20 sont éluées avec un gradient linéaire de 0.5 M NaCl. La chromatographie de perméation sur gel Bio-Sil 250 (Bio-Rad) en milieu tampon Tris-HCl 50 mM pH 7.5, DTT 1 mM, glycérol 20%, NaCl 0.1 M permet finalement d'obtenir une protéine pure à plus de 97%. Elle ne présente aucune protéine contaminante en PAGE-SDS. Cette 25 pureté est confirmée par l'unicité de la séquence NH₂-terminale. Son poids moléculaire dans cette technique est de 57 000. Les différentes étapes de purification de la cobyric acid synthase avec leur facteur de purification et leur rendement sont portées sur le tableau ci-dessous.

30 Tableau: Purification de la cobyric acid synthase.

	Etape de purification	vol (ml)	Protéines (mg)	Activité spé.	Rendement ^a	Facteur de purification ^a
				U /mg a b A B		
5	Extrait brut	8.5	203	114 / 118	-	-
	MonoQ 10/10	8.0	35.5	388 / 425	60	3.4
	Phényl-superoxe	8.0	3.23	1988 / 2021	28	17
	MonoQ 5/5	1.0	1.20	4549 / 4085	24	40
	Bio-Sil 250	0.75	0.88	4992 / N.D.	19	44

10 a/ Avec l'acide Ado-cobyrique a,c-diamide comme substrat

b/ Avec l'acide Ado-cobyrique pentaamide comme substrat

ND = Non Déterminé

Le très haut degré de pureté de la protéine purifiée ainsi que la constance du ratio des activités d'amidation des acides
 15 cobyriques diamide et pentaamide tout au long du procédé de purification de la protéine (voir tableau ci-dessus) indiquent sans ambiguïté qu'une seule et même protéine est responsable des quatre activités d'amidation du noyau corrine aux positions b, d, e, et g.

c) Séquence NH₂-terminale de la cobyric acid synthase de
 20 *Pseudomonas denitrificans* et identification du gène de structure de *Pseudomonas denitrificans* codant pour cette activité.

La séquence NH₂-terminale de la cobyric acid synthase de *Pseudomonas denitrificans* a été déterminée comme décrit précédemment. Seize résidus ont été identifiés:

25 Thr-Arg-Arg-Ile-Met-Leu-Gln-Gly-Thr-Gly-Ser-Asp-Val-Gly-Lys-Ser

La séquence NH₂-terminale de la protéine COBQ (figure 47) correspond exactement à cette séquence, si ce n'est que dans la séquence présentée sur la figure 47, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la

méthionine aminoterminal est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids moléculaire de la cobyric acid synthase purifiée est estimé par électrophorèse PAGE-SDS à 57 000. La protéine COBQ a un poids moléculaire déduit de sa séquence de 52 000 (figure 47). Les correspondances entre les séquences NH₂-terminales et les poids moléculaires indiquent clairement que la protéine COBQ correspond à la cobyric acid synthase. Le gène cobQ est le gène de structure de la cobyric acid synthase.

6.1.9. Identification de la protéine COB0 codée par le gène cob0.

a) dosage de l'activité cob(I)alamin adenosyltransferase (EC 2.5.1.17).

Cet exemple illustre le dosage d'une activité enzymatique directement liée à la voie de biosynthèse des cobalamines. Il s'agit de la cob(I)alamin adenosyltransferase (EC 2.5.1.17). Cette enzyme a été mise en évidence dans les cellules bactériennes (Ohta et al., 1976, Brady et al., 1962) et animales (Fenton et al., 1978). Elle a été purifiée à partir de *Clostridium tetanomorphum* (Vitols et al., 1966).

Les fractions contenant l'activité cob(I)alamin adenosyltransferase (environ 20 unités) sont incubées de façon anaérobie à 30°C durant 15 min à l'abri de la lumière dans 1 ml de tampon Tris/HCl 0.2 M pH 8.0 en présence de 5 mM DTT, 400 µM [8-¹⁴C] ATP (2.5 µCi/µmol), 800 µM MnCl₂, 50 µM hydroxocobalamine ou diaquacobinamide et 3 mg KBH₄. La réaction est ensuite stoppée par chauffage à 80°C durant 10 min et, après centrifugation à 15000 x g durant 5 min, 200 µl de surnageant sont analysés par CLHP (Gimsing et al., 1986, Jacobsen et al., 1986).

L'unité d'activité enzymatique est définie comme la quantité d'enzyme nécessaire pour générer 1 nmol d'adénosylcorrinoïde par min dans ces conditions.

b) Purification de l'activité cob(I)alamin adenosyltransferase.

A partir du dosage décrit à l'exemple 6.1.9 a), la purification de la cob(I)alamin adenosyltransferase de *Pseudomonas denitrificans* est réalisée comme décrit ci-dessous.

Dans une expérience typique de purification, 10 g de cellules humides de la souche SC510 Rif^r dans laquelle le gène cobO a été amplifié sont suspendues dans 20 ml de tampon Tris-HCl 0.2 M pH 8.0 et soniquées durant 40 min à 4°C. L'extrait brut est ensuite récupéré par centrifugation durant 1 h à 50 000 x g et dessalé sur colonnes PD10 (Pharmacia) équilibrées avec du tampon 50 mM Tris-HCl pH 8.0, 5 mM DTT (tampon A). La solution protéique est ensuite fractionnée (280 mg de protéines à chaque chromatographie) sur une colonne Mono Q HR 10/10 (Pharmacia) à l'aide d'un gradient de 0 à 0.5 M KCl dans le tampon A, puis les fractions contenant l'activité sont poolées, concentrées par ultrafiltration et chromatographiées sur une colonne de Phenyl-Superose HR 10/10 (Pharmacia) dans un gradient linéaire décroissant de sulfate d'ammonium (de 1,7 à 0 M), la colonne étant équilibrée dans le tampon 0.1 M Tris-HCl pH 8.0, 5 mM DTT. Pour parfaire la purification, la protéine est finalement chromatographiée, après concentration par ultrafiltration, sur une colonne Bio-Sil 250 (Bio-Rad) éluée avec le tampon 50 mM Tris-HCl pH 7.5, 0.1 M NaCl, 5 mM DTT.

Après cette étape, l'enzyme est pure à plus de 95%. Elle ne présente aucune protéine contaminante en PAGE-SDS. Son poids moléculaire dans cette technique est de 28 000. Ce degré de pureté est confirmé par l'unicité de la séquence NH₂-terminale. Les différentes étapes de purification de la cob(I)alamin adenosyltransferase, avec leur facteur de purification et leur rendement sont portées sur le tableau ci-dessous pour les deux substrats suivants: diaquacobinamide (a) et hydroxocobalamine (b). Ces résultats démontrent l'absence de spécificité de cette enzyme pour la nature du substrat corrinoïde. D'autre part, tous les corrinoïdes de la voie de biosynthèse entre l'acide cobyrinique diamide et la B₁₂ ont été isolés (Blanche et al., résultats non publiés) sous leur forme native, et se sont révélés être sous forme coenzyme. Ceci démontre que le substrat naturel

de la cob(I)alamin adenosyltransferase est l'acide cobyrinique a,c-diamide.

Tableau: Purification de la cob(I)alamin adenosyltransferase.

5	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. U /mg a b A B	Rendement ^a	Facteur de purification ^a
	Extrait brut ^c	100	1400	5.4 / 3.4	-	-
10	MonoQ 10/10	90	140	34.9 / 14.1	65	6.5
	Phényl-Superoose	30	15.9	84.5 / 49.5	18	16
	Bio-Sil 250	6.5	2.9	182.4/ 88.7	7.0	34

c/ après dessalage sur PD10

c) Séquence NH2-terminale de la cob(I)alamin adenosyltransferase de Pseudomonas denitrificans et identification du gène de structure de Pseudomonas denitrificans codant pour cette activité.

La séquence NH2-terminale de la cob(I)alamin adenosyltransferase de Pseudomonas denitrificans purifiée comme décrit à l'exemple 6.1.9 b) a été déterminée comme décrit précédemment. 13 résidus ont été identifiés:

Ser-Asp-Glu-Thr-?-Val-Gly-Gly-Glu-Ala-Pro-Ala-Lys-Lys

La séquence NH2-terminale de la protéine COBO (figure 47) correspond exactement à la séquence NH2-terminale de la cob(I)alamin adenosyltransferase, si ce n'est que dans la séquence présentée sur la figure 47, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la méthionine aminoterminal est certainement excisée in vivo par la

méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids moléculaire de la cob(I)alamin adenosyltransferase purifiée est estimé par électrophorèse PAGE-SDS à 28 000. La protéine COB0 a un poids moléculaire déduit de sa séquence de 24 000 (figure 47). Les correspondances entre les séquences NH₂-terminales et les poids moléculaires indiquent clairement que la protéine COB0 correspond à la cob(I)alamin adenosyltransferase. Le gène cob0 est le gène de structure de la cob(I)alamin adenosyltransferase.

6.1.10. Identification de la protéine COBN codée par le gène cobN.

a) Mise en évidence de l'activité de transformation de l'acide hydrogenobyrique a,c-diamide en acide cobyrique a,c-diamide.

Cet exemple illustre la mise en évidence d'une activité enzymatique directement liée à la voie de biosynthèse des cobalamines qui n'a jusqu'à ce jour jamais été décrite. Il s'agit de l'activité de transformation de l'acide hydrogenobyrique a,c-diamide en acide cobyrique a,c-diamide.

Cette activité est mise en évidence, entre autres, par l'expérience typique suivante. Un extrait brut de la souche SC510 Rif^r est obtenu par sonication de 10 g de cellules humides dans 20 ml de tampon Tris/HCl 0.2 M pH 8.0, puis élimination des débris cellulaires par centrifugation durant 1 h à 50 000 x g. 1000 mg de protéines de cet extrait sont incubés 1 h à 30 °C avec de l'acide hydrogénobyrique diamide marqué au carbone 14 (32 nmol; 50 µCi/µmol) dans 40 ml de tampon Tris/HCl 0.2 M pH 8.0 contenant 7 mM ATP, 200 µM CoCl₂. La réaction est arrêtée par addition de 7.5 ml KH₂PO₄ 1M et 6 ml KCN 0.3 M suivi du chauffage pendant 10 min à 80 °C. Après centrifugation à 15000 x g durant 15 min, l'analyse CLHP du surnageant montre: (1) la formation durant l'incubation de 19.2 nmol d'acide cobyrique a,c-diamide ayant la même radioactivité spécifique que l'acide hydrogénobyrique a,c-diamide de départ et (2) la disparition

d'une quantité correspondante de ce dernier. Pour confirmer qu'il s'agit bien d'acide cobyrinique a,c-diamide, le produit est purifié par CLHP puis estérifié dans du méthanol contenant 5% d'acide sulfurique (18h, 20°C). L'authenticité de l'ester pentaméthylrique de l'acide cobyrinique a,c-diamide produit est démontré par TLC (par rapport à un échantillon de référence) et spectrométrie de masse. Notons que dans des conditions similaires d'incubation où le marquage radioactif est introduit non pas dans l'acide hydrogénobyrrinique a,c-diamide mais dans le cobalt (en utilisant le cobalt 57), de l'acide cobyrinique a,c-diamide marqué au cobalt 57 est biosynthétisé et les mêmes conclusions ont pu être tirées. L'acide hydrogénobyrrinique a,c-diamide marqué au carbone 14 est obtenu de la manière suivante: l'acide hydrogénobyrrinique est biosynthétisé in vitro en utilisant le [methyl-14C]SAM puis transformé en acide hydrogénobyrrinique a,c-diamide et purifié par CLHP comme décrit à l'exemple 6.1.2.

Cette étude démontre que l'insertion du cobalt a lieu au niveau de l'acide hydrogénobyrrinique a,c-diamide chez *P.denitrificans*. Dans les conditions décrites, l'acide hydrogénobyrrinique n'est pas substrat de la chélation enzymatique par le cobalt.

b) Dosage et purification d'une protéine de la souche SC510 Rif^r impliquée dans la transformation de l'acide hydrogenobyrrinique a,c-diamide en acide cobyrinique a,c-diamide.

La fraction à doser (0.5 à 2 unités) est incubée pendant 60 min à 30 °C avec 50 µl d'extrait brut de la souche SC510 Rif^r obtenu comme décrit ci-dessus, 7 mM ATP, 200 µM CoCl₂, 7 µM acide hydrogénobyrrinique a,c-diamide marqué au carbone 14 (50 µCi/µmol) dans 400 µl de tampon Tris/HCL 0.1 M pH 8.0. La réaction est arrêtée par addition de 75 µl KH₂PO₄ 1 M et 60 µl KCN 0.3 M, puis chauffage pendant 10 min à 80°C. Après centrifugation à 15000 x g durant 15 min, le surnageant est analysé par CLHP, pour quantifier l'acide cobyrinique a,c-diamide formé (Cf exemple 9). L'unité d'activité enzymatique est définie comme la quantité d'enzyme nécessaire pour

généraler 1 nmol d'acide cobyrique a,c-diamide par h dans ces conditions. Dans ces conditions, il apparaît que les extraits de la souche SC510 Rifr dans laquelle on a introduit le plasmide pXL1909 (voir exemple 4.5.2) présentent une activité entre 20 et 50 fois plus élevée que les extraits de la souche SC510 Rifr. C'est sur cette base qu'est purifiée une protéine, seule responsable de cette amplification d'activité.

Dans une expérience typique de purification, 10 g de cellules humides de la souche SC510 Rifr dans laquelle on a introduit le plasmide pXL1909 sont suspendues dans 20 ml de tampon Tris/HCl 0.2 M pH 8.0 et soniquées durant 30 min à 4°C. L'extrait brut est ensuite récupéré par centrifugation durant 1 h à 50 000 x g et dessalé sur colonnes PD10 (Pharmacia) équilibrées avec du tampon 0.1 M Tris/HCl pH 8.0 (tampon A). La solution protéique est ensuite fractionnée (213 mg de protéines à chaque chromatographie) sur une colonne Mono Q HR 10/10 (Pharmacia) à l'aide d'un gradient de 0 à 0.5 M KCl dans le tampon A, puis les fractions contenant l'activité sont poolées, concentrées par ultrafiltration, dessalées sur colonnes PD10 (Pharmacia) équilibrées avec le tampon 0.1 M Tris/HCl pH 7.2 (tampon B), et chromatographiées sur une colonne Mono Q HR 10/10 (Pharmacia) à l'aide d'un gradient de 0 à 0.5 M KCl dans le tampon B. Les fractions contenant l'activité sont poolées, concentrées par ultrafiltration, dessalées sur colonnes PD10 (Pharmacia) équilibrées avec le tampon B, et chromatographiées sur une colonne Mono Q HR 5/5 (Pharmacia) à l'aide d'un gradient de 0 à 0.5 M KCl dans le tampon B. Pour parfaire la purification, la protéine est finalement chromatographiée sur une colonne Bio-Sil 250 (Bio-Rad) éluée avec 20 mM potassium phosphate-50 mM sodium sulfate, pH 6.8. Après cette étape, l'enzyme est pure à plus de 95%. Elle ne présente aucune protéine contaminante en PAGE-SDS. Son poids moléculaire dans cette technique est de 135 000. Ce degré de pureté est confirmé par l'unicité de la séquence NH2-terminale. Les différentes étapes de purification de la protéine de la souche SC510 Rifr impliquée dans la transformation de l'acide hydrogenobyrique a,c-diamide en acide

cobyrinique a,c-diamide, avec leur facteur de purification et leur rendement sont portées sur le tableau ci-dessous.

5 Tableau: Purification d'une protéine de la souche SC510 Rifr impliquée dans la transformation de l'acide hydrogenobyrinique a,c-diamide en acide cobyrinique a,c-diamide.

10		Etape de purification	vol (ml)	Protéines (mg)	Activité spé. (u/mg de protéines)	Rendement	Facteur de purification ¹
		Extrait brut	31.5	1278	0.23	-	-
		MonoQ 10/10	44	79.2	2.4	64	10
		MonoQ 10/10	21	33.6	6.8	78	30
		MonoQ 5/5	3	6.6	16.0	36	70
15		Bio-Sil 250	1.8	5.9	16.3	33	71

c) Séquence NH2-terminale de la protéine impliquée dans la transformation de l'acide hydrogenobyrinique a,c-diamide en acide cobyrinique a,c-diamide de Pseudomonas denitrificans et identification du gène de structure de Pseudomonas denitrificans
20 codant pour cette activité.

La séquence NH2-terminale de cette protéine purifiée comme décrit à l'exemple 6.1.10b) a été déterminée comme décrit précédemment. Six résidus ont été identifiés:

His-Leu-Leu-Leu-Ala-Gln

25 La séquence NH2-terminale de la protéine COBN (figure 47) correspond exactement à la séquence NH2-terminale de la protéine purifiée, si ce n'est que dans la séquence présentée sur la figure 47, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la méthionine aminotermi-

est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids moléculaire de la protéine purifiée est estimé par électrophorèse PAGE-SDS à 135 000. La protéine COBN a un poids moléculaire déduit de sa séquence de 138 000 (figure 5 47). Les correspondances entre les séquences NH₂-terminales et les poids moléculaires indiquent clairement que la protéine COBN correspond à la protéine impliquée dans la transformation de l'acide hydrogenobyrique a,c-diamide en acide cobyrique a,c-diamide. Le gène cobN est donc le gène de structure de cette protéine.

10 6.1.11. Identification de la protéine COBP codée par le gène CobP.

a) Dosage de l'activité cobinamide kinase.

Cet exemple illustre le dosage d'une activité enzymatique de la voie de biosynthèse des cobalamines qui n'a jusqu'à ce jour jamais été
15 étudiée. Il s'agit de l'activité cobinamide kinase. Elle catalyse la phosphorylation ATP-dépendante du groupement hydroxyl du résidu (R)-1-amino-2-propanol du Ado-cobinamide pour générer le cobinamide phosphate.

La fraction à doser est incubée à l'obscurité à 30°C pendant 60 min
20 dans 500 µl de tampon 0.1 M Tris-HCl pH 8.8 contenant 1 mM EDTA, 1 mM ATP, 2.5 mM MgCl₂, et 16 µM Ado-cobinamide (Blanche et al., 1989). La réaction est stoppée en ajoutant 500 µl d'une solution aqueuse 20 mM de cyanure de potassium. Après chauffage à 80°C durant 10 min et centrifugation à 5000 x g durant 10 min, le cobinamide phosphate
25 formé présent dans le surnageant est dosé par CLHP (Cf exemple 9) en utilisant le gradient linéaire simplifié suivant: de 25% à 30% de B dans A en 15 min, puis 30% à 100% de B en 12 min, et 3 min à 100% de B.

L'unité d'activité est définie comme la quantité d'enzyme nécessaire
30 pour générer 1 nmol de cobinamide phosphate à partir du cobinamide par h dans ces conditions.

b) Dosage de l'activité cobinamide phosphate guanylyltransferase.

Cet exemple illustre le dosage d'une activité enzymatique de la voie de biosynthèse des cobalamines qui n'a jusqu'à ce jour jamais été étudiée. Il s'agit de l'activité cobinamide phosphate
5 guanylyltransferase. Elle catalyse l'addition de la partie GMP d'une molécule de GTP sur l'Ado-cobinamide phosphate, générant ainsi une molécule de GDP-cobinamide et libérant une molécule de pyrophosphate.

Cette activité est dosée dans les mêmes conditions que la cobinamide
10 kinase si ce n'est que l'Ado-cobinamide phosphate (16 μ M) (Blanche et al., 1989) et le GTP (2 mM) remplacent l'Ado-cobinamide et l'ATP respectivement, durant l'incubation.

L'unité d'activité est définie comme la quantité d'enzyme nécessaire pour générer 1 nmol de GDP-cobinamide à partir du cobinamide
15 phosphate par h dans ces conditions.

c) Purification de la cobinamide kinase.

A partir du dosage décrit à l'exemple 6.1.11a), la purification de la cobinamide kinase de *Pseudomonas denitrificans* est réalisée comme décrit ci-dessous.

20 Dans une expérience typique de purification, 5 g de cellules humides de SC510 Rifr dans laquelle on a introduit le plasmide pXL623 (voir exemple 4.5.2) sont soniqués dans 20 ml de tampon Tris 0.1 M pH 7.6 (tampon A). Après centrifugation (50 000 x g durant 1 h) et dialyse 4 h contre le tampon A, 4.5 ml du rétentat sont injectés sur Mono Q HR 10/10
25 (Pharmacia), équilibrée avec le tampon A. Les protéines sont éluées avec un gradient linéaire de 0.4 M NaCl et les fractions actives regroupées sont passées à travers une colonne de PD-10 (Pharmacia) équilibrée dans du Tris-HCl 30 mM-phosphate de potassium 5 mM-chlorure de calcium 5 μ M pH 7.6 (tampon B). La solution protéique
30 est fractionnée sur une colonne Bio-Gel HPHT (Bio-Rad), équilibrée dans le tampon B et éluée avec un gradient de 5 à 350 mM de phosphate de

potassium. Les fractions actives sont regroupées et amenées à 500 mM en sulfate d'ammonium, puis fractionnées sur une Phényl-Superose HR 5/5 (Pharmacia), éluée avec un gradient décroissant de sulfate d'ammonium. la fraction contenant l'activité est finalement
5 repurifiée sur une colonne Mono Q HR 5/5 dans le Tris-HCl à pH 7.3. Après cette étape, la protéine est pure à plus de 97%. Elle ne présente aucune protéine contaminante en PAGE-SDS. Cette pureté est confirmée par l'unicité de la séquence NH₂-terminale. Son poids moléculaire dans cette technique est de 20 000. Les différentes
10 étapes de purification de la cobinamide kinase avec leur facteur de purification et leur rendement sont portées sur le tableau A.

Les fractions contenant l'activité cobinamide kinase présentent aussi l'activité cobinamide phosphate guanylyltransferase. D'autre part, comme le montrent les résultats présentés dans le tableau ci-dessus, le
15 ratio de ces deux activités reste constant dans les fractions tout au long de la purification. Enfin, la protéine purifiée présente un très haut degré de pureté dépassant les 97%. L'ensemble de ces résultats indiquent donc sans ambiguïté qu'une seule et même protéine est responsable des deux activités successives que sont la
20 cobinamide kinase et la cobinamide phosphate guanylyltransferase de la voie de biosynthèse des cobalamines chez *Pseudomonas denitrificans*.

d) Séquence NH₂-terminale de la cobinamide kinase-cobinamide phosphate guanylyltransferase de *Pseudomonas denitrificans* et
25 identification du gène de structure de *Pseudomonas denitrificans* codant pour cette activité.

La séquence NH₂-terminale de la cobinamide kinase-cobinamide phosphate guanylyltransferase de *Pseudomonas denitrificans* a été déterminée comme décrit précédemment. Dix résidus ont été identifiés:

30 Ser-Ser-Leu-Ser-Ala-Gly-Pro-Val-Leu-Val

La séquence NH₂-terminale de la protéine COBP (figure 47) correspond exactement à cette séquence, si ce n'est que dans la séquence présentée sur la figure 47, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la
5 méthionine aminoterminal est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids moléculaire de la cobinamide kinase-cobinamide phosphate guanylyltransferase purifiée est estimé par électrophorèse PAGE-SDS à 20 000. La protéine COBP a un poids moléculaire déduit de sa
10 séquence de 19 500 (figure 47). Les correspondances entre les séquences NH₂-terminales et les poids moléculaires indiquent clairement que la protéine COBP correspond à la cobinamide kinase-cobinamide phosphate guanylyltransferase. Le gène cobP est le
15 gène de structure de la cobinamide kinase-cobinamide phosphate guanylyltransferase.

6.2 - Détermination des propriétés de protéines COB par mesures d'intermédiaires de biosynthèse accumulés

Cet exemple illustre comment il est possible d'attribuer à une protéine COB de Pseudomonas denitrificans une activité
20 enzymatique. Cette activité est attribuée d'après des données obtenues concernant les intermédiaires de biosynthèse accumulés chez le ou les mutants Cob bloqués dans l'étape en question. En effet si un mutant accumule un intermédiaire de biosynthèse, il est très probable que ce mutant est bloqué dans l'étape qui a pour substrat l'intermédiaire en
25 question.

6.2.1. Propriétés des protéines COBC et COBD

Les mutants Cob G643 (Agrobacterium tumefaciens) et G572 (Pseudomonas putida) déjà décrits dans les exemples 1 et 4 sont bloqués dans l'étape correspondant à la protéine COBC. En effet ces
30 deux mutants ne sont pas complémentés par les insertions inactivantes de transposons Tn₅ qui se trouvent dans le gène cobC. Les deux souches G643 et G572, ainsi que les souches parentes non mutées [C58-C9 Rif^r et KT 2440 Rif^r (Cameron et al., 1989)] ont été cultivées en milieu

PS4' pour les *A.tumefaciens*, et PS4'' pour les *P.putida* (PS4' et PS4'' correspondent à du milieu PS4 contenant respectivement 100 et 1000 fois moins de cobalt que le PS4 décrit précédemment) 3 jours comme cela est décrit ci-dessus. Du $^{57}\text{CoCl}_2$ a été ajouté aux cultures (2,5 $\mu\text{Ci}/0,1 \mu\text{m}$ pour une culture de 25 ml). Les corrinoïdes intracellulaires ont été isolés sous leur forme native et identifiés par leur comportement CLHP. Les souches parentes n'accumulent pas de corrinoïdes autres que du coenzyme B_{12} . Les deux mutants, G643 et G572, accumulent de l'acide cobyrique adénosylé dans des proportions respectives de 11 % et 6 %. Ces proportions en % sont calculées par rapport au niveau de coenzyme B_{12} synthétisé par la souche parente. Outre l'acide cobyrique, le mutant G643 accumule de l'acide cobyrique pentaamide dans une proportion de 2 %; l'acide cobyrique pentaamide est l'intermédiaire qui précède l'acide cobyrique. L'étude de ces mutants fait ressortir qu'ils sont bloqués après l'acide cobyrique. Tous ces mutants Cob sont soit bloqués entre l'uro'gen III et le cobinamide, soit entre le cobinamide et les cobalamines. Les mutants G643 et G572 sont bloqués entre l'uro'gen III et le cobinamide. Or si ces mutants sont bloqués avant le cobinamide et accumulent tous les deux de l'acide cobyrique, les protéines pour lesquelles ils codent ne peuvent intervenir que dans l'étape enzymatique (appelée cobinamide synthase) qui catalyse l'amidation de l'acide cobyrique par un résidu aminopropanol pour donner le cobinamide; elles peuvent aussi éventuellement intervenir dans la synthèse du substrat de la réaction qui apporte l'aminopropanol, si ce n'est l'aminopropanol lui-même. Le gène cobC code pour une protéine qui est soit la cobinamide synthase, soit une de ses sous-unités.

Le mutant Cob G634 d'*Agrobacterium tumefaciens* qui est bloqué dans l'étape correspondant au gène cobD a été analysé de la même manière. Ce mutant n'est pas complémenté par les insertions inactivationnelles dans le gène cobD (exemple 4.1). Le seul corrinoïde intracellulaire trouvé chez ce mutant est de l'acide cobyrique adénosylé. Comme les mutants précédents, ce mutant code pour une protéine intervenant dans la transformation de l'acide cobyrique en cobinamide ou bien éventuellement dans la synthèse de l'autre substrat

de la réaction.

Ces deux gènes différents (cobC et cobD) codent pour deux protéines qui interviennent dans la même étape.

6.2.2. Propriétés des protéines COBF à COBM

- 5 Les mutants d'Agrobacterium tumefaciens déjà décrits dont on connaît, d'après l'étude décrite à l'exemple 4.2, dans quels gènes chacun d'entre eux est bloqué, ont été étudiés. Ce sont les mutants : G612 (cobF), G615 (cobG), G616 (cobH), G613 (cobI), G611 (cobJ), G620 (cobK), G638 (cobL) et G609 (cobM) ; nous avons figuré entre
- 10 parenthèses le gène de Pseudomonas denitrificans responsable de la complémentation de ces mutants (exemple 5) qui correspond donc au gène muté chez ce mutant. Ces mutants ont été cultivés en milieu PS4 comme cela est décrit précédemment avec du cobalt marqué. Après quatre jours d'incubation, les mutants ont été analysés pour leur contenu
- 15 intracellulaire en corrinoïdes et descobaltocorrinoïdes (Cf exemples 6.1.2 et 9).

Tableau : Intermédiaires accumulés par des mutants d'Agrobacterium tumefaciens bloqués dans les gènes du fragment 8,7 kb de Pseudomonas denitrificans

p Souches	Descobaltocorrinoïdes intracellulaires en % ¹			Corrinoïdes intracellulaires en % des cobalamines	Gène muté
	HAB	HABM	HABD		
5	C58-C9*	100	100	100	-
	G612	< 5	< 5	64	<u>cobF</u>
	G615	< 5	< 5	84	<u>cobG</u>
	G616	35	< 10	< 10	<u>cobH</u>
	G613	< 5	< 5	57	<u>cobI</u>
10	G611	< 5	< 5	65	<u>cobJ</u>
	G620	12	< 5	< 10	<u>cobK</u>
	G638	< 5	< 5	47	<u>cobL</u>
	G609	< 5	< 5	33	<u>cobM</u>

HAB : acide hydrogénobyrique
HABM : acide hydrogénobyrique monoamide
15 HABD : acide hydrogénobyrique diamide

* en fait il s'agit de la souche C58-C9 Rif^rNal^r déjà décrite (Cameron et al., 1989)

¹les valeurs sont exprimées en % des mêmes intermédiaires accumulés chez la souche parente non mutée C58-C9 Rif^rNal^r.

20 Ces résultats montrent que tous les mutants n'accumulent aucun corrinoïde (à l'exception du mutant inactivé dans le gène cobF, G612, qui lui, accumule du cobinamide mais à un niveau faible équivalent à 2,2 % des cobalamines synthétisées par la souche non mutée). Toutefois, certains mutants (G612, G615 et G616) ont des
25 niveaux de cobalamines qui représentent plus de 10 % du niveau de cobalamines de la souche parente. Il est probable que tous ces mutants sont bloqués au moins avant l'acide cobyrique diamide. Tous ces mutants accumulent de l'acide hydrogénobyrique et de l'acide hydrogénobyrique diamide en quantités inférieures à la souche non

mutée; ils sont donc très probablement bloqués avant l'acide hydrogénobyrique. Nous pouvons conclure que tous les gènes cobF à cobG codent pour des protéines qui interviennent avant l'acide hydrogénobyrique. On sait que le mutant G613 est muté dans le gène cobI qui code pour la SP₂MT, intervenant bien avant l'acide hydrogénobyrique. Pour ce mutant, les résultats du présent exemple, concernant l'accumulation d'intermédiaires, sont en parfait accord avec l'étape inactivée chez ce mutant, à savoir que ce mutant n'accumule aucun intermédiaire après l'acide hydrogénobyrique à un niveau supérieur à ce que l'on observe avec la souche non mutée. Ce résultat est pour les gènes cobF, cobJ, cobL et cobM cohérent avec ceux de l'exemple 6.4 où il est proposé que ces gènes codent pour des protéines qui catalysent des transferts de méthyl SAM dépendants et donc qui interviennent avant l'acide hydrogénobyrique. A l'exception de cobI qui est le gène de structure de la SP₂MT ces gènes interviennent après le précorrène-3. En effet comme ce ne sont ni les gènes de structure de la SUMT et de la SP₂MT ils interviennent forcément après, c'est-à-dire après le précorrène-3 (tous les gènes cob décrits dans la présente invention interviennent entre l'uro'gène III et les cobalamines). Ces gènes cobF à cobH et cobJ à cobM codent pour des enzymes qui interviennent entre le précorrène-3 et l'acide hydrogénobyrique.

6.2.3. Propriétés des protéines COBS et COBT

Le mutant G2035 décrit dans les exemples 1 et 4.3 est bloqué dans l'étape correspondant à la protéine COBS. Le mutant G2037 décrit dans l'exemple 1 est bloqué dans l'étape correspondant à la protéine COBT. Ces souches ainsi que la souche parente (*Agrobacterium tumefaciens* C58C9Rifr) sont cultivées en milieu PS4' (il s'agit du milieu PS4 où la concentration en chlorure de cobalt est 100 fois plus faible que dans le milieu PS4) en présence de cobalt radioactif ⁵⁷CoCl₂ pendant 3 jours, leur contenu intracellulaire en descobaltocorrinoïdes est analysé ainsi que le contenu en corrinoïdes, comme cela a déjà été décrit ci-dessus (Cf exemple 6.2.2). Les souches

G2035 et G2037 n'accumulent pas de corrinoïdes et de larges concentrations (supérieures à celles observées avec la souche parente) en acide hydrogénobyrrinique et acide hydrogénobyrrinique mono et diamide ne sont présentes qu'avec la souche G2035. Ce mutant est
5 probablement bloqué dans une étape située après l'acide hydrogénobyrrinique diamide et avant l'acide cobyrinique diamide. Par conséquent le gène cobS coderait pour une des enzymes impliquées dans la transformation de l'acide hydrogénobyrrinique diamide en acide cobyrinique diamide; cette protéine peut donc intervenir soit dans
10 l'insertion du cobalt, soit dans la réduction du cobalt de l'acide cobyrinique a,c-diamide non adénosylé. Par contre, le mutant G2037 serait bloqué dans une étape située en amont de l'acide hydrogénobyrrinique. Le gène cobT coderait pour une protéine impliquée dans une étape enzymatique en amont de l'acide hydrogénobyrrinique et
15 en aval du précorrène-3 (d'autres gènes de structure codant pour les enzymes impliqués en aval du précorrène-3 ont déjà été identifiés). Une autre possibilité pour la protéine COBT est qu'elle intervienne, comme cela est proposé à l'exemple 5, comme protéine liant le cobalt, et/ou comme protéine qui interagit avec d'autre(s) protéine(s) par sa
20 partie acide.

6.2.4. Propriétés de la protéine COBV

Les mutants G2039 et G2040 décrits dans les exemples 1 et 4.4 sont bloqués dans l'étape correspondant à la protéine COBV. Ces souches ainsi que la souche parente sont cultivées en milieu PS4'
25 pendant 3 jours en présence de cobalt radioactif $^{57}\text{CoCl}_2$, puis leur contenu intracellulaire en descobaltocorrinoïdes est analysé et le contenu en corrinoïdes est déterminé, comme cela est décrit dans l'exemple 9. Les souches G2039 et G2040 accumulent de l'acide cobyrinique, du cobinamide, du cobinamide phosphate et du GDP-cobinamide.
30 Ces mutants sont probablement bloqués dans une étape enzymatique en aval de la GDP-cobinamide. Le gène cobV coderait pour une enzyme impliquée dans la transformation de la GDP-cobinamide en cobalamine, voir figure 5. Ce résultat est en parfait accord avec l'activité

cobalamin (5'-phosphate) synthase de la protéine COBV qui possède l'Ado-GDP-cobinamide comme substrat.

6.3 - Détermination de l'activité de protéines COB par études d'affinité vis-à-vis du SAM

5 Cet exemple illustre comment il est possible, à partir de protéines COB purifiées de Pseudomonas denitrificans, de mettre en évidence in vitro une activité de fixation du SAM. Si une protéine COB possède une telle activité, cela signifie que cette protéine COB est une méthyltransférase de la voie et qu'elle intervient dans un des
10 transferts des huit groupements méthyl qui se produisent entre l'uro'gen III et l'acide cobyrinique.

6.3.1. Test d'affinité pour le SAM sur une protéine purifiée

15 Le test repose sur le principe suivant lequel les méthyltransférases de la voie de biosynthèse des cobalamines ont certainement un site de fixation du SAM. Ce site doit être mis en évidence par une affinité du SAM plus élevée que pour toute protéine qui ne fixe pas le SAM spécifiquement. Après incubation de la protéine à étudier en présence d'un excès de SAM radioactif, celle-ci est
20 séparée du SAM libre par une chromatographie de perméation de gel. La radioactivité retrouvée dans la fraction du poids moléculaire de la protéine correspond au SAM fixé durant l'incubation. La chromatographie est effectuée à 2°C pour limiter au maximum la libération du SAM lié durant la séparation.

25 La protéine (environ 10 µg) est incubée durant 10 minutes à 30°C dans 200 µl de Tris-HCl 0,1 M pH 7,7 avec 5 nmoles de [méthyl-³H]-SAM (1 µCi). Après incubation, 100 µl du mélange sont immédiatement injectés sur une colonne de TSK-125 (Bio-Rad) éluée à 1 ml/minute avec le mélange de sulfate de sodium 50
30 mM/dihydrogénophosphate de sodium 20 mM pH 6,8 préconisé par le distributeur de cette colonne. Des fractions de 0,5 ml sont collectées et comptées en scintillation liquide. Les temps de rétention de la protéine et du SAM sont obtenus directement d'après l'enregistrement

de l'absorbance de l'éluat à 280 nm.

6.3.2. Etude in vitro de la fixation du SAM sur les protéines COBA et COBF de Pseudomonas denitrificans

a) Purification des protéines COBF et COBA

5 La protéine COBF de Pseudomonas denitrificans est purifiée comme cela est décrit ci-après. Dans une expérience typique de purification, 5 g de cellules humides de la souche SC150 Rif^r dans laquelle a été introduit le plasmide pXL1546 (voir ex.7.3) obtenues après culture en milieu PS4 sont resuspendues dans 30 ml de Tris-HCl
10 0,1 M pH 7,7 et soniquées durant 15 minutes à 4°C. L'extrait brut est ensuite récupéré par centrifugation 1 heure à 50 000 g et le surnageant est passé sur une colonne de DEAE-Sephadex (1 ml de gel) pour éliminer les composés tétrapyrroliques présents. 10 mg de
15 protéines (0,7 ml) de cet extrait sont ensuite injectés sur une colonne de MonoQ HR 5/5 équilibrée avec le même tampon. Les protéines sont éluées avec un gradient linéaire de KCl (0 à 0,25 M). La protéine COBF est éluee avec 0,20 M de KCl. Elle est diluée deux fois avec du Tris-HCl 0,1 M pH 7,7 et purifiée une seconde fois sur une MonoQ HR 5/5. L'électrophorèse en PAGE-SDS avec révélation au bleu de Comassie
20 est utilisée pour visualiser la protéine. Cette technique montre d'autre part que COBF est d'une pureté de 95 % environ après cette étape de purification. La séquence NH₂-terminale de la protéine purifiée a été réalisée comme cela est décrit précédemment . Deux séquences NH₂-terminales apparaissent en même temps sur chaque cycle
25 de dégradation; ce sont les séquences suivantes dans les proportions indiquées :

Séquence 1 (34 % d'abondance)

1	2	3	4	5	6	7	8	9	10	11
Ala	Glu	Ala	Gly	Met	Arg	Lys	Ile	Leu	Ile	Ile

30 Séquence 2 (66 % d'abondance)

1	2	3	4	5	6	7	8	9	10	11
Met	Arg	Lys	Ile	Leu	Ile	Ile	Gly	Ile	Gly	Ser

La séquence 1 correspond à la séquence NH₂-terminale de la protéine COBF qui est donnée sur la figure 16, si ce n'est que la

méthionine aminoterminal est excisée suivant des règles déjà énoncées (Hirel et al., 1989) par la méthionine aminopeptidase (Ben Bassat et Bauer, 1989). La séquence 2, présente en quantité la plus importante, correspond à la même protéine mais dont l'initiation de traduction se
5 serait faite non pas au codon ATG d'initiation de la traduction que nous avons supposé, mais à celui situé 5 codons en aval sur la phase codante (figure 16). En effet les acides aminés de cette séquence sont exactement ceux que l'on trouve sur la séquence de la protéine COBF à partir de la deuxième méthionine (acide aminé n° 6) de cette séquence
10 (figure 16). Dans ce cas la méthionine aminoterminal n'est pas excisée ce qui confirme les règles déjà énoncées (Hirel et al., 1989). Il y a chez la souche SC510 Rif^r portant le plasmide pXL1546, deux initiations de traduction, d'une part celle correspondant au codon méthionine positionné à la bonne distance, sur notre construction, de
15 la séquence de Shine et Delgarno et d'autre part celle qui se fait au deuxième codon méthionine se trouvant sur la séquence du gène cobF présenté à la figure 16. De cela, il ressort que probablement la protéine COBF commence non pas à la méthionine indiquée sur la figure 16 mais à celle se trouvant 5 acides aminés plus loin.

20 De toute manière, ce résultat montre qu'il s'agit bien de la protéine COBF qui est exprimée et que celle-ci est exprimée avec une forme allongée de 4 acides aminés. Lors de la purification les deux formes protéiques sont purifiées. Dans cet exemple, nous appellerons protéine COBF purifiée le mélange de ces deux protéines
25 purifiées.

La protéine COBA de Pseudomonas denitrificans est purifiée comme cela est décrit précédemment (Blanche et al., 1989).

b) Fixation du SAM

La fixation du SAM sur ces deux protéines est étudiée
30 comme cela est décrit précédemment à l'exemple 6.3.1.a). La sérum albumine bovine et la protéine COBH purifiée sont utilisées comme contrôles négatifs. Pour les protéines COBA et COBF on observe un pic de radioactivité à la sortie de la colonne de TSK-125 au temps de sortie de ces protéines (figure 20). Dans ce test, la protéine COBI

présente la même propriété de fixation du SAM. A l'opposé, il n'y a pas de tels pics de radioactivité avec la BSA et la protéine COBH. Ce test met en évidence la fixation in vitro du SAM sur les protéines COBA, COBI et COBF. Ces résultats montrent que COBA, COBI et COBF sont des SAM-méthyltransférases. Ce résultat est tout à fait en accord avec les activités de COBA et COBI puisqu'il s'agit respectivement de la SUMT et de la SP₂MT de Pseudomonas denitrificans. La protéine COBF est donc probablement une SAM méthyltransférase de la voie de biosynthèse des cobalamines. Ce test confirme que COBF est une méthyltransférase.

10 6.4 - Détermination de l'activité de protéines COB par études d'homologies de séquences

Cet exemple illustre comment par des comparaisons entre les séquences de diverses protéines COB de Pseudomonas denitrificans il est possible de trouver les protéines COB qui sont des SAM-méthyltransférases de la voie de biosynthèse des cobalamines.

15 Les protéines COBI et COBA sont toutes les deux des SAM méthyltransférases de la voie de biosynthèse. Ces deux protéines ont été comparées suivant le programme de Kanehisa, 1984. Cette comparaison fait ressortir trois régions de forte homologie (figure 20 21). Dans chacune de ces régions il y a plus de 45 % d'homologie stricte entre les deux protéines. Trois régions de forte homologie entre COBA et CYSG sont également présentées (figure 22); ce sont les mêmes régions de COBA qui présentent une forte homologie avec COBI. Ces régions de fortes homologies entre COBA, CYSG et COBI présentent 25 de l'homologie avec d'autres protéines COB. Il s'agit des protéines COBF, COBJ, COBL et COBM (figure 23). En ce qui concerne la région 1, les protéines COBF, COBL et COBM présentent des homologies significatives par rapport à toutes les protéines de Genpro, qui est une extraction protéique de Genbank (version 59) augmentée des parties 30 codantes putatives supérieures à 200 acides aminés, suivant le programme de Kanehisa (1984). En ce qui concerne la région 2, les protéines COBJ, COBL et COBM présentent des homologies significatives par rapport à toutes les protéines de Genpro (version 59). En ce qui concerne la troisième région d'homologie, COBJ, COBL et COBM

présentent des homologies significatives par rapport à toutes les protéines de Genpro (version 59). Les comparaisons de séquences permettent donc de mettre en évidence que quatre protéines COBF, COBJ, COBL et COBM présentent des homologies significatives avec les régions conservées des séquences de trois types de méthyltransférases COBA, COBI et COBF. Les protéines COBG, COBH et COBK ne présentent pas d'homologies significatives avec les régions conservées des méthylases. La protéine COBF ne présente d'homologie significative avec les autres protéines que dans la région 1. Ces homologies doivent correspondre probablement au fait que toutes ces protéines sont des méthyltransférases. Ce résultat recoupe les données biochimiques décrites sur COBF concernant la capacité qu'a cette protéine à lier le SAM in vitro (exemple 6.3). Ces homologies, d'une part, permettent de confirmer que COBF est une SAM méthyltransférase de la voie de biosynthèse des cobalamines et d'autre part mettent en évidence que COBJ, COBL et COBM pourraient être des SAM méthyltransférases de la voie de biosynthèse des cobalamines. Ces résultats montrent également l'homologie qui existe entre les protéines COB de P.denitrificans et les protéines isofonctionnelles d'autres microorganismes.

20 **EXEMPLE 6(B) - Purification et clonage du gène de structure de la SUMT de Methanobacterium ivanovii.**

Cet exemple illustre comment il est possible d'obtenir, dans d'autres microorganismes, des enzymes COB et des gènes cob correspondants à ceux identifiés chez P. denitrificans.

25 **6(B).1. Purification de la SUMT de Methanobacterium ivanovii.**

Cet exemple décrit la purification de la SUMT de Methanobacterium ivanovii et l'étude de ses propriétés catalytiques. La souche de Methanobacterium ivanovii DSM2611 est cultivée comme cela est décrit (Souillard et al., 1988). 12 g de cellules humides sont obtenus. Celles ci sont resuspendues dans 80 ml de tampon Tris/HCl 0.1 M pH 7.6 contenant 5 mM DTT et 1 mM EDTA et soniquées pendant 1h30 à 4 °C puis centrifugées durant 1h à 50 000 g. L'extrait est ensuite débarrassé de composés tétrapyrroliques libres par passage à travers une petite

colonne de DEAE-Sephadex A25 montée dans le même tampon. Les protéines précipitant entre 55 et 75% de saturation en sulfate d'ammonium sont solubilisées dans un tampon Tris/HCl 0.1M pH 7.5, 0.5 mM DTT, 1.7 M sulfate d'ammonium et injectées sur une colonne de Phényl-Superose HR 5 10/10 (Pharmacia France/SA), éluée avec un gradient décroissant (de 1.7 M à 0 M en sulfate d'ammonium). Les fractions actives sont passées sur une colonne de Séphadex G-25 équilibrée avec du tampon Tris/HCl 0.1 M, pH 7.5, 0.5 mM DTT, glycérol 25% (tampon A) puis injectées sur une Mono Q HR 5/5 (Pharmacia France SA) équilibrée avec du tampon A et éluée avec 10 un gradient de KCl de 0 à 0.3 M; cette étape est répétée une seconde fois dans les mêmes conditions. Une chromatographie de perméation de gel sur Bio-Sil TSK-250 (BioRad France SA) de la fraction active de l'étape précédente permet d'obtenir une protéine homogène en PAGE-SDS et en RP-CLHP (C-18 µBondapak). Les différentes étapes de purification avec 15 leur rendement ainsi que leur facteur de purification sont décrites sur le tableau ci-dessous.

Comme il est montré dans ce tableau, le facteur de purification total est de plus de 4 500. Des propriétés de l'enzyme pure ont été étudiées selon des méthodes déjà décrites (Blanche et al., 1989). Cette 20 enzyme a bien une activité SUMT, c'est à dire, qu'elle catalyse bien le transfert de deux groupements méthyls, SAM-dépendant, en C-2 et en C-7 de l'uro'gen III. Le poids moléculaire de l'enzyme estimé par perméation de gel est de 60,000 +/- 1,500 tandis que par PAGE-SDS il est de 29,000 ce qui montre clairement qu'il s'agit d'une enzyme homodimérique. Dans 25 des conditions déjà décrites (Blanche et al., 1989) l'enzyme à un Km pour l'uro'gen III de 52 +/- 8 nM. De plus, cette enzyme ne présente pas d'inhibition par son substrat à des concentrations inférieures à 20 µM, alors que la SUMT de *Pseudomonas denitrificans* présente une inhibition par l'uro'gen III à une concentration supérieure à 2 µM (Blanche et al., 30 1989).

Tableau : Purification de la SUMT de *M. ivanovii*.

5	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. (u/mg de protéines)	Rendement	Facteur de purification ¹
	Extrait brut	92	731	0.337	-	-
	55-75% SA	7.1	153	1.215	76	3.6
	Phényl-superoxe	9.5	8.34	15.35	52	46
	MonoQ 5/5	1.0	0.252	422	43	1252
10	Bio-Sil TSK	1.0	0.061	1537	38	4561

1/ calculé d'après le rendement en protéines.

La Vmax de la SUMT de *M. ivanovii* a été déterminée. Celle-ci est de 1537 U/mg de protéines. Cette valeur est supérieure à celle trouvée pour la SUMT de *P. denitrificans* déjà déterminée dans les conditions optimales de la réaction (compte tenu de son inhibition par l'uro'gen III), 489 U/mg de protéines (Blanche et al., 1989).

6(B).2. Clonage du gène de structure de la SUMT de *M. ivanovii* chez *E. coli*.

6(B).2.1. Clonage d'un fragment interne au gène de structure de la SUMT de *M. ivanovii*. Pour ce faire, on procède de la manière suivante : 200 picomoles de la SUMT de *M. ivanovii* sont utilisées pour le séquençage NH2-terminal de la protéine comme cela est décrit précédemment. En outre, un fragment peptidique obtenu par une digestion trypsique de la protéine est lui aussi soumis à un séquençage de sa partie NH2-terminale. Les séquences obtenues sont présentées sur la figure 48. Les oligonucléotides sens et antisens respectivement 946, 923 et 947 (voir figure 48) sont synthétisés comme cela est décrit précédemment; ces oligonucléotides contiennent à leur extrémité 5' un site de restriction qui est soit EcoRI pour les oligonucléotides sens soit HindIII pour l'oligonucléotide antisens. Ces oligonucléotides sont utilisés pour une expérience d'amplification enzymatique de l'ADN (Saiki et al., 1988), comme cela est schématisé sur la figure 48.B.

L'ADN génomique de *M. ivanovii* est préparé de la manière suivante: 0.4 g de cellules de *M. ivanovii* (DSM 2611) sont lavées avec une solution 0.15 M NaCl. Les cellules sont ensuite incubées dans 4 ml d'une solution de sucrose 25 %, Tris/HCl 50 mM pH 8, lysozyme 40 mg puis 2 à 3 h à 50°C après addition de 40 mg de protéinase K et de 5 ml d'une solution SDS 0.2 %, EDTA 0.1 M pH 8. L'ADN est ensuite extrait au phénol-chloroforme (50 %-50 %) 2 fois, puis 2 fois au chloroforme et ensuite précipité par l'isopropanol et repris dans 3 ml de TNE (Tris/HCl 10 mM pH 8, EDTA 1 mM, Na-Cl 100 mM).

10 L' Amplification enzymatique de l'ADN de *M. ivanovii* est effectuée suivant le protocole de Saiki et al., 1988, dans un volume de 0.1 ml avec 600 ng d'ADN génomique de *M. ivanovii*, en utilisant les amorces 946 et 947 (réaction 1) ou 923 et 947 (réaction 2). Le tampon utilisé pour cette réaction est MgCl₂ 1mM, KCl 50 mM, gélatine 0.001 %
15 et chaque dNTP à une concentration de 0.2 mM; pour chaque réaction d'amplification, 10 mg de chaque oligonucléotide sont utilisés ainsi que 2.5 unités de Taq DNA polymérase (Cetus Corporation). L'amplification est réalisée sur 30 cycles dans le Perkin-Elmer Cetus DNA Amplification system; au cours de chaque cycle, l'ADN est dénaturé 1 min à 95 °C, les
20 amorces oligonucléotidiques sont hybridées à l'ADN simple brin 2 min à 38 °C et les néobrans sont polymérisés pendant 3 min à 72 °C. Les produits d'amplification sont ensuite extraits au chloroforme et subissent ensuite une précipitation éthanolique; ils peuvent ensuite être visualisés après migration sur gel d'acrylamide puis être digérés
25 par les enzymes de restriction telles que EcoRI et HindIII.

Dans le cas de la réaction 1, deux fragments sont observés : à 615 pb ainsi qu'à 240 pb. En ce qui concerne la réaction 2, deux fragments sont aussi observés : à 630 et 170 pb. La totalité du produit d'une réaction d'amplification enzymatique entre les oligonucléotides
30 946-947 est séparée par migration sur gel d'acrylamide; le fragment de 615 pb est purifié comme cela est décrit précédemment. Ce fragment est ensuite digéré par EcoRI et HindIII afin de rendre les extrémités du fragment cohésives. Ce fragment est ensuite ligaturé avec de l'ADN de la forme répllicative du phage M13mp19. La ligature est transformée dans *E.*
35 *coli* TG1. Six clones recombinants, contenant un insert de 615 pb, sont

analysés par séquençage avec le primer universel -20 (Pharmacia SA, France). Comme il est montré sur la figure 49, lorsque l'on séquence l'ADN simple brin des phages recombinants qui contiennent l'insert de 615 pb, il doit être observé, en aval du site EcoRI, une séquence non
5 dégénérée correspondant à celle de l'oligonucléotide 946 suivie, dans la même phase, par une séquence qui code pour les acides aminés LITLKAVNVLK?ADVVL (? signifie qu'à cette position, le résidu n'a pas pu être déterminé); cette séquence correspond à celle qui, dans la séquence NH2-terminale de la SUMT, suit les acides aminés correspondant à
10 l'oligonucléotide 946 (voir figure 48). Pour deux clones, il a effectivement été observé, après le site EcoRI, une séquence pouvant coder pour la région NH2-terminale de la SUMT de *Methanobacterium ivanovii*, celle-ci commençant par l'enchaînement Pro-Gly-Asp-Pro-Glu-Leu qui sont les acides aminés codée par une séquence contenant
15 l'oligonucléotide 946. Cette observation montre que ces deux formes répliquatives recombinantes contiennent un insert qui correspond à un fragment interne au gène de structure de la SUMT de *Methanobacterium ivanovii*. La forme répliquative portant ce fragment interne au gène de structure de *M. ivanovii* est appelée pG10.

20 6(B).2.2. Clonage du gène de structure de la SUMT de *Methanobacterium ivanovii*.

L'ADN génomique de *Methanobacterium ivanovii* est digéré par plusieurs enzymes de restriction (digestions simples ou doubles). Après digestion, les fragments sont séparés par électrophorèse sur gel
25 d'agarose, puis ils sont transférés sur une membrane de nylon comme cela est décrit précédemment. Après dénaturation des fragments ainsi transférés et préhybridation, on procède à une hybridation avec la forme répliquative pG10 comme sonde marquée au 32P comme cela est décrit précédemment. Il est ainsi trouvé qu'un fragment issu d'une digestion
30 EcoRI-BglII de 3.2 kb de *Methanobacterium ivanovii* hybride avec la sonde (voir Figure 50). 40 µg d'ADN génomique de *M. ivanovii* sont ensuite digérés par EcoRI et BglII puis séparés par migration sur gel d'agarose. Les fragments ayant une taille comprise entre 3 et 3.5 kb sont électroélués comme cela est décrit précédemment. Les fragments ainsi

purifiés sont mis à ligaturer avec le vecteur pBKS+ (Stratagene Cloning Systems, La Jolla) digéré par BamHI-EcoRI. La ligature est transformée dans *E. coli* DH5 α (Gibco BRL). Les transformants sont sélectionnés sur milieu LB supplémenté avec de l'ampicilline et du X-gal. 800 colonies
5 blanches sont repiquées sur filtre; après croissance puis lyse des bactéries on procède à une hybridation sur colonies selon la technique de Gr $\ddot{u}$ nstein et Hogness (1975). La sonde utilisée est la forme réplivative pG10 marquée au 32P. Un seul clone positif après ce test d'hybridation avec la sonde est trouvé. L'ADN plasmidique de ce clone
10 est appelé pXL1809 (voir figure 56). Une digestion de cet ADN par EcoRI-XbaI permet de visualiser, comme attendu, un insert de 3.2 kb. Le plasmide pXL1809 est séquencé sur les 2 brins par la technique de Chen et Seeburg (1985). Une séquence de 955 bases est obtenue (figure 51). Une analyse des phases ouvertes nous conduit à identifier une phase
15 ouverte de la base 34 (ATG) à la base 729 (TGA). Cette phase ouverte code pour une protéine dont la séquence est présentée sur la figure 52. Cette protéine a un poids moléculaire de 24 900 (voir figure 53) ce qui est proche du poids moléculaire de la protéine purifiée à partir de *M. ivanovii*. La séquence NH₂-terminale de cette protéine est exactement
20 celle déterminée pour la SUMT de *M. ivanovii* purifiée (voir figure 48 et figure 52). Ces observations établissent sans ambiguïté que le gène cloné et séquencé est bien le gène de structure de la SUMT de *M. ivanovii*. Comme cette activité est supposée intervenir dans la biosynthèse des corrinoïdes chez toutes les bactéries, ce gène est
25 désigné par gène corA et la protéine codée par ce même gène protéine CORA. Le profil d'hydrophobicité de la protéine CORA de *M. ivanovii*, réalisé à partir du programme de Hopp et Woods (1981) montre qu'il s'agit, comme cela est attendu, d'une protéine hydrophile, comme cela est présenté sur la figure 54. La protéine CORA de *M. ivanovii* montre un
30 degré d'homologie stricte de plus de 40 % vis à vis de COBA de *P. denitrificans* (figure 53). Cette homologie s'étend sur presque la totalité des deux protéines puisqu'elle concerne les résidus 3 à 227 de CORA de *M. ivanovii* et les résidus 17 à 251 de COBA de *P. denitrificans*. Cette homologie reflète les homologies structurales qu'il existe entre
35 deux protéines catalysant la même réaction. Ce sont les mêmes régions

qui sont les plus conservées entre CORA et COBA de *P. denitrificans* que celle qui sont conservées entre COBA de *P. denitrificans* et CYSG d'E. coli (figures 22).

EXEMPLE 7 - Expression de protéines COB

7.1 - Expression chez *Pseudomonas denitrificans*

5 Cet exemple illustre que l'amplification d'un gène de structure d'une protéine COB de *Pseudomonas denitrificans* chez *Pseudomonas denitrificans* conduit à l'amplification de l'activité de la protéine COB.

7.1.1 - Expression de la protéine COBA

10 Le plasmide pXL557 correspond au plasmide pXL59 dans lequel le fragment BqlII-EcoRV (respectivement aux positions 80 et 2394 sur la séquence de la figure 7) de 2,4 kb du fragment de 5,4 kb a été cloné. Ce fragment contient les gènes cobA et cobE.

15 Le plasmide pXL545 contient uniquement le gène cobE. Sa construction a été décrite dans l'exemple 4.1.

20 Ces deux plasmides ont été introduits par transfert conjugatif chez SC510 Rif^r. Les souches SC510 Rif^r, SC510 Rif^r pXL59, SC510 Rif^r pXL557 et SC510 Rif^r pXL545 ont été cultivées en milieu PS4. A 4 jours, les cultures ont été arrêtées et les activités SUMT ont été dosées suivant un protocole standard déjà décrit (F. Blanche et al., 1989). Les activités sont portées ci-dessous.

Tableau : Activité SUMT de SC510 Rif^r et de quelques unes de ses dérivées

5

Souche	SUMT dosée nmole/h/mg de protéines
SC510 Rif ^r	0.05
SC510 Rif ^r pXL59	0.04
SC510 Rif ^r pXL557	2.10
SC510 Rif ^r pXL545	0.05

10

15

Il ressort clairement de ces résultats que seul le plasmide pXL557 entraîne chez SC510 Rif^r une nette augmentation de l'activité SUMT (un facteur 50). Cette augmentation résulte de l'amplification de cobA et non de cobE puisque le plasmide pXL545, qui ne permet que l'amplification de cobE, ne provoque pas d'augmentation d'activité SUMT. Ce résultat confirme que cobA est le gène de structure de la SUMT de Pseudomonas denitrificans. Ce résultat montre que l'on peut obtenir une amplification de l'activité SUMT chez Pseudomonas denitrificans par amplification du gène de structure de la SUMT de Pseudomonas denitrificans.

7.1.2 - Expression de la protéine COBI

20

25

Un fragment provenant du fragment d'ADN de 8,7 kb, contenant le gène de structure de la SP₂MT (cobI), est cloné sur un plasmide à large spectre d'hôte chez les bactéries gram-négatives, puis ce plasmide est ensuite introduit par conjugaison chez Pseudomonas denitrificans SC510 Rif^r. L'activité S-adénosyl-L-méthionine:précorrine-2 méthyltransférase de la souche est ensuite mesurée par rapport à celle de la souche portant le vecteur.

30

Le fragment BamHI-BamHI-SstI-SstI de 1,9 kb contenant les gènes cobH et cobI est purifié à partir du fragment de 8,7 kb. Des "linkers" XbaI et EcoRI sont placés respectivement aux extrémités BamHI et SstI après que celles-ci aient été remplies avec l'ADN polymérase du bactériophage T4. Le fragment est ensuite inséré entre les sites XbaI et EcoRI du plasmide à large spectre d'hôte pXL59. Il

porte la résistance à la kanamycine. Le plasmide ainsi obtenu est nommé le pXL1148 (figure 24).

5 Par ailleurs un plasmide voisin a été construit : le fragment BamHI-BamHI-SstI de 1,5 kb contenant seulement le gène cobH en entier et la partie 5' du gène cobI a été purifié à partir du fragment de 8,7 kb. Des "linkers" XbaI et EcoRI ont été ajoutés aux sites BamHI et SstI respectivement après que ceux-ci aient été remplis ou digérés avec l'ADN polymérase du phage T4. Ce fragment a ensuite été inséré entre les sites EcoRI et XbaI du pXL59 pour donner le
10 plasmide pXL1149. Les plasmides pXL1148 et pXL1149 ne diffèrent que par la présence sur pXL1148 du fragment SstI-SstI de 0,3 kb qui contient l'extrémité 3' du gène cobI. pXL1148 possède le gène de structure entier de cobI au contraire de pXL1149. Les deux plasmides contiennent le gène cobH.

15 Ces deux plasmides ont été introduits par conjugaison chez SC510 Rif^r. Les souches SC510 Rif^r, SC510 Rif^r pXL59, SC510 Rif^r pXL1148 et SC510 Rif^r pXL1149 sont cultivées en milieu PS4. Après 4 jours de culture, les cellules sont récoltées et les activités SP₂MT sont dosées comme cela est décrit dans l'exemple 6.1.3 a).

20 Le résultat de ces dosages est porté ci-dessous avec les activités de SP₂MT définies comme à l'exemple 6.1.3 a).

Tableau : Activités SP₂MT de diverses souches dérivées de Pseudomonas denitrificans

25

Souche	Activité SP ₂ MT ¹ en %
SC510 Rif ^r	< 5
SC510 Rif ^r pXL59	< 5
SC510 Rif ^r pXL1148	75
SC510 Rif ^r pXL1149	< 5

¹pour 500 µg d'extrait brut introduit dans le test.

L'activité est exprimée en % comme cela est défini à l'exemple 6.1.3 a).

Seul le plasmide pXL1148 apporte une augmentation sensible de l'activité SP₂MT. Au contraire le plasmide pXL1149 ne donne pas de résultats différents de ceux observés avec les contrôles SC510 Rif^r et SC510 Rif^r pXL59. Le pXL1148 est le seul plasmide à contenir le gène cobI et il est le seul à amplifier l'activité SP₂MT; ce résultat confirme que le gène de structure de la SP₂MT de Pseudomonas denitrificans est le gène cobI. De plus, si les protéines totales de ces différentes souches sont séparées par électrophorèse en conditions dénaturantes (PAGE-SDS à 10 % d'acrylamide) on observe spécifiquement dans le cas du pXL1148 la présence d'une bande qui correspond à une protéine ayant un poids moléculaire de 25 000 environ (figure 25). Le poids moléculaire de cette protéine correspond à celui de la protéine COBI. Le plasmide pXL1148 permet d'obtenir chez Pseudomonas denitrificans la surproduction de la protéine COBI.

7.1.3 - Expression de COBF

L'expression est obtenue en positionnant en amont du gène cobF le promoteur P_{trp} de E.coli et le site de fixation des ribosomes du gène cII du bactériophage lambda. L'expression ainsi obtenue est beaucoup plus élevée que celle observée par simple amplification génique grâce au même plasmide multicopie.

Le fragment EcoRI-BamHI-BamHI de 2 kb environ du pXL1496 (exemple 7.2.1) est purifié (figure 26). Ce fragment contient le promoteur P_{trp} de E.coli et le site de fixation des ribosomes du gène cII du bactériophage lambda en amont du gène cobF. En aval du gène cobF se trouve le terminateur de l'opéron rrnB de E.coli. Ce fragment est cloné aux sites EcoRI-BamHI du plasmide pKT230 pour donner le pXL1546 (figure 26). Le pKT230 est un plasmide du groupe d'incompatibilité Q qui se réplique chez presque toutes les bactéries gram-négatives (Bagdasarian et al., 1981); ce plasmide porte la résistance à la kanamycine. Le plasmide pXL1546 et pKT230 sont

introduits, par conjugaison, chez SC510 Rif^r. Les souches SC510 Rif^r, SC510 Rif^r pKT230 et SC510 Rif^r pXL1546 sont cultivées en milieu PS4 comme cela est décrit précédemment. Après quatre jours de culture, les protéines totales des différentes souches sont analysées en PAGE-SDS à 10 %. Comme cela est montré à la figure 27, on observe dans l'extrait de SC510 Rif^r pXL1546 une protéine d'un poids moléculaire de 32 000 environ qui est surexprimée; cette protéine co-migre avec la protéine qui est surexprimée par E.coli B pXL1496 (exemple 7.2.1). De plus cette protéine est spécifiquement exprimée dans la souche SC510 Rif^r contenant le pXL1546 où elle représente au moins 20 % des protéines totales. Par contre on n'observe pas cette protéine dans les protéines totales des souches SC510 Rif^r et SC510 Rif^r pKT230. Cette protéine surexprimée est donc la protéine COBF.

7.1.4 - Expression de COBH

Cet exemple décrit l'amplification d'un fragment d'ADN de Pseudomonas denitrificans contenant le gène cobH. La protéine qui est codée par ce gène est purifiée; il s'agit de la protéine COBH. Le plasmide pXL1149, décrit à l'exemple 7.1.2, ne contient sur l'insert d'ADN provenant du fragment de 8,7 kb que le gène cobH en entier. Chez SC510 Rif^r, ce plasmide, contrairement au vecteur, entraîne la surexpression d'une protéine de poids moléculaire 22 000 (figure 25).

7.1.5 - Expression de COBV

Cet exemple décrit l'amplification de l'activité cobalamin (5'-phosphate) synthase par un plasmide ne portant que cobV (pXL699, voir figure 38). L'activité cobalamin (5'-phosphate) synthase est amplifiée chez SC877Rif^r par le plasmide pXL699 d'un facteur 50 par rapport à la même souche avec le vecteur pXL435, pXL1303, pXL1324 ou pKT230. Ce plasmide ne contient sur son insert que cobV en entier plus les parties 5' terminales d'ORF18 et de cobU. Il est certain que dans une telle souche (SC877Rif^r pXL699) la protéine COBV est surexprimée; cette surexpression est d'un facteur 50 par rapport à l'expression de la souche SC877Rif^r.

7.1.6. Expression de la protéine CORA

Le fragment EcoRI-BamHI-BamHI de 1.5 kb du pXL1832 (voir exemple 7.2.4) contenant le promoteur Ptrp puis le RBS cII du bactériophage λ , le gène de structure de la SUMT de *M. ivanovii* et la région terminatrice de l'opéron rrnB de *E. coli* est cloné aux sites EcoRI-BamHI du pKT230 (Bagdasarian et al., 1981). De cette manière, le plasmide pXL1841 est obtenu (voir figure 56). Ce plasmide est mobilisé chez *P. denitrificans* SC510 Rif^r comme cela est décrit précédemment. Un transconjugant est étudié plus en détail. Cette souche est cultivée en milieu PS4 et l'activité SUMT des extraits bactériens est dosée en même temps que celle de la souche contrôle SC510 Rif^r pXL435 (Cameron et al., 1989). Les activités de ces souches sont présentées ci-dessous.

Souche	Activité spécifique SUMT en pmol/h/mg de protéines
15 SC510 Rif ^r pXL435	50-100
SC510 Rif ^r pXL1841	1700

Ce résultat montre clairement qu'il y a expression de l'activité SUMT de *M. ivanovii* chez *P. denitrificans* grâce au plasmide pXL1841 puisque l'activité SUMT de la souche SC510 Rif^r pXL1841 est nettement supérieure 20 à celle de SC510 Rif^r pXL435.

7.2 - Expression chez *E. coli*

Cet exemple illustre comment une protéine COB de *Pseudomonas denitrificans* peut être surproduite chez *E. coli*.

7.2.1 - Expression de COBF

25 La surproduction est obtenue en positionnant en amont du gène cob le promoteur Ptrp de *E. coli* et le site de fixation des ribosomes du gène cII du bactériophage lambda. Le fragment EcoRI-XhoI de 2250 pb du fragment EcoRI de 8,7 kb (aux positions respectives 0 et

2250 sur la séquence présentée à la figure 8) a été cloné dans le phage M13mp19 (Norrand et al., 1983) entre les sites EcoRI et SalI. Le plasmide ainsi construit est nommé pXL1405. Un site NdeI a été introduit par mutagenèse dirigée de manière à ce que les trois dernières bases (ATG) de ce site de restriction constituent le site d'initiation de traduction du gène cobF. Cela revient à modifier les trois bases qui précèdent l'ATG du gène cobF GAA (le G est à la position 733 sur la séquence présentée sur la figure 8) en CAT. Le fragment NdeI-SphI-SphI (figure 26) contenant le gène cobF est ensuite purifié; ce fragment de 1,5 kb est ensuite cloné entre les sites NdeI-SphI du plasmide pXL694 (Denèfle et al., 1987). Le plasmide ainsi construit est nommé pXL1496 (figure 26). Sur le fragment EcoRI-NdeI de 120 pb (qui provient du pXL694) qui précède le gène cobF sont présents des signaux de régulation de l'expression génétique de E.coli. Ces signaux sont constitués par la région [-40+1] du promoteur P_{trp} de E.coli, puis par 73 pb qui contiennent le site de fixation des ribosomes du gène cII du bactériophage λ (Denèfle et al., 1987). En aval du gène cobF se trouvent les terminateurs de l'opéron rrnB de E.coli (sur le fragment HindIII-BamHI). Le plasmide pXL1496 a été introduit par transformation dans la souche de E.coli (Monod et Wollman, 1947). L'expression du gène cobF a été étudiée comme cela est déjà décrit (Denèfle et al., 1987) dans des conditions où le promoteur P_{trp} est soit réprimé (présence de tryptophane), soit non réprimé (absence de tryptophane). Le milieu où l'expression a été réalisée est du milieu minimum M9 (Miller, 1972) supplémenté par 0,4 % de glucose, 0,4 % de casaminoacides, 10 mM thiamine et 40 µg/ml de tryptophane dans le cas où l'on veut réprimer le promoteur P_{trp}. La souche E.coli B pXL1496 a été cultivée à 37°C dans le milieu décrit ci-dessus avec 100 µg d'ampicilline. Comme il est montré sur la figure 28, l'absence de tryptophane entraîne l'expression d'une protéine d'un poids moléculaire de 32 000. En effet dans l'extrait de protéines totales de E.coli B pXL1496 analysé en PAGE-SDS (figure 28) on observe clairement une protéine d'un poids moléculaire de 32 000 qui représente entre 1 et 4 % des protéines totales. Cette protéine est présente en quantités nettement moins importantes dans l'extrait des protéines totales de

E.coli B pXL1496 cultivée dans les mêmes conditions, mais en présence de tryptophane. Le poids moléculaire de la protéine qui est exprimée dans ces conditions est proche du poids moléculaire de la protéine COBF déduit de la séquence en acides aminés de la protéine qui est 28 927 (figure 16). La protéine qui est ainsi exprimée chez E.coli est la protéine COBF.

7.2.2 - Expression de COBT

La surproduction est obtenue en fusionnant à l'extrémité 5' du gène cob le promoteur lac et les trois premiers codons de lacZ de E.coli.

Le site EcoRI situé à la position 2624 sur la séquence présentée sur la figure 32 du fragment de 4.8 kb contient le quatrième codon du gène cobT. Le fragment EcoRI-XbaI de 3,5 kb du pXL837 (voir figure 36) est cloné aux sites EcoRI et XbaI du pTZ18R ou pTZ19R (Pharmacia) pour 15 générer les pXL1874 ou pXL1875 respectivement; ces deux plasmides différent par l'orientation du gène cobT tronqué vis à vis du promoteur de l'opéron lactose d'E. coli (Plac). Plac est en amont de cobT sur pXL1874 tandis que c'est le contraire sur le pXL1875. Le clonage aux sites EcoRI-XbaI du pTZ18R du fragment EcoRI-XbaI du 20 pXL837 permet de réaliser une fusion de protéine entre les 4 premiers acides aminés de la β -galactosidase d'E. coli et le gène cobT à partir de son 4ème codon. L'expression de ce gène lacZ' 'cobT est sous le contrôle des signaux d'expression de lacZ. Les plasmides pXL1874, pXL1875, pTZ18R sont introduits par transformation dans la souche de 25 E.coli BL21. L'expression du gène cobT est étudié comme cela est déjà décrit (Maniatis et al., 1989).

Comme le montre la figure 42B, une protéine d'un poids moléculaire de 72 000 n'est exprimée qu'avec le pXL1874 et représente, dans l'extrait de protéines totales de BL21, pXL1874 analysé en PAGE-SDS, 1 à 4 % des 30 protéines totales. Le poids moléculaire de la protéine qui est exprimée dans ces conditions est proche du poids moléculaire de la protéine COBT déduit de la séquence en acides aminés qui est de 70 335 sur la figure 40. Cette expérience montre clairement qu'à partir du site EcoRI situé sur le quatrième codon du gène cobT une phase ouverte

compatible avec celle trouvée pour le gène cobT peut être exprimée.

7.2.3 - Expression d'une protéine COBS tronquée.

Un site BamHI est situé au niveau du 45ème codon du gène COBS. Le fragment BamHI-BamHI de 1.2 kb contenant la partie 3' du gène cobS et des séquences en aval de ce gène est excisé du pXL843 pour
5 être cloné au site BamHI du plasmide pET-3b (Rosenberg et al., 1987) pour générer le pXL1937. Le fragment BamHI est orienté de manière à ce que la partie tronquée du gène cobS soit fusionnée, en phase, avec les 12 premiers codons de la protéine majeure de capsid du bactériophage
10 T7 ou gène 10 (Rosenberg et al., -1987). Ce gène hybride est sous le contrôle du promoteur ϕ 10 du bactériophage T7. Le plasmide pXL1937 ainsi que le pET-3b sont introduits par transformation dans la E. coli BL21 pLysS (W. Studier, communication personnelle). Après réisolement sur milieu sélectif, les 2 souches sont cultivées en milieu liquide L
15 jusqu'à une D₀ 610 nm de 1; à ce stade là le milieu est ajusté à une concentration en IPTG (isopropyl β -galactoside) de 1 mM afin d'induire l'expression de la polymérase du batériophage T7 (Rosenberg et al., 1987). La culture est ensuite incubée 3 h à 37°C, puis des lysats bactériens sont préparés. Les protéines totales des bactéries ainsi
20 cultivées sont séparées par PAGE en conditions dénaturantes. Comme on le voit sur la figure 42A, il y a spécifiquement surexpression d'une protéine de 33 000 avec la culture BL21 pLysS pXL1937. Ce poids moléculaire est tout à fait compatible avec le poids moléculaire attendu pour la protéine fusion (33 kD). Cette expérience montre
25 clairement qu'à partir du site BamHI situé au niveau du 45ème codon du gène cobS, une phase ouverte compatible avec celle trouvée pour le gène cobS peut être surexprimée.

7.2.4. Expression de la protéine CORA

Les oligonucléotides suivants ont été synthétisés comme cela est
30 décrit précédemment :

oligonucléotide 1277

5' GGC CGA ATT CAT ATG GTA GTT TAT TTA 3'

-----1 2 3 4 5 (1 à 5; 5

EcoRI -----

premiers codons

NdeI

de la SUMT de M.ivanovii)

5 oligonucléotide 1278

5' GGC CGA GCT CTA TTA CAT AAT T

-----=====

SstI

(= séquence figurant sur la figure 51, positions 926 à 915 sur le brin
10 complémentaire du brin codant)

L'oligonucléotide 1277 possède les séquences de reconnaissance des
enzymes de restriction EcoRI et NdeI. Les trois dernières bases du site
NdeI (ATG), qui correspondent à un codon d'initiation de traduction,
sont directement suivies par les codons 2 à 5 du gène de structure de la
15 SUMT de M. ivanovii tels qu'ils figurent dans la séquence présentée sur
la figure 52. L'oligonucléotide 1278 contient la séquence de
reconnaissance de SstI, suivie directement de la séquence TATTACATAATT
qui correspond à une séquence présente sur le fragment de 955 pb
contenant le gène corA présenté sur la figure 51; cette séquence se
20 trouve à la position 926 à 915 (voir figure 51) sur le brin
complémentaire du brin codant pour la protéine CORA. Les deux
oligonucléotides 1277 et 1278 contiennent donc dans leur partie 3' des
séquences correspondant respectivement au brin codant du gène corA et au
brin complémentaire en aval de ce gène. Ces deux oligonucléotides
25 peuvent être utilisés pour réaliser une expérience d'amplification
enzym-atique avec le plasmide pXL1809 comme matrice. Cette expérience
permet d'obtenir un fragment de 910 pb contenant le gène corA de M.
ivanovii possédant un site NdeI au niveau de l'ATG du gène corA et un
site SstI à l'autre extrémité du fragment après la fin du gène corA.
30 L'amplification enzymatique est réalisée comme précédemment décrit pour
l'amplification enzymatique faite sur l'ADN génomique de M. ivanovii, si
ce n'est que la matrice est constituée par 10 ng d'ADN du plasmide
pXL1809; les températures utilisées sont les mêmes, mais seulement 20

cycles d'amplification sont réalisés. Comme cela est précédemment décrit, les produits d'amplification sont digérés par NdeI et SstI avant d'être séparés par migration sur gel d'agarose. Comme attendu, un fragment d'une taille 910 pb est effectivement visualisé. Ce fragment 5 est purifié comme cela est déjà décrit. Ce fragment est cloné aux sites NdeI et SstI du pXL694 (Denèfle et al., 1987). Le plasmide résultant nommé pXL1832 est décrit sur la figure 56. Sur ce plasmide de la même manière que cela est décrit à l'exemple 7.2, le gène de structure de la SUMT de *M. ivanovii* est précédé par le site de fixation des ribosomes du 10 gène cII du bactériophage λ . En amont de ce RBS se trouve le promoteur Ptrp. Le plasmide pXL1832 est introduit dans *E. coli* B5548 qui est une souche de *E. coli* portant la mutation cysG44 (Cossart et Sanzey, 1982) par transformation. Les activités SUMT des souches *E. coli* B5548 pUC13 et *E. coli* B5548 pXL1832 sont dosées sur des extraits obtenus à partir 15 de cellules cultivées en milieu LB supplémenté en ampicilline. Le dosage de l'activité SUMT est réalisé comme cela a déjà été décrit (Blanche et al., 1989). Les résultats de ce dosage sont donnés ci-dessous.

Souche	Activité spécifique SUMT en pmol/h/mg de protéines
20 <i>E. coli</i> B5548 pUC13	5.9
<i>E. coli</i> B5548 pXL1832	310

Les résultats présentés sur le tableau ci-dessus montrent clairement qu'il y a expression d'une activité SUMT chez la souche *E. coli* B5548 lorsque celle-ci contient une plasmide pXL1832 qui exprime la SUMT de *M. 25 ivanovii*. La SUMT de *M. ivanovii* peut donc être exprimée chez *E. coli*.

EXEMPLE 8 - Amplification de la production de cobalamines par les techniques d'ADN recombinant

8.1 - Amplification chez P.denitrificans

Cet exemple illustre comment une amélioration de la 30 production de cobalamines est obtenue chez Pseudomonas denitrificans

SC510 Rif^r par amplification de gènes cob de Pseudomonas denitrificans SC510.

5 8.1.1 Amélioration de la production de cobalamines chez Pseudomonas denitrificans par levée d'une étape limitante dans la biosynthèse des cobalamines

Cet exemple illustre comment la productivité en cobalamines de souches de Pseudomonas denitrificans peut être améliorée par amplification de gènes cob de Pseudomonas denitrificans. Cette amélioration résulte de la levée d'une étape limitante de la
10 voie de biosynthèse.

Le plasmide pXL367 est décrit à l'exemple 4.2 (figure 13). Ce plasmide correspond au pRK290 (Ditta et al., 1981) dans lequel le fragment EcoRI de 8,7 kb a été inséré. Ce plasmide pXL367 procure une amélioration de la biosynthèse de cobalamines chez la souche SC510
15 Rif^r. Les souches SC510 Rif^r, SC510 Rif^r pRK290 et SC510 Rif^r pXL367 sont cultivées en erlenmeyer dans du milieu PS4 suivant les conditions décrites dans les protocoles expérimentaux. On observe une amélioration du titre de production due à la présence du plasmide pXL367. En effet la souche SC510 Rif^r pXL367 produit 30 % de plus de
20 cobalamines que les souches SC510 Rif^r et SC510 Rif^r pRK290. Cette amélioration n'est pas due à l'amplification de n'importe quels gènes de Pseudomonas denitrificans mais à l'amplification spécifique de gènes portés par le fragment EcoRI de 8,7 kb. En effet le plasmide pXL723 décrit sur la figure 11 ne donne aucune amélioration et le même
25 titre de production est observé avec ce plasmide qu'avec les souches SC510 Rif^r et SC510 Rif^r pRK290.

30 8.1.2 Amélioration de la production de coenzyme B₁₂ chez Pseudomonas denitrificans par levée de deux étapes limitantes dans la biosynthèse des cobalamines

Cet exemple illustre comment la productivité en cobalamines de souches de Pseudomonas denitrificans peut être améliorée par amplification de gènes cob de Pseudomonas denitrificans. Cette amélioration résulte de la levée de deux étapes limitantes de la

voie de biosynthèse.

Le fragment ClaI-EcoRV de 2,4 kb issu du fragment de 5,4 kb (contenant les gènes cobA et cobE) est cocloné avec le fragment EcoRI de 8,7 kb sur le plasmide à large spectre d'hôte pXL203. Le plasmide ainsi construit est appelé pXL525 (figure 29). Ce plasmide est introduit chez SC510 Rif^r par conjugaison. La souche SC510 Rif^r pXL525 produit 20 % de cobalamines de plus que SC510 Rif^r pXL367. L'amplification des gènes cobA et cobE permet de lever une nouvelle étape limitante chez SC510 Rif^r dans la biosynthèse des cobalamines. La souche SC510 Rif^r de Pseudomonas denitrificans est améliorée dans le présent exemple par la levée successive de deux étapes limitantes. Cet exemple montre que la levée de deux étapes limitantes dans la biosynthèse des cobalamines peut conduire à des améliorations supplémentaires de production.

8.2 - Amélioration de la productivité de cobalamines chez Agrobacterium tumefaciens

Cet exemple illustre l'amélioration de la production de cobalamines d'une souche productrice de cobalamines par l'amplification de gènes cob de Pseudomonas denitrificans SC510.

La souche utilisée est une souche de bactérie gram-négative ; il s'agit d'une souche d'Agrobacterium tumefaciens.

Les plasmides décrits dans les exemples 4.2 et 8.1, pXL367 et pXL525 ainsi que le vecteur pRK290 (Ditta et al., 1981) et le plasmide pXL368 (figure 29), sont introduits par transfert conjugatif chez la souche d'Agrobacterium tumefaciens C58-C9 Rif^r (Cameron et al., 1989). Les souches C58-C9 Rif^r, C58-C9 Rif^r pRK290, C58-C9 Rif^r pXL367, C58-C9 Rif^r pXL368 et C58-C9 Rif^r pXL525 sont cultivées en milieu PS4 à 30°C comme cela est décrit précédemment. Les cobalamines produites sont dosées comme cela est décrit précédemment. Les titres de productions sont portés sur le tableau ci-dessous.

Tableau : Titres de vitamine B₁₂ produite par différentes souches d'Agrobacterium tumefaciens recombinantes.

5

Souche	Vitamine B ₁₂ en mg/l
C58-C9 Rif ^r	0.4
C58-C9 Rif ^r pRK290	0.4
C58-C9 Rif ^r pXL367	0.8
C58-C9 Rif ^r pXL368	0.8
C58-C9 Rif ^r pXL525	1.2

10 Comme il apparaît clairement sur le tableau ci-dessus, la
production de cobalamines est améliorée chez la souche d'Agrobacterium
tumefaciens utilisée. Deux plasmides différents améliorent la
production de cobalamines de la souche d'Agrobacterium tumefaciens
utilisée: pXL367 et pXL368. Ces plasmides contiennent respectivement
le fragment EcoRI de 8,7 kb (gènes cobF à cobM) et le fragment
15 ClaI-EcoRV de 2,4 kb (gène cobE et cobA). Séparément, ils améliorent
la production de cobalamines d'Agrobacterium tumefaciens C58-C9 Rif^r
d'un facteur 2; ce résultat montre qu'il est possible d'améliorer la
production de cobalamines d'une souche d'Agrobacterium tumefaciens en
amplifiant des fragments portant des gènes cob de Pseudomonas
20 denitrificans. Dans le cas présent on peut parler d'amélioration
hétérologue, c'est-à-dire d'amélioration de la production de
cobalamines d'une souche moyennant l'amplification de gènes cob d'une
autre souche.

25 Les améliorations de production de cobalamines apportées
par les différents fragments de Pseudomonas denitrificans contenant
des gènes cob sont cumulables, c'est-à-dire qu'en mettant sur le même
plasmide les deux fragments qui sont séparément clonés sur le pXL367
et le pXL368, on observe des améliorations additives. Le plasmide
pXL525 apporte chez Agrobacterium tumefaciens C58-C9 Rif^r une
30 amélioration de production supérieure à celle apportée par chacun des
fragments clonés séparément sur le même vecteur.

8.3 - Amélioration de la productivité de cobalamines chez Rhizobium meliloti

Cet exemple décrit l'amélioration de la production de cobalamines d'une autre souche productrice de cobalamines.

5 Le plasmide décrit à l'exemple 8.2, pXL368, ainsi que le vecteur pRK290 (Ditta et al., 1981) sont introduits par transfert conjugatif chez la souche Rhizobium meliloti 102F34 Rif^r (Leong et al., 1982). Les transconjugants: 102F34 Rif^r, 102F34 Rif^r pRK290 et 102F34 Rif^r pXL368 sont cultivés en milieu PS4 à 30°C comme cela est
10 décrit précédemment. Les cobalamines produites sont dosées comme cela est décrit précédemment. Les titres de productions sont portés sur le tableau ci-dessous.

Tableau : Titres des cobalamines produites par différentes souches de Rhizobium meliloti recombinantes

15

Souche	Vitamine B ₁₂ en mg/l
102F34 Rif ^r	0.4
102F34 Rif ^r pRK290	0.4
102F34 Rif ^r pXL368	0.8

20

Comme il apparaît clairement sur le tableau ci-dessus, la production de cobalamines est améliorée chez la souche de Rhizobium meliloti utilisée. Le plasmide pXL368 améliore la production de cobalamines de la souche de Rhizobium meliloti utilisée. Ce plasmide
25 contient le fragment ClaI-EcoRV de 2,4 kb (gènes cobA et cobE); il améliore la production de cobalamines de Rhizobium meliloti 102F34 Rif^r d'un facteur 2. Ce résultat montre qu'il est possible d'améliorer la production de cobalamines d'une souche de Rhizobium meliloti en amplifiant des fragments portant des gènes cob de Pseudomonas denitrificans. Dans le cas présent on peut parler d'amélioration
30 hétérologue, c'est-à-dire d'amélioration de la production de cobalamines d'une souche moyennant l'amplification de gènes cob d'une

autre souche.

EXEMPLE 9 - Dosage des corrinoïdes et des descobaltocorrinoïdes dans les moûts et les cellules de souches productrices de corrinoïdes

5 Cet exemple illustre comment il est possible d'identifier et de doser les différents corrinoïdes et descobaltocorrinoïdes produits par différentes souches productrices de cobalamines. Ce dosage permet entre autre de doser le coenzyme B₁₂.

10 Les moûts (ou les cellules seules) sont cyanurés comme cela a déjà été décrit (Renz, 1971). Après centrifugation, une aliquote du surnageant est passée à travers une colonne de DEAE-Sephadex qui est ensuite lavée extensivement avec du dihydrogenophosphate de potassium 1M. Les fractions recueillies sont regroupées et dessalées sur une cartouche Sep-Päk C-18 (Waters). Après évaporation et resuspension dans l'eau (100 µl à 1 ml suivant la
15 quantité de corrinoïdes présents), les corrinoïdes sont identifiés et dosés par CLHP sur une colonne de Nucléosil C-18 (Macherey-Nagel). La colonne est éluée à 1 ml/mn avec un gradient d'acétonitrile (de 0 % à 100 %) dans du tampon phosphate de potassium 0,1 M contenant 10 mM de KCN.

20 Les corrinoïdes sont visualisés en détection U.V. à 371 nm et/ou par la détection spécifique du ⁵⁷Co (si la culture a été effectuée en présence de ⁵⁷CoCl₂) à l'aide d'un détecteur Berthold LB 505. Ils sont donc identifiés et par comparaison de leur temps de rétention avec des étalons. De même, les descobaltocorrinoïdes (acide
25 hydrogénobyrique, acide hydrogénobyrique monoamide et acide hydrogénobyrique diamide) sont visualisés en détection U.V. à 330 nm. Par cette technique les intermédiaires suivants sont séparés: l'acide cobyrique, l'acide cobyrique monoamide, l'acide cobyrique diamide, l'acide cobyrique triamide, l'acide cobyrique tétraamide, l'acide cobyrique pentaamide, l'acide cobyrique, le
30 cobinamide, le cobinamide phosphate, le GDP-cobinamide, la vitamine B₁₂-phosphate et la vitamine B₁₂. Les formes adénosylées de ces produits sont aussi séparées et dosées par cette technique. Pour ce

faire, l'étape initiale de cyanuration est supprimée et la colonne CLHP est éluée avec du tampon dépourvu de KCN. Nous donnons à la figure 31 les temps de rétention de différents étalons séparés par ce système et identifié à la sortie de la colonne par absorbance U.V.

- 5 Un échantillon de la souche SC510 Rif^R a été déposé le 30 janvier 1990 au Centraal Bureau voor Schimmelcultures à Baarn (Pays-Bas) où il a été enregistré sous la référence CBS 103.90.

REVENDICATIONS

1. Séquence d'ADN partiellement purifiée caractérisée en ce qu'elle code pour un polypeptide impliqué dans la biosynthèse des cobalamines et/ou des cobamides, ladite séquence d'ADN étant caractérisée en ce qu'elle est choisie parmi:
 - les gènes cobA, cobB, cobC, cobD, cobE, cobF, cobG, cobH, cobI, cobJ, cobK, cobL, cobM, cobN, cobO, cobP, cobQ,
10 cobR, cobS, cobT, cobU, cobV, cobW, cobX, corA, présentés aux figures 15, 16, 40, 41, 47 et 52, ou
 - les homologues de ces séquences résultant de la dégénérescence du code génétique.
2. Gène partiellement purifié contenant une séquence d'ADN selon la revendication 1.
3. ADN recombinant caractérisé en ce qu'il contient au moins une séquence d'ADN selon la revendication 1 ou 2.
4. ADN recombinant selon la revendication 3, caractérisé en ce que lesdites séquences d'ADN sont placées sous le
20 contrôle de signaux d'expression.
5. ADN recombinant selon la revendication 4, caractérisé en ce que les signaux d'expression peuvent être homologues ou hétérologues de la séquence d'ADN.
6. ADN recombinant selon l'une quelconque des revendications 3 à 5, caractérisé en ce qu'il fait partie d'un plasmide d'expression.

7. Plasmide caractérisé en ce qu'il contient au moins une séquence d'ADN selon la revendication 1, et des séquences permettant leur expression.

8. Plasmide selon la revendication 7, caractérisé en ce qu'il contient:

- un ADN recombinant selon l'une quelconque des revendications 3 à 6,
- un système de répllication,
- au moins un marqueur de sélection.

10 9. Plasmide selon la revendication 7, caractérisé en ce qu'il est choisi parmi:

- le plasmide pXL1500 contenant les gènes cobA, cobB, cobC et cobE,
- les plasmides pXL723 et pXL302 contenant les gènes cobC et cobD,
- le plasmide pXL1397 contenant les gènes cobB et cobC,
- les plasmides pXL368 et pXL557 contenant les gènes cobA et cobE,
- les plasmides pXL545 et pXL545 Ω contenant le gène cobE,
- les plasmides pXL367 et pXL253 contenant les gènes cobF, cobG, cobH, cobI, cobJ, cobK, cobL, cobM,
- le plasmide pXL1148 contenant les gènes cobH et cobI,
- le plasmide pXL1149 contenant le gène cobH,
- les plasmides pXL1496 et pXL1546 contenant le gène cobF,
- le plasmide pXL525 contenant les gènes cobA, cobE et cobF à cobM, le plasmide pXL843 contenant les gènes cobX,

cobS, et cobT, le plasmide pXL699 contenant le gène cobV, le plasmide pXL1324 contenant le gène cobU, les plasmides pXL618 et pXL623 contenant respectivement les gènes cobQ et cobP, le plasmide pXL593 contenant le gène cobP et cobW et le plasmide pXL1909 contenant les gènes cobP, cobW, cobN, et cobO.

10. Cellule isolée dans laquelle a été introduite une séquence d'ADN selon l'une quelconque des revendications 1 à 6, ou un plasmide selon l'une quelconque des
10 revendications 7 à 9.

11. Polypeptide caractérisé en ce qu'il est impliqué dans la biosynthèse des cobalamines et/ou des cobamides, ledit polypeptide étant caractérisé en ce qu'il est codé par une séquence d'ADN selon l'une quelconque des revendications 1 à 6.

12. Polypeptide selon la revendication 11, caractérisé en ce qu'il intervient dans la transformation du précorrène-3 en 5'-déoxy 5'-adénosyl(Ado) acide cobyrinique a,c-diamide.

13. Polypeptide selon la revendication 12, caractérisé en
20 ce qu'il contient tout ou partie des séquences peptidiques COBB, COBF, COBG, COBG, COBH, COBJ, COBK, COBL, COBM, COBN, COBS et COBT présentées aux figures 15, 16, 40 et 41.

14. Polypeptide selon la revendication 12, caractérisé en ce qu'il catalyse le transfert d'un groupement méthyl aux positions C1, C5, C11, C15 ou C17 intervenant entre le précorrène-3 et l'acide cobyrinique a,c-diamide.

15. Polypeptide selon la revendication 14, caractérisé en ce qu'il contient tout ou partie des séquences peptidiques COBF, COBJ, COBL et COBM présentées à la figure 16.
16. Polypeptide selon la revendication 11, caractérisé en ce qu'il intervient dans la transformation de l'acide cobyrique en cobinamide.
17. Polypeptide selon la revendication 16, caractérisé en ce qu'il contient tout ou partie des séquences peptidiques CODC et COBD présentées à la figure 15.
- 10 18. Polypeptide selon la revendication 11, caractérisé en ce qu'il possède une activité S-adénosyl-L-méthionine: précorrène-2 méthyl transférase (SP2MT).
19. Polypeptide selon la revendication 18, caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBI présentée à la figure 16.
20. Polypeptide selon la revendication 11, caractérisé en ce qu'il possède une activité acide cobyrique et/ou hydrogénobyrique a,c-diamide synthase.
21. Polypeptide selon la revendication 20, caractérisé en
20 ce qu'il contient tout ou partie de la séquence peptidique COBB présentée à la figure 15.
22. Polypeptide selon la revendication 11, caractérisé en ce qu'il possède une activité précorrène-8x mutase.

23. Polypeptide selon la revendication 22, caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBH présentée à la figure 16.
24. Polypeptide selon la revendication 11, caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBE présentée à la figure 15.
25. Polypeptide selon la revendication 11, caractérisé en ce qu'il possède une activité nicotinate-nucléotide: diméthylbenzimidazole phosphoribosyltransférase.
- 10 26. Polypeptide selon la revendication 25, caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBU présentée à la figure 41.
27. Polypeptide selon la revendication 11, caractérisé en ce qu'il possède une activité cobalamine-(5'-phosphate) synthase.
28. Polypeptide selon la revendication 27, caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBV présentée à la figure 41.
29. Polypeptide selon la revendication 11, caractérisé en
20 ce qu'il possède une activité cobyric acid synthase.
30. Polypeptide selon la revendication 29, caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBQ présentée à la figure 47.
31. Polypeptide selon la revendication 11, caractérisé en ce qu'il possède une activité précorrin-6x reductase.

32. Polypeptide selon la revendication 31, caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBK présentée à la figure 16.
33. Polypeptide selon la revendication 11, caractérisé en ce qu'il intervient dans la transformation du cobinamide en GDP-cobinamide.
34. Polypeptide selon la revendication 33, caractérisé en ce qu'il possède une activité cobinamide kinase et cobinamide phosphate guanylyltransferase.
- 10 35. Polypeptide selon la revendication 34, caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBP présentée à la figure 47.
36. Polypeptide selon la revendication 11, caractérisé en ce qu'il contient tout ou partie des séquences peptidiques COBS, COBT et COBX présentées à la figure 40.
37. Polypeptide selon la revendication 11, caractérisé en ce qu'il est choisi parmi les protéines CORA, COBB, COBC, COBD, COBE, COBF, COBG, COBH, COBI, COBJ, COBK, COBL, COBM, COBN, COBP, COBQ, COBS, COBT, COBU, COBV, COBW et COBX
20 présentées aux figures 15, 16, 40, 41, 47 et 52.
38. Procédé de production des polypeptides selon les revendications 11 à 37, caractérisé en ce que:
- on introduit dans une cellule hôte une séquence d'ADN selon les revendications 1 à 6 ou un plasmide selon les revendications 7 à 9 contenant une telle séquence,

- on cultive cette cellule recombinante dans des conditions d'expression de ladite séquence, et
- on récupère les polypeptides produits.

39. Procédé selon la revendication 38, caractérisé en ce que la cellule hôte peut être choisie parmi les procaryotes, les cellules animales ou végétales.

40. Procédé selon la revendication 39, caractérisé en ce que la cellule hôte est une archaebactérie ou une eubactérie.

10 41. Procédé selon la revendication 40, caractérisé en ce que la cellule hôte est *E.coli*, *Pseudomonas denitrificans*, *Rhizobium melitoti*, *Agrobacterium tumefaciens*, ou *Salmonella typhimurium*.

42. Procédé permettant d'augmenter la production de cobalamines et/ou de cobamides ou de leurs précurseurs caractérisé en ce que:

- on introduit dans un microorganisme producteur de ces composés, ou potentiellement producteur de ces composés, une ou plusieurs séquences d'ADN selon l'une quelconque des
- 20 revendications 3 à 6,
- on cultive le microorganisme ainsi obtenu dans des conditions de synthèse des cobalamines et/ou des cobamides et d'expression de ladite séquence,
 - on récupère les cobalamines et/ou les cobamides ou leurs précurseurs produits.

43. Procédé selon la revendication 42, caractérisé en ce que l'on introduit dans le microorganisme un plasmide selon les revendications 7 à 9 contenant de telles séquences.
44. Procédé selon la revendication 43, caractérisé en ce que la séquence d'ADN introduite dans le microorganisme code pour un polypeptide catalysant une étape limitante de la biosynthèse des cobalamines et/ou des cobamides.
45. Procédé selon la revendication 43, caractérisé en ce que les séquences d'ADN introduites dans le microorganisme
10 codent pour des polypeptides catalysant des étapes limitantes de la biosynthèse des cobalamines et/ou des cobamides.
46. Procédé selon l'une quelconque des revendications 42 à 45, caractérisé en ce que le microorganisme est choisi parmi *P. denitrificans*, *R. meliloti*, et *A. tumefaciens*.
47. Procédé selon la revendication 46, caractérisé en ce que le microorganisme est *P. denitrificans*.
48. Procédé selon la revendication 47, caractérisé en ce que le microorganisme est *P. denitrificans* SC510 Rif^R.
- 20 49. Procédé selon l'une quelconque des revendications 42 à 48, caractérisé en ce que l'on introduit dans le microorganisme un plasmide selon la revendication 9.
50. Procédé selon les revendications 42 à 49, caractérisé en ce que l'on introduit dans la souche *P. denitrificans* SCS10 Rif^R le plasmide pXL525.

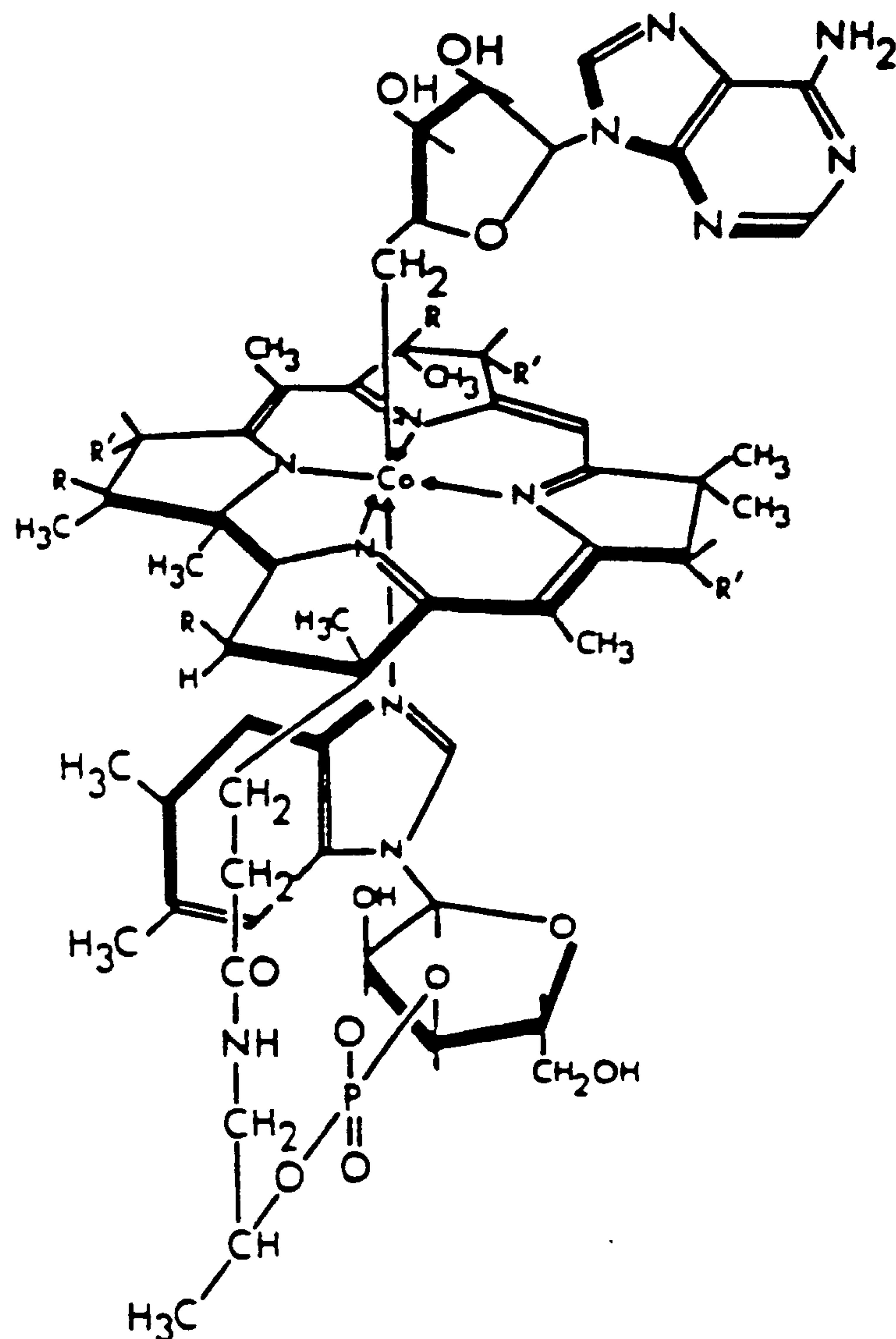
51. Procédé selon l'une quelconque des revendications 42 à 50, caractérisé en ce que les cobalamines et/ou les cobamides produits sont récupérés par:

- solubilisation,
- conversion en forme cyano, et
- purification.

52. Procédé selon l'une quelconque des revendications 42 à 51, caractérisé en ce que les cobalamides est le coenzyme B12.

10 53. Procédé selon l'une quelconque des revendications 42 à 51, caractérisé en ce que le précurseur est choisi parmi les descobaltocorrinoïdes et les corrinoïdes.

1 / 151



Adenosylcobalamine R = CH₂CONH₂, R' = CH₂CH₂CONH₂.

FIGURE 1

.. 2 / 151

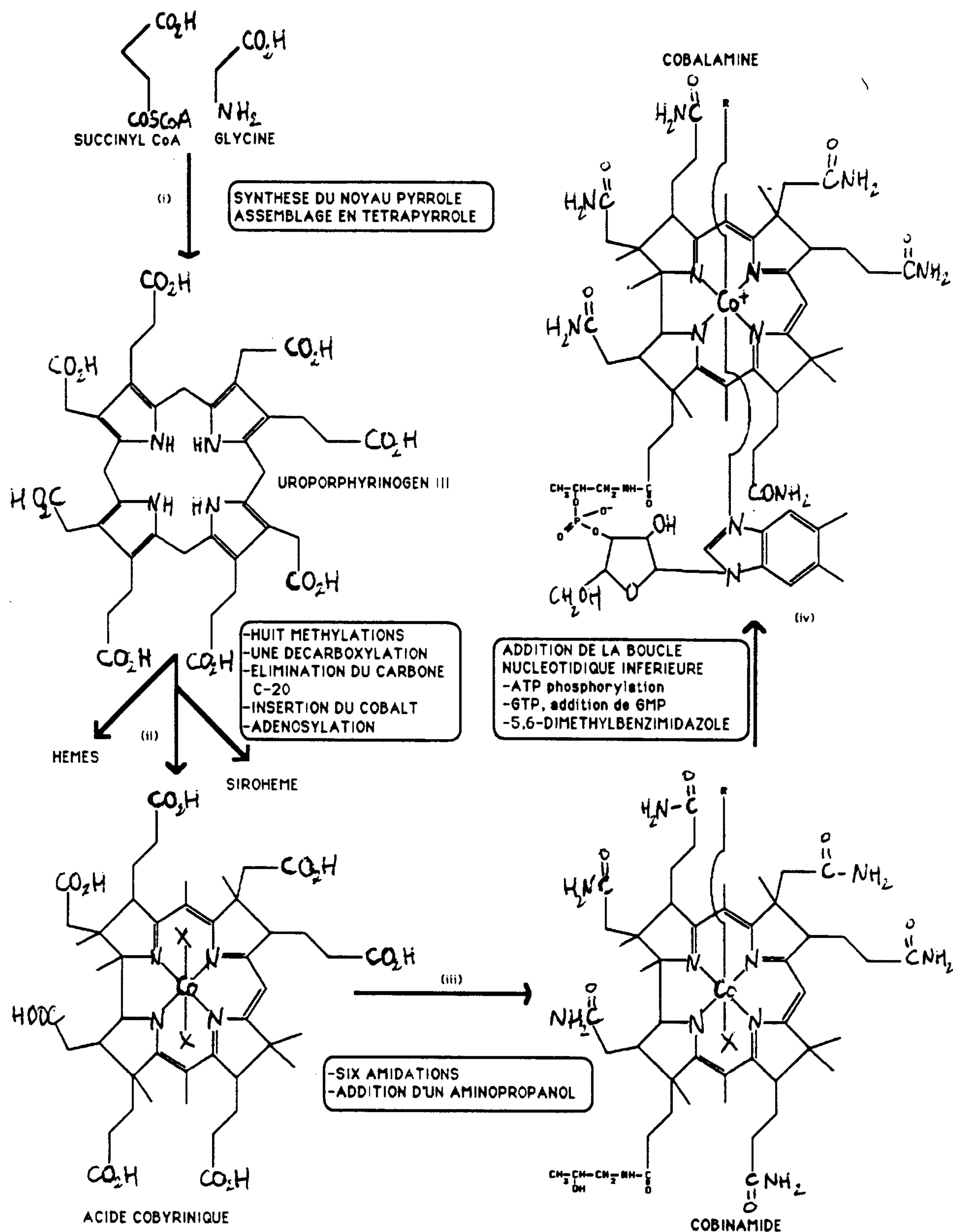
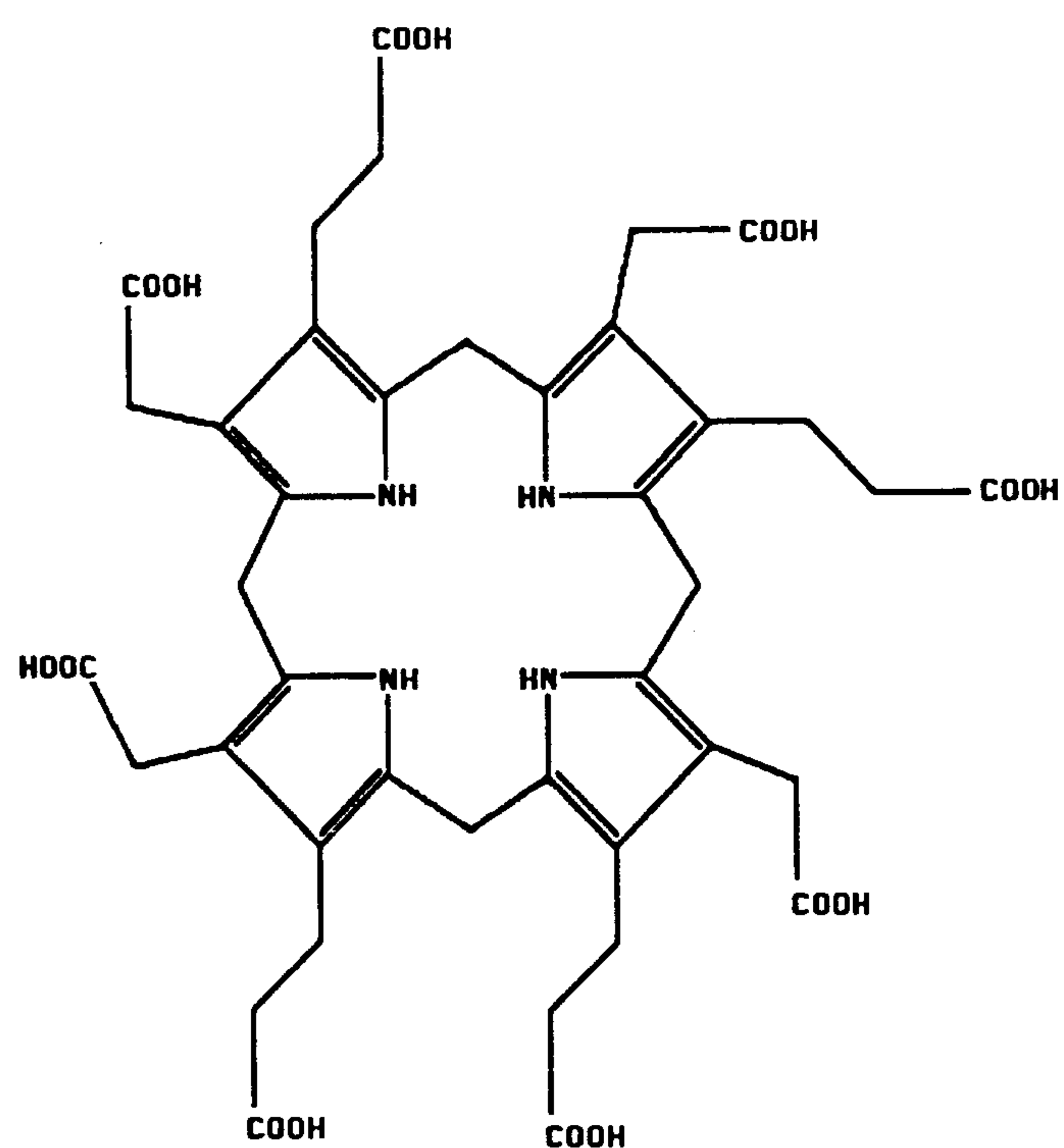
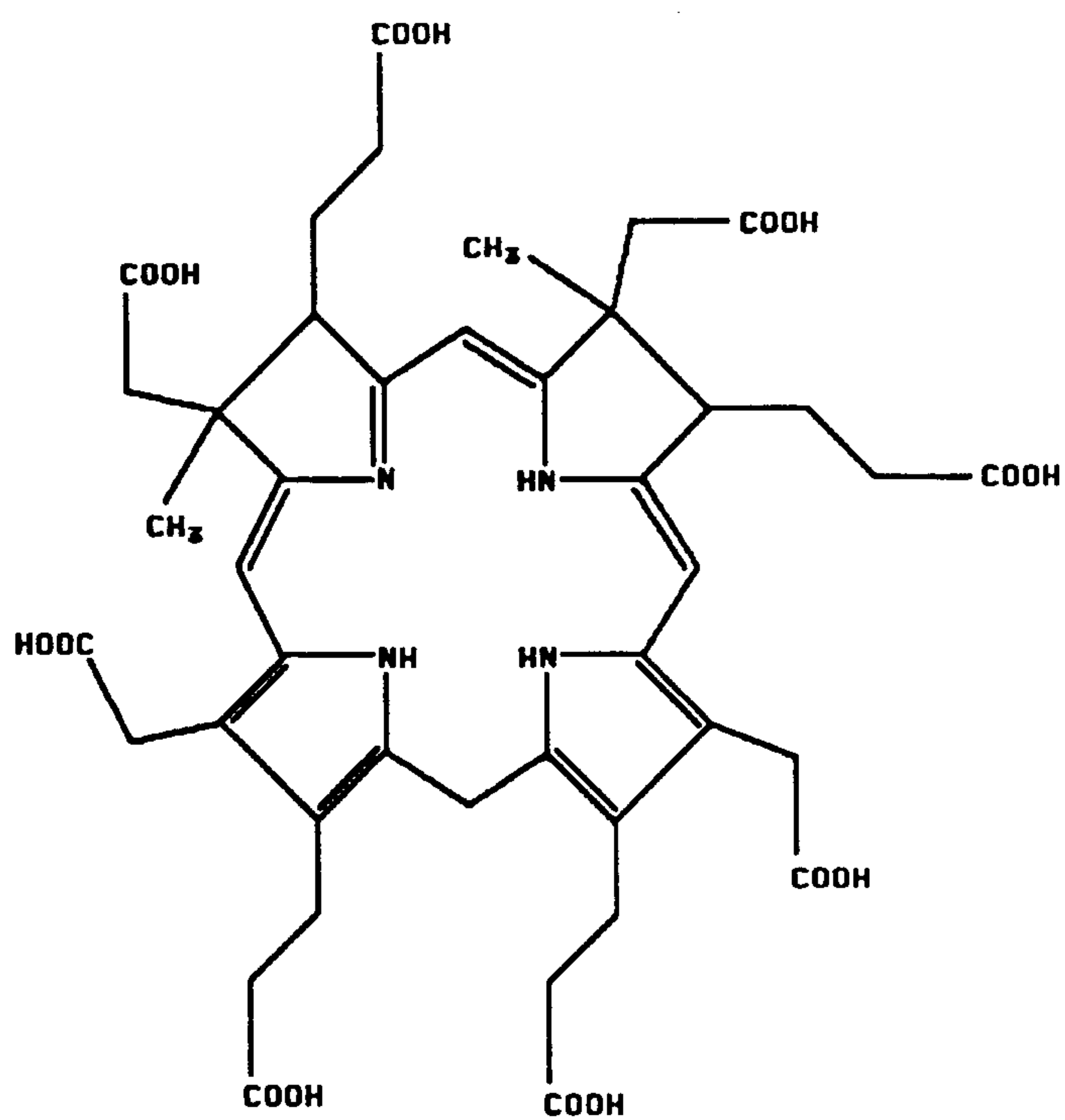


FIGURE 2

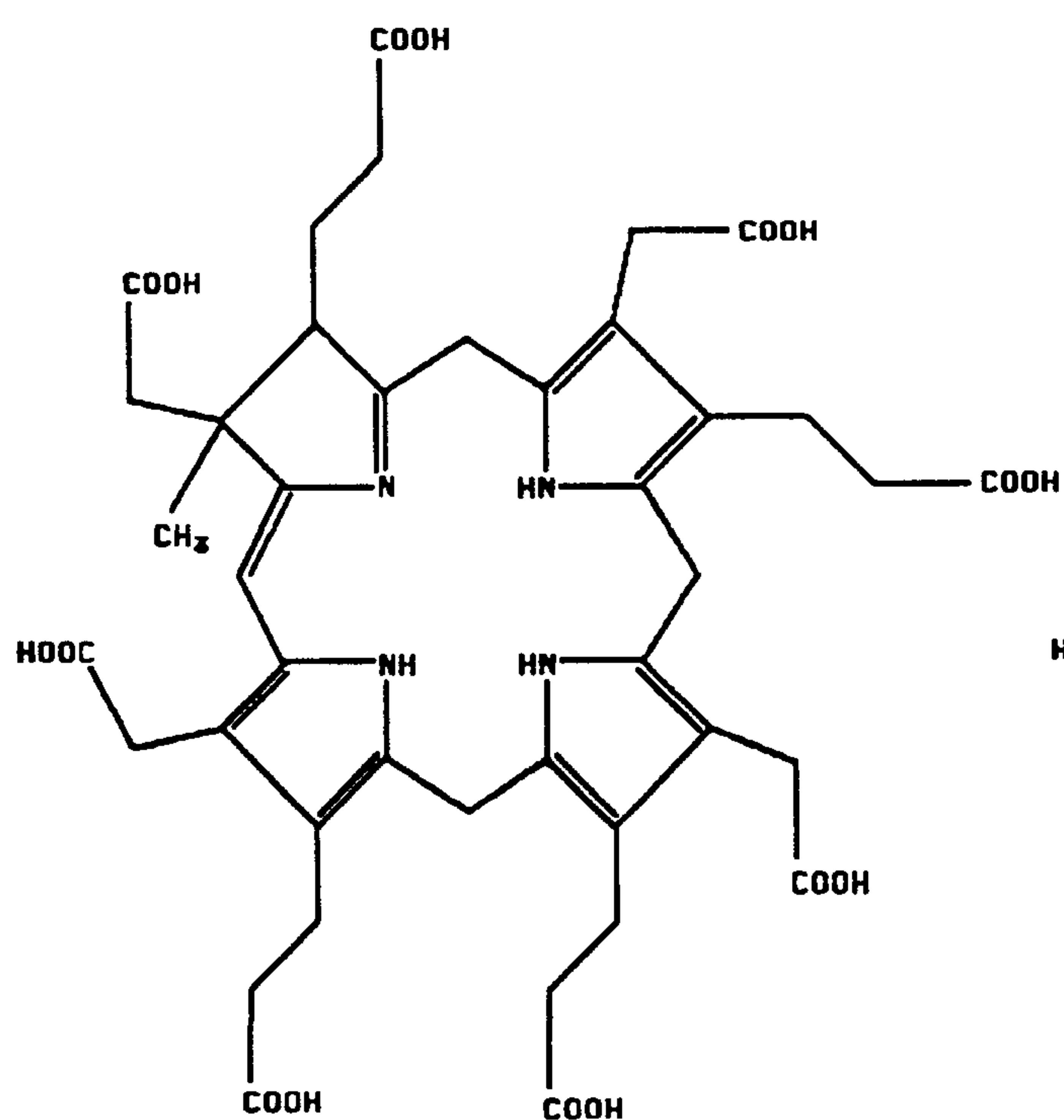
3/151



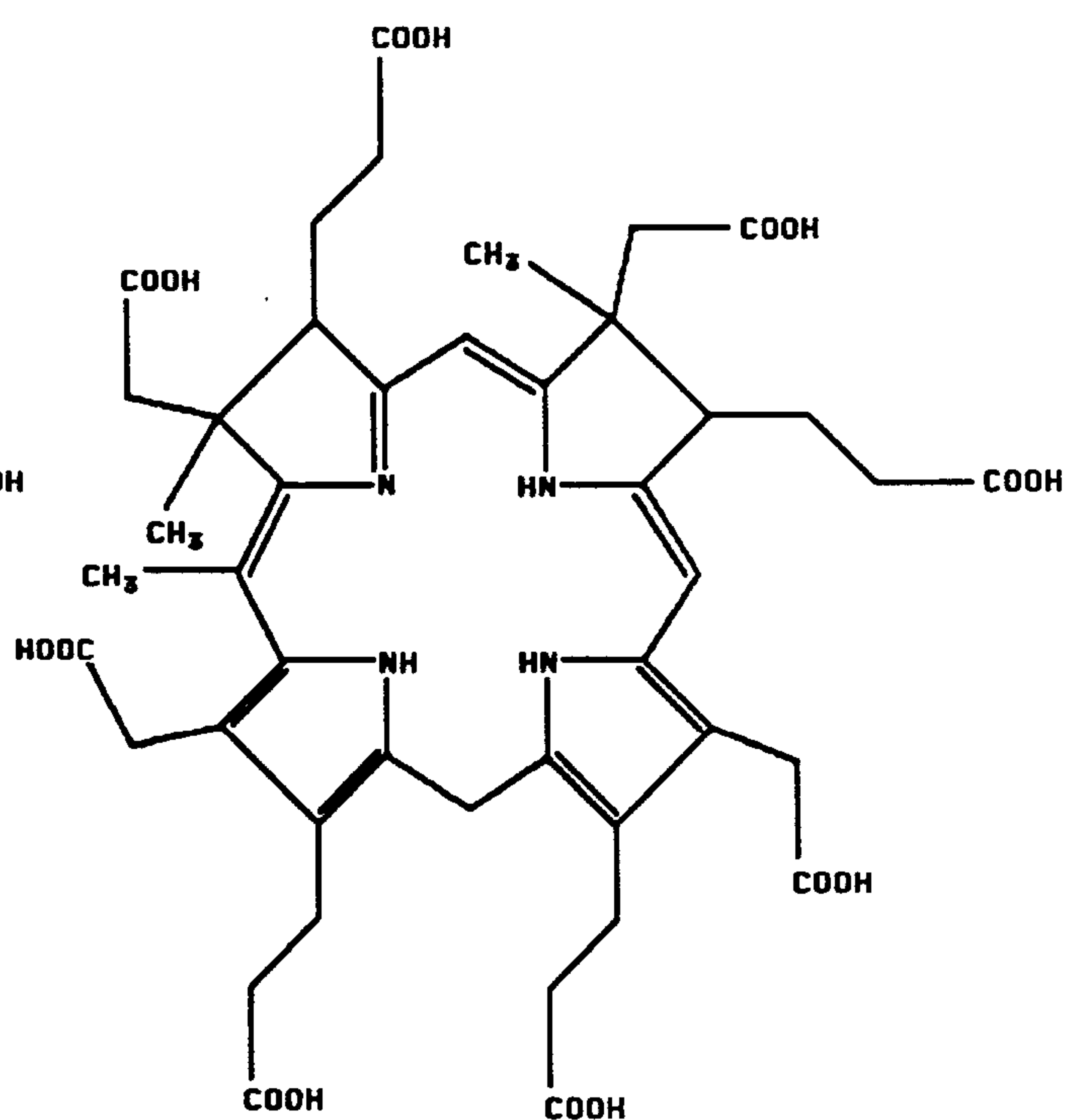
uro'gen III



précorrine-2



précorrine-1



précorrine-3

FIGURE 3

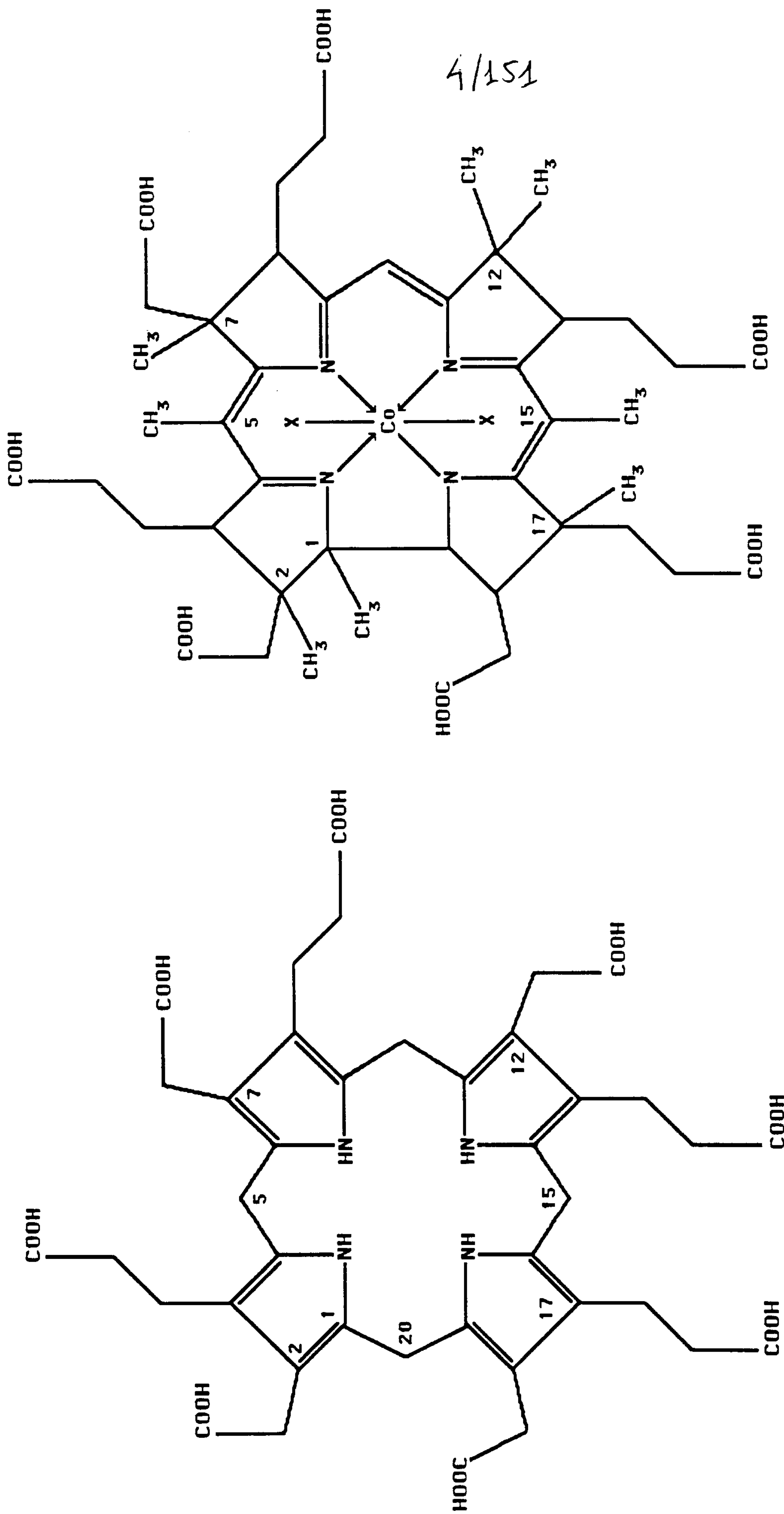


FIGURE 4

5 / 151

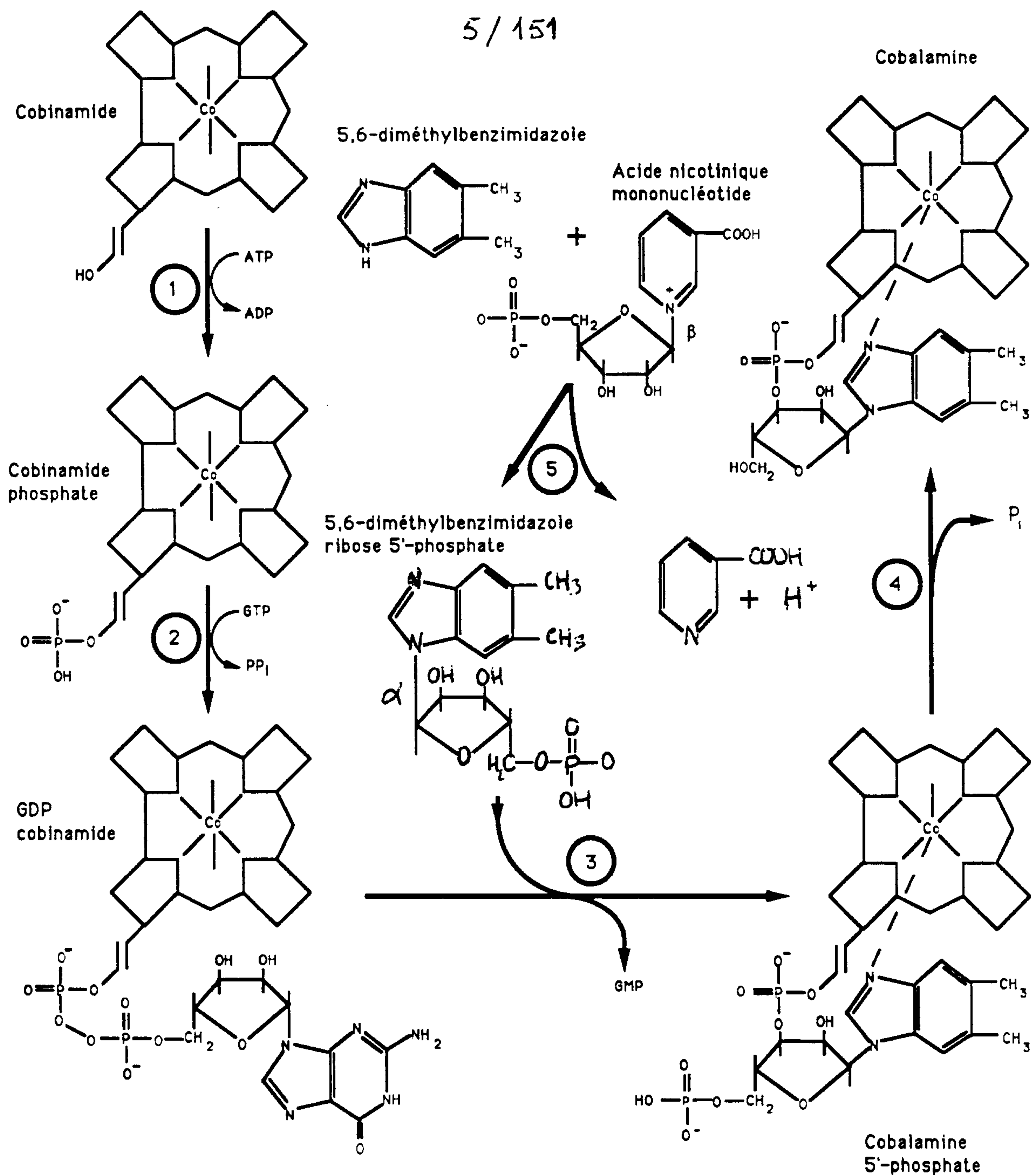


FIGURE 5

10	20	30	40	50	60
GGGCTGCAGG	TCGACTCTAG	AATCGATGAA	GCCTGCGATG	AAGGCGGCGA	CGAACAGGAA
CCCGACGTCC	AGCTGAGATC	TTAGCTACTT	CGGACGCTAC	TTCCGCCGCT	GCTTGTCTTT
70	80	90	100	110	120
GGCGAGCAGG	TGGAAGGCGA	GATCTTGAC	GGCGGGGACT	CGAGAGGAGA	GCTGTCAGGC
CCGCTCGTCC	ACCTTCCGCT	CTAGAACGTG	CCGCCCCCTGA	GCTCTCCTCT	CGACAGTCCG
130	140	150	160	170	180
GGGATTTTCC	GCCTTGTGTC	AGAGCCCGGC	GCGATTTGCA	AAGCCTTCTG	TCGCGGTGTT
CCCTAAAAGG	CGGAACACAG	TCTCGGGCCG	CGCTAAACGT	TTCCGAAGAC	AGCGCCACAA
190	200	210	220	230	240
GCTGTCCATG	CAGGTGTCGA	AATTGAAAAA	CCGACAAAGA	TTCAAAGCCT	TGTTCCAGCT
CGACAGGTAC	GTCCACAGCT	TTAACTTTTT	GGCTGTTTCT	AAGTTTCGGA	ACAAGGTCGA
250	260	270	280	290	300
CGCTGTCTTT	CTGGATGGAG	GCGCTCTCGC	CCGCATGGTG	CCGAAGAAGG	GCTGTCCTTG
GCGACAGAAA	GACCTACCTC	CGCGAGAGCG	GGCGTACCAC	GGCTTCTTCC	CGACAGGAAC
310	320	330	340	350	360
CGATACGGTA	GGCGGATGAC	GATCTTCCTC	AAACGCGACA	TGGCGATGGC	GCAATCCGGT
GCTATGCCAT	CCGCCTACTG	CTAGAAGGAG	TTTGCGCTGT	ACCGCTACCG	CGTTAGGCCA
370	380	390	400	410	420
TTGACCGGCC	TTCCGCGCTC	CGGTAAAAAT	GAAGGATATG	CGACGGCGTC	CGCTTTGGCG
AACTGGCCGG	AAGGCGCGAG	GCCATTTTTA	CTTCTATAC	GCTGCCGCAG	GCGAAACCGC
430	440	450	460	470	480
GACTGAAAGA	GCGTCCGGTG	CGGCCGACCC	AGTCAGGGGG	GCATCAGCCG	GTGCTGTCCA
CTGACTTTCT	CGCAGGCCAC	GCCGGCTGGG	TCAGTCCCCC	CGTAGTCGGC	CACGACAGGT
490	500	510	520	530	540
GATCGGCCGG	GACGGATCGT	CCCAGCCGGC	GCTTCGTAA	GGAGAACAAC	GAAGGGAGCC
CTAGCCGGCC	CTGCCTAGCA	GGGTCGGCCG	CGAAGCAATT	CCTCTTGTTG	CTTCCCTCGG
550	560	570	580	590	600
GGCCGCCGAT	GCCATCGGGC	CAACACTCTG	CACAGACGAC	GAAAGCAGGA	GCCGGGCTGG
CCGGCGGCTA	CGGTAGCCCC	GTTGTGAGAC	GTGTCTGCTG	CTTTCGTCCT	CGGCCCCGACC
610	620	630	640	650	660
TGCTCGGGCT	CGGCTGCGAG	CGTCGCACGC	CGGCCGAAGA	GGTGATCGCC	CTTGCCGAGC
ACGAGCCCGA	GCCGACGCTC	GCAGCGTGCG	GCCGGCTTCT	CCACTAGCGG	GAACGGCTCG
670	680	690	700	710	720
GTGCGCTTGC	CGATGCCGGT	GTTGCGCCCC	GCGATCTGCG	GCTGGTCGCC	TCGCTCGATG
CACGCGAACG	GCTACGGCCA	CAACGCGGGC	CGCTAGACGC	CGACCAGCGG	AGCGAGCTAC
730	740	750	760	770	780
CTCGCGCCGA	GGAGCCGGCG	ATCCTGGCGG	CCGCTCAGCA	TTTCGCGGTT	CCGGCCGCGT
GAGCGCGGCT	CCTCGGCCGC	TAGGACCGCC	GGCGAGTCGT	AAAGCGCCAA	GGCCGGCGCA
790	800	810	820	830	840
TCTACGATGC	CGCCACGCTC	GAAGCCGAAG	CTTCCCGGCT	CGCCAACCCG	TCCGAGATCG
AGATGCTACG	GCGGTGCGAG	CTTCGGCTTC	GAAGGGCCGA	GCGGTTGGGC	AGGCTCTAGC
850	860	870	880	890	900
TCTTTGCCTA	CACGGGTTGT	CATGGCGTTG	CCGAGGGTGC	AGCGCTCGTC	GGCGCCGGTC
AGAAACGGAT	GTGCCCAACA	GTACCGCAAC	GGCTCCACG	TCGCGAGCAG	CCGCGGCCAG
910	920	930	940	950	960
GCGAAGCCGT	GCTGATTGTG	CAGAAGATCG	TCTCCGCCCA	TGCGACGGCC	GCACTTGCCG
CGCTTCGGCA	CGACTAACAC	GTCTTCTAGC	AGAGGCGGGT	ACGCTGCCGG	CGTGAACGGC

FIGURE 7.1

8/151

970	980	990	1000	1010	1020
GGCCGGCGAC	CTTGCGCGCC	GAAAAGCGCA	TCCAGGCGGC	GGAGGCTGTC	TGATGCATTC
CCGGCCGCTG	GAACGCGCGG	CTTTTCGCGT	AGGTCCGCCG	CCTCCGACAG	ACTACGTAAG
1030	1040	1050	1060	1070	1080
TTATGTTGTT	GAATTGAATC	AATCTTTTGC	CCGGGGTTTC	TCTCAAGTGG	AATCCGGTTC
AATACAACAA	CTTAACCTAG	TTAGAAAACG	GGCCCCAAAG	AGAGTTCACC	TTAGGCCAAG
1090	1100	1110	1120	1130	1140
TTTAGAGAGC	GCGTCAGGCG	TGCCGTTGGG	TGGCGCCGAA	ATACAGGTGG	GACAGCACGC
AAATCTCTCG	CGCAGTCCGC	ACGGCAACCC	ACCGCGGCTT	TATGTCCACC	CTGTCGTGCG
1150	1160	1170	1180	1190	1200
ATGATCGACG	ACCTCTTTGC	CGGATTGCCG	GCGCTCGAAA	AAGGTTCCGT	CTGGCTGGTC
TACTAGCTGC	TGGAGAAACG	GCCTAACGGC	CGCGAGCTTT	TTCCAAGCCA	GACCGACCAG
1210	1220	1230	1240	1250	1260
GGCGCCGGCC	CCGGCGATCC	CGGCCTGTTG	ACGCTGCATG	CGGCCAATGC	GCTGCGCCAG
CCGCGGCCGG	GGCCGCTAGG	GCCGGACAAC	TGCGACGTAC	GCCGGTTACG	CGACGCGGTC
1270	1280	1290	1300	1310	1320
GCGGATGTGA	TCGTGCATGA	TGCGCTGGTC	AACGAGGATT	GCCTGAAGCT	CGCGCGGCCG
CGCCTACACT	AGCACGTACT	ACGCGACCAG	TTGCTCCTAA	CGGACTTCGA	GCGCGCCGGC
1330	1340	1350	1360	1370	1380
GGCGCCGTGC	TGGAGTTTGC	GGGCAAGCGT	GGCGGCAAGC	CGTCGCCGAA	GCAGCGCGAC
CCGCGGCACG	ACCTCAAACG	CCCGTTTCGA	CCGCCGTTTC	GCAGCGGCTT	CGTCGCGCTG
1390	1400	1410	1420	1430	1440
ATCTCGCTTC	GCCTCGTCGA	ACTCGCGCGC	GCCGGCAACC	GGGTGCTGCG	CCTCAAAGGC
TAGAGCGAAG	CGGAGCAGCT	TGAGCGCGCG	CGGCCGTTGG	CCCACGACGC	GGAGTTTCCG
1450	1460	1470	1480	1490	1500
GGCGATCCCT	TCGTCTTCGG	TCGCGGTGGC	GAGGAGGCGC	TGACGCTGGT	CGAACACCAG
CCGCTAGGGA	AGCAGAAGCC	AGCGCCACCG	CTCCTCCGCG	ACTGCGACCA	GCTTGTGGTC
1510	1520	1530	1540	1550	1560
GTGCCGTTCC	GAATCGTGCC	CGGCATCACC	GCCGGTATCG	GCGGGCTTGC	CTATGCCGGC
CACGGCAAGG	CTTAGCACGG	GCCGTAGTGG	CGGCCATAGC	CGCCCGAACG	GATACGGCCG
1570	1580	1590	1600	1610	1620
ATTCCCCTGA	CCCATCGCGA	GGTCAACCAC	GCGGTCACTT	TCCTGACTGG	CCATGATTCC
TAAGGGCACT	GGGTAGCGCT	CCAGTTGGTG	CGCCAGTGAA	AGGACTGACC	GGTACTAAGG
1630	1640	1650	1660	1670	1680
TCCGGCCTGG	TGCCGGATCG	CATCAACTGG	CAGGGCATCG	CCAGCGGCTC	GCCTGTCATC
AGGCCGGACC	ACGGCCTAGC	GTAGTTGACC	GTCCCGTAGC	GGTCGCCGAG	CGGACAGTAG
1690	1700	1710	1720	1730	1740
GTCATGTACA	TGGCGATGAA	ACATATCGGC	GCGATCACCG	CCAACCTCAT	TGCCGGCGGC
CAGTACATGT	ACCGCTACTT	TGTATAGCCG	CGCTAGTGCC	GGTTGGAGTA	ACGGCCGCCG
1750	1760	1770	1780	1790	1800
CGCTCGCCGG	ACGAACCGGT	CGCCTTCGTC	TGCAACGCCG	CGACGCCGCA	GCAGGCGGTG
GCGAGCGGCC	TGCTTGGCCA	GCGGAAGCAG	ACGTTGCGGC	GCTGCGGCGT	CGTCCGCCAC
1810	1820	1830	1840	1850	1860
CTGGAAACGA	CGCTTGCGCG	TGCAGAGGCC	GATGTTGCGG	CGGCAGGGCT	GGAGCCGCCG
GACCTTTGCT	GCGAACGCGC	ACGTCTCCGG	CTACAACGCC	GCCGTCCCGA	CCTCGGCGGC
1870	1880	1890	1900	1910	1920
GCGATCGTCG	TCGTGCGCGA	GGTGGTGCGG	CTGCGCGCAG	CGCTCGACTG	GATCGGCGCG

FIGURE 7.2

9/131

CGCAGCAGC AGCAGCCGCT CCACCACGCC GACGCGCGTC GCGAGCTGAC CTAGCCGCGC

1930	1940	1950	1960	1970	1980
CTGGACGGGC	GCAAGCTTGC	CGCCGACCCG	TTCGCCAATC	GCATTCTCAG	GAACCCGGCA
GACCTGCCCCG	CGTTCGAACG	GCGGCTGGGC	AAGCGGTTAG	CGTAAGAGTC	CTTGGGCCGT
1990	2000	2010	2020	2030	2040
TGAGCGGATT	GCTGATTGCC	GCACCCGCGT	CCGGCTCCGG	CAAGACGACG	GTGACGCTCG
ACTCGCCTAA	CGACTAACGG	CGTGGGCGCA	GGCCGAGGCC	GTTCTGCTGC	CACTGCGAGC
2050	2060	2070	2080	2090	2100
GGCTGATGCG	CGCCCTGAAG	AGGCGCGGCG	TGGCGATCGC	GCCCGGCAAG	GCGGGGCCGG
CCGACTACGC	GCGGGACTTC	TCCGCGCCGC	ACCGCTAGCG	CGGGCCGTTT	CGCCCCGGCC
2110	2120	2130	2140	2150	2160
ACTATATCGA	TCCCGCTTTC	CACGCGGCAG	CGACCGGCGA	GCCCTGCTTC	AACTACGACC
TGATATAGCT	AGGGCGAAAG	GTGCGCCGTC	GCTGGCCGCT	CGGGACGAAG	TTGATGCTGG
2170	2180	2190	2200	2210	2220
CCTGGGCGAT	GCGCCCGGAA	CTGCTGCTTG	CCAATGCGTC	GCATGTGGCC	TCCGGCGGGC
GGACCCGCTA	CGCGGGCCTT	GACGACGAAC	GGTTACGCAG	CGTACACCGG	AGGCCGCCCG
2230	2240	2250	2260	2270	2280
GCACATTGAT	CGTCGAGGCG	ATGATGGGAC	TGCATGACGG	TGCTGCCGAC	GGCTCGGGAA
CGTGTAACCTA	GCAGCTCCGC	TACTACCCTG	ACGTACTGCC	ACGACGGCTG	CCGAGCCCTT
2290	2300	2310	2320	2330	2340
CGCCAGCGGA	CCTCGCCGCG	ACGCTGAACC	TTGCGGTCAT	TCTGGTGGTC	GATTGCGCCC
GCGGTCGCCT	GGAGCGGCGC	TGCGACTTGG	AACGCCAGTA	AGACCACCAG	CTAACGCGGG
2350	2360	2370	2380	2390	2400
GCATGTCCCA	GTCGGTTGCC	GCCCTCGTGC	GCGGCTATGC	GGATCATCGC	GACGATATCC
CGTACAGGGT	CAGCCAACGG	CGGGAGCACG	CGCCGATACG	CCTAGTAGCG	CTGCTATAGG
2410	2420	2430	2440	2450	2460
GGGTGGTTGG	CGTCATCCTC	AACAAGGTCTG	GCAGCGATCG	GCATGAAATG	ATGCTGCGCG
CCCACCAACC	GCAGTAGGAG	TTGTTCCAGC	CGTCGCTAGC	CGTACTTTAC	TACGACGCGC
2470	2480	2490	2500	2510	2520
ATGCGCTCGG	CAAGGTGCGC	ATGCCTGTCT	TCGGCGTGCT	CCGGCAGGAC	AGCGCATTGC
TACGCGAGCC	GTTCCACGCG	TACGGACAGA	AGCCGCACGA	GGCCGTCCTG	TCGCGTAACG
2530	2540	2550	2560	2570	2580
AACTGCCGGA	GCGCCATCTC	GGGCTCGTGC	AGGCGGGCGA	ACACTCAGCG	CTTGAGGGCT
TTGACGGCCT	CGCGGTAGAG	CCCGAGCACG	TCCGCCCGCT	TGTGAGTCGC	GAACCTCCCGA
2590	2600	2610	2620	2630	2640
TCATCGAGGC	GGCGGCCGCG	CGGGTCGAGG	CTGCCTGCGA	TCTCGACGCC	ATCCGCCTGA
AGTAGCTCCG	CCGCCGGCGC	GCCCAGCTCC	GACGGACGCT	AGAGCTGCGG	TAGGCGGACT
2650	2660	2670	2680	2690	2700
TCGCGACGAT	TTTCCCGCAG	GTGCCCCGCG	CGGCCGATGC	CGAGCGTTTG	CGGCCGCTCG
AGCGCTGCTA	AAAGGGCGTC	CACGGGCGCC	GCCGGCTACG	GCTCGCAAAC	GCCGGCGAGC
2710	2720	2730	2740	2750	2760
GTCAGCGCAT	CGCGGTCGCG	CGCGATATCG	CCTTTGCCTT	CTGCTACGAG	CACCTGCTTT
CAGTCGCGTA	GCGCCAGCGC	GCGCTATAGC	GGAAACGGAA	GACGATGCTC	GTGGACGAAA
2770	2780	2790	2800	2810	2820
ACGGCTGGCG	GCAAGGCGGC	GCGGAGATTT	CCTTCTTCTC	GCCGCTCGCC	GACGAGGGGC
TGCCGACCGC	CGTTCCGCCG	CGCCTCTAAA	GGAAGAAGAG	CGGCGAGCGG	CTGCTCCCCG
2830	2840	2850	2860	2870	2880

FIGURE 7.3

10/151

CGGATGCGGC AGCCGATGCC GTCTATCTTC CGGGGGGTTA TCCGGAGCTG CATGCGGGGC
GCCTACGCCG TCGGCTACGG CAGATAGAAG GCCCCCAAT AGGCCTCGAC GTACGCCCCG

2890 2900 2910 2920 2930 2940
AGCTGAGCGC CGCCGCCCGA TTCCGTTCGG GCATGCATTC CGCGGCGGAA CGCGGCGCCC
TCGACTCGCG GCGGCGGGCT AAGGCAAGGC CGTACGTAAG GCGCCGCCTT GCGCCGCGGG

2950 2960 2970 2980 2990 3000
GCATCTTCGG CGAGTGCGGC GGCTATATGG TGCTCGGCGA AGGGCTTGTC GCTGCCGATG
CGTAGAAGCC GCTCACGCCG CCGATATAACC ACGAGCCGCT TCCCGAACAG CGACGGCTAC

3010 3020 3030 3040 3050 3060
GCACACGCTA CGACATGCTC GGCCTGCTGC CGCTCGTAAC CAGTTTTGCC GAGCGCAGGC
CGTGTGCGAT GCTGTACGAG CCGGACGACG GCGAGCATTG GTCAAAACGG CTCGCGTCCG

3070 3080 3090 3100 3110 3120
GGCACCTCGG CTATCGCCGC GTCGTGCCTG TCGACAACGC CTTCTTCGAT GGACCCATGA
CCGTGGAGCC GATAGCGGCG CAGCACGGAC AGCTGTTGCG GAAGAAGCTA CCTGGGTACT

3130 3140 3150 3160 3170 3180
CGGCGCACGA ATTCCACTAT GCGACCATCG TCGCCGAAGG GGCGGCCGAT CGGCTGTTTG
GCCGCGTGCT TAAGGTGATA CGCTGGTAGC AGCGGCTTCC CCGCCGGCTA GCCGACAAAC

3190 3200 3210 3220 3230 3240
CGGTCAGCGA CGCCGCCGGC GAGGATCTCG GCCAGGCGGG CCTCCGGCGC GGCCCTGTCC
GCCAGTCGCT GCGGCGGCCG CTCCTAGAGC CGGTCCGCCC GGAGGCCGCG CCGGGACAGC

3250 3260 3270 3280 3290 3300
CCGGTTCCTT CATGCATCTG ATCGACGTCG CAGGTGCTGC ATGAGCGCAC CGATCGTTCA
GGCCAAGGAA GTACGTAGAC TAGCTGCAGC GTCCACGACG TACTCGCGTG GCTAGCAAGT

3310 3320 3330 3340 3350 3360
TGGTGGCGGC ATCACCAGAG CCGCAGCGCG CTATGGCGGC CGGCCTGAAG ACTGGCTCGA
ACCACCGCCG TAGTGGCTCC GGCCTCGCGC GATACCGCCG GCCGGACTTC TGACCGAGCT

3370 3380 3390 3400 3410 3420
TCTGTCGACC GGCATCAATC CATGCCCCGT CGCCTTGCCC GCGGTCCCTG AGCGCGCCTG
AGACAGCTGG CCGTAGTTAG GTACGGGGCA GCGGAACGGG CGCCAGGGAC TCGCGCGGAC

3430 3440 3450 3460 3470 3480
GCACCGGCTG CCGGATCGGC AGACGGTAGA TGATGCGCGG AGCGCCGCCG CCGACTACTA
CGTGGCCGAC GGCCTAGCCG TCTGCCATCT ACTACGCGCC TCGCGGCGGC GGCTGATGAT

3490 3500 3510 3520 3530 3540
CCGCACCAAC GGCGTGCTGC CTTTGCCGGT GCCGGGCACC CAGTCGGTGA TCCAGCTCCT
GGCGTGTTG CCGCACGACG GAAACGGCCA CGGCCCGTGG GTCAGCCACT AGGTCGAGGA

3550 3560 3570 3580 3590 3600
GCCACGTCTT GCTCCGGCCA ACAGGCACGT CGCGATTTTC GGGCCGACCT ATGGCGAGTA
CGGTGCAGAA CGAGGCCGGT TGTCCGTGCA GCGCTAAAAG CCCGGCTGGA TACCGCTCAT

3610 3620 3630 3640 3650 3660
TGCCCGCGTG CTTGAAGCGG CCGGCTTTGC TGTCGATCGC GTCGCGGATG CCGACGCGCT
ACGGGCGCAC GAACTTCGCC GGCCGAAACG ACAGCTAGCG CAGCGCCTAC GGCTGCGCGA

3670 3680 3690 3700 3710 3720
CACGGCCGAA CATGGGCTTG TCATCGTCGT CAACCCCAAC AACCCGACCG GCCGCGCCTT
GTGCCGGCTT GTACCCGAAC AGTAGCAGCA GTTGGGGTTG TTGGGCTGGC CGGCGCGGAA

3730 3740 3750 3760 3770 3780
GGCGCCGGCG GAGCTTCTGG CGATCGCCGC AAGGCAGAAG GCGAGCGGCG GACTGCTGCT
CCGCGGCCGC CTCGAAGACC GCTAGCGGCG TTCCGTCTTC CGCTCGCCGC CTGACGACGA

FIGURE 7.4

11/151

3790	3800	3810	3820	3830	3840
GGTCGATGAG	GCCTTCGGCG	ATCTTGAGCC	GCAACTGAGT	GTCGCTGGTC	ACGCGTCAGG
CCAGCTACTC	CGGAAGCCGC	TAGAACTCGG	CGTTGACTCA	CAGCGACCAG	TGCGCAGTCC
3850	3860	3870	3880	3890	3900
GCAAGGCAAC	CTCATCGTCT	TCCGCTCCTT	CGGCAAGTTC	TTCGGCCTTG	CGGGCCTGCG
CGTTCCGTTG	GAGTAGCAGA	AGGCGAGGAA	GCCGTTCAAG	AAGCCGGAAC	GCCCGGACGC
3910	3920	3930	3940	3950	3960
CCTCGGCTTC	GTCGTTGCGA	CCGAGCCAGT	GCTTGCAATC	TTTGCCGATT	GGCTCGGTCC
GGAGCCGAAG	CAGCAACGCT	GGCTCGGTCA	CGAACGTAGG	AAACGGCTAA	CCGAGCCAGG
3970	3980	3990	4000	4010	4020
CTGGGCTGTC	TCCGGCCCCG	CGTTGACGAT	CTCGAAAGCG	CTGATGCAGG	GCGATACGAA
GACCCGACAG	AGGCCGGGCC	GCAACTGCTA	GAGCTTTCGC	GACTACGTCC	CGCTATGCTT
4030	4040	4050	4060	4070	4080
GGCGATCGCG	GCGGGCATCC	TCGAGCGTCG	CGCCGGCCTC	GATGCGGCTC	TCGATGGGGC
CCGCTAGCGC	CGCCCGTAGG	AGCTCGCAGC	GCGGCCGGAG	CTACGCCGAG	AGCTACCCCG
4090	4100	4110	4120	4130	4140
AGGGCTCAAC	CGTATCGGCG	GCACGGGGCT	ATTCGTGCTG	GTCGAGCATC	CCAGGGCAGC
TCCCGAGTTG	GCATAGCCGC	CGTGCCCCGA	TAAGCACGAC	CAGCTCGTAG	GGTCCCGTCG
4150	4160	4170	4180	4190	4200
TCTGCTGCAG	GAGCGGCTCT	GCGAGGCCCA	TATTCTCACG	CGCAAGTTCG	ACTATGCCCC
AGACGACGTC	CTCGCCGAGA	CGCTCCGGGT	ATAAGAGTGC	GCGTTCAAGC	TGATACGGGG
4210	4220	4230	4240	4250	4260
GACCTGGCTC	AGGGTCGGTC	TTGCGCCTGA	CGCGGCTGGT	GACCGACGGC	TGGCGGACGC
CTGGACCGAG	TCCAGCCAG	AACGCGGACT	GCGCCGACCA	CTGGCTGCCG	ACCGCCTGCG
4270	4280	4290	4300	4310	4320
GCTTGCCCCG	ATGGAGCTCT	GAGGTGTCGG	AGACGATCCT	GCTCATTCTC	GCGCTGGCGC
CGAACGGGCG	TACCTCGAGA	CTCCACAGCC	TCTGCTAGGA	CGAGTAAGAG	CGCGACCGCG
4330	4340	4350	4360	4370	4380
TGGTGATCGA	CCGCGTTGTC	GGCGATCCGG	ACTGGCTCTG	GGCGCGCGTG	CCGCATCCGG
ACCACTAGCT	GGCGCAACAG	CCGCTAGGCC	TGACCGAGAC	CCGCGCGCAC	GGCGTAGGCC
4390	4400	4410	4420	4430	4440
TCGTGTTTTT	CGGCAAGGCC	ATCGGCTTTT	TCGACGCGCG	GCTGAACCGG	GAGGACCTCG
AGCACAAAAA	GCCGTTCCGG	TAGCCGAAAA	AGCTGCGCGC	CGACTTGGCC	CTCCTGGAGC
4450	4460	4470	4480	4490	4500
AGGATAGCGC	GCGCAAATTT	CGTGGCGTCG	TCGCGATCCT	TTTGTTGCTT	GGCATCAGCG
TCCTATCGCG	CGCGTTTAAA	GCACCGCAGC	AGCGCTAGGA	AAACAACGAA	CCGTAGTCGC
4510	4520	4530	4540	4550	4560
CCTGGTTCGG	CCATCTGCTG	CATCGCCTGT	TCGCCGTCCT	CGGACCGCTC	GGCTTTCTGC
GGACCAAGCC	GGTAGACGAC	GTAGCGGACA	AGCGGCAGGA	GCCTGGCGAG	CCGAAAGACG
4570	4580	4590	4600	4610	4620
TCGAGGCGGT	TCTGGTCGCG	GTCTTCCTGG	CACAGAAGAG	CCTCGCCGAT	CACGTGCGTC
AGCTCCGCCA	AGACCAGCGC	CAGAAGGACC	GTGTCTTCTC	GGAGCGGCTA	GTGCACGCAG
4630	4640	4650	4660	4670	4680
GCGTGGCCGG	GGGCTTGCGA	CAGGGCGGGC	TGGAAGGCGG	GCGTGCCGCC	GTGTGATGA
CGCACCGGCC	CCCGAACGCT	GTCCCGCCCG	ACCTTCCGCC	CGCACGGCGG	CACAGCTACT
4690	4700	4710	4720	4730	4740
TCGTTGGTCG	CGATCCAAAG	ACGCTCGACG	AGCCGGCGGT	CTGCCGTGCC	GCGATCGAAA
AGCAACCAGC	GCTAGGTTTC	TGCGAGCTGC	TCGGCCGCCA	GACGGCACGG	CGCTAGCTTT

FIGURE 7.5

12/151

4750	4760	4770	4780	4790	4800
GCCTTGCCGA	GAATTTCTCC	GACGGCGTCG	TGGCGCCGGC	CTTCTGGTAC	GCGGTTGCCG
CGGAACGGCT	CTTAAAGAGG	CTGCCGCAGC	ACCGCGGCCG	GAAGACCATG	CGCCAACGGC
4810	4820	4830	4840	4850	4860
GCCTGCCGGG	GCTTCTTGCC	TACAAGATGC	TGAACACCGC	CGATTGATG	ATCGGCCACA
CGGACGGCCC	CGAAGAACGG	ATGTTCTACG	ACTTGTGGCG	GCTAAGCTAC	TAGCCGGTGT
4870	4880	4890	4900	4910	4920
AGTCGCCGAA	ATATCTGCAC	TTCGGCTGGG	CCTCGGCCCG	ACTCGACGAT	CTCGCCAACC
TCAGCGGCTT	TATAGACGTG	AAGCCGACCC	GGAGCCGGGC	TGAGCTGCTA	GAGCGGTTGG
4930	4940	4950	4960	4970	4980
TGCCGGCAGC	GAGGCTCTCG	ATCCTTTTGA	TCTCAGCCGG	TGCGCTGATC	CATCGTGCGC
ACGGCCGTCG	CTCCGAGAGC	TAGGAAAACT	AGAGTCGGCC	ACGCGACTAG	GTAGCACC GC
4990	5000	5010	5020	5030	5040
CCAGCGCCGC	CAAGGATGCG	CTGACCGTGG	CCCTTCGCGA	CCATGGCCTG	CACCGCTCGC
GGTCGCGGCG	GTTCTTACGC	GACTGGCACC	GGGAAGCGCT	GGTACCGGAC	GTGGCGAGCG
5050	5060	5070	5080	5090	5100
CGAACTCCGG	CTGGCCGGAA	GCGGCCATGG	CCGGCGCGCT	CGATCTGCAG	CTTGCCGGTC
GCTTGAGGCC	GACCGGCCTT	CGCCGGTACC	GGCCGCGCGA	GCTAGACGTC	GAACGGCCAG
5110	5120	5130	5140	5150	5160
CGCGGATCTA	TGGCGGCGTC	AAGGTCAGCG	AACCTATGAT	CAACGGTCCG	GGCCGAGCGG
GCGCCTAGAT	ACCGCCGCAG	TTCCAGTCGC	TTGGATACTA	GTTGCCAGGC	CCGGCTCGCC
5170	5180	5190	5200	5210	5220
TTGCAACAAG	CGAAGACATC	GACGCCGGTA	TTGCTGTATT	TTATGGCGCC	TGTACGGTCA
AACGTTGTTC	GCTTCTGTAG	CTGCGGCCAT	AACGACATAA	AATACCGCGG	ACATGCCAGT
5230	5240	5250	5260	5270	5280
TGGCCGGGTT	TGTTCTTGCA	ATCGCAATGA	TTTGATCGCG	GAAGTTGACC	TTCGCATTAA
ACCGGCCCAA	ACAAGAACGT	TAGCGTTACT	AAACTAGCGC	CTTCAACTGG	AAGCGTAATT
5290	5300	5310	5320	5330	5340
GACTCTGCTT	TCCATATGTA	TTAAGATCGT	ATCATATTCG	ATCAGTTATT	CTCCTGGAAC
CTGAGACGAA	AGGTATACAT	AATTCTAGCA	TAGTATAAGC	TAGTCAATAA	GAGGACCTTG
5350	5360	5370	5380	5390	5400
GTTTGGTTCC	ACCGGTACGT	GTTCTGCTTC	CCGGAGAGAG	AAGCATGCGC	AAAAGCTT
CAAACCAAGG	TGGCCATGCA	CAAGCAGAAG	GGCCTCTCTC	TTCGTACGCG	TTTTCGAA

FIGURE 7.6

13/151

10	20	30	40	50	60
GAATTCGCCA	GCGCCTACAT	GGCTGACCTC	AAGCAGTTCC	TCGTGGCCCA	GAAGAACGAG
CTTAAGCGGT	CGCGGATGTA	CCGACTGGAG	TTCGTCAAGG	AGCACCGGGT	CTTCTTGCTC
70	80	90	100	110	120
GGCCGGCAGA	TTTTCCTCTG	CGGGCCTGAG	TATTTTCGCG	CGCTCGACCT	GACGCCGCTC
CCGGCCGTCT	AAAAGGGAGC	GCCCGGACTC	ATAAAAGCGC	GCGAGCTGGA	CTGCGGCGAG
130	140	150	160	170	180
GACAAGGTGC	GCGTGGTCAT	TCTCGGCCAG	GATCCCTATC	ACGGTGACGG	CCAGGCGCAT
CTGTTCCACG	CGCACCAGTA	AGAGCCGGTC	CTAGGGATAG	TGCCACTGCC	GGTCCGCGTA
190	200	210	220	230	240
GGGCTCTGCT	TCAGCGTTTCG	CCCCGGTGTC	CGGACGCCGC	CGTCGCTGGT	CAACATCTAC
CCCGAGACGA	AGTCGCAAGC	GGGGCCACAG	GCCTGCGGCG	GCAGCGACCA	GTTGTAGATG
250	260	270	280	290	300
AAGGAACTGA	ATACCGATCT	CGGTATTCCG	CCGGCGCGTC	ACGGTTTTCT	CGAAAGCTGG
TTCCTTGACT	TATGGCTAGA	GCCATAAGGC	GGCCGCGCAG	TGCCAAAAGA	GCTTTCGACC
310	320	330	340	350	360
GCAAGGCAGG	GCGTGCTGCT	TTTGAACAGC	GTGCTGACGG	TAGAGCGCGG	GAACGTGCGT
CGTTCCGTCC	CGCACGACGA	AAACTTGTCG	CACGACTGCC	ATCTCGCGCC	CTTGACGCA
370	380	390	400	410	420
CACACCAGGG	TCACGGTTGG	GAAAAGTTCA	CGGATGCGAT	CATCCGTGCG	GTCAACGAGG
GTGTGGTCCC	AGTGCCAACC	CTTTTCAAGT	GCCTACGCTA	GTAGGCACGC	CAGTTGCTCC
430	440	450	460	470	480
CCGAGCATCC	CGTCGTCTTC	ATGCTTTGGG	GCTCCTATGC	GCAGAAGAAG	GCGGCCTTCG
GGCTCGTAGG	GCAGCAGAAG	TACGAAACCC	CGAGGATACG	CGTCTTCTTC	CGCCGGAAGC
490	500	510	520	530	540
TCGACCGCTC	GCGCCATCTT	GTCCTGAGGG	CACCACATCC	GTCGCCGCTC	TCAGCCCATT
AGCTGGCGAG	CGCGGTAGAA	CAGGACTCCC	GTGGTGTAGG	CAGCGGCGAG	AGTCGGGTAA
550	560	570	580	590	600
CCGGCTTTCT	CGGCTGCCGG	CATTTTTCCT	AGGCCAATGC	CTTCCTCGAA	AGCAAAGGCT
GGCCGAAAGA	GCCGACGGCC	GTAAAAAGGG	TCCGGTTACG	GAAGGAGCTT	TCGTTTCCGA
610	620	630	640	650	660
TCGATCCGAT	CGACTGGCGG	CTGCCGGAAG	ATCCGGCTGC	GGACATCAAC	TGAAGGCTTG
AGCTAGGCTA	GCTGACCGCC	GACGGCCTTT	TAGGCCGACG	CCTGTAGTTG	ACTTCCGAAC
670	680	690	700	710	720
GCGCGAATGA	CGGCTTTGTC	GTCGCCCTGA	GGTCTTGCCCT	TGGCGGCGGC	GATCCGCCTA
CGCGCTTACT	GCCGAAACAG	CAGCGGGACT	CCAGAACGGA	ACCGCCGCCG	CTAGGCGGAT
730	740	750	760	770	780
AGACGCCCGA	ACGAAATGGC	GGAGGCGGGC	ATGCGCAAAA	TTCTGATCAT	CGGCATCGGT
TCTGCGGGCT	TGCTTTACCG	CCTCCGCCCG	TACGCGTTTT	AAGACTAGTA	GCCGTAGCCA
790	800	810	820	830	840
TCGGGCAATC	CCGAACACAT	GACCGTGACG	GCGATCAACG	CGCTGAACTG	CGCCGACGTG
AGCCCGTTAG	GGCTTGTGTA	CTGGCACGTC	CGCTAGTTGC	GCGACTTGAC	GCGGCTGCAC
850	860	870	880	890	900
CTCTTTATCC	CGACCAAGGG	AGCGAAGAAG	ACCGAGCTTG	CCGAAGTGCG	CCGCGACATC
GAGAAATAGG	GCTGGTTCCC	TCGCTTCTTC	TGGCTCGAAC	GGCTTCACGC	GGCGCTGTAG
910	920	930	940	950	960
TGCGCCCGCT	ACGTCACGCG	CAAGGACAGC	CGCACCGTCG	AGTTCGCGGT	GCCCGTGCGG
ACGCGGGCGA	TGCAGTGCGC	GTTCTGTGTC	GCGTGGCAGC	TCAAGCGCCA	CGGGCACGCC

FIGURE 8.1

14/151

970	980	990	1000	1010	1020
CGCACCGAAG	GCGTCAGCTA	TGACGGCAGC	GTCGATGACT	GGCACGCCCA	GATCGCTGGG
GCGTGCTTC	CGCAGTCGAT	ACTGCCGTCG	CAGCTACTGA	CCGTGCGGGT	CTAGCGACCC
1030	1040	1050	1060	1070	1080
ATTTACGAAG	CGCTTCTATC	GAAGGAGTTG	GGCGAAGAGG	GAAGTGGCGC	GTTTCTCGTC
TAAATGCTTC	GCGAAGATAG	CTTCCTCAAC	CCGCTTCTCC	CTTGACCGCG	CAAAGAGCAG
1090	1100	1110	1120	1130	1140
TGGGGCGACC	CGATGCTCTA	TGACAGCACC	ATTCGCATCG	TCGAGCGGGT	CAAGGCACGC
ACCCCGCTGG	GCTACGAGAT	ACTGTCGTGG	TAAGCGTAGC	AGCTCGCCCA	GTTCCGTGCG
1150	1160	1170	1180	1190	1200
GGTGAGGTCG	CCTTCGCCTA	CGACGTCATT	CCCGGGATCA	CCAGTCTGCA	GGCGCTTTGC
CCACTCCAGC	GGAAGCGGAT	GCTGCAGTAA	GGGCCCTAGT	GGTCAGACGT	CCGCGAAACG
1210	1220	1230	1240	1250	1260
GCCAGCCACC	GCATTCCGCT	GAACCTCGTC	GGCAAGCCGG	TGGAGATCAC	CACGGGGCGT
CGGTGCGTGG	CGTAAGGCGA	CTTGGAGCAG	CCGTTCGGCC	ACCTCTAGTG	GTGCCCCGCA
1270	1280	1290	1300	1310	1320
CGGCTGCACG	AAAGCTTTCC	CGAGAAGAGC	CAGACCTCGG	TCGTCATGCT	CGATGGCGAA
GCCGACGTGC	TTTCGAAAGG	GCTCTTCTCG	GTCTGGAGCC	AGCAGTACGA	GCTACCGCTT
1330	1340	1350	1360	1370	1380
CAGGCGTTTC	AGCGGGTCGA	GGACCCGGAG	GCGGAGATCT	ATTGGGGCGC	CTATCTCGGC
GTCCGCAAAG	TCGCCCAGCT	CCTGGGCCTC	CGCCTCTAGA	TAACCCCGCG	GATAGAGCCG
1390	1400	1410	1420	1430	1440
ACGCGGGATG	AGATCGTCAT	TTCCGGCCGC	GTGGCTGAGG	TGAAGGACCG	GATCCTTGAA
TGCGCCCTAC	TCTAGCAGTA	AAGGCCGGCG	CACCGACTCC	ACTTCCTGGC	CTAGGAACTT
1450	1460	1470	1480	1490	1500
ACGCGGGCGG	CGGCGCGCGC	GAAGATGGGA	TGGATCATGG	ACATCTATCT	CCTGCGCAAG
TGCGCCCGCC	GCCGCGCGCG	CTTCTACCCT	ACCTAGTACC	TGTAGATAGA	GGACGCGTTC
1510	1520	1530	1540	1550	1560
GGCGCCGACT	TCGACGAGTG	ACGGGGAGGG	CCGATCTGCG	TCGTGTTTGA	TCTCACTCAA
CCGCGGCTGA	AGCTGCTCAC	TGCCCCCTCC	GGCTAGACGC	AGCACAAACT	AGAGTGAGTT
1570	1580	1590	1600	1610	1620
GGTTTGCGGC	TGTGTTATAG	CGTCTTAAGA	GGCTTCTTCA	GGGAGGAGAA	CCTCAAGTGA
CCAAACGCCG	ACACAATATC	GCAGAATTCT	CCGAAGAAGT	CCCTCCTCTT	GGAGTTCACT
1630	1640	1650	1660	1670	1680
TGACGGATTT	GATGACCAGC	TGCGCCCTTC	CATTGACCGG	AGATGCCGGC	ACCGTCGCTT
ACTGCCTAAA	CTACTGGTCG	ACGCGGGAAG	GTAAGTGGCC	TCTACGGCCG	TGGCAGCGAA
1690	1700	1710	1720	1730	1740
CGATGCGCCG	CGGCGCCTGC	CCGTCCTTGG	CAGAGCCGAT	GCAGACCGGC	GACGGCCTGC
GCTACGCGGC	GCCGCGGACG	GGCAGGAACC	GTCTCGGCTA	CGTCTGGCCG	CTGCCGGACG
1750	1760	1770	1780	1790	1800
TCGTGAGGGT	GAGGCCAACG	GATGACAGCC	TGACGCTGCC	GAAGGTCATT	GCCCTTGCCA
AGCACTCCCA	CTCCGGTTGC	CTACTGTGCG	ACTGCGACGG	CTTCCAGTAA	CGGGAACGGT
1810	1820	1830	1840	1850	1860
CGGCTGCCGA	GCGCTTCGGC	AATGGCATCA	TCGAGATTAC	CGCGCGCGGA	AACCTGCAGC
GCCGACGGCT	CGCGAAGCCG	TTACCGTAGT	AGCTCTAATG	GCGCGCGCCT	TTGGACGTCG
1870	1880	1890	1900	1910	1920
TTCGCGGCCT	GAGCGCGGCT	TCGGTGCCAA	GGCTGGCGCA	GGCGATCGGC	GATGCGGAGA

FIGURE 8.2

15/151

AAGCGCCGGA CTCGCGCCGA AGCCACGGTT CCGACCGCGT CCGCTAGCCG CTACGCCTCT

1930 1940 1950 1960 1970 1980
 TCGCCATTGC CGAGGGGCTC GCGATCGAGG TGCCGCCCCCT GGCCGGCATC GACCCGGACG
 AGCGGTAACG GCTCCCCGAG CGCTAGCTCC ACGGCGGGGA CCGGCCGTAG CTGGGCCTGC

1990 2000 2010 2020 2030 2040
 AGATCGCCGA TCCGCGGCCG ATTGCCACTG AGCTTCGTGA AGCGTTGGAT GTGCGCCAGG
 TCTAGCGGCT AGGCGCCGGC TAACGGTGAC TCGAAGCACT TCGCAACCTA CACGCGGTCC

2050 2060 2070 2080 2090 2100
 TGCCGTTGAA GCTTGACCCC AAATTATCCG TCGTCATCGA TAGCGGTGGC CGGTTTGCTC
 ACGGCAACTT CGAACGTGGG TTTAATAGGC AGCAGTAGCT ATCGCCACCG GCCAAACCAG

2110 2120 2130 2140 2150 2160
 TCGGCGCTGT CGTCGCCGAC ATTCGCCTTC AGGCGGTTTC GACTGTCGCG GGGGTGGCCT
 AGCCGCGACA GCAGCGGCTG TAAGCGGAAG TCCGCCAAAG CTGACAGCGC CCCCACCGGA

2170 2180 2190 2200 2210 2220
 GGGTGCTGTC GCTTGGCGGC ACGTCAACGA AGGCATCGAG CGTCGGGACG TTGGCCGGCA
 CCCACGACAG CGAACCGCCG TGCAGTTGCT TCCGTAGCTC GCAGCCCTGC AACC GGCCGT

2230 2240 2250 2260 2270 2280
 ACGCGGTCGT GCCGGCCCTG ATCACCATTTC TCGAGAACT GGCGAGCCTG GGCACGACGA
 TGCGCCAGCA CGGCCGGGAC TAGTGGTAAG AGCTCTTTGA CCGCTCGGAC CCGTGCTGCT

2290 2300 2310 2320 2330 2340
 TGCGCGGGCG CGATCTGGAC CCGTCGGAAA TCCGCGCGCT CTGTGCTGT GAGACATCGT
 ACGCGCCCGC GCTAGACCTG GGCAGCCTTT AGGCGCGCGA GACAGCGACA CTCTGTAGCA

2350 2360 2370 2380 2390 2400
 CCGAACGCCC GGCCGCTCCG CGTTCGGCCG CAATACCCGG CATTGATGCG CTGGGTAACG
 GGCTTGCGGG CCGGCGAGGC GCAAGCCGGC GTTATGGGCC GTAAGTACGC GACCCATTGC

2410 2420 2430 2440 2450 2460
 CCGACACCGT TCTCGGCCTC GGTCTGGCCT TTGCTCAGGT GGAGGCCGCC GCGCTGGCAT
 GGCTGTGGCA AGAGCCGGAG CCAGACCGGA AACGAGTCCA CCTCCGGCGG CGCGACCGTA

2470 2480 2490 2500 2510 2520
 CCTACCTGCA TCAGGTCCAG GCGCTTGGCG CCAATGCGAT CCGGCTTGCG CCCGGGCACG
 GGATGGACGT AGTCCAGGTC CGCGAACCGC GGTACGCTA GGCCGAACGC GGGCCCGTGC

2530 2540 2550 2560 2570 2580
 CCTTCTTCGT CCTCGGCCTT TGCCCCGAGA CCGCGGCTGT GGCGCAGAGC CTGGCAGCGT
 GGAAGAAGCA GGAGCCGGAA ACGGGGCTCT GGCGCCGACA CCGCGTCTCG GACCGTCGCA

2590 2600 2610 2620 2630 2640
 CACACGGTTT TCGCATTGCC GAGCAGGATC CGCGCAATGC GATCGCCACC TGCGCCGGCA
 GTGTGCCAAA AGCGTAACGG CTCGTCCTAG GCGCGTTACG CTAGCGGTGG ACGCGGCCGT

2650 2660 2670 2680 2690 2700
 GCAAGGGTTG CGCCTCGGCG TGGATGGAAA CCAAGGGCAT GGCCGAGCGC CTCGTCGAGA
 CGTTCCCAAC GCGGAGCCGC ACCTACCTTT GGTTCGCGTA CCGGCTCGCG GAGCAGCTCT

2710 2720 2730 2740 2750 2760
 CGGCGCCGGA ATTGCTCGAC GGGTCGCTCA CCGTGCATCT CTCCGGCTGC GCCAAGGGCT
 GCCGCGGCCT TAACGAGCTG CCCAGCGAGT GGCACGTAGA GAGGCCGACG CGGTTCCCGA

2770 2780 2790 2800 2810 2820
 GCGCCCGGCC GAAGCCGTCC GAACTGACGC TTGTGCGTGC GCCATCAGGA TACGGGCTTG
 CGCGGGCCGG CTTGCGCAGG CTTGACTGCG AACAGCCACG CGGTAGTCCT ATGCCCGAAC

2830 2840 2850 2860 2870 2880

FIGURE 8.3

16/151

TCGTAAATGG	GGCTGCCAAT	GGCTTGCCAA	GCGCCTACAC	CGATGAGAAT	GGAATGGGAT
AGCATTTACC	CCGACGGTTA	CCGAACGGTT	CGCGGATGTG	GCTACTCTTA	CCTTACCCTA
2890	2900	2910	2920	2930	2940
CCGCCCTTGC	CCGGCTCGGC	CGGCTGGTGC	GGCAAAACAA	AGACGCTGGC	GAATCGGCGC
GGCGGGAACG	GGCCGAGCCG	GCCGACCACG	CCGTTTTGTT	TCTGCGACCG	CTTAGCCGCG
2950	2960	2970	2980	2990	3000
AGTCCTGTCT	TACACGGCTC	GGAGCTGCGC	GCGTCTCGGC	AGCGTTCGAA	CAGGGATAGA
TCAGGACAGA	ATGTGCCGAG	CCTCGACGCG	CGCAGAGCCG	TCGCAAGCTT	GTCCCTATCT
3010	3020	3030	3040	3050	3060
CATGCCTGAG	TATGATTACA	TTCGCGATGG	CAACGCCATC	TACGAGCGTT	CCTTCGCCAT
GTACGGACTC	ATACTAATGT	AAGCGCTACC	GTTGCGGTAG	ATGCTCGCAA	GGAAGCGGTA
3070	3080	3090	3100	3110	3120
CATCCGCGCC	GAGGCCGATC	TGTCGCGCTT	CTCCGAAGAG	GAAGCGGATC	TGGCTGTGCG
GTAGGCGCGG	CTCCGGCTAG	ACAGCGCGAA	GAGGCTTCTC	CTTCGCCTAG	ACCGACACGC
3130	3140	3150	3160	3170	3180
CATGGTGCAC	GCCTGCGGTT	CCGTCGAGGC	GACCAGGCAG	TTCGTGTTTT	CTCCCGATTT
GTACCACGTG	CGGACGCCAA	GGCAGCTCCG	CTGGTCCGTC	AAGCACAAAA	GAGGGCTAAA
3190	3200	3210	3220	3230	3240
CGTAAGCTCG	GCCCGTGCGG	CGCTGAAAGC	CGGTGCGCCG	ATCCTCTGCG	ATGCCGAGAT
GCATTGAGC	CGGGCACGCC	GCGACTTTTC	GCCACGCGGC	TAGGAGACGC	TACGGCTCTA
3250	3260	3270	3280	3290	3300
GGTTGCGCAC	GGTGTACCCC	GCGCCCGTCT	GCCGGCCGGC	AACGAGGTGA	TCTGCACGCT
CCAACGCGTG	CCACAGTGGG	CGCGGGCAGA	CGGCCGGCCG	TTGCTCCACT	AGACGTGCGA
3310	3320	3330	3340	3350	3360
GCGCGATCCT	CGCACGCCCG	CACTTGCGGC	CGAGATCGGC	AACACCCGCT	CCGCCGCAGC
CGCGCTAGGA	GCGTGCGGGC	GTGAACGCCG	GCTCTAGCCG	TTGTGGGCGA	GGCGGCGTCG
3370	3380	3390	3400	3410	3420
CCTGAAGCTC	TGGAGCGAGC	GGCTGGCCGG	TTCGGTGGTC	GCGATCGGCA	ACGCGCCGAC
GGAATTGAG	ACCTCGCTCG	CCGACCGGCC	AAGCCACCAG	CGCTAGCCGT	TGCGCGGCTG
3430	3440	3450	3460	3470	3480
GGCGTTGTTC	TTCTTCTTGG	AAATGCTGCG	CGACGGCGCG	CCGAAGCCGG	CGGCAATCCT
CCGCAACAAG	AAGGAGAACC	TTTACGACGC	GCTGCCGCGC	GGCTTCGGCC	GCCGTTAGGA
3490	3500	3510	3520	3530	3540
CGGCATGCCC	GTCGGTTTCG	TCGGTGCGGC	GGAATCGAAG	GATGCGCTGG	CCGAGAACTC
GCCGTACGGG	CAGCCAAAGC	AGCCACGCCG	CCTTAGCTTC	CTACGCGACC	GGCTCTTGAG
3550	3560	3570	3580	3590	3600
CTATGGCGTT	CCCTTCGCCA	TCGTGCGCGG	CCGCCTCGGC	GGGAGTGCCA	TGACGGCGGC
GATACGCAA	GGGAAGCGGT	AGCACGCGCC	GGCGGAGCCG	CCCTCACGGT	ACTGCCGCCG
3610	3620	3630	3640	3650	3660
AGCGCTTAAC	TCGCTCGCGA	GGCCGGGCCT	GTGAGCGGCG	TCGGCGTGGG	GCGCCTGATC
TCGCGAATTG	AGCGAGCGCT	CCGGCCCCGA	CACTCGCCGC	AGCCGCACCC	CGCGGACTAG
3670	3680	3690	3700	3710	3720
GGTGTTGGGA	CCGGCCCCGG	TGATCCGGAA	CTTTTGACGG	TCAAGGCGGT	GAAGGCGCTC
CCACAACCCT	GGCCGGGGCC	ACTAGGCCTT	GAAAACTGCC	AGTTCCGCCA	CTTCCGCGAG
3730	3740	3750	3760	3770	3780
GGGCAAGCCG	ATGTGCTTGC	CTATTTTCGCC	AAGGCCGGGC	GAAGCGGTAA	CGGCCGCGCG
CCCGTTTCGGC	TACACGAACG	GATAAAGCGG	TTCCGGCCCC	CTTCGCCATT	GCCGGCGCGC

FIGURE 8.4

17/151

3790	3800	3810	3820	3830	3840
GTGGTCGAGG	GTCTGCTGAA	GCCCGATCTT	GTCGAGCTGC	CGCTATACTA	TCCGGTGACG
CACCAGCTCC	CAGACGACTT	CGGGCTAGAA	CAGCTCGACG	GCGATATGAT	AGGCCACTGC
3850	3860	3870	3880	3890	3900
ACCGAAATCG	ACAAGGACGA	TGGCGCCTAC	AAGACCCAGA	TCACCGACTT	CTACAATGCG
TGGCTTTAGC	TGTTCCCTGCT	ACCGCGGATG	TTCTGGGTCT	AGTGGCTGAA	GATGTTACGC
3910	3920	3930	3940	3950	3960
TCGGCCGAAG	CGGTAGCGGC	GCATCTTGCC	GCCGGGCGCA	CGGTGCGCGT	GCTCAGTGAA
AGCCGGCTTC	GCCATCGCCG	CGTAGAACGG	CGGCCCGCGT	GCCAGCGGCA	CGAGTCACTT
3970	3980	3990	4000	4010	4020
GGCGACCCGC	TGTTCTATGG	TTCCTACATG	CATCTGCATG	TGCGGCTCGC	CAATCGTTTC
CCGCTGGGCG	ACAAGATACC	AAGGATGTAC	GTAGACGTAC	ACGCCGAGCG	GTTAGCAAAG
4030	4040	4050	4060	4070	4080
CCGGTCGAGG	TGATCCCCGG	CATTACCGCC	ATGTCCGGCT	GTTGGTCGCT	TGCCGGCCTG
GGCCAGCTCC	ACTAGGGGCC	GTAATGGCGG	TACAGGCCGA	CAACCAGCGA	ACGGCCGGAC
4090	4100	4110	4120	4130	4140
CCGCTGGTGC	AGGGCGACGA	CGTGCTCTCG	GTGCTTCCGG	GCACCATGGC	CGAGGCCGAG
GGCGACCACG	TCCCGCTGCT	GCACGAGAGC	CACGAAGGCC	CGTGGTACCG	GCTCCGGCTC
4150	4160	4170	4180	4190	4200
CTCGGCCGCA	GGCTTGCGGA	TACCGAAGCC	GCCGTGATCA	TGAAGGTCGG	GCGCAATTTG
GAGCCGGCGT	CCGAACGCCT	ATGGCTTCGG	CGGCACTAGT	ACTTCCAGCC	CGCGTTAAAC
4210	4220	4230	4240	4250	4260
CCGAAGATCC	GTCGGGCGCT	CGCTGCCTCC	GGCCGTCTCG	ACCAGGCCGT	CTATGTCGAA
GGCTTCTAGG	CAGCCCGCGA	GCGACGGAGG	CCGGCAGAGC	TGGTCCGGCA	GATACAGCTT
4270	4280	4290	4300	4310	4320
CGCGGCACGA	TGAAGAACGC	GGCGATGACG	GCTCTTGCGG	AAAAGGCCGA	CGACGAGGCG
GCGCCGTGCT	ACTTCTTGCG	CCGCTACTGC	CGAGAACGCC	TTTTCCGGCT	GCTGCTCCGC
4330	4340	4350	4360	4370	4380
CCCTATTTCT	CGCTGGTGCT	CGTTCCCGGC	TGGAAGGACC	GACCATGACC	GGTACGCTCT
GGGATAAAGA	GCGACCACGA	GCAAGGGCCG	ACCTTCCTGG	CTGGTACTGG	CCATGCGAGA
4390	4400	4410	4420	4430	4440
ATGTCGTCGG	TACCGGACCG	GGCAGCGCCA	AGCAGATGAC	GCCGGAAACG	GCGGAAGCCG
TACAGCAGCC	ATGGCCTGGC	CCGTCGCGGT	TCGTCTACTG	CGGCCTTTGC	CGCCTTCGGC
4450	4460	4470	4480	4490	4500
TTGCGGCCGC	TCAGGAGTTT	TACGGCTACT	TTCCCTATCT	CGACCGGCTG	AACCTCAGAC
AACGCCGGCG	AGTCCTCAAA	ATGCCGATGA	AAGGGATAGA	GCTGGCCGAC	TTGGAGTCTG
4510	4520	4530	4540	4550	4560
CGGATCAGAT	CCGTGTCGCC	TCGGACAACC	GCGAGGAGCT	CGATCGGGCA	CAGGTCGCGC
GCCTAGTCTA	GGCACAGCGG	AGCCTGTTGG	CGCTCCTCGA	GCTAGCCCGT	GTCCAGCGCG
4570	4580	4590	4600	4610	4620
TGACGCGGGC	TGCGGCAGGC	GTGAAGGTCT	GCATGGTCTC	CGGTGGCGAT	CCCGGTGTCT
ACTGCGCCCG	ACGCCGTCCG	CACTTCCAGA	CGTACCAGAG	GCCACCGCTA	GGGCCACAGA
4630	4640	4650	4660	4670	4680
TTGCCATGGC	GGCCGCCGTC	TGCGAGGCGA	TCGACAAGGG	ACCGGCGGAA	TGGAAGTCGG
AACGGTACCG	CCGGCGGCAG	ACGCTCCGCT	AGCTGTTCCC	TGGCCGCCTT	ACCTTCAGCC
4690	4700	4710	4720	4730	4740
TTGAACTGGT	GATCACGCCC	GGCGTGACCG	CGATGCTCGC	CGTTGCCGCC	CGCATCGGCG
AACTTGACCA	CTAGTGCGGG	CCGCACTGGC	GCTACGAGCG	GCAACGGCGG	GCGTAGCCGC

FIGURE 8.5

18/151

4750	4760	4770	4780	4790	4800
CGCCGCTCGG	TCATGATTTC	TGTGCGATCT	CGCTTTCCGA	CAATCTGAAG	CCCTGGGAAG
GCGGCGAGCC	AGTACTAAAG	ACACGCTAGA	GCGAAAGGCT	GTTAGACTTC	GGGACCCTTC
4810	4820	4830	4840	4850	4860
TCATCACCCG	GCGTCTCAGG	CTGGCGGCGG	AAGCGGGCTT	CGTCATTGCC	CTCTACAATC
AGTAGTGGGC	CGCAGAGTCC	GACCGCCGCC	TTCGCCCCGA	GCAGTAACGG	GAGATGTTAG
4870	4880	4890	4900	4910	4920
CGATCAGCAA	GGCGCGGCCC	TGGCAGCTCG	GTGAGGCCTT	CGAGCTTCTG	CGCAGCGTTC
GCTAGTCGTT	CCGCGCCGGG	ACCGTCGAGC	CACTCCGGAA	GCTCGAAGAC	GCGTCGCAAG
4930	4940	4950	4960	4970	4980
TGCCGGCAAG	CGTTCCGGTC	ATCTTCGGCC	GTGCGGGCCG	GCGGCCGGAC	GAACGGATCG
ACGGCCGTTC	GCAAGGCCAG	TAGAAGCCGG	CACGCCGGCC	CGCCGGCCTG	CTTGCCTAGC
4990	5000	5010	5020	5030	5040
CGGTGATGCC	GCTCGGCGAG	GCCGATGCCA	ACCGCGCCGA	CATGGCGACC	TGCGTCATCA
GCCACTACGG	CGAGCCGCTC	CGGCTACGGT	TGGCGCGGCT	GTACCGCTGG	ACGCAGTAGT
5050	5060	5070	5080	5090	5100
TCGGCTCGCC	GGAGACGCGC	ATCGTCGAGC	GCGACGGCCA	ACCCGATCTC	GTCTACACAC
AGCCGAGCGG	CCTCTGCGCG	TAGCAGCTCG	CGCTGCCGGT	TGGGCTAGAG	CAGATGTGTG
5110	5120	5130	5140	5150	5160
CGCGCTTCTA	TGCAGGGGCG	AGCCAGTGAG	CGATGCGGTT	GAGTGCCTCG	TCGCAACTGC
GCGCGAAGAT	ACGTCCCCGC	TCGGTCACTC	GCTACGCCAA	CTCACGGAGC	AGCGTTGACG
5170	5180	5190	5200	5210	5220
CGACCGTCGG	CACGTCCGCG	GGCTTGCGCC	GCTCGACCAT	GATCACCTCG	ATGCCGAGCC
GCTGGCAGCC	GTGCAGGCGC	CCGAACGCGG	CGAGCTGGTA	CTAGTGGAGC	TACGGCTCGG
5230	5240	5250	5260	5270	5280
GGCGCGCTGC	GGCAATCTTG	CCGTAGGTGG	CGCTGCCACC	GCTGTTCTTG	GCGACGATCA
CCGCGCGACG	CCGTTAGAAC	GGCATCCACC	GCGACGGTGG	CGACAAGAAC	CGCTGCTAGT
5290	5300	5310	5320	5330	5340
CATCGATCTG	CCGACTCCTG	AGCAACGCGG	CTTCGTGCGC	TTCCGCAAAG	GGACCGGTTC
GTAGCTAGAC	GGCTGAGGAC	TCGTTGCGCC	GAAGCAGCCG	AAGGCGTTTC	CCTGGCCAGC
5350	5360	5370	5380	5390	5400
CCAGGATCGC	CTCCTGGTCG	GGCAGATTAA	GCGGCGGCGT	CACCGGATCG	ACGCTGCGGA
GGTCCTAGCG	GAGGACCAGC	CCGTCTAATT	CGCCGCCGCA	GTGGCCTAGC	TGCGACGCCT
5410	5420	5430	5440	5450	5460
TGACGTAGCT	GTGCTGCGGC	GCGACCTCGA	AGTGGAAAGC	TTCCTGTCTG	CCTATCGCCA
ACTGCATCGA	CACGACGCCG	CGCTGGAGCT	TCACCTTTTC	AAGGACAGCT	GGATAGCGGT
5470	5480	5490	5500	5510	5520
GGAAGACGCG	GCGTCGCCGA	TCACCGAGCG	CGCTGACGGC	CTCGACAACG	CTATCGACAG
CCTTCTGCGC	CGCAGCGGCT	AGTGGCTCGC	GCGACTGCCG	GAGCTGTTGC	GATAGCTGTC
5530	5540	5550	5560	5570	5580
CAGTCCAGCG	GTCGCCAGGC	AGGGGCACCC	ATTCCGGTCG	GCGGAGGGCG	ATAAGCGCAA
GTCAGGTCGC	CAGCGGTCCG	TCCCCGTGGG	TAAGGCCAGC	CGCCTCCCGC	TATTCGCGTT
5590	5600	5610	5620	5630	5640
CGCCGGTTCT	TTGCGCTGCG	TCCGCGGCGT	TGTGCGAAAT	GCGTGCGGCA	AAGGGGTGCG
GCGGCCAAGA	AACGCGACGC	AGGCGCCGCA	ACACGCTTTA	CGCACGCCGT	TTCCCCACGC
5650	5660	5670	5680	5690	5700
TCGCATCGAC	CAGCAGCGCG	ATGTTTTTCG	CATGCACGAA	ATGCGCCAGC	CCATCGGCGC

FIGURE 8.6

19/151

AGCGTAGCTG GTCGTCGCGC TACAAAAGCA GTACGTGCTT TACGCGGTCG GGTAGCCGCG

5710 5720 5730 5740 5750 5760
 CGCCAAAGCC GCCGATGCGC GTCTTGACCG GCTGCGGCCG CGGGTCCGCG GTGCGGCCGG
 GCGGTTTCGG CCGCTACGCG CAGAACTGGC CGACGCCGGC GCCCAGGCGC CACGCCGGCC

5770 5780 5790 5800 5810 5820
 CCAGCGAGAT GGCGGTGTCG TAGCGGACAT CTTCCGGCCAA GCGGCGCGCG AGTTCGCGTG
 GGTCGCTCTA CCGCCACAGC ATCGCCTGTA GAAGCCGGTT CGCCGCGCGC TCAAGCGCAC

5830 5840 5850 5860 5870 5880
 CCTCGGTGGT GCCACCCAGA ATCAGAATAC GAGGTTTTTC CATGGCTGAC GTGTCGAACA
 GGAGCCACCA CCGTGGGTCT TAGTCTTATG CTCCAAAAAG GTACCGACTG CACAGCTTGT

5890 5900 5910 5920 5930 5940
 GCGAACCCGC CATAGTCTCC CCCTGGCTGA CCGTCATCGG TATCGGTGAG GATGGTGTAG
 CGCTTGGGCG GTATCAGAGG GGGACCGACT GGCAGTAGCC ATAGCCACTC CTACCACATC

5950 5960 5970 5980 5990 6000
 CGGGTCTCGG CGACGAGGCC AAGCGGCTGA TCGCCGAAGC GCCGGTCGTC TACGGCGGCC
 GCCCAGAGCC GCTGCTCCGG TTCGCCGACT AGCGGCTTCG CGGCCAGCAG ATGCCGCCGG

6010 6020 6030 6040 6050 6060
 ATCGTCATCT GGAGCTCGCC GCCTCCCTCA TCACCGGCGA AGCGCACAAT TGGCTAAGCC
 TAGCAGTAGA CCTCGAGCGG CGGAGGGAGT AGTGGCCGCT TCGCGTGTTA ACCGATTCGG

6070 6080 6090 6100 6110 6120
 CCCTCGAACG CTCGGTCGTC GAGATCGTCG CGCGTCGCGG CAGCCCGGTG GTGGTGCTTG
 GGGAGCTTGC GAGCCAGCAG CTCTAGCAGC GCGCAGCGCC GTCGGGCCAC CACCACGAAC

6130 6140 6150 6160 6170 6180
 CCTCGGGCGA CCCGTTCTTC TTCGGCGTCG GCGTGACGCT GGCGCGCCGC ATCGCCTCGG
 GGAGCCCGCT GGGCAAGAAG AAGCCGCAGC CGCACTGCGA CCGCGCGGCG TAGCGGAGCC

6190 6200 6210 6220 6230 6240
 CCGAAATACG CACGCTTCCG GCGCCGTCGT CGATCAGTCT TGCCGCCTCG CGCCTCGGCT
 GGCTTTATGC GTGCGAAGGC CGCGGCAGCA GCTAGTCAGA ACGGCGGAGC GCGGAGCCGA

6250 6260 6270 6280 6290 6300
 GGGCGCTGCA GGATGCGACG CTCGTCTCCG TACATGGGCG GCCGCTGGAT CTGGTGCGAC
 CCCGCGACGT CCTACGCTGC GAGCAGAGGC ATGTACCCGC CGGCGACCTA GACCACGCTG

6310 6320 6330 6340 6350 6360
 CGCATTTGCA TCCGGGGGCG CGTGTGCTTA CGCTCACGTC GGACGGTGCG GGTCCGCGAG
 GCGTAAACGT AGGCCCCCGC GCACACGAAT GCGAGTGCAG CCTGCCACGC CCAGGCGCTC

6370 6380 6390 6400 6410 6420
 ACCTTGCCGA GCTTCTGGTT TCAAGCGGCT TCGGTCAGTC GCGACTGACC GTGCTCGAAG
 TGGAACGGCT CGAAGACCAA AGTTCGCCGA AGCCAGTCAG CGCTGACTGG CACGAGCTTC

6430 6440 6450 6460 6470 6480
 CGCTGGGCGG CGCCGGCGAA CGGGTGACGA CGCAGATCGC CGCGCGCTTC ATGCTCGGCC
 GCGACCCGCC GCGGCCGCTT GCCCACTGCT GCGTCTAGCG GCGCGCGAAG TACGAGCCGG

6490 6500 6510 6520 6530 6540
 TCGTGCAATC TTTGAACGTC TGCGCCATTG AGGTGGCGGC CGACGAGGGC GCGCGCATCC
 AGCACGTAGG AACTTGCAG ACGCGGTAAC TCCACCGCCG GCTGCTCCCG CGCGCGTAGG

6550 6560 6570 6580 6590 6600
 TGCCGCTTGC CGCCGGCCGC GACGATGCGC TGTTCTGAACA TGACGGGCAG ATCACCAAGC
 ACGGCGAACG GCGGCCGGCG CTGCTACGCG ACAAGCTTGT ACTGCCCGTC TAGTGGTTTC

6610 6620 6630 6640 6650 6660

FIGURE 8.7

20/151

GCGAGGTGCG	GGCGCTGACG	CTGTCTGGCAC	TCGCACCGCG	CAAGGGCGAA	CTGCTATGGG
CGCTCCACGC	CCGCGACTGC	GACAGCCGTG	AGCGTGGCGC	GTTCCCGCTT	GACGATACCC
6670	6680	6690	6700	6710	6720
ACATCGGCGG	CGGCTCCGGC	TCGATCGGCA	TCGAATGGAT	GCTCGCCGAT	CCGACCATGC
TGTAGCCGCC	GCCGAGGCCG	AGCTAGCCGT	AGCTTACCTA	CGAGCGGCTA	GGCTGGTACG
6730	6740	6750	6760	6770	6780
AGGCGATCAC	CATCGAGGTT	GAGCCGGAGC	GGGCAGCGCG	CATCGGCCGC	AACGCGACGA
TCCGCTAGTG	GTAGCTCCAA	CTCGGCCTCG	CCCGTCGCGC	GTAGCCGGCG	TTGCGCTGCT
6790	6800	6810	6820	6830	6840
TGTTCTGGCGT	GCCCGGGCTG	ACGGTTGTCTG	AAGGCGAGGC	GCCGGCGGCG	CTTGCCGGCC
ACAAGCCGCA	CGGGCCCGAC	TGCCAACAGC	TTCCGCTCCG	CGGCCGCCGC	GAACGGCCGG
6850	6860	6870	6880	6890	6900
TGCCACAACC	GGACGCGATC	TTCATCGGCG	GCGGCGGCAG	CGAAGACGGC	GTCATGGAAG
ACGGTGTGTTG	CCTGCGCTAG	AAGTAGCCGC	CGCCGCCGTC	GCTTCTGCCG	CAGTACCTTC
6910	6920	6930	6940	6950	6960
CAGCGATCGA	GGCGCTCAAG	TCAGGCGGAC	GGCTGGTTGC	CAACGCGGTG	ACGACGGACA
GTCGCTAGCT	CCGCGAGTTC	AGTCCGCCTG	CCGACCAACG	GTTGCGCCAC	TGCTGCCTGT
6970	6980	6990	7000	7010	7020
TGGAAGCGGT	GCTGCTCGAT	CATCACGCGC	GGCTCGGCGG	TTCGCTGATC	CGCATCGATA
ACCTTCGCCA	CGACGAGCTA	GTAGTGCGCG	CCGAGCCGCC	AAGCGACTAG	GCGTAGCTAT
7030	7040	7050	7060	7070	7080
TCGCGCGTGC	AGGACCCATC	GGCGGCATGA	CCGGCTGGAA	GCCGGCCATG	CCGGTCACCC
AGCGCGCACG	TCCTGGGTAG	CCGCCGTACT	GGCCGACCTT	CGGCCGGTAC	GGCCAGTGGG
7090	7100	7110	7120	7130	7140
AATGGTTCGTG	GACGAAGGGC	TAAAGCAGTT	CCAGCGAAAG	TGTGACGCGG	TTTTGCGTCC
TTACCAGCAC	CTGCTTCCCG	ATTTCTGTC	GGTCGCTTTC	AACTGCGCC	AAAACGCAGG
7150	7160	7170	7180	7190	7200
GGAAGTGC	AAGAAAAAGA	AAGAGTAACC	TATGACGGTA	CATTTTCATCG	GCGCCGGCCC
CCTTGACGCG	TTCTTTTCT	TTCTCATTGG	ATACTGCCAT	GTAAAGTAGC	CGCGGCCGGG
7210	7220	7230	7240	7250	7260
GGGAGCCGCA	GACCTGATCA	CGGTGCGTGG	TCGCGACCTG	ATCGGGCGCT	GCCCGGTCTG
CCCTCGGCGT	CTGGACTAGT	GCCACGCACC	AGCGCTGGAC	TAGCCCGCGA	CGGGCCAGAC
7270	7280	7290	7300	7310	7320
CCTTTACGCC	GGCTCGATCG	TCTCGCCGGA	GCTGCTGCGA	TATTGCCCGC	CGGGCGCCCC
GGAAATGCGG	CCGAGCTAGC	AGAGCGGCCT	CGACGACGCT	ATAACGGGCG	GCCCGCGGGC
7330	7340	7350	7360	7370	7380
CATTGTCGAT	ACGGCGCCGA	TGTCCCTCGA	CGAGATCGAG	GCGGAGTATG	TGAAGGCCGA
GTAACAGCTA	TGCCGCGGCT	ACAGGGAGCT	GCTCTAGCTC	CGCCTCATAC	ACTTCCGGCT
7390	7400	7410	7420	7430	7440
AGCCGAAGGG	CTCGACGTGG	CGCGGCTTCA	TTCGGGCGAC	CTTTCGGTCT	GGAGTGCTGT
TCGGCTTCCC	GAGCTGCACC	GCGCCGAAGT	AAGCCCGCTG	GAAAGCCAGA	CCTCACGACA
7450	7460	7470	7480	7490	7500
GGCCGAACAG	ATCCGCCGGC	TCGAGAAGCA	TGGCATCGCC	TATACGATGA	CGCCGGGCGT
CCGGCTTGTC	TAGGCGGCCG	AGCTCTTCGT	ACCGTAGCGG	ATATGCTACT	GCGGCCCGCA
7510	7520	7530	7540	7550	7560
TCCTTCCTTT	GCGGCGGCGG	CTTCAGCGCT	CGGTGCGGAA	TTGACCATTC	CGGCCGTGGC
AGGAAGGAAA	CGCCGCCGCC	GAAGTCGCGA	GCCAGCGCTT	AACTGGTAAG	GCCGGCACCG

FIGURE 8.8

21/151

7570	7580	7590	7600	7610	7620
CCAGAGCCTG	GTGCTGACCC	GCGTTTCGGG	CCGCGCCTCG	CCGATGCCGA	ACTCAGAAAC
GGTCTCGGAC	CACGACTGGG	CGCAAAGCCC	GGCGCGGAGC	GGCTACGGCT	TGAGTCTTTG
7630	7640	7650	7660	7670	7680
GCTTTCCGCT	TTCGGCGCTA	CGGGATCGAC	GCTGGCAATC	CACCTTGCGA	TCCATGCGCT
CGAAAGGCGA	AAGCCGCGAT	GCCCTAGCTG	CGACCGTTAG	GTGGAACGCT	AGGTACGCGA
7690	7700	7710	7720	7730	7740
TCAGCAGGTG	GTCGAGGAAC	TGACGCCGCT	CTACGGTGCC	GACTGCCCCG	TCGCCATCGT
AGTCGTCCAC	CAGCTCCTTG	ACTGCGGCGA	GATGCCACGG	CTGACGGGCC	AGCGGTAGCA
7750	7760	7770	7780	7790	7800
CGTCAAGGCC	TCCTGGCCGG	ACGAACGCGT	GGTGCGCGGC	ACGCTCGGTG	ACATCGCCGC
GCAGTTCCGG	AGGACCGGCC	TGCTTGCGCA	CCACGCGCCG	TGCGAGCCAC	TGTAGCGGCG
7810	7820	7830	7840	7850	7860
CAAGGTGGCG	GAAGAGCCGA	TCGAGCGCAC	GGCGCTGATC	TTCGTGCGTC	CGGGGCTCGA
GTTCCACCGC	CTTCTCGGCT	AGCTCGCGTG	CCGCGACTAG	AAGCAGCCAG	GCCCCGAGCT
7870	7880	7890	7900	7910	7920
AGCCTCCGAT	TTCCGTGAAA	GCTCGCTCTA	CGATCCCGCC	TATCAGCGGC	GCTTCAGAGG
TCGGAGGCTA	AAGGCACTTT	CGAGCGAGAT	GCTAGGGCGG	ATAGTCGCCG	CGAAGTCTCC
7930	7940	7950	7960	7970	7980
GCGCGGCGAA	TAGGCCGCAC	TCCCTCGGGG	GTCGGCCTAA	GTTTCCCGCT	GAGAGGGTTT
CGCGCCGCTT	ATCCGGCGTG	AGGGAGCCCC	CAGCCGGATT	CAAAGGGCGA	CTCTCCCAA
7990	8000	8010	8020	8030	8040
TGAAACCTAT	TCTGCCGGTT	CTTCGCGCGG	CGGCCGCTGC	TTGAGCGGGA	CGCCGCGCTT
ACTTTGGATA	AGACGGCCAA	GAAGCGCGCC	GCCGGCGACG	AACTCGCCCT	GCGGCGCGAA
8050	8060	8070	8080	8090	8100
TTCCTCGACG	CGGTGCGGGT	AGAGCGCTGC	CTGTCCAAGC	AGCATCAGCG	TCACCGGCGT
AAGGAGCTGC	GCCAGCGCCA	TCTCGCGACG	GACAGGTTTC	TCGTAGTCGC	AGTGGCCGCA
8110	8120	8130	8140	8150	8160
GGTGGCGACG	ACGAAGACGA	TGATCAGGAT	TTCGTGGAAT	ACCCAGCGGC	TCTGCAGCAC
CCACCGCTGC	TGCTTCTGCT	ACTAGTCCTA	AAGCACCTTA	TGGGTCGCCG	AGACGTCGTG
8170	8180	8190	8200	8210	8220
GGCAAAGCAG	ATGATAGAGG	CGGCGCAGAT	CATCAGTACG	CCGCCGCTGG	TCGCCAGCGT
CCGTTTCGTC	TACTATCTCC	GCCGCGTCTA	GTAGTCATGC	GGCGGCGACC	AGCGGTCGCA
8230	8240	8250	8260	8270	8280
CGGTGCGTGC	AGGCGCTCGT	AGAAGCTGGT	GAACCGGAGC	AAGCCGACGG	AGCCGATCAG
GCCACGCACG	TCCGCGAGCA	TCTTCGACCA	CTTGGCCTCG	TTCGGCTGCC	TCGGCTAGTC
8290	8300	8310	8320	8330	8340
CGCCACTGCG	GCGCCGAGGA	CGGTGAGCCC	GCAGACGAGA	ACGGCTGCCC	AGACGGGAAG
GCGGTGACGC	CGCGGCTCCT	GCCACTCGGG	CGTCTGCTCT	TGCCGACGGG	TCTGCCCTTC
8350	8360	8370	8380	8390	8400
GTCGGTGAGG	TGGCTCATTC	GATGATCTCC	CCGCGCATCA	GGAACCTGCC	GAAGGCGATC
CAGCCACTCC	ACCGAGTAAG	CTACTAGAGG	GGCGCGTAGT	CCTTGAACGG	CTTCCGCTAG
8410	8420	8430	8440	8450	8460
GACGAGACGA	AGCCGATCAA	AGCCACGATC	AGGGCGGACT	CGAAATAGAG	CGAGTTGGCC
CTGCTCTGCT	TCGGCTAGTT	TCGGTGCTAG	TCCCGCCTGA	GCTTTATCTC	GCTCAACCGG
8470	8480	8490	8500	8510	8520
GTGCGGATGC	CGAAGGTCAA	GAGCATCAGC	ATGGCGTTGA	TATAGAGCGT	GTCGAGGCCG
CACGCCTACG	GCTTCCAGTT	CTCGTAGTCG	TACCGCAACT	ATATCTCGCA	CAGCTCCGGC

FIGURE 8.9

22/151

8530	8540	8550	8560	8570	8580
AGGATACGGT	CCTGGGCGCG	CGGTCCCCTC	ACCATGCGAT	AGAAGGCAAA	AGCCATCGCC
TCCTATGCCA	GGACCCGCGC	GCCAGGGGAG	TGGTACGCTA	TCTTCCGTTT	TCGGTAGCGG
8590	8600	8610	8620	8630	8640
AGGCCGAGCA	TGATCTGGGC	AATCAGGATC	GACCAGATGA	TTGAAAGTTC	CATCATCCGA
TCCGGCTCGT	ACTAGACCCG	TTAGTCCTAG	CTGGTCTACT	AACTTTCAAG	GTAGTAGGCT
8650	8660	8670	8680	8690	8700
ATATCTCCTT	CAGGGCGGTC	TCATAGCGCT	TGACCGTATC	GAGCCAGATG	TCCTCGTTCT
TATAGAGGAA	GTCCCGCCAG	AGTATCGCGA	ACTGGCATAG	CTCGGTCTAC	AGGAGCAAGA
8710	8720	8730	8740	8750	8760
CCATGTCGAG	CACGTGGAAG	AGCAGGGACT	TGCGGCCGCG	ATCCGGGGAA	TTC
GGTACAGCTC	GTGCACCTTC	TCGTCCCTGA	ACGCCGGCGC	TAGGCCCTT	AAG

FIGURE 8.10

23 / 151

SEQUENCE : FRAGMENT DE 5398 BP DE 1 A 1200 LONGUEUR = 1200

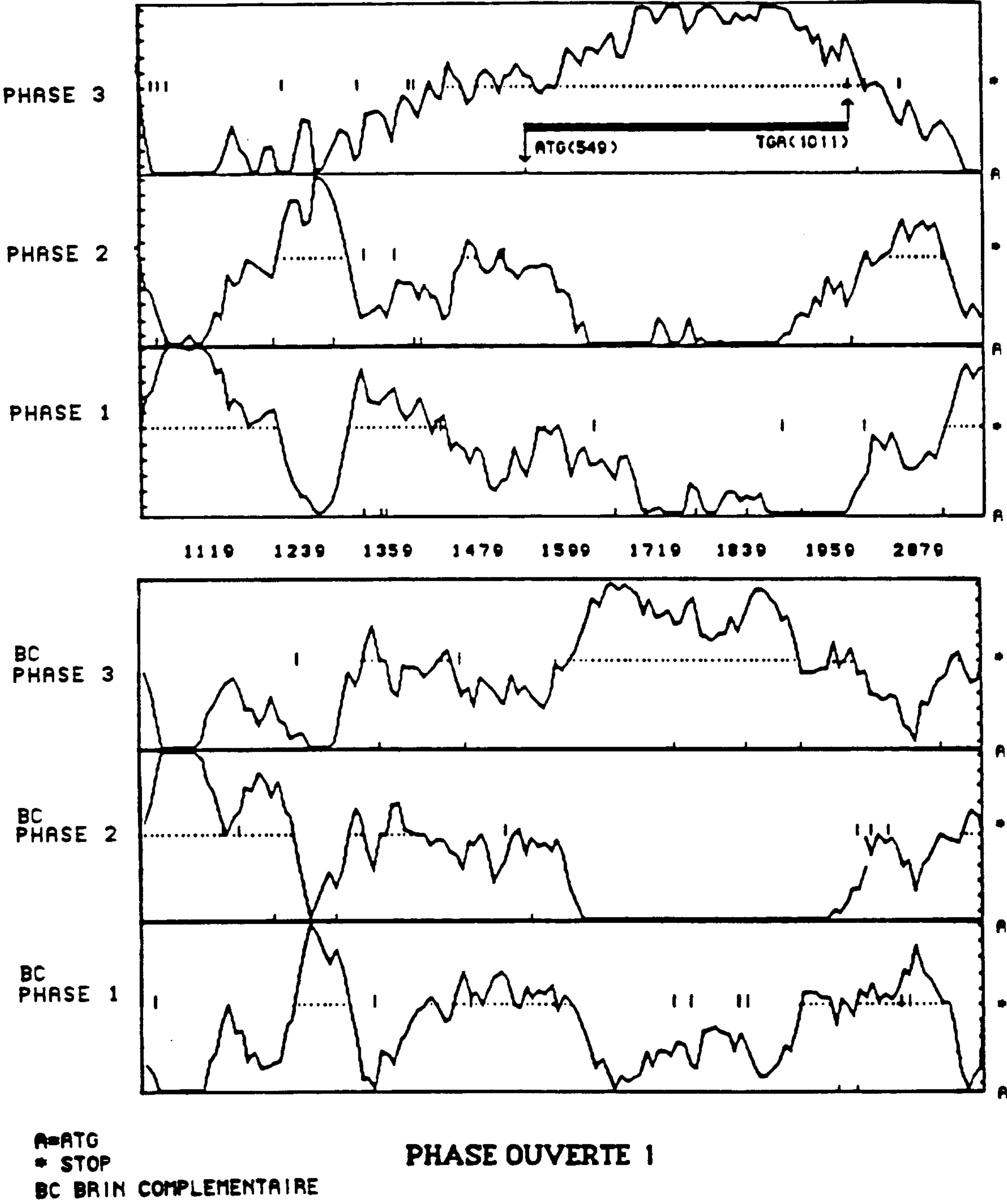
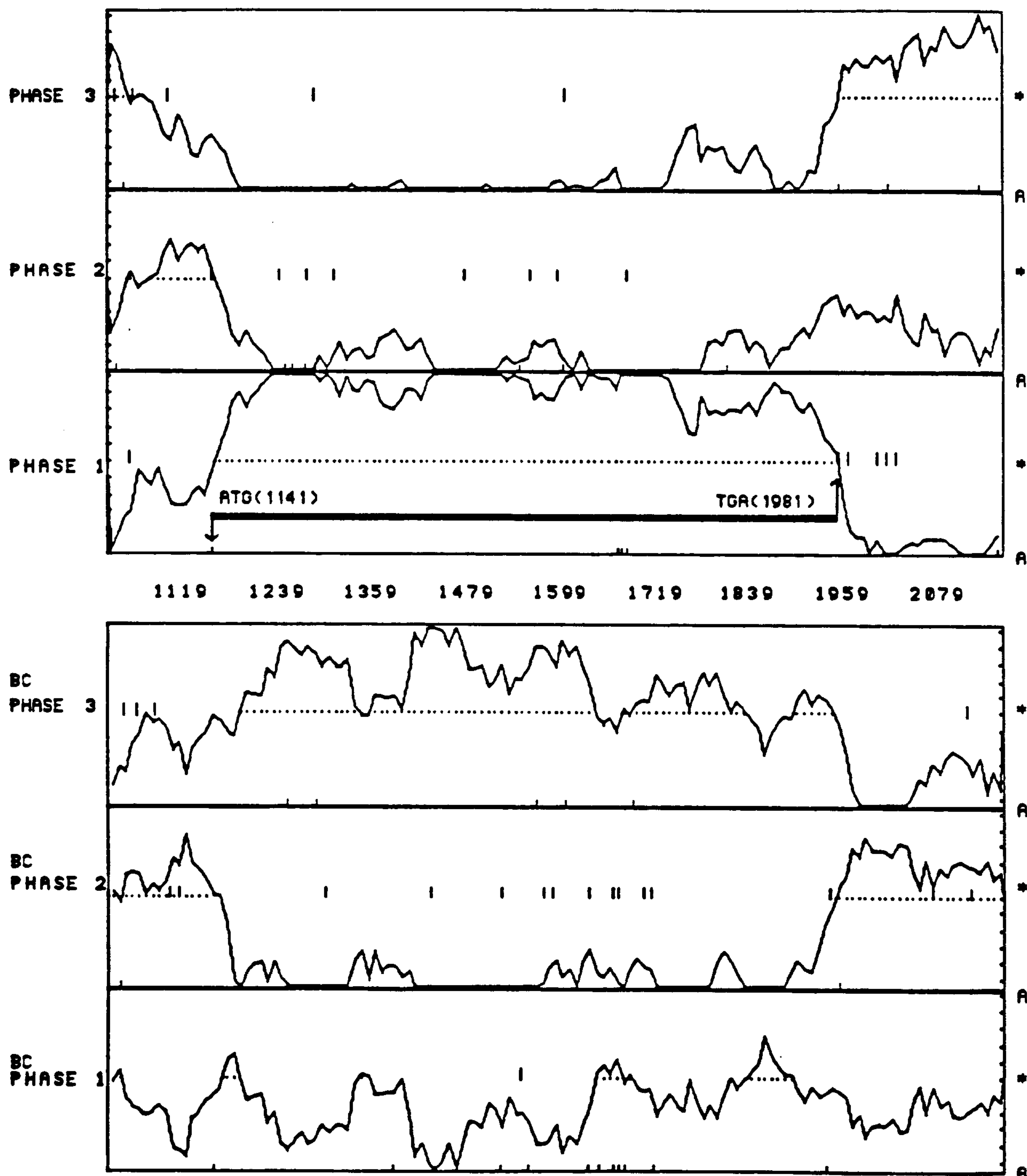


FIGURE 9.1

24/151

SEQUENCE: FRAGMENT DE 5398 BP DE 1000 A 2200 LONGUEUR = 1201 BP



A=ATG
 *: STOP
 BC: BRIN COMPLEMENTAIRE

PHASE OUVERTE 2

FIGURE 9.2

FEUILLE DE REMPLACEMENT

SEQUENCE : FRAGMENT DE 5398 BP DE 1800 A 3400 LONGUEUR = 1601

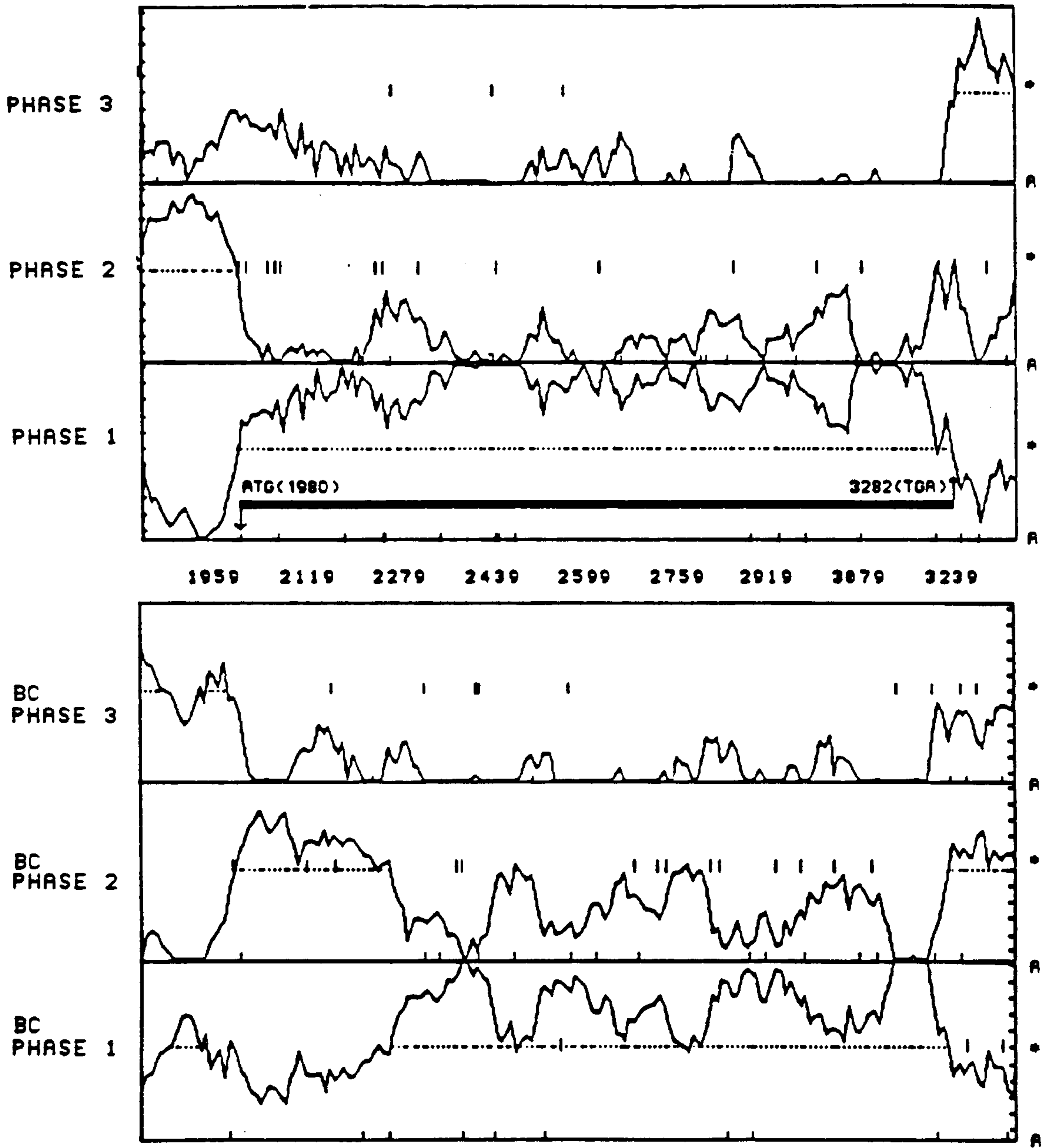


FIGURE 9.3

SEQUENCE : FRAGMENT DE 5398 BP DE 3000 A 4500 LONGUEUR = 1501

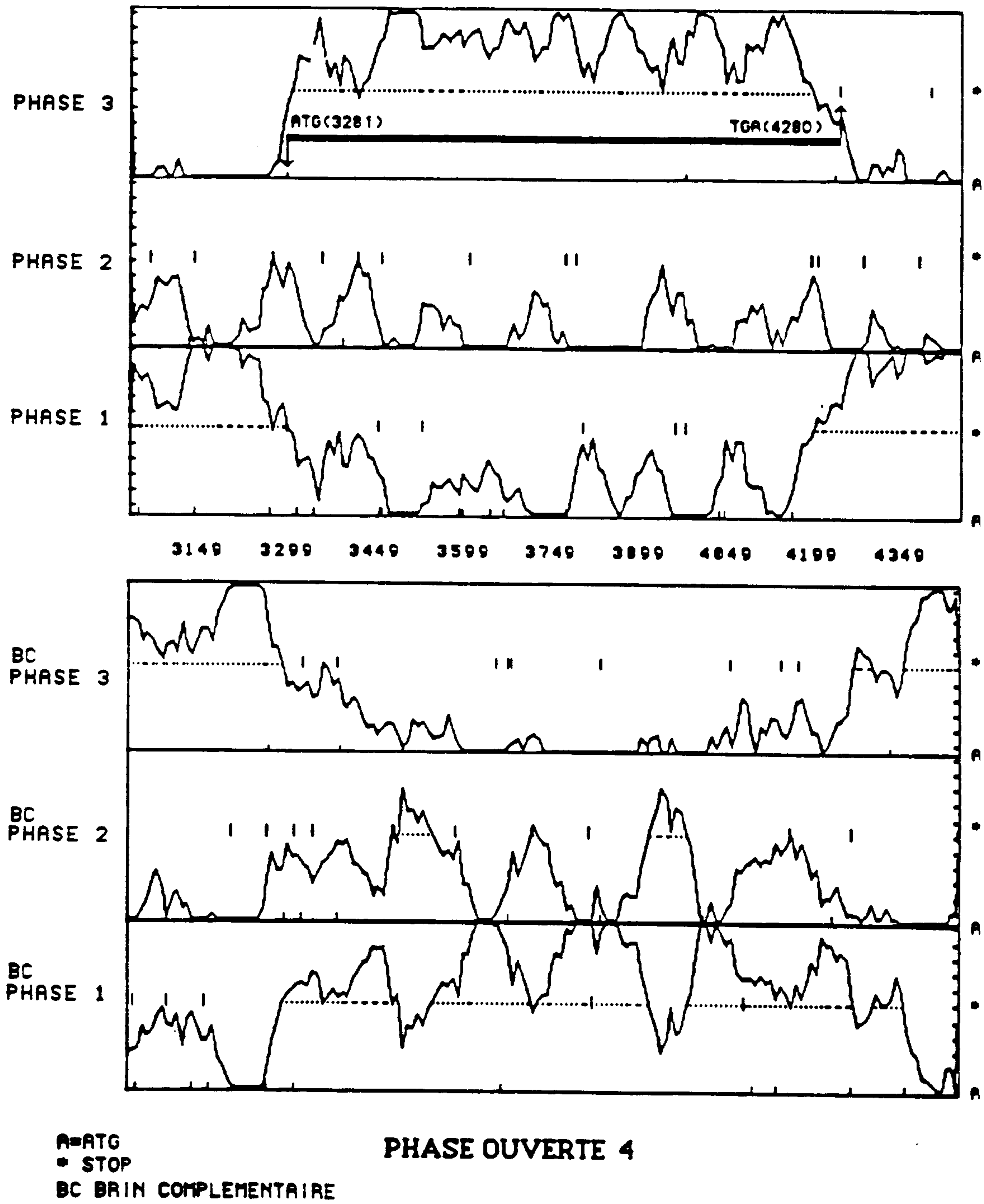


FIGURE 9.4

27 / 151

SEQUENCE : FRAGMENT DE 5398 BP DE 3800 A 5398 LONGUEUR = 1599

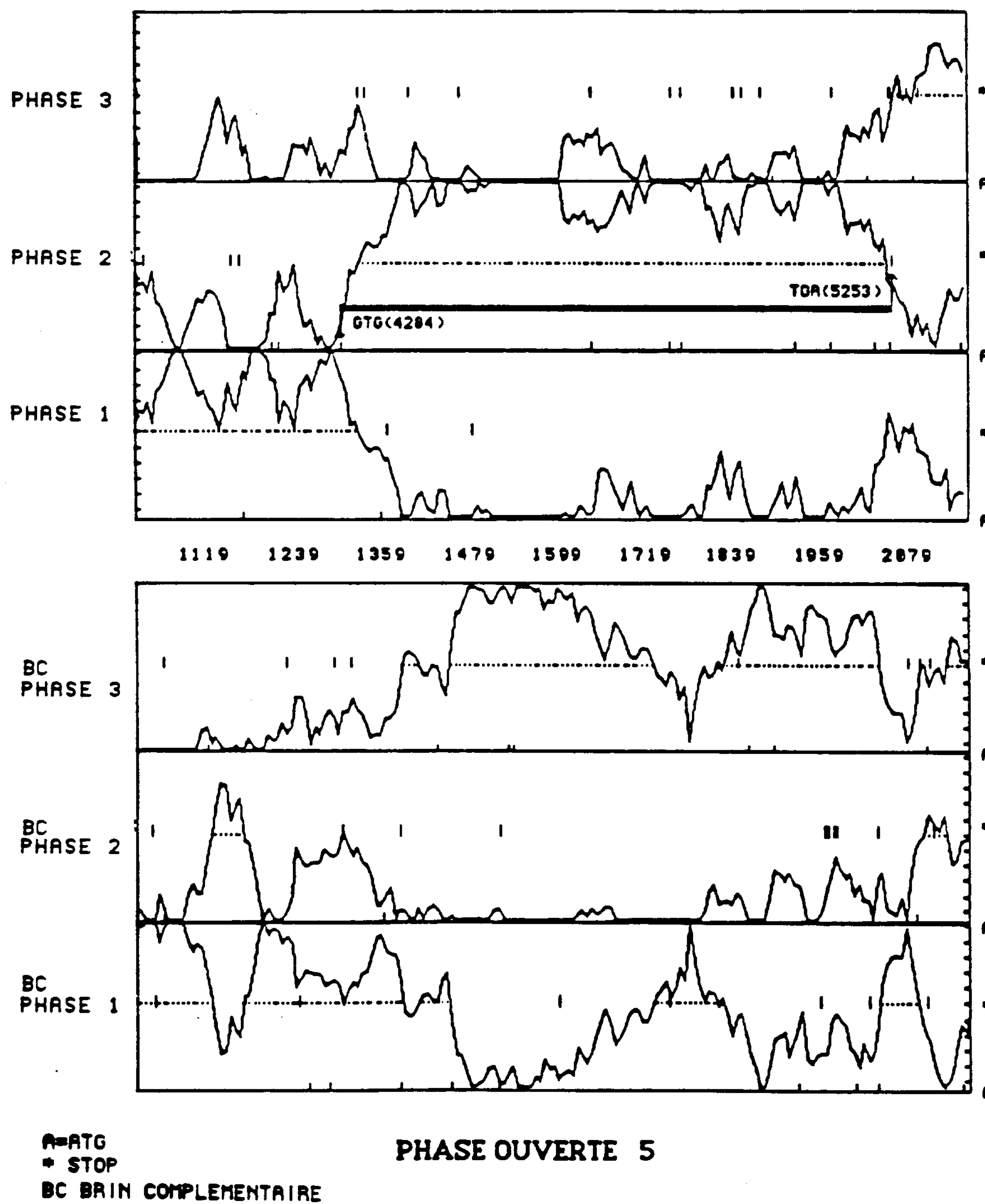
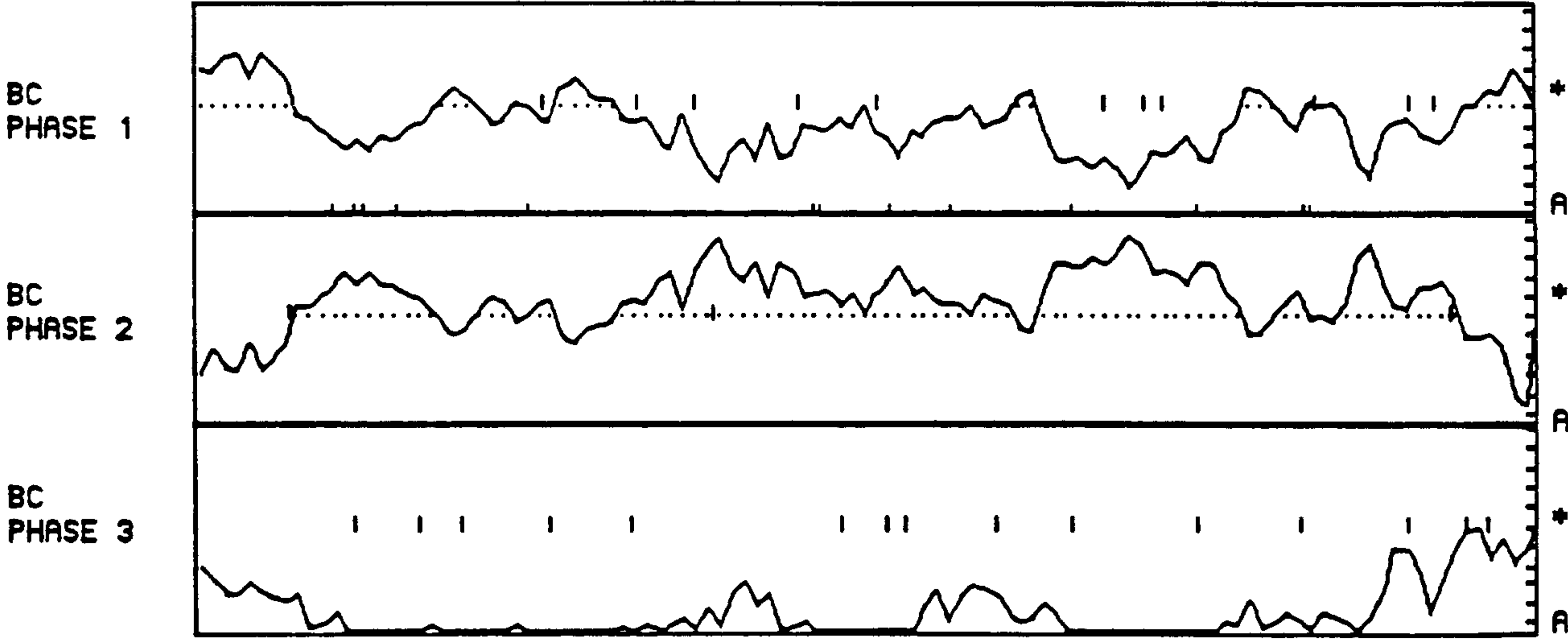
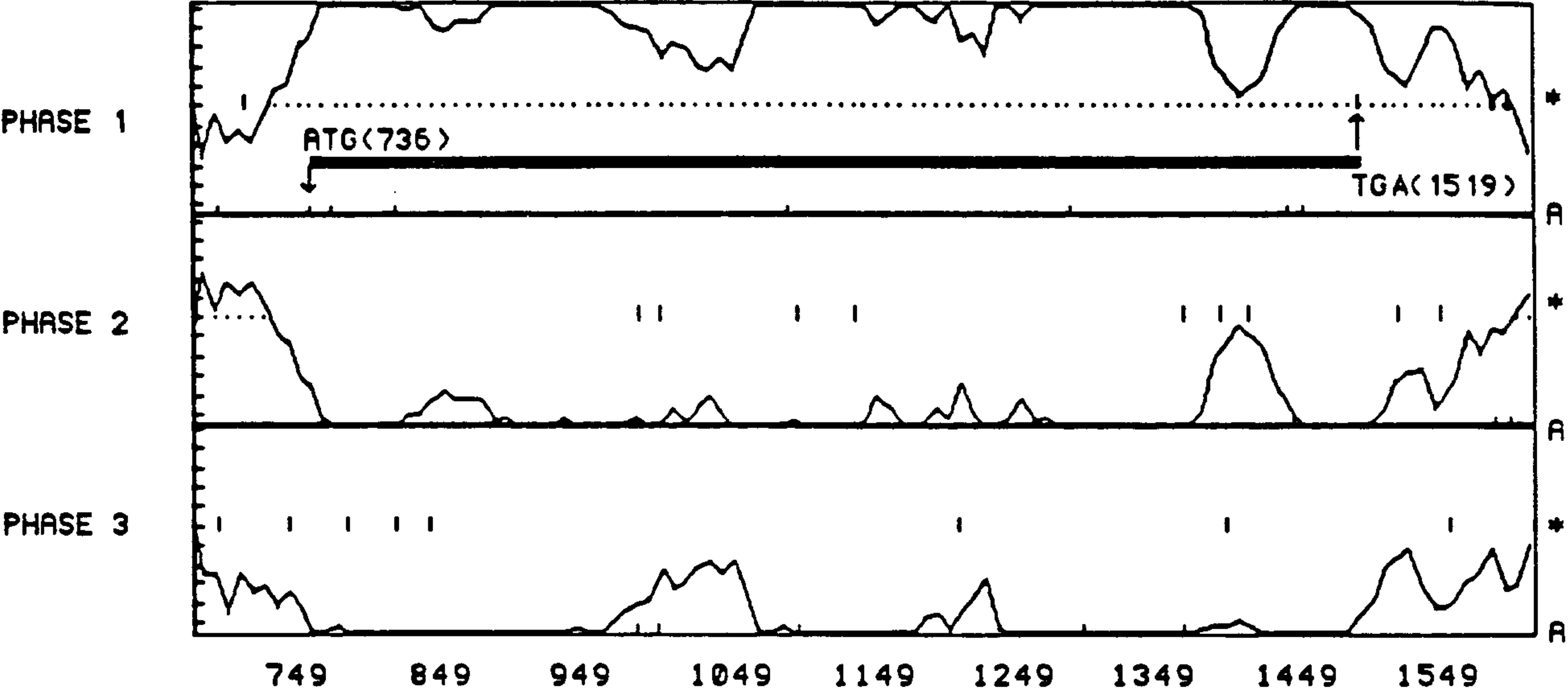


FIGURE 9.5

28/151

SEQUENCE : FRAGMENT DE 8753 BP DE 650 A 1650 LONGUEUR = 1001



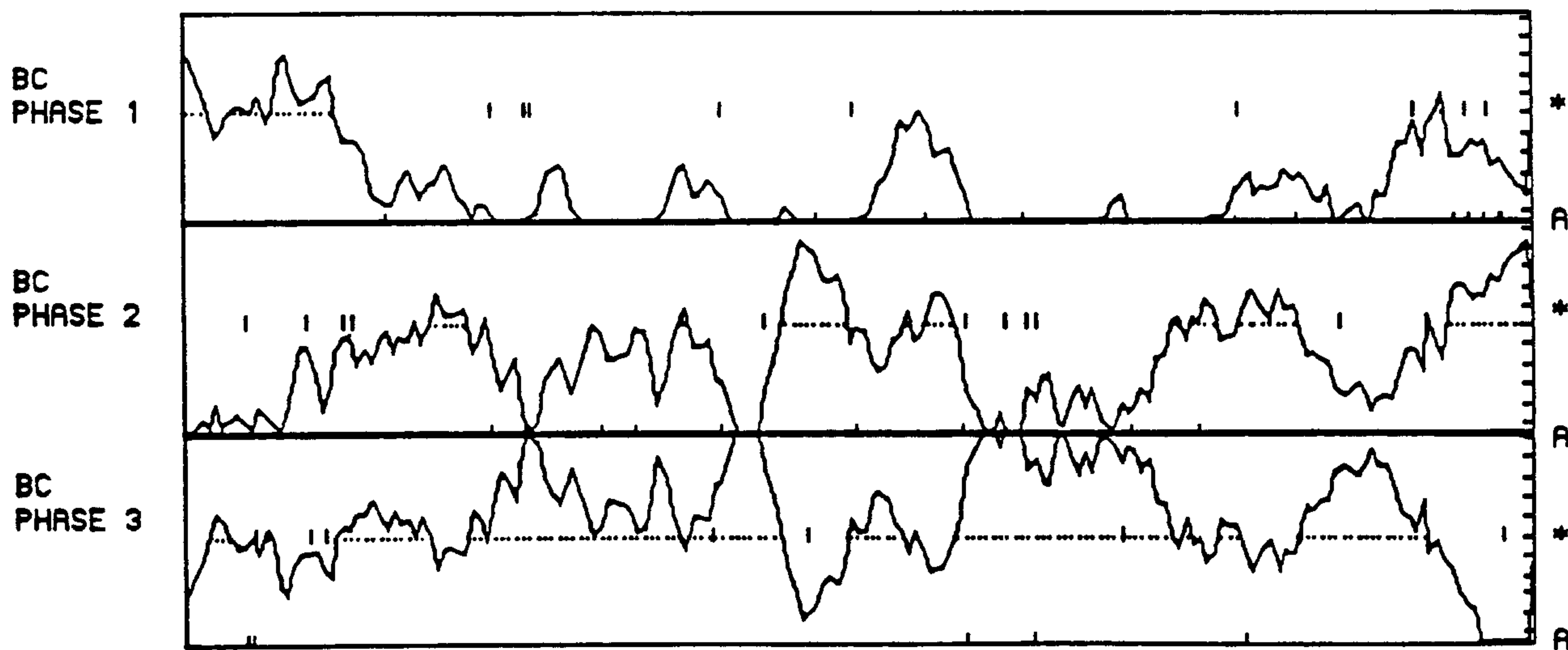
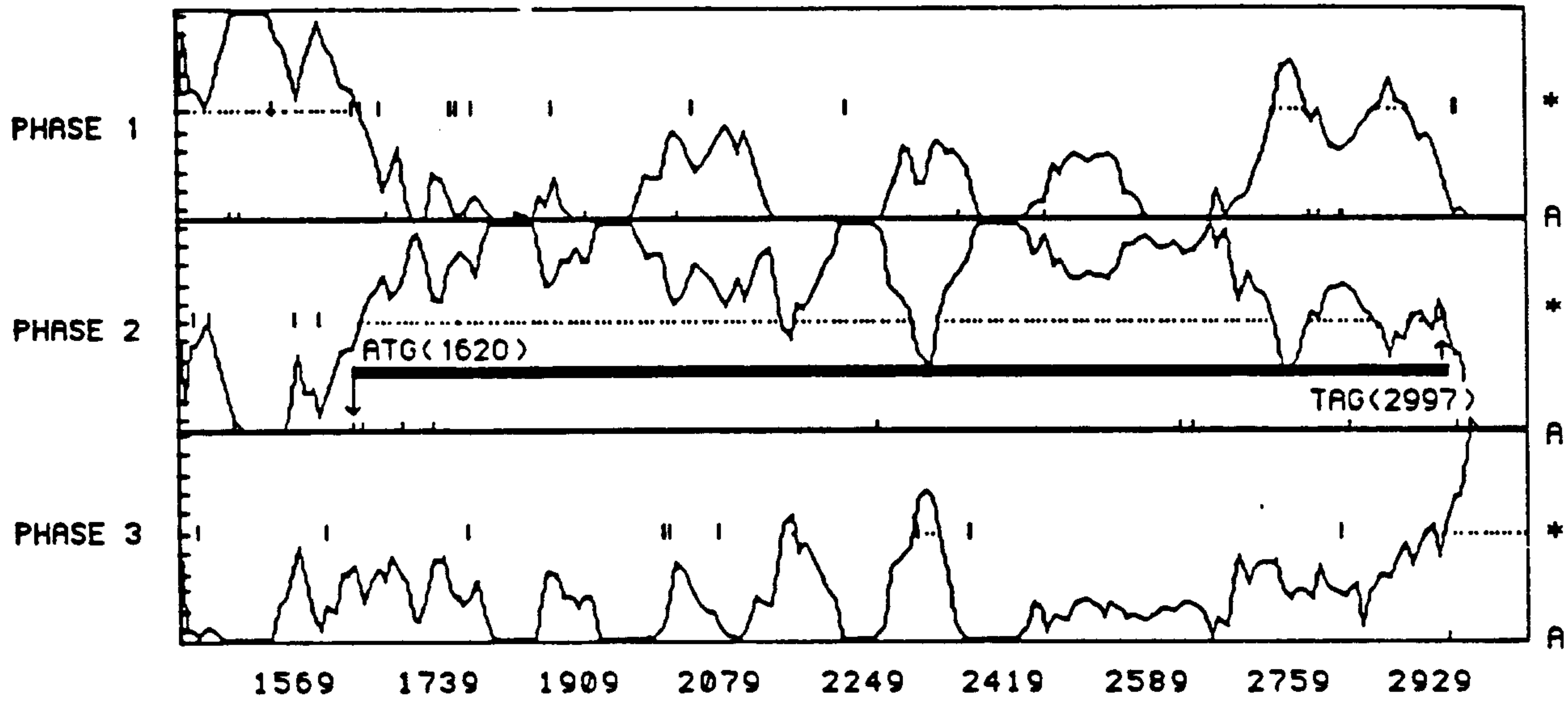
PHASE OUVERTE 6

A=ATG
* STOP
BC BRIN COMPLEMENTAIRE

FIGURE 10.1

29/151

SEQUENCE : FRAGMENT DE 8753 BP DE 1400 A 3100 LONGUEUR = 1701



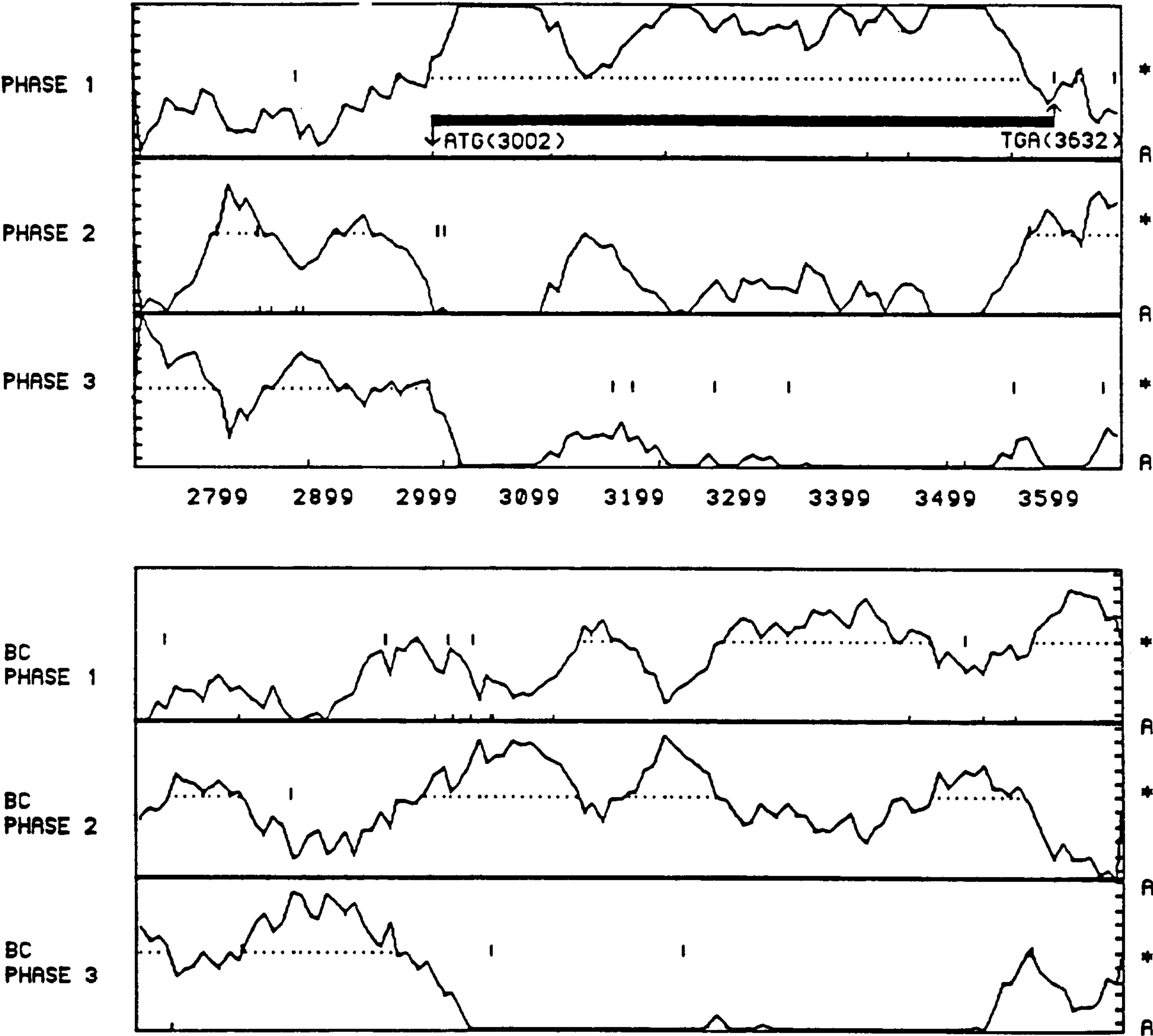
A=ATG
* STOP
BRIN COMPLEMENTAIRE

PHASE OUVERTE 7

FIGURE 10.2

PL. 30/151

SEQUENCE : FRAGMENT DE 8753 BP DE 2700 A 3700 LONGUEUR = 1001



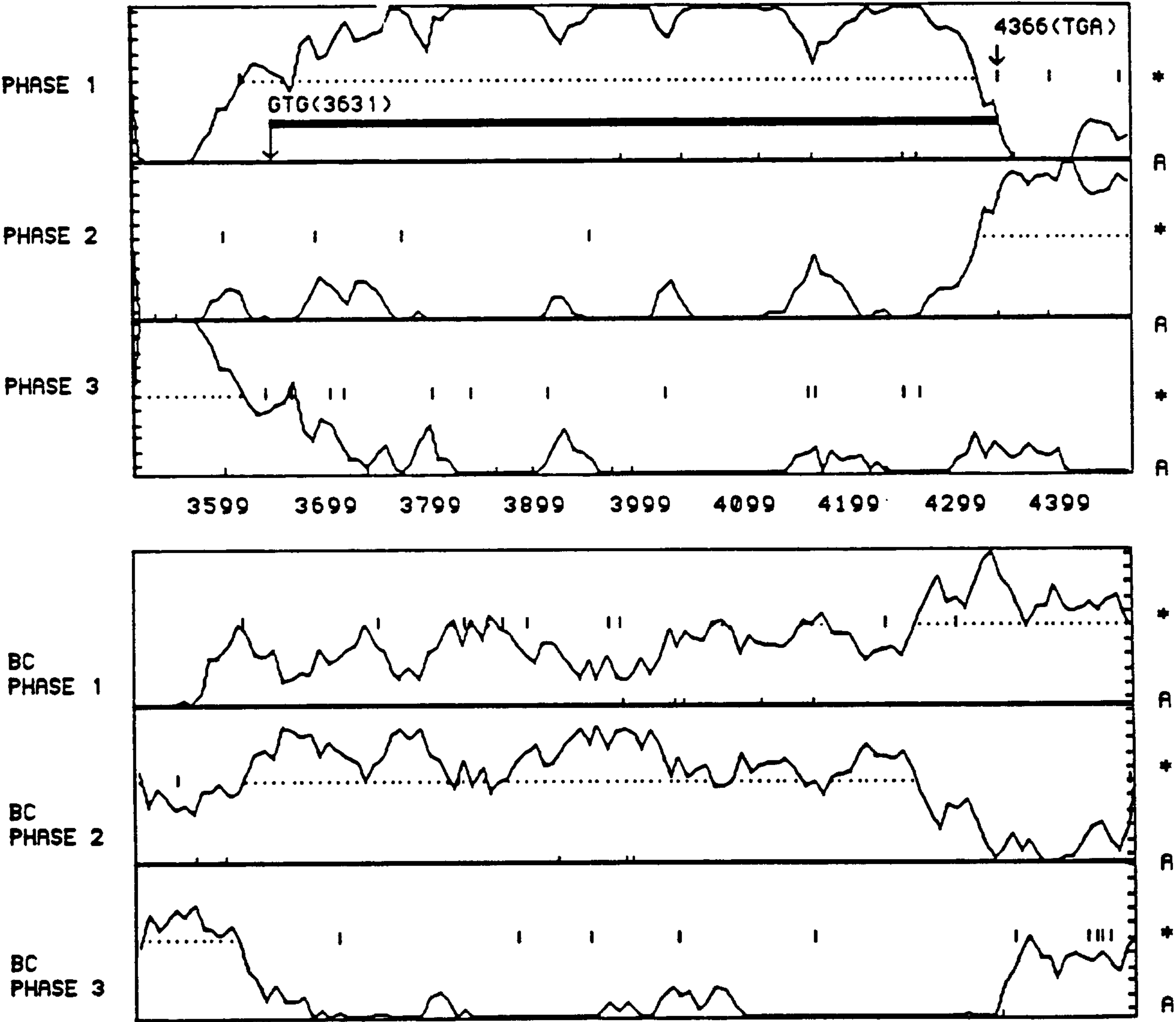
PHASE OUVERTE 8

A=ATG
* STOP
BC BRIN COMPLEMENTAIRE

FIGURE 10.3

PL. 31/151

SEQUENCE : FRAGMENT DE 8753 BP DE 3500 A 4500 LONGUEUR = 1001



PHASE OUVERTE 9

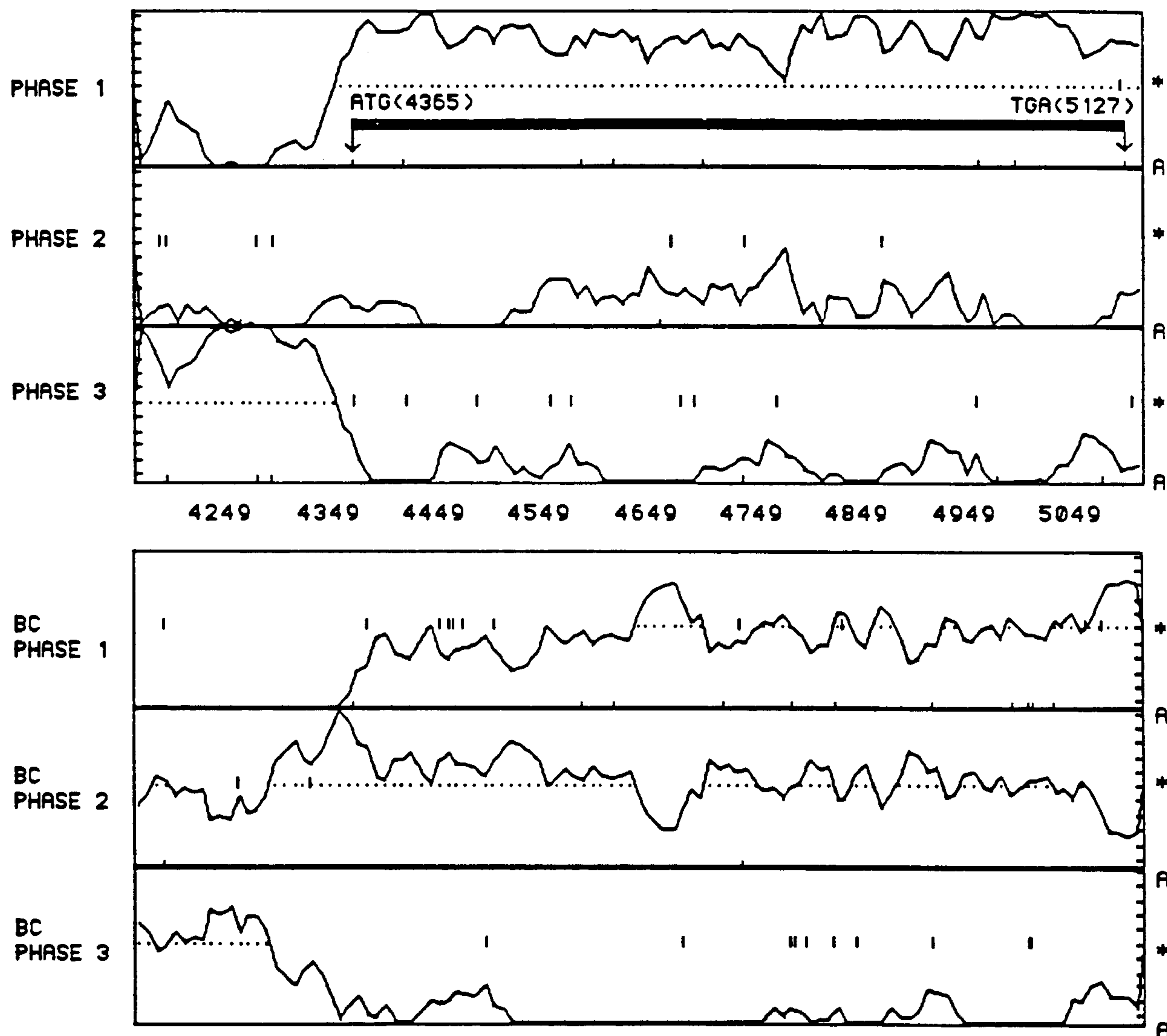
FIGURE 10.4

A=ATG
* STOP
BC BRIN COMPLEMENTAIRE

FIGURE 10.4

PL. 32 / 151

SEQUENCE : FRAGMENT DE 8753 BP DE 4150 A 5150 LONGUEUR = 1001



PHASE OUVERTE 10

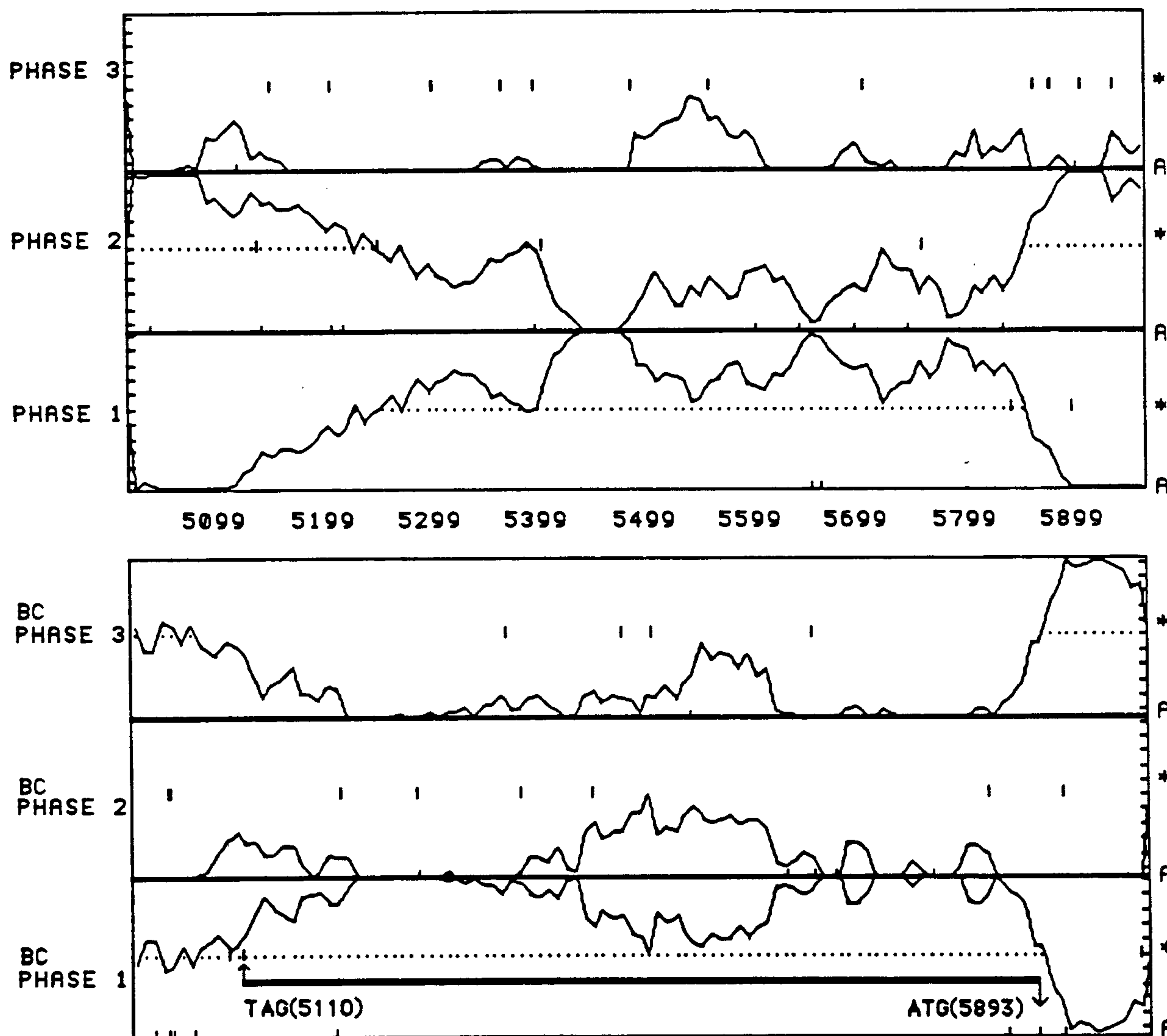
A=ATG
 * STOP
 BC BRIN COMPLEMENTAIRE

FIGURE 10.5

PL. 33/151

SEQUENCE : 253TOT

DE 5000 A 6000 LONGUEUR = 1001



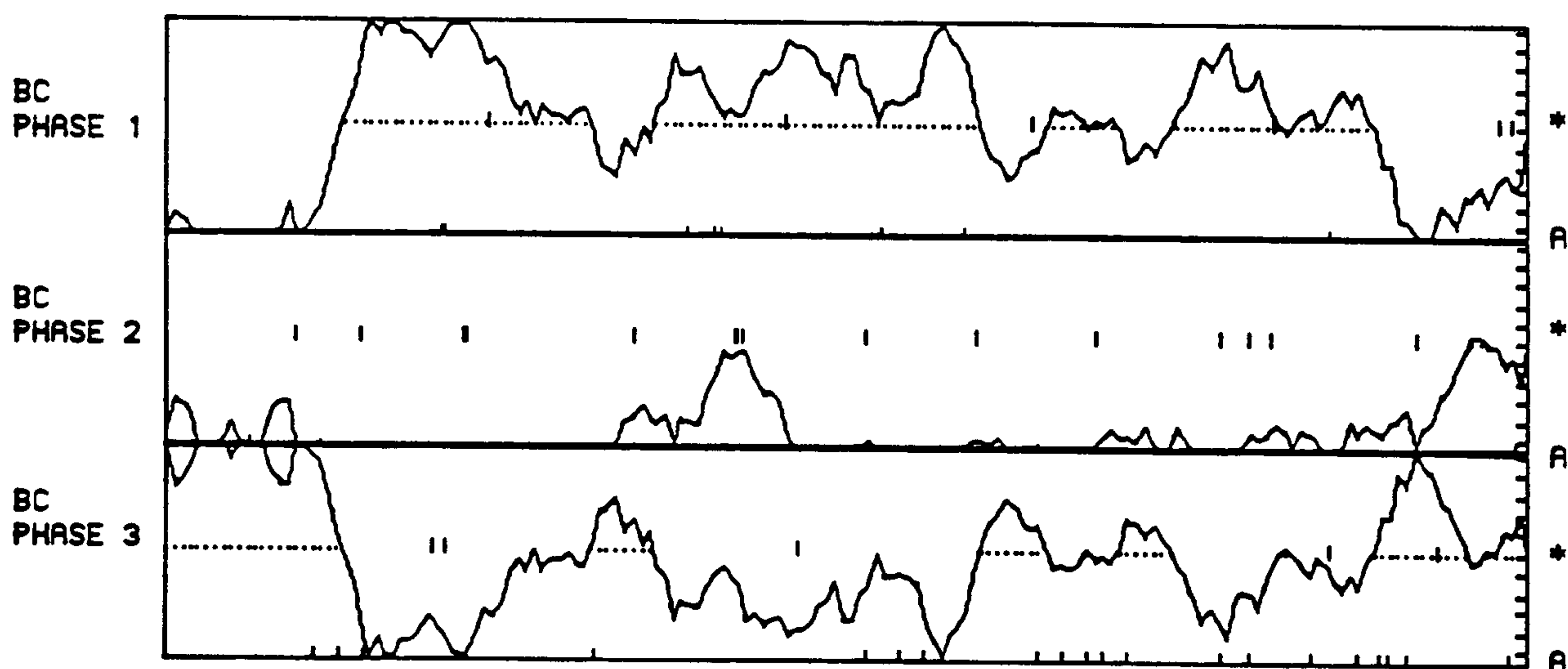
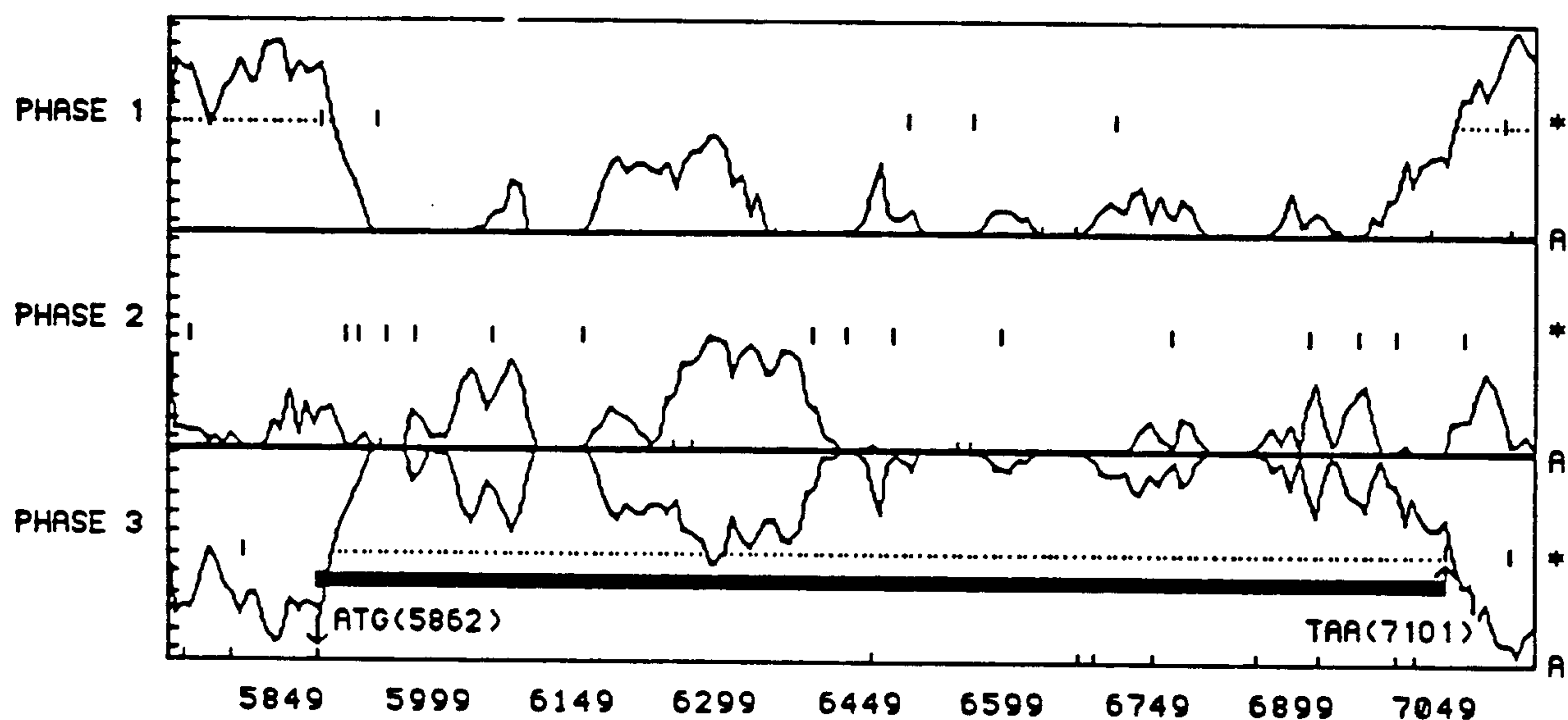
PHASE OUVERTE 11

FIGURE 10.6

FEUILLE DE REMPLACEMENT

PL. 34/151

SEQUENCE : FRAGMENT DE 8753 BP DE 5700 A 7200 LONGUEUR = 1501



PHASE OUVERTE 12

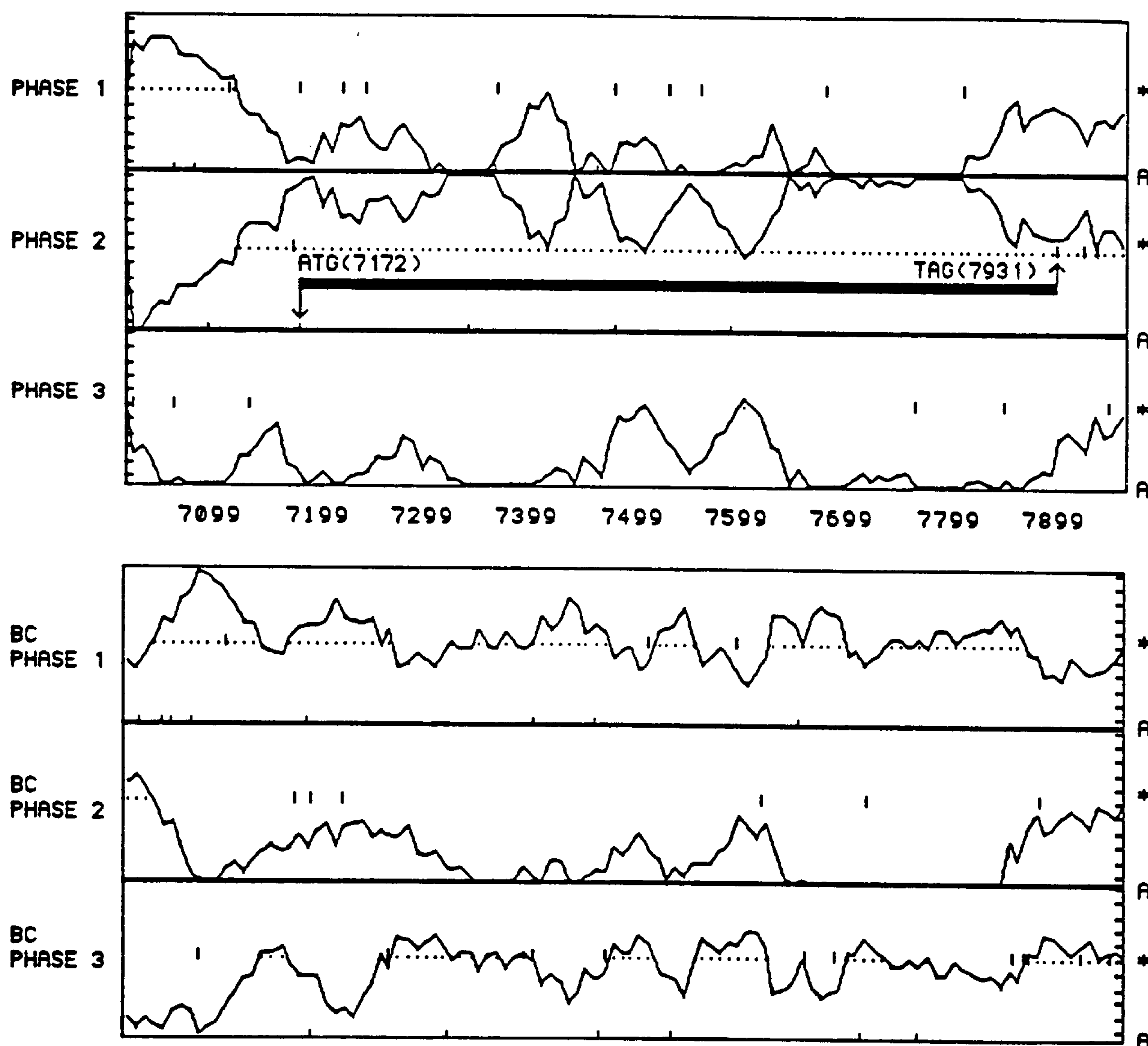
A=ATG
 * STOP
 BC BRIN COMPLEMENTAIRE

FIGURE 10.7

FEUILLE DE REMPLACEMENT

PL. 35 / 151.

SEQUENCE : FRAGMENT DE 8753 BP DE 7000 A 8000 LONGUEUR = 1001



PHASE OUVERTE 13

A=ATG
 * STOP
 BC BRIN COMPLEMENTAIRE

FIGURE 10.8

FEUILLE DE REMPLACEMENT

36 / 151

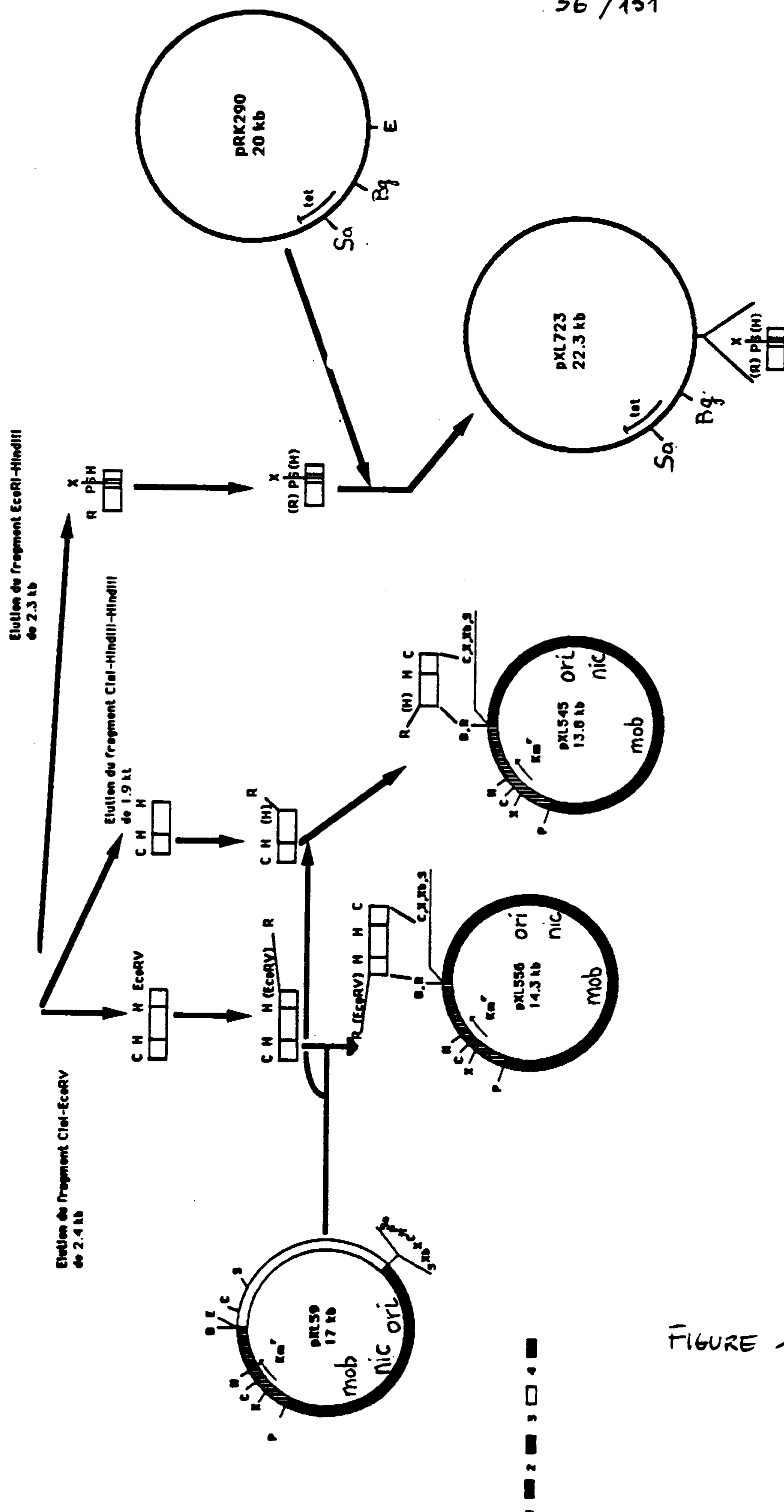


FIGURE 11

37 / 151

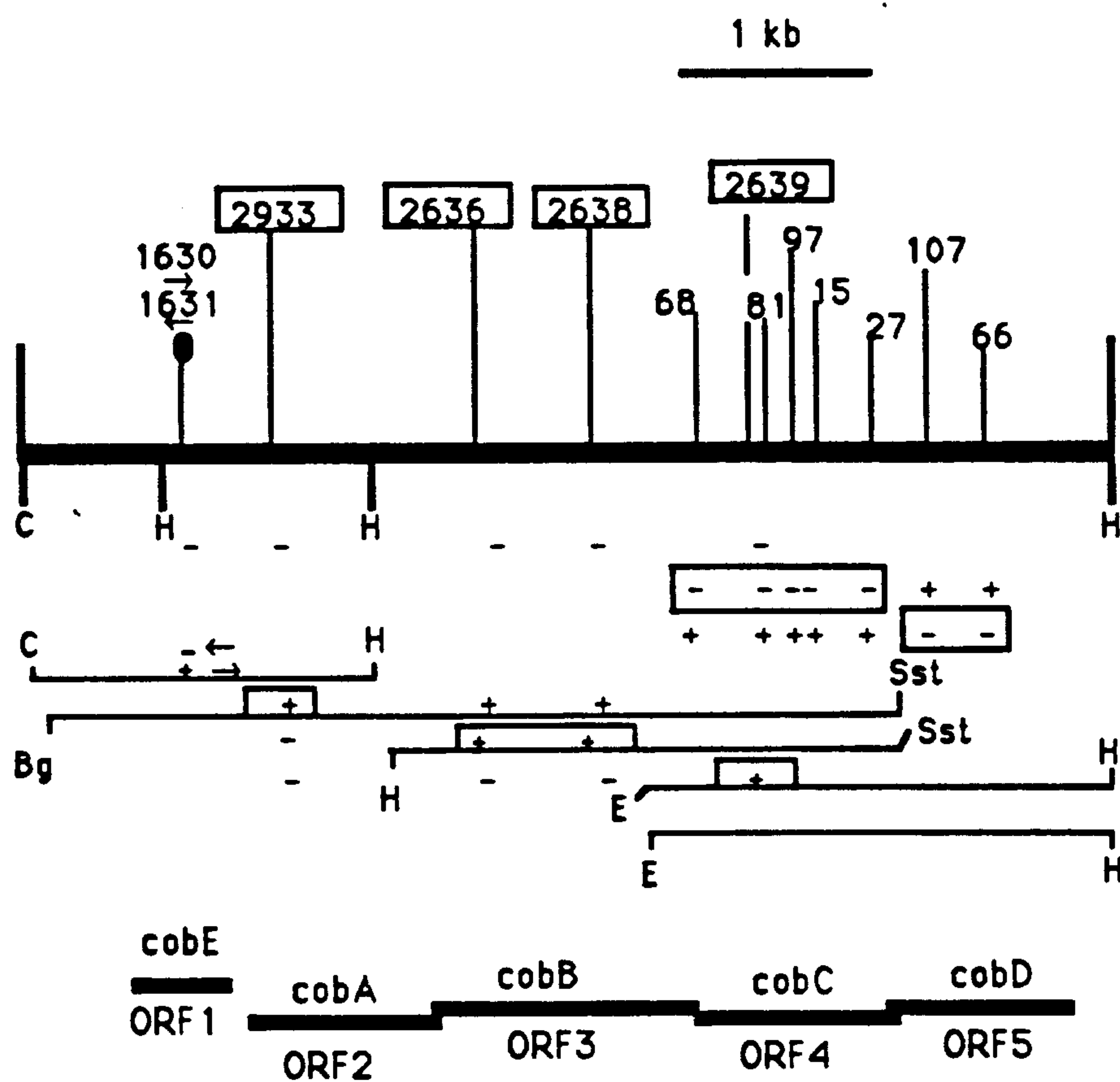


FIGURE 12

38 1151

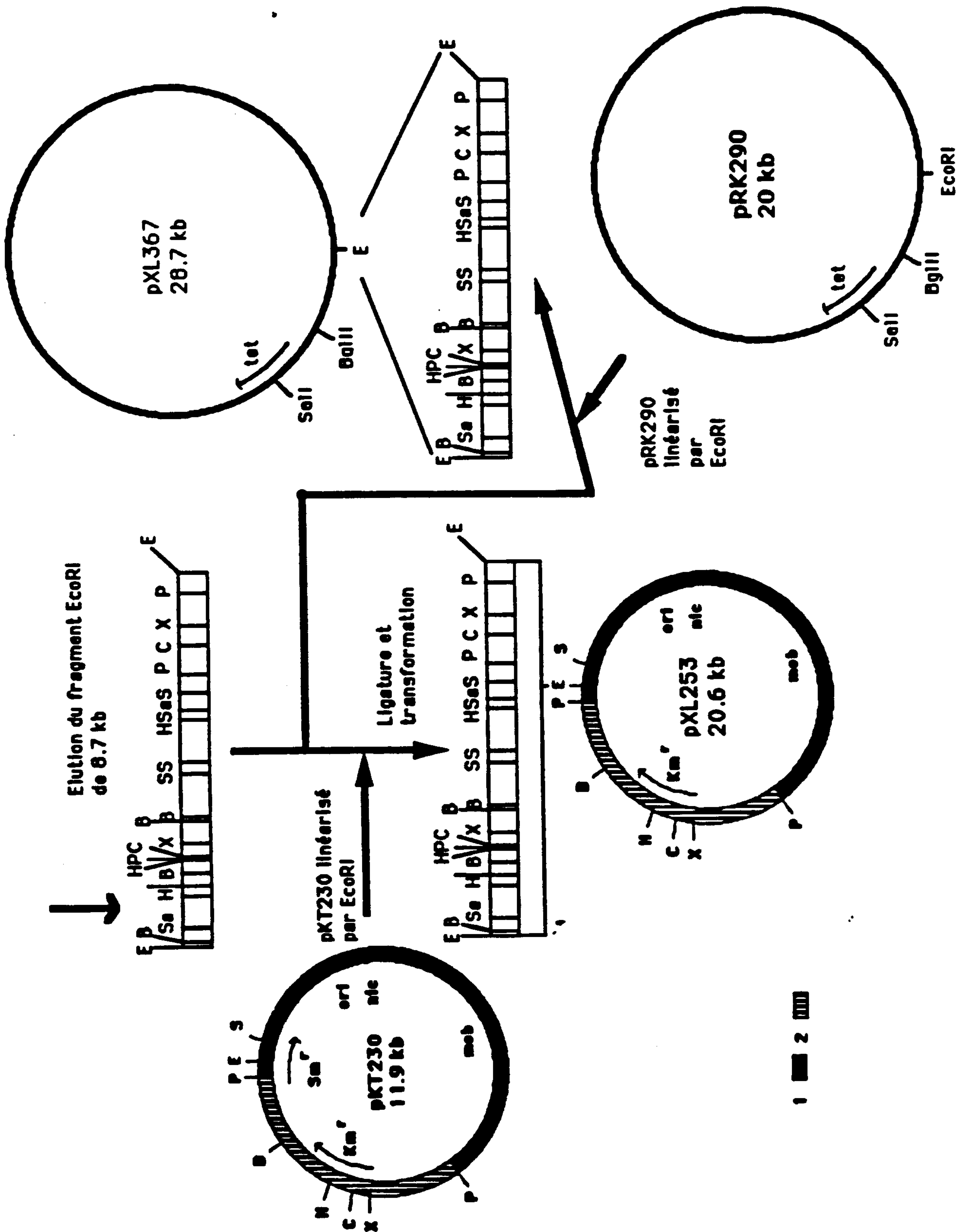
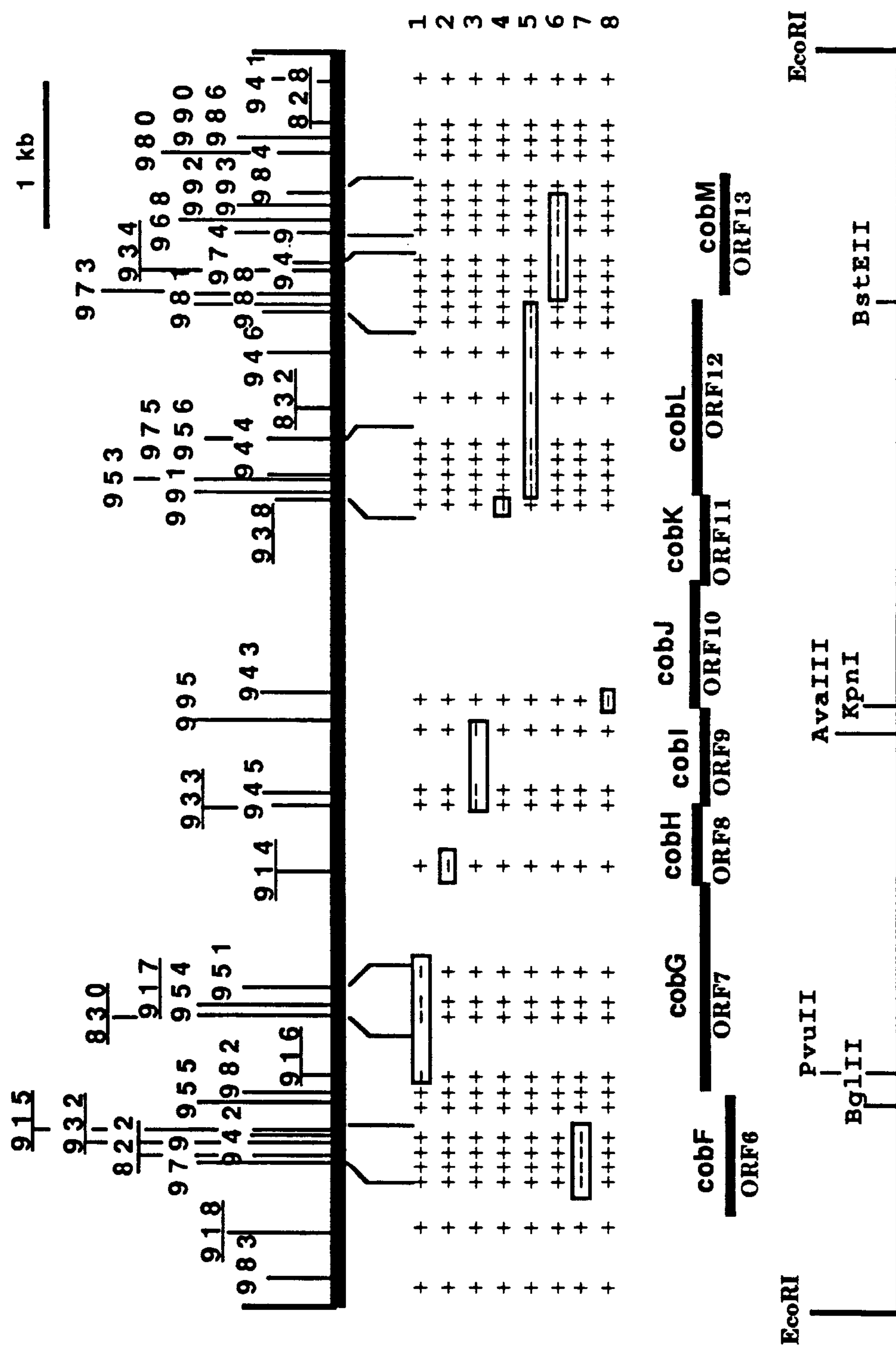


FIGURE 13



40 / 151

GENE cobA ET PROTEINE COBASEQUENCE DU FRAGMENT ClaI-HindIII-HindIII-HindIII DE 5396 BP, DE 1141 A 1980

```

MetIleAspAspLeuPheAlaGlyLeuProAlaLeuGluLysGlySerValTrpLeuValGlyAlaGlyProGly
ATGATCGACGACCTCTTTGCCGGATTGCCGGCGCTCGAAAAAGGTTCCGGTCTGGCTGGTCGGCGCCGGCCCCGGC
1141      1151      1161      1171      1181      1191      1201
AspProGlyLeuLeuThrLeuHisAlaAlaAsnAlaLeuArgGlnAlaAspValIleValHisAspAlaLeuVal
GATCCCGGCCTGTTGACGCTGCATGCGGCCAATGCGCTGCGCCAGGCGGATGTGATCGTGCATGATGCGCTGGTC
1216      1226      1236      1246      1256      1266      1276
AsnGluAspCysLeuLysLeuAlaArgProGlyAlaValLeuGluPheAlaGlyLysArgGlyGlyLysProSer
AACGAGGATTGCCTGAAGCTCGCGCGGCCGGCGCCGTGCTGGAGTTTGGCGGCAAGCGTGGCGGCAAGCCGTCG
1291      1301      1311      1321      1331      1341      1351
ProLysGlnArgAspIleSerLeuArgLeuValGluLeuAlaArgAlaGlyAsnArgValLeuArgLeuLysGly
CCGAAGCAGCGCGACATCTCGCTTCGCCTCGTCGAACTCGCGCGCGCCGGCAACCGGGTGCTGCGCCTCAAAGGC
1366      1376      1386      1396      1406      1416      1426
GlyAspProPheValPheGlyArgGlyGlyGluGluAlaLeuThrLeuValGluHisGlnValProPheArgIle
GGCGATCCCTTCGTCTTCGGTCGCGGTGGCGAGGAGGCGCTGACGCTGGTCGAACACCAGGTGCCGTTCCGAATC
1441      1451      1461      1471      1481      1491      1501
ValProGlyIleThrAlaGlyIleGlyGlyLeuAlaTyrAlaGlyIleProValThrHisArgGluValAsnHis
GTGCCCCGGCATCACCGCCGGTATCGGCGGGCTTGCCCTATGCCGGCATTCCCGTGACCCATCGCGAGGTCAACCAC
1516      1526      1536      1546      1556      1566      1576
AlaValThrPheLeuThrGlyHisAspSerSerGlyLeuValProAspArgIleAsnTrpGlnGlyIleAlaSer
GCGGTCACTTTCTGACTGGCCATGATTCTCCGGCCTGGTGCCGGATCGCATCAACTGGCAGGGCATCGCCAGC
1591      1601      1611      1621      1631      1641      1651
GlySerProValIleValMetTyrMetAlaMetLysHisIleGlyAlaIleThrAlaAsnLeuIleAlaGlyGly
GGCTCGCCTGTCATCGTCATGTACATGGCGATGAAACATATCGGCGCGATCACCGCCAACCTCATTGCCGGCGGC
1666      1676      1686      1696      1706      1716      1726
ArgSerProAspGluProValAlaPheValCysAsnAlaAlaThrProGlnGlnAlaValLeuGluThrThrLeu
CGCTCGCCGGACGAACCGGTGCGCTTCGTCTGCAACGCCGCGACGCCGCGAGGCGGTGCTGGAAACGACGCTT
1741      1751      1761      1771      1781      1791      1801
AlaArgAlaGluAlaAspValAlaAlaAlaGlyLeuGluProProAlaIleValValValGlyGluValValArg
GCGCGTGCGAGAGGCCGATGTTGCGGCGGCGAGGGCTGGAGCCGCCGGCGATCGTCGTCGTCGGCGAGGTGGTGCGG
1816      1826      1836      1846      1856      1866      1876
LeuArgAlaAlaLeuAspTrpIleGlyAlaLeuAspGlyArgLysLeuAlaAlaAspProPheAlaAsnArgIle
CTGCGCGCAGCGCTCGACTGGATCGGCGCGCTGGACGGGCGCAAGCTTGCCGCCGACCCGTTCCGCAATCGCATT
1891      1901      1911      1921      1931      1941      1951
LeuArgAsnProAla***
CTCAGGAACCCGGCATGA
1966      1976      1986      1996      2006      2016      2026

```

FIGURE 15.1

41/151

NOM = COBA

PREMIER RESIDU = 1
DERNIER RESIDU = 280

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	8	2.86	1176.56	4.02
2	LEU	L	31	11.07	3505.48	11.99
3	ILE	I	16	5.71	1809.28	6.19
4	MET	M	4	1.43	524.16	1.79
5	VAL	V	27	9.64	2674.89	9.15
6	SER	S	8	2.86	696.24	2.38
7	PRO	P	19	6.79	1843.95	6.31
8	THR	T	10	3.57	1010.50	3.46
9	ALA	A	41	14.64	2912.64	9.96
10	TYR	Y	2	0.71	326.12	1.12
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	2.50	959.42	3.28
13	GLN	Q	6	2.14	768.36	2.63
14	ASN	N	9	3.21	1026.36	3.51
15	LYS	K	8	2.86	1024.72	3.51
16	ASP	D	15	5.36	1725.45	5.90
17	GLU	E	13	4.64	1677.52	5.74
18	CYS	C	2	0.71	206.02	0.70
19	TRP	W	3	1.07	558.24	1.91
20	ARG	R	19	6.79	2965.90	10.15
21	GLY	G	32	11.43	1824.64	6.24
22	-	-	0	0.00	0.00	0.00

RESIDUS = 280
POIDS MOLECULAIRE = 29234.
INDEX DE POLARITE (%) = 34.
POINT ISOELECTRIQUE = 7.51
DO 260 (1mg/ml) = 0.464 DO 280 (1mg/ml) = 0.652

PROFIL D'HYDROPHILICITE DE LA PROTEINE COBA
COBA DE 1 A 280

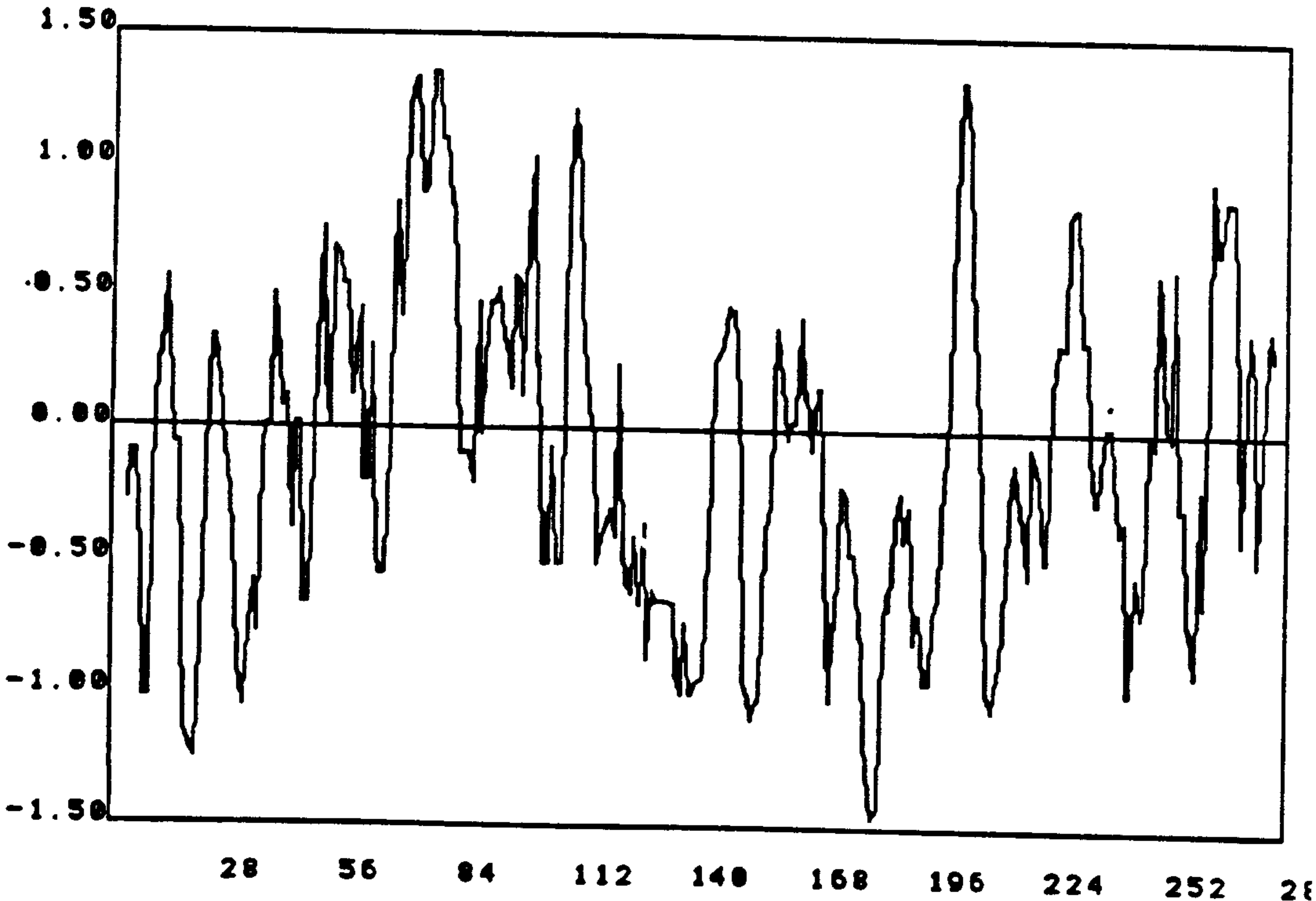


FIGURE 15.2

42/151

GENE cobB ET PROTEINE COBBSEQUENCE DU FRAGMENT ClaI-HindIII-HindIII-HindIII DE 5396 BP, DE 1980 A 3281

MetSerGlyLeuLeuIleAlaAlaProAlaSerGlySerGlyLysThrThrValThrLeuGlyLeuMetArgAla
 ATGAGCGGATTGCTGATTGCCGCACCCGCGTCCGGCTCCGGCAAGACGACGGTGACGCTCGGGCTGATGCGCGCC
 1980 1990 2000 2010 2020 2030 2040
 LeuLysArgArgGlyValAlaIleAlaProGlyLysAlaGlyProAspTyrIleAspProAlaPheHisAlaAla
 CTGAAGAGGCGCGGCGTGGCGATCGCGCCCGGCAAGGCGGGCCGGACTATATCGATCCCGCTTCCACGCGGCA
 2055 2065 2075 2085 2095 2105 2115
 AlaThrGlyGluProCysPheAsnTyrAspProTrpAlaMetArgProGluLeuLeuLeuAlaAsnAlaSerHis
 GCGACCGGCGAGCCCTGCTTCAACTACGACCCCTGGGCGATGCGCCCGGAAGTCTGCTTGCCAATGCGTGCAT
 2130 2140 2150 2160 2170 2180 2190
 ValAlaSerGlyGlyArgThrLeuIleValGluAlaMetMetGlyLeuHisAspGlyAlaAlaAspGlySerGly
 GTGGCCTCCGGCGGGCGCACATTGATCGTCGAGGCGATGATGGGACTGCATGACGGTGCTGCCGACGGCTCGGGA
 2205 2215 2225 2235 2245 2255 2265
 ThrProAlaAspLeuAlaAlaThrLeuAsnLeuAlaValIleLeuValValAspCysAlaArgMetSerGlnSer
 ACGCCAGCGGACCTCGCCGCGACGCTGAACCTTGCGGTCATTCTGGTGGTCGATTGCGCCCGCATGTCCCAGTCG
 2280 2290 2300 2310 2320 2330 2340
 ValAlaAlaLeuValArgGlyTyrAlaAspHisArgAspAspIleArgValValGlyValIleLeuAsnLysVal
 GTTGCCGCCCTCGTGCGGGCTATGCGGATCATCGCGACGATATCCGGGTGGTTGGCGTCATCCTCAACAAGGTC
 2355 2365 2375 2385 2395 2405 2415
 GlySerAspArgHisGluMetMetLeuArgAspAlaLeuGlyLysValArgMetProValPheGlyValLeuArg
 GGCAGCGATCGGCATGAAATGATGCTGCGCGATGCGCTCGGCAAGGTGCGCATGCCTGTCTTCGGCGTGCTCCGG
 2430 2440 2450 2460 2470 2480 2490
 GlnAspSerAlaLeuGlnLeuProGluArgHisLeuGlyLeuValGlnAlaGlyGluHisSerAlaLeuGluGly
 CAGGACAGCGCATTGCAACTGCCGAGCGCCATCTCGGGCTCGTGACGGCGGGCAACACTCAGCGCTTGAGGGC
 2505 2515 2525 2535 2545 2555 2565
 PheIleGluAlaAlaAlaAlaArgValGluAlaAlaCysAspLeuAspAlaIleArgLeuIleAlaThrIlePhe
 TTCATCGAGGCGGCGCGCGGGTCGAGGCTGCCTGCGATCTCGACGCCATCCGCCTGATCGCGACGATTTTC
 2580 2590 2600 2610 2620 2630 2640
 ProGlnValProAlaAlaAlaAspAlaGluArgLeuArgProLeuGlyGlnArgIleAlaValAlaArgAspIle
 CCGCAGGTGCCCGCGGCGCGCGGATGCCGAGCGTTTGCGGCCGCTCGGTCAGCGCATCGCGGTGCGCGCGATATC
 2655 2665 2675 2685 2695 2705 2715
 AlaPheAlaPheCysTyrGluHisLeuLeuTyrGlyTrpArgGlnGlyGlyAlaGluIleSerPhePheSerPro
 GCCTTTGCCTTCTGCTACGAGCACCTGCTTTACGGCTGGCGGCAAGGCGGCGCGGAGATTTCTTCTTCTCGCCG
 2730 2740 2750 2760 2770 2780 2790
 LeuAlaAspGluGlyProAspAlaAlaAlaAspAlaValTyrLeuProGlyGlyTyrProGluLeuHisAlaGly
 CTCGCCGACGAGGGGCGGATGCGGCAGCCGATGCCGTCTATCTTCCGGGGGGTTATCCGGAGCTGCATGCGGGG
 2805 2815 2825 2835 2845 2855 2865
 GlnLeuSerAlaAlaAlaArgPheArgSerGlyMetHisSerAlaAlaGluArgGlyAlaArgIlePheGlyGlu
 CAGCTGAGCGCCGCGCCGATTCCGTTCCGGCATGCATTCCGCGGCGGAACGCGGCGCCCGCATCTTCGGCGAG
 2880 2890 2900 2910 2920 2930 2940
 CysGlyGlyTyrMetValLeuGlyGluGlyLeuValAlaAlaAspGlyThrArgTyrAspMetLeuGlyLeuLeu
 TGCGGCGGCTATATGGTGCTCGGCGAAGGGCTTGTCGCTGCCGATGGCACACGCTACGACATGCTCGGCCTGCTG
 2955 2965 2975 2985 2995 3005 3015
 ProLeuValThrSerPheAlaGluArgArgArgHisLeuGlyTyrArgArgValValProValAspAsnAlaPhe
 CCGCTCGTAACCAGTTTTGCGGAGCGCAGGCGGCACCTCGGCTATCGCCGCGTCGTGCCTGTCGACAACGCCTTC
 3030 3040 3050 3060 3070 3080 3090
 PheAspGlyProMetThrAlaHisGluPheHisTyrAlaThrIleValAlaGluGlyAlaAlaAspArgLeuPhe
 TTCGATGGACCCATGACGGCGCACGAATTCCACTATGCGACCATCGTCGCCGAAGGGGCGGCGGATCGGCTGTTT
 3105 3115 3125 3135 3145 3155 3165
 AlaValSerAspAlaAlaGlyGluAspLeuGlyGlnAlaGlyLeuArgArgGlyProValAlaGlySerPheMet
 GCGGTCAGCGACGCCGCGGCGAGGATCTCGGCCAGGCGGGCCTCCGGCGCGGCCCTGTCGCCGGTTTCCTTCATG
 3180 3190 3200 3210 3220 3230 3240
 HisLeuIleAspValAlaGlyAlaAla***
 CATCTGATCGACGTCGCAGGTGCTGCATGA
 3255 3265 3275 3285 3295 3305 3315

FIGURE 15.3

43/151

NOM = COBB

PREMIER RESIDU = 1
DERNIER RESIDU = 434

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	17	3.92	2500.19	5.47
2	LEU	L	45	10.37	5088.60	11.14
3	ILE	I	17	3.92	1922.36	4.21
4	MET	M	14	3.23	1834.56	4.02
5	VAL	V	31	7.14	3071.17	6.72
6	SER	S	19	4.38	1653.57	3.62
7	PRO	P	21	4.84	2038.05	4.46
8	THR	T	12	2.76	1212.60	2.65
9	ALA	A	76	17.51	5399.04	11.82
10	TYR	Y	11	2.53	1793.66	3.93
11	*	*	0	0.00	0.00	0.00
12	HIS	H	14	3.23	1918.84	4.20
13	GLN	Q	9	2.07	1152.54	2.52
14	ASN	N	5	1.15	570.20	1.25
15	LYS	K	5	1.15	640.45	1.40
16	ASP	D	28	6.45	3220.84	7.05
17	GLU	E	21	4.84	2709.84	5.93
18	CYS	C	5	1.15	515.05	1.13
19	TRP	W	2	0.46	372.16	0.81
20	ARG	R	34	7.83	5307.40	11.62
21	GLY	G	48	11.06	2736.96	5.99
22	-	-	0	0.00	0.00	0.00

RESIDUS = 434
POIDS MOLECULAIRE = 45676.
INDEX DE POLARITE (%) = 34.
POINT ISOELECTRIQUE = 6.47
DO 260 (1mg/ml) = 0.351 DO 280 (1mg/ml) = 0.529

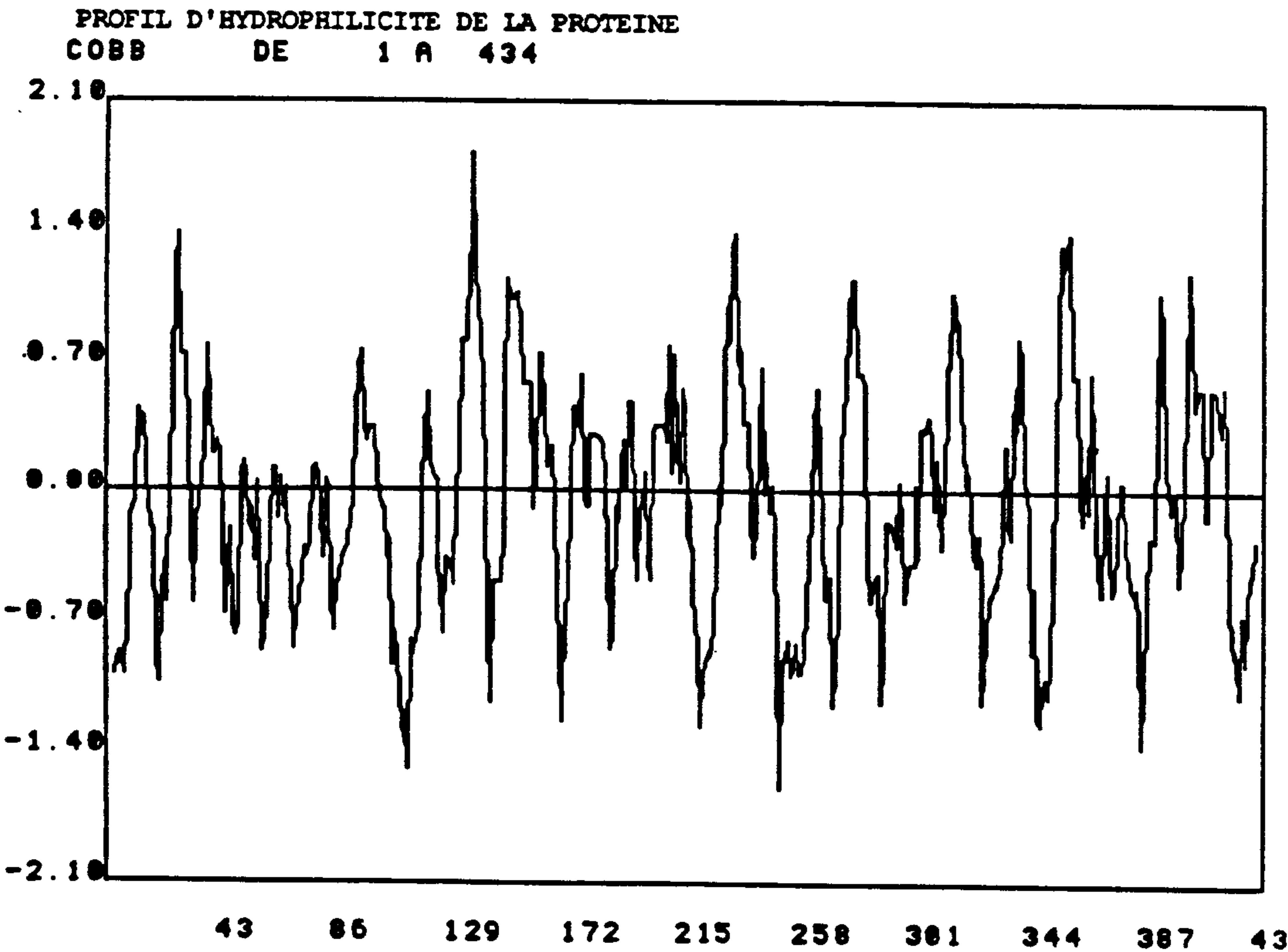


FIGURE 15.4

44/151

GENE cobC ET PROTEINE COBCSEQUENCE DU FRAGMENT ClaI-HindIII-HindIII-HindIII DE 5396 BP, DE 13281 A 4279

MetSerAlaProIleValHisGlyGlyGlyIleThrGluAlaAlaAlaArgTyrGlyGlyArgProGluAspTrp
 ATGAGCGCACCGATCGTTCATGGTGGCGGCATCACCGAGGCCGACGCGCGCTATGGCGGCCGGCCTGAAGACTGG
 3281 3291 3301 3311 3321 3331 3341
 LeuAspLeuSerThrGlyIleAsnProCysProValAlaLeuProAlaValProGluArgAlaTrpHisArgLeu
 CTCGATCTGTGACCGGCATCAATCCATGCCCCGTCGCCTTGCCCGCGGTCCCTGAGCGCGCCTGGCACCGGCTG
 3356 3366 3376 3386 3396 3406 3416
 ProAspArgGlnThrValAspAspAlaArgSerAlaAlaAlaAspTyrTyrArgThrAsnGlyValLeuProLeu
 CCGGATCGGCAGACGGTAGATGATGCGCGGAGCGCCGCGCCGACTACTACCGCACCAACGGCGTGCTGCCTTTG
 3431 3441 3451 3461 3471 3481 3491
 ProValProGlyThrGlnSerValIleGlnLeuLeuProArgLeuAlaProAlaAsnArgHisValAlaIlePhe
 CCGGTGCCGGGCACCCAGTCGGTGATCCAGCTCCTGCCACGTCTTGCTCCGGCCAACAGGCACGTCGCGATTTTC
 3506 3516 3526 3536 3546 3556 3566
 GlyProThrTyrGlyGluTyrAlaArgValLeuGluAlaAlaGlyPheAlaValAspArgValAlaAspAlaAsp
 GGGCCGACCTATGGCGAGTATGCCCCGCGTGCTTGAAGCGGCCGGCTTTGCTGTCGATCGCGTCGCGGATGCCGAC
 3581 3591 3601 3611 3621 3631 3641
 AlaLeuThrAlaGluHisGlyLeuValIleValValAsnProAsnAsnProThrGlyArgAlaLeuAlaProAla
 GCGCTCACGGCCGAACATGGGCTTGTCATCGTCGTCAACCCCAACAACCCGACCGGCCGCGCCTTGCGCGCCGGCG
 3656 3666 3676 3686 3696 3706 3716
 GluLeuLeuAlaIleAlaAlaArgGlnLysAlaSerGlyGlyLeuLeuLeuValAspGluAlaPheGlyAspLeu
 GAGCTTCTGGCGATCGCCGCAAGGCAGAAGGCGAGCGGCGGACTGCTGCTGGTCGATGAGGCCTTCGGCGATCTT
 3731 3741 3751 3761 3771 3781 3791
 GluProGlnLeuSerValAlaGlyHisAlaSerGlyGlnGlyAsnLeuIleValPheArgSerPheGlyLysPhe
 GAGCCGCAACTGAGTGTGCTGCTCACGCGTCAGGGCAAGGCAACCTCATCGTCTTCCGCTCCTTCGGCAAGTTC
 3806 3816 3826 3836 3846 3856 3866
 PheGlyLeuAlaGlyLeuArgLeuGlyPheValValAlaThrGluProValLeuAlaSerPheAlaAspTrpLeu
 TTCGGCCTTGCGGGCCTGCGCCTCGGCTTCGTCGTTGCGACCGAGCCAGTGCTTGATCCTTTGCCGATTGGCTC
 3881 3891 3901 3911 3921 3931 3941
 GlyProTrpAlaValSerGlyProAlaLeuThrIleSerLysAlaLeuMetGlnGlyAspThrLysAlaIleAla
 GGTCCCTGGGCTGTCTCCGGCCCCGGCGTTGACGATCTCGAAAGCGCTGATGCAGGGCGATACGAAGGCGATCGCG
 3956 3966 3976 3986 3996 4006 4016
 AlaGlyIleLeuGluArgArgAlaGlyLeuAspAlaAlaLeuAspGlyAlaGlyLeuAsnArgIleGlyGlyThr
 GCGGGCATCCTCGAGCGTCGCGCCGGCCTCGATGCGGCTCTCGATGGGGCAGGGCTCAACCGTATCGGCGGCACG
 4031 4041 4051 4061 4071 4081 4091
 GlyLeuPheValLeuValGluHisProArgAlaAlaLeuLeuGlnGluArgLeuCysGluAlaHisIleLeuThr
 GGGCTATTCGTGCTGGTCGAGCATCCAGGGCAGCTCTGCTGCAGGAGCGGCTCTGCGAGGCCCATATTCTCACG
 4106 4116 4126 4136 4146 4156 4166
 ArgLysPheAspTyrAlaProThrTrpLeuArgValGlyLeuAlaProAspAlaAlaGlyAspArgArgLeuAla
 CGCAAGTTCGACTATGCCCCGACCTGGCTCAGGGTCGGTCTTGCGCCTGACGCGGCTGGTGACCGACGGCTGGCG
 4181 4191 4201 4211 4221 4231 4241
 AspAlaLeuAlaArgMetGluLeu***
 GACGCGCTTGCCCCGCATGGAGCTCTGA
 4256 4266 4276 4286 4296 4306 4316

FIGURE 15.5

45/151

NOM = COBC
PREMIER RESIDU = 1
DERNIER RESIDU = 333

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	11	3.30	1617.77	4.62
2	LEU	L	43	12.91	4862.44	13.90
3	ILE	I	13	3.90	1470.04	4.20
4	MET	M	3	0.90	393.12	1.12
5	VAL	V	24	7.21	2377.68	6.79
6	SER	S	11	3.30	957.33	2.74
7	PRO	P	23	6.91	2232.15	6.38
8	THR	T	14	4.20	1414.70	4.04
9	ALA	A	56	16.82	3978.24	11.37
10	TYR	Y	6	1.80	978.36	2.80
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	2.10	959.42	2.74
13	GLN	Q	8	2.40	1024.48	2.93
14	ASN	N	8	2.40	912.32	2.61
15	LYS	K	5	1.50	640.45	1.83
16	ASP	D	19	5.71	2185.57	6.25
17	GLU	E	15	4.50	1935.60	5.53
18	CYS	C	2	0.60	206.02	0.59
19	TRP	W	5	1.50	930.40	2.66
20	ARG	R	25	7.51	3902.50	11.15
21	GLY	G	35	10.51	1995.70	5.70
22	-	-	0	0.00	0.00	0.00

RESIDUS = 333
POIDS MOLECULAIRE = 34992.
INDEX DE POLARITE (%) = 34.
POINT ISOELECTRIQUE = 6.72
DO 260 (1mg/ml) = 0.670 DO 280 (1mg/ml) = 0.998

PROFIL D'HYDROPHILICITE DE LA PROTEINE
COBC DE 1 A 333

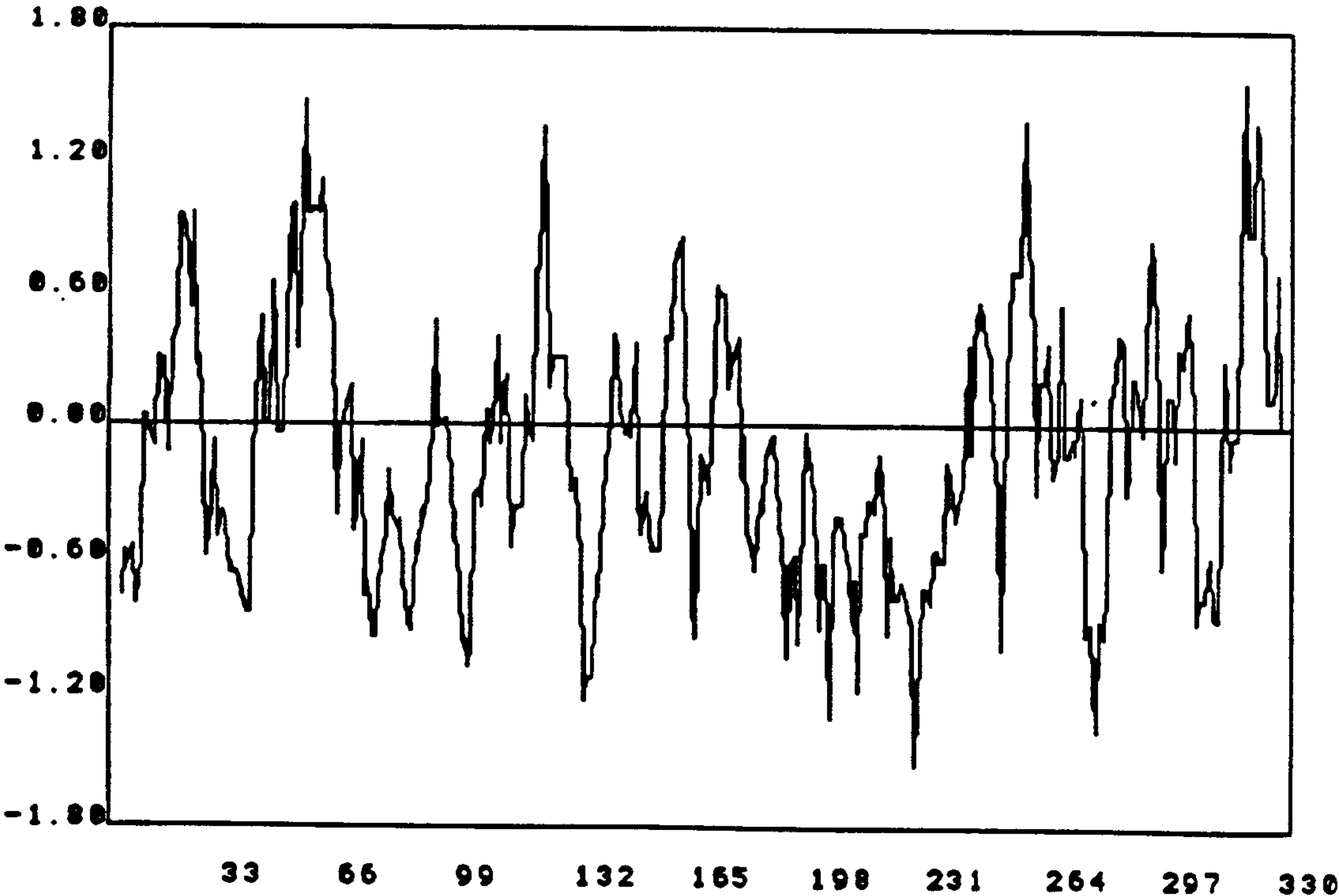


FIGURE 15.6

46/151

GENE cobD ET PROTEINE COBDSEQUENCE DU FRAGMENT ClaI-HindIII-HindIII-HindIII DE 5396 BP, DE 4284 A 5252

```

MetSerGluThrIleLeuLeuIleLeuAlaLeuAlaLeuValIleAspArgValValGlyAspProAspTrpLeu
GTGTCGGAGACGATCCTGCTCATTCTCGCGCTGGCGCTGGTGATCGACCGCGTTGTCGGCGATCCGGACTGGCTC
4284      4294      4304      4314      4324      4334      4344
TrpAlaArgValProHisProValValPhePheGlyLysAlaIleGlyPhePheAspAlaArgLeuAsnArgGlu
TGGGCGCGCGTGCCGCATCCGGTCGTGTTTTTCGGCAAGGCCATCGGCTTTTTTCGACGCGCGGCTGAACCGGGAG
4359      4369      4379      4389      4399      4409      4419
AspLeuGluAspSerAlaArgLysPheArgGlyValValAlaIleLeuLeuLeuLeuGlyIleSerAlaTrpPhe
GACCTCGAGGATAGCGCGCGCAAATTTCTGTCGGCGTCGTCGCGATCCTTTTGTTGCTTGGCATCAGCGCCTGGTTC
4434      4444      4454      4464      4474      4484      4494
GlyHisLeuLeuHisArgLeuPheAlaValLeuGlyProLeuGlyPheLeuLeuGluAlaValLeuValAlaVal
GGCCATCTGCTGCATCGCCTGTTTCGCCGTCCTCGGACCGCTCGGCTTTCTGCTCGAGGCGGTTCTGGTCGCGGTC
4509      4519      4529      4539      4549      4559      4569
PheLeuAlaGlnLysSerLeuAlaAspHisValArgArgValAlaGlyGlyLeuArgGlnGlyGlyLeuGluGly
TTCCTGGCACAGAAGAGCCTCGCCGATCACGTGCGTCGCGTGCGCGGGGGCTTGCGACAGGGCGGGCTGGAAGGC
4584      4594      4604      4614      4624      4634      4644
GlyArgAlaAlaValSerMetIleValGlyArgAspProLysThrLeuAspGluProAlaValCysArgAlaAla
GGGCGTGCCGCCGTGTCGATGATCGTTGGTCGCGATCCAAAGACGCTCGACGAGCCGGCGGTCTGCCGTGCCGCG
4659      4669      4679      4689      4699      4709      4719
IleGluSerLeuAlaGluAsnPheSerAspGlyValValAlaProAlaPheTrpTyrAlaValAlaGlyLeuPro
ATCGAAAGCCTTGCCGAGAATTTCTCCGACGGCGTCGTGGCGCCGGCCTTCTGGTACGCGGTTGCCGGCCTGCCG
4734      4744      4754      4764      4774      4784      4794
GlyLeuLeuAlaTyrLysMetLeuAsnThrAlaAspSerMetIleGlyHisLysSerProLysTyrLeuHisPhe
GGGCTTCTTGCTACAAGATGCTGAACACCGCCGATTTCGATGATCGGCCACAAGTCGCCGAAATATCTGCACTTC
4809      4819      4829      4839      4849      4859      4869
GlyTrpAlaSerAlaArgLeuAspAspLeuAlaAsnLeuProAlaAlaArgLeuSerIleLeuLeuIleSerAla
GGCTGGGCCTCGGCCGACTCGACGATCTCGCCAACCTGCCGGCAGCGAGGCTCTCGATCCTTTTGATCTCAGCC
4884      4894      4904      4914      4924      4934      4944
GlyAlaLeuIleHisArgGlyAlaSerAlaAlaLysAspAlaLeuThrValAlaLeuArgAspHisGlyLeuHis
GGTGCGCTGATCCATCGTGCGCCAGCGCCGCCAAGGATGCGCTGACCGTGCCCTTCGCGACCATGGCCTGCAC
4959      4969      4979      4989      4999      5009      5019
ArgSerProAsnSerGlyTrpProGluAlaAlaMetAlaGlyAlaLeuAspLeuGlnLeuAlaGlyProArgIle
CGCTCGCCGAACCTCCGGCTGGCCGGAAGCGGCCATGGCCGGCGCGCTCGATCTGCAGCTTGCCGGTCCGCGGATC
5034      5044      5054      5064      5074      5084      5094
TyrGlyGlyValLysValSerGluProMetIleAsnGlyProGlyArgAlaValAlaThrSerGluAspIleAsp
TATGGCGGCGTCAAGGTCAGCGAACCTATGATCAACGGTCCGGGCCGAGCGGTTGCAACAAGCGAAGACATCGAC
5109      5119      5129      5139      5149      5159      5169
AlaGlyIleAlaValPheTyrGlyAlaCysThrValMetAlaGlyPheValLeuAlaIleAlaMetIle***
GCCGGTATTGCTGTATTTTATGGCGCCTGTACGGTCATGGCCGGGTTTGTCTTGCAATCGCAATGATTGA
5184      5194      5204      5214      5224      5234      5244

```

FIGURE 15.7

47/151

NOM = COBD
PREMIER RESIDU = 1
DERNIER RESIDU = 323

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	14	4.33	2058.98	6.02
2	LEU	L	45	13.93	5088.60	14.89
3	ILE	I	18	5.57	2035.44	5.96
4	MET	M	8	2.48	1048.32	3.07
5	VAL	V	27	8.36	2674.89	7.83
6	SER	S	17	5.26	1479.51	4.33
7	PRO	P	15	4.64	1455.75	4.26
8	THR	T	6	1.86	606.30	1.77
9	ALA	A	51	15.79	3623.04	10.60
10	TYR	Y	5	1.55	815.30	2.39
11	*	*	0	0.00	0.00	0.00
12	HIS	H	9	2.79	1233.54	3.61
13	GLN	Q	3	0.93	384.18	1.12
14	ASN	N	6	1.86	684.24	2.00
15	LYS	K	9	2.79	1152.81	3.37
16	ASP	D	18	5.57	2070.54	6.06
17	GLU	E	11	3.41	1419.44	4.15
18	CYS	C	2	0.62	206.02	0.60
19	TRP	W	6	1.86	1116.48	3.27
20	ARG	R	20	6.19	3122.00	9.14
21	GLY	G	33	10.22	1881.66	5.51
22	-	-	0	0.00	0.00	0.00

RESIDUS = 323
POIDS MOLECULAIRE = 34175.
INDEX DE POLARITE (%) = 31.
POINT ISOELECTRIQUE = 8.00
DO 260 (1mg/ml) = 0.789 DO 280 (1mg/ml) = 1.150

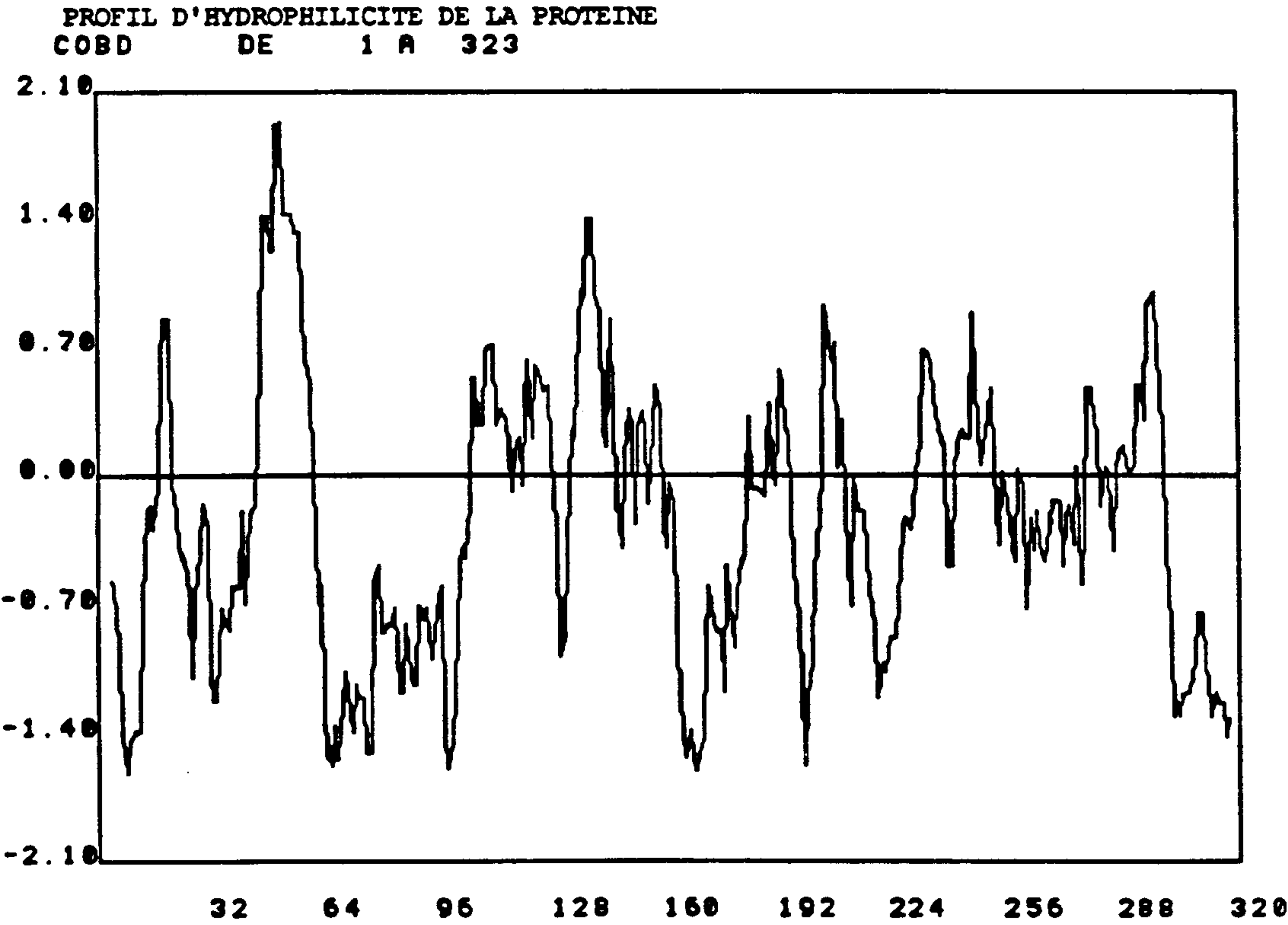


FIGURE 15.8

48 / 151

GENE cobE ET PROTEINE COBESEQUENCE DU FRAGMENT ClaI-HindIII-HindIII-HindIII DE 5396 BP, DE 549 A 1010

```

MetProSerGlyGlnHisSerAlaGlnThrThrLysAlaGlyAlaGlyLeuValLeuGlyLeuGlyCysGluArg
ATGCCATCGGGCCAACACTCTGCACAGACGACGAAAGCAGGAGCCGGGCTGGTGCTCGGGCTCGGCTGCGAGCGT
549      559      569      579      589      599      609
ArgThrProAlaGluGluValIleAlaLeuAlaGluArgAlaLeuAlaAspAlaGlyValAlaProGlyAspLeu
CGCACGCCGGCCGAAGAGGTGATCGCCCTTGCCGAGCGTGCGCTTGCCGATGCCGGTGTTGCGCCCGGCGATCTG
624      634      644      654      664      674      684
ArgLeuValAlaSerLeuAspAlaArgAlaGluGluProAlaIleLeuAlaAlaAlaGlnHisPheAlaValPro
CGGCTGGTCGCCTCGCTCGATGCTCGCGCCGAGGAGCCGGCGATCCTGGCGGCCGCTCAGCATTTCGCGGTTCCG
699      709      719      729      739      749      759
AlaAlaPheTyrAspAlaAlaThrLeuGluAlaGluAlaSerArgLeuAlaAsnProSerGluIleValPheAla
GCCGCGTTCTACGATGCCGCCACGCTCGAAGCCGAAGCTTCCCGGCTCGCCAACCCGTCCGAGATCGTCTTTGCC
774      784      794      804      814      824      834
TyrThrGlyCysHisGlyValAlaGluGlyAlaAlaLeuValGlyAlaGlyArgGluAlaValLeuIleValGln
TACACGGGTTGTCATGGCGTTGCCGAGGGTGCAGCGCTCGTCGGCGCCGGTCGCGAAGCCGTGCTGATTGTGCAG
849      859      869      879      889      899      909
LysIleValSerAlaHisAlaThrAlaAlaLeuAlaGlyProAlaThrLeuArgAlaGluLysArgIleGlnAla
AAGATCGTCTCCGCCCATGCGACGGCCGCACTTGCCGGGCCGCGACCTTGCGCGCCGAAAAGCGCATCCAGGCG
924      934      944      954      964      974      984
AlaGluAlaVal***
GCGGAGGCTGTCTGA
999      1009      1019      1029      1039      1049      1059

```

FIGURE 15.9

49/151

NOM = COBE

PREMIER RESIDU = 1

DERNIER RESIDU = 154

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	3	1.95	441.21	2.85
2	LEU	L	15	9.74	1696.20	10.96
3	ILE	I	6	3.90	678.48	4.38
4	MET	M	1	0.65	131.04	0.85
5	VAL	V	12	7.79	1188.84	7.68
6	SER	S	6	3.90	522.18	3.37
7	PRO	P	7	4.55	679.35	4.39
8	THR	T	7	4.55	707.35	4.57
9	ALA	A	41	26.62	2912.64	18.82
10	TYR	Y	2	1.30	326.12	2.11
11	*	*	0	0.00	0.00	0.00
12	HIS	H	4	2.60	548.24	3.54
13	GLN	Q	5	3.25	640.30	4.14
14	ASN	N	1	0.65	114.04	0.74
15	LYS	K	3	1.95	384.27	2.48
16	ASP	D	4	2.60	460.12	2.97
17	GLU	E	13	8.44	1677.52	10.84
18	CYS	C	2	1.30	206.02	1.33
19	TRP	W	0	0.00	0.00	0.00
20	ARG	R	9	5.84	1404.90	9.08
21	GLY	G	13	8.44	741.26	4.79
22	-	-	0	0.00	0.00	0.00

RESIDUS = 154

POIDS MOLECULAIRE = 15478.

INDEX DE POLARITE (%) = 34.

POINT ISOELECTRIQUE = 5.61

DO 260 (1mg/ml) = 0.113 DO 280 (1mg/ml) = 0.154

PROFIL D'HYDROPHILICITE DE LA PROTEINE
COBE DE 1 A 154

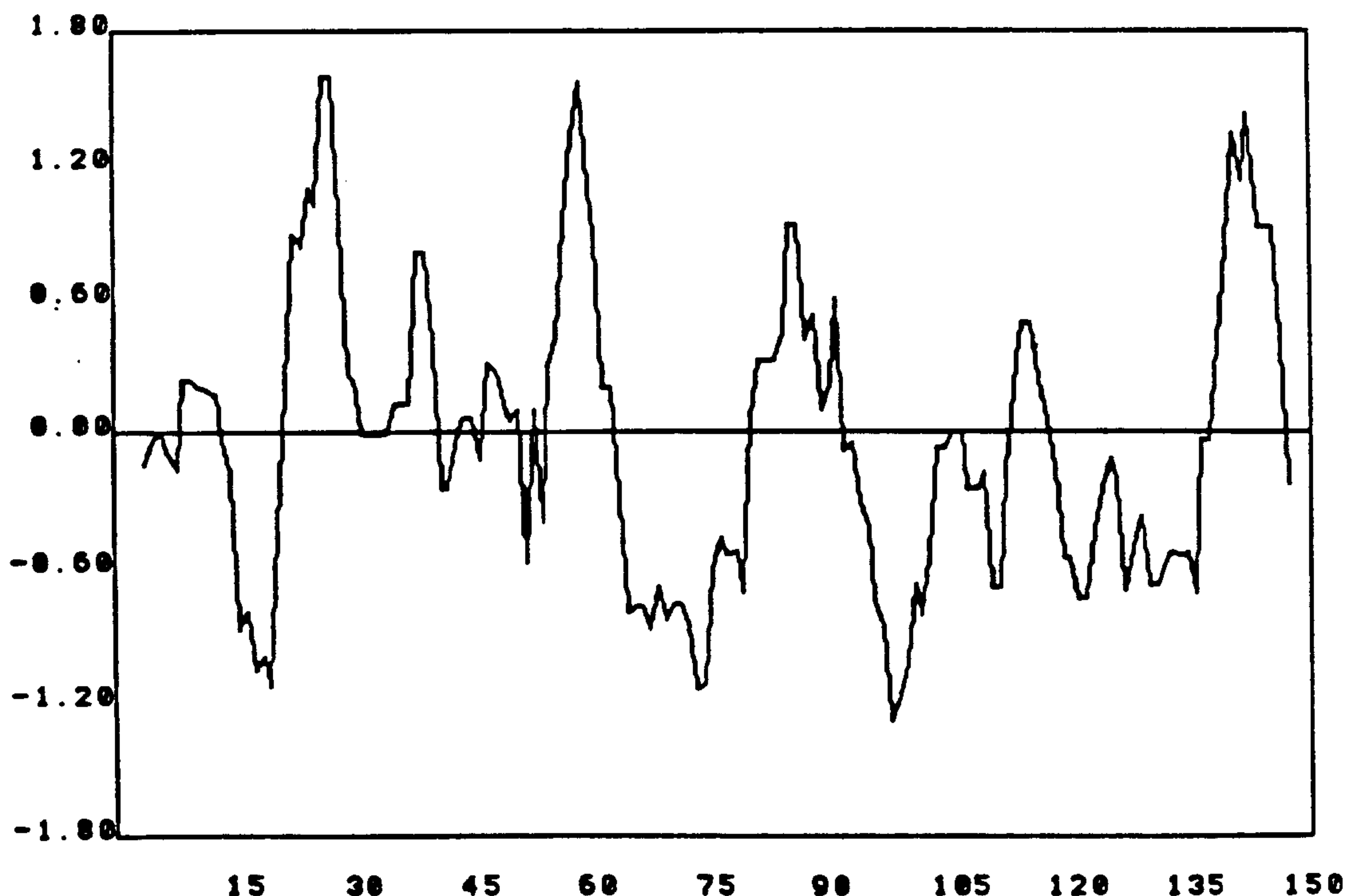


FIGURE 15.10

50 / 151

GENE cobF ET PROTEINE COBF

SEQUENCE DU FRAGMENT DE 8753 BP, DE 736 A 1521

```

MetAlaGluAlaGlyMetArgLysIleLeuIleIleGlyIleGlySerGlyAsnProGluHisMetThrValGln
ATGGCGGAGGCGGGCATGCGCAAAATTCTGATCATCGGCATCGGTTCGGGCAATCCCGAACACATGACCGTGCAG
736      746      756      766      776      786      796
AlaIleAsnAlaLeuAsnCysAlaAspValLeuPheIleProThrLysGlyAlaLysLysThrGluLeuAlaGlu
GCGATCAACGCGCTGAACGCGCCGACGTGCTCTTTATCCCGACCAAGGGAGCGAAGAAGACCGAGCTTGCCGAA
811      821      831      841      851      861      871
ValArgArgAspIleCysAlaArgTyrValThrArgLysAspSerArgThrValGluPheAlaValProValArg
GTGCGCCGCGACATCTGCGCCCGCTACGTACGCGCAAGGACAGCCGACCGTCGAGTTCGCGGTGCCCGTGCGG
886      896      906      916      926      936      946
ArgThrGluGlyValSerTyrAspGlySerValAspAspTrpHisAlaGlnIleAlaGlyIleTyrGluAlaLeu
CGCACCGAAGGCGTCAGCTATGACGGCAGCGTCGATGACTGGCACGCCAGATCGCTGGGATTTACGAAGCGCTT
961      971      981      991      1001      1011      1021
LeuSerLysGluLeuGlyGluGluGlyThrGlyAlaPheLeuValTrpGlyAspProMetLeuTyrAspSerThr
CTATCGAAGGAGTTGGGCGAAGAGGGAAGTGGCGCGTTTCTCGTCTGGGGCGACCCGATGCTCTATGACAGCACC
1036      1046      1056      1066      1076      1086      1096
IleArgIleValGluArgValLysAlaArgGlyGluValAlaPheAlaTyrAspValIleProGlyIleThrSer
ATTCGCATCGTCGAGCGGGTCAAGGCACGCGGTGAGGTGCGCTTCGCCTACGACGTCATTCCCGGGATCACCAGT
1111      1121      1131      1141      1151      1161      1171
LeuGlnAlaLeuCysAlaSerHisArgIleProLeuAsnLeuValGlyLysProValGluIleThrThrGlyArg
CTGCAGGCGCTTTGCGCCAGCCACCGCATTCCGCTGAACCTCGTCGGCAAGCCGGTGGAGATCACCACGGGGCGT
1186      1196      1206      1216      1226      1236      1246
ArgLeuHisGluSerPheProGluLysSerGlnThrSerValValMetLeuAspGlyGluGlnAlaPheGlnArg
CGGCTGCACGAAAGCTTTCCCGAGAAGAGCCAGACCTCGGTTCGTCATGCTCGATGGCGAACAGGCGTTTCAGCGG
1261      1271      1281      1291      1301      1311      1321
ValGluAspProGluAlaGluIleTyrTrpGlyAlaTyrLeuGlyThrArgAspGluIleValIleSerGlyArg
GTCGAGGACCCGGAGGCGGAGATCTATTGGGGCGCCTATCTCGGCACGCGGGATGAGATCGTCATTTCGCGCCGC
1336      1346      1356      1366      1376      1386      1396
ValAlaGluValLysAspArgIleLeuGluThrArgAlaAlaAlaArgAlaLysMetGlyTrpIleMetAspIle
GTGGCTGAGGTGAAGGACCGGATCCTTGAAACGCGGGCGGGCGCGCGCGAAGATGGGATGGATCATGGACATC
1411      1421      1431      1441      1451      1461      1471
TyrLeuLeuArgLysGlyAlaAspPheAspGlu***
TATCTCCTGCGCAAGGGCGCCGACTTCGACGAGTGA
1486      1496      1506      1516

```

FIGURE 16.1

51/151

PROTEINE COBF

PREMIER RESIDU = 1
DERNIER RESIDU = 261

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	7	2.68	1029.49	3.56
2	LEU	L	19	7.28	2148.52	7.43
3	ILE	I	21	8.05	2374.68	8.21
4	MET	M	7	2.68	917.28	3.17
5	VAL	V	22	8.43	2179.54	7.53
6	SER	S	12	4.60	1044.36	3.61
7	PRO	P	9	3.45	873.45	3.02
8	THR	T	14	5.36	1414.70	4.89
9	ALA	A	27	10.34	1918.08	6.63
10	TYR	Y	8	3.07	1304.48	4.51
11	*	*	0	0.00	0.00	0.00
12	HIS	H	4	1.53	548.24	1.90
13	GLN	Q	6	2.30	768.36	2.66
14	ASN	N	4	1.53	456.16	1.58
15	LYS	K	12	4.60	1537.08	5.31
16	ASP	D	16	6.13	1840.48	6.36
17	GLU	E	23	8.81	2967.92	10.26
18	CYS	C	3	1.15	309.03	1.07
19	TRP	W	4	1.53	744.32	2.57
20	ARG	R	21	8.05	3278.10	11.33
21	GLY	G	22	8.43	1254.44	4.34

RESIDUS = 261
POIDS MOLECULAIRE = 28927.
INDEX DE POLARITE (%) = 43.
POINT ISOELECTRIQUE = 5.70
DO 260 (1mg/ml) = 0.705 DO 280 (1mg/ml) = 1.097

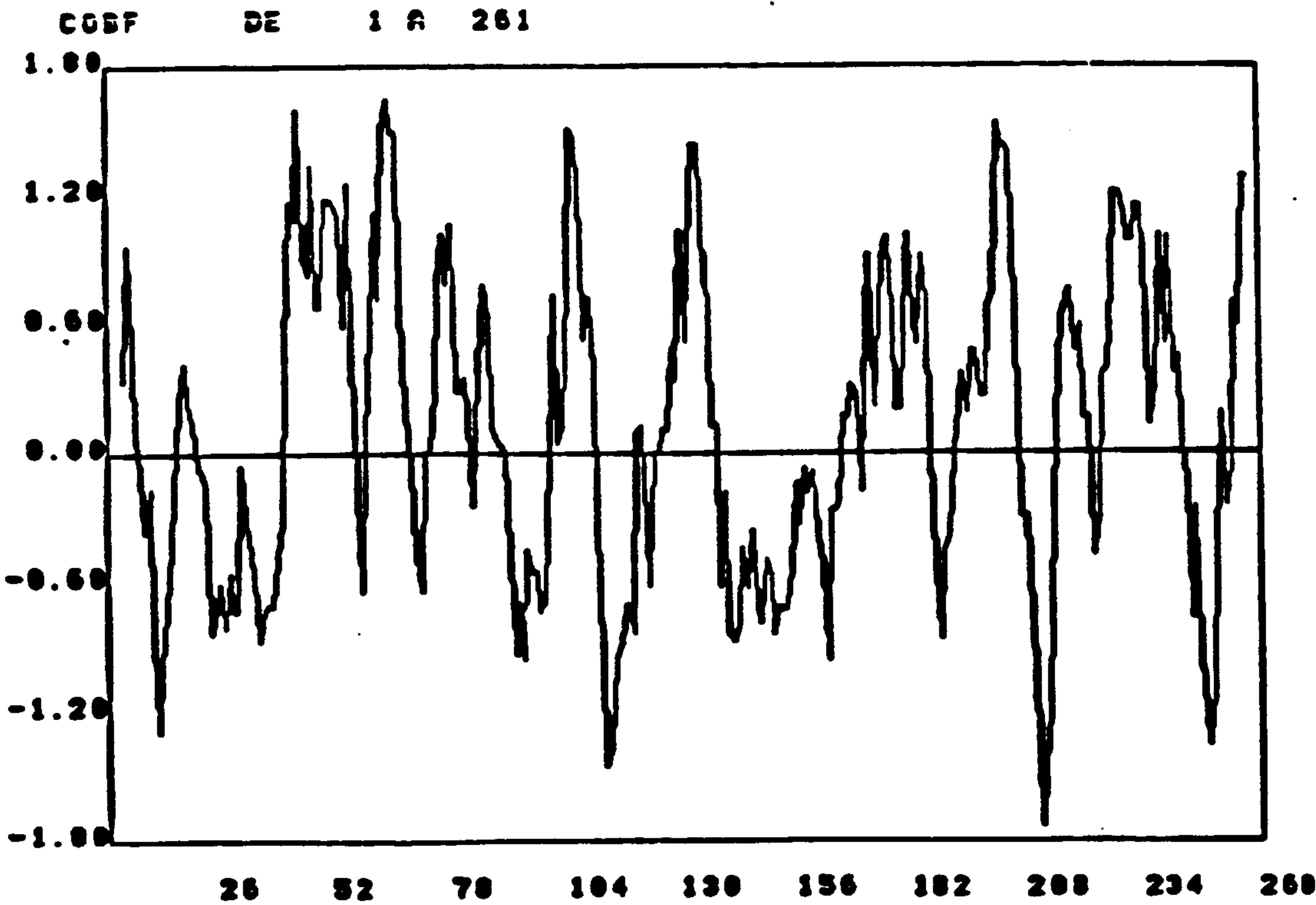


FIGURE 16.2

GENE cobG ET PROTEINE COBG

52/151

SEQUENCE DU FRAGMENT DE 8753 BP, DE 1620 A 2999

```

MetThrAspLeuMetThrSerCysAlaLeuProLeuThrGlyAspAlaGlyThrValAlaSerMetArgArgGly
ATGACGGATTGATGACCAGCTGCGCCCTTCCATTGACCGGAGATGCCGGCACCGTCGCTTCGATGCGCCGCGGC
1620      1630      1640      1650      1660      1670      1680
AlaCysProSerLeuAlaGluProMetGlnThrGlyAspGlyLeuLeuValArgValArgProThrAspAspSer
GCCTGCCCCGTCCTTGGCAGAGCCGATGCAGACCGGCGACGGCCTGCTCGTGAGGGTGAGGCCAACGGATGACAGC
1695      1705      1715      1725      1735      1745      1755
LeuThrLeuProLysValIleAlaLeuAlaThrAlaAlaGluArgPheGlyAsnGlyIleIleGluIleThrAla
CTGACGCTGCCGAAGGTCATTGCCCTTGCCACGGCTGCCGAGCGCTTCGGCAATGGCATCATCGAGATTACCGCG
1770      1780      1790      1800      1810      1820      1830
ArgGlyAsnLeuGlnLeuArgGlyLeuSerAlaAlaSerValProArgLeuAlaGlnAlaIleGlyAspAlaGlu
CGCGGAAACCTGCAGCTTCGCGGCCTGAGCGCGGCTTCGGTGCCAAGGCTGGCGCAGGCGATCGGCGATGCGGAG
1845      1855      1865      1875      1885      1895      1905
IleAlaIleAlaGluGlyLeuAlaIleGluValProProLeuAlaGlyIleAspProAspGluIleAlaAspPro
ATCGCCATTGCCGAGGGGCTCGCGATCGAGGTGCCGCCCTGGCCGGCATCGACCCGGACGAGATCGCCGATCCG
1920      1930      1940      1950      1960      1970      1980
ArgProIleAlaThrGluLeuArgGluAlaLeuAspValArgGlnValProLeuLysLeuAlaProLysLeuSer
CGGCCGATTGCCACTGAGCTTCGTGAAGCGTTGGATGTGCGCCAGGTGCCGTTGAAGCTTGACCCAAATTATCC
1995      2005      2015      2025      2035      2045      2055
ValValIleAspSerGlyGlyArgPheGlyLeuGlyAlaValValAlaAspIleArgLeuGlnAlaValSerThr
GTCGTCATCGATAGCGGTGGCCGGTTTGGTCTCGGCGCTGTGCTCGCCGACATTCGCCTTCAGGCGGTTTCGACT
2070      2080      2090      2100      2110      2120      2130
ValAlaGlyValAlaTrpValLeuSerLeuGlyGlyThrSerThrLysAlaSerSerValGlyThrLeuAlaGly
GTCGCGGGGGTGGCCTGGGTGCTGTCGCTTGGCGGCACGTCAACGAAGGCATCGAGCGTCGGGACGTTGGCCGGC
2145      2155      2165      2175      2185      2195      2205
AsnAlaValValProAlaLeuIleThrIleLeuGluLysLeuAlaSerLeuGlyThrThrMetArgGlyArgAsp
AACGCGGTGCTGCCGGCCCTGATCACCATTCTCGAGAACTGGCGAGCCTGGGCACGACGATGCGCGGGCGCGAT
2220      2230      2240      2250      2260      2270      2280
LeuAspProSerGluIleArgAlaLeuCysArgCysGluThrSerSerGluArgProAlaAlaProArgSerAla
CTGGACCCGTCGGAAATCCGCGCGCTCTGTGCTGTGAGACATCGTCCGAACGCCCGGCCGCTCCGCGTTCGGCC
2295      2305      2315      2325      2335      2345      2355
AlaIleProGlyIleHisAlaLeuGlyAsnAlaAspThrValLeuGlyLeuGlyLeuAlaPheAlaGlnValGlu
GCAATACCCGGCATTTCATGCGCTGGGTAACGCCGACACCGTTCTCGGCCTCGGTCTGGCCTTTGCTCAGGTGGAG
2370      2380      2390      2400      2410      2420      2430
AlaAlaAlaLeuAlaSerTyrLeuHisGlnValGlnAlaLeuGlyAlaAsnAlaIleArgLeuAlaProGlyHis
GCCGCCGCGCTGGCATCTACCTGCATCAGGTCCAGGCGCTTGGCGCCAATGCGATCCGGCTTGCGCCCGGGCAC
2445      2455      2465      2475      2485      2495      2505
AlaPhePheValLeuGlyLeuCysProGluThrAlaAlaValAlaGlnSerLeuAlaAlaSerHisGlyPheArg
GCCTTCTTCGTCCTCGGCCTTTGCCCCGAGACCGCGGCTGTGGCGCAGAGCCTGGCAGCGTCACACGGTTTTCGC
2520      2530      2540      2550      2560      2570      2580
IleAlaGluGlnAspProArgAsnAlaIleAlaThrCysAlaGlySerLysGlyCysAlaSerAlaTrpMetGlu
ATTGCCGAGCAGGATCCGCGCAATGCGATCGCCACCTGCGCCGGCAGCAAGGGTTGCGCCTCGGCGTGATGGAA
2595      2605      2615      2625      2635      2645      2655
ThrLysGlyMetAlaGluArgLeuValGluThrAlaProGluLeuLeuAspGlySerLeuThrValHisLeuSer
ACCAAGGGCATGGCCGAGCGCCTCGTCGAGACGGCGCCGGAATTGCTCGACGGGTCGCTCACCGTGCATCTCTCC
2670      2680      2690      2700      2710      2720      2730
GlyCysAlaLysGlyCysAlaArgProLysProSerGluLeuThrLeuValGlyAlaProSerGlyTyrGlyLeu
GGCTGCGCCAAGGGCTGCGCCCGGCCGAAGCCGTCCGAAGTACGCTTGTCGGTGCGCCATCAGGATACGGGCTT
2745      2755      2765      2775      2785      2795      2805
ValValAsnGlyAlaAlaAsnGlyLeuProSerAlaTyrThrAspGluAsnGlyMetGlySerAlaLeuAlaArg
GTCGTAAATGGGGCTGCCAATGGCTTGCCAAGCGCCTACACCGATGAGAATGGAATGGGATCCGCCCTTGCCCGG
2820      2830      2840      2850      2860      2870      2880
LeuGlyArgLeuValArgGlnAsnLysAspAlaGlyGluSerAlaGlnSerCysLeuThrArgLeuGlyAlaAla
CTCGGCCGGCTGGTGCGGCAAAACAAAGACGCTGGCGAATCGGCGCAGTCCTGTCTTACACGGCTCGGAGCTGCG
2895      2905      2915      2925      2935      2945      2955
ArgValSerAlaAlaPheGluGlnGly***
CGCGTCTCGGCAGCGTTCTGAACAGGGATAG
2970      2980      2990      3000

```

FIGURE 16.3

53/151

PROTEINE COBG

PREMIER RESIDU = 1

DERNIER RESIDU = 459

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	7	1.53	1029.49	2.21
2	LEU	L	56	12.20	6332.48	13.57
3	ILE	I	21	4.58	2374.68	5.09
4	MET	M	8	1.74	1048.32	2.25
5	VAL	V	31	6.75	3071.17	6.58
6	SER	S	32	6.97	2784.96	5.97
7	PRO	P	26	5.66	2523.30	5.41
8	THR	T	27	5.88	2728.35	5.85
9	ALA	A	78	16.99	5541.12	11.88
10	TYR	Y	3	0.65	489.18	1.05
11	*	*	0	0.00	0.00	0.00
12	HIS	H	5	1.09	685.30	1.47
13	GLN	Q	13	2.83	1664.78	3.57
14	ASN	N	10	2.18	1140.40	2.44
15	LYS	K	10	2.18	1280.90	2.75
16	ASP	D	19	4.14	2185.57	4.68
17	GLU	E	24	5.23	3096.96	6.64
18	CYS	C	10	2.18	1030.10	2.21
19	TRP	W	2	0.44	372.16	0.80
20	ARG	R	29	6.32	4526.90	9.70
21	GLY	G	48	10.46	2736.96	5.87
22	-	-	0	0.00	0.00	0.00

RESIDUS = 459

POIDS MOLECULAIRE = 46661.

INDEX DE POLARITE (%) = 37.

POINT ISOELECTRIQUE = 6.41

DO 260 (1mg/ml) = 0.215 DO 280 (1mg/ml) = 0.315

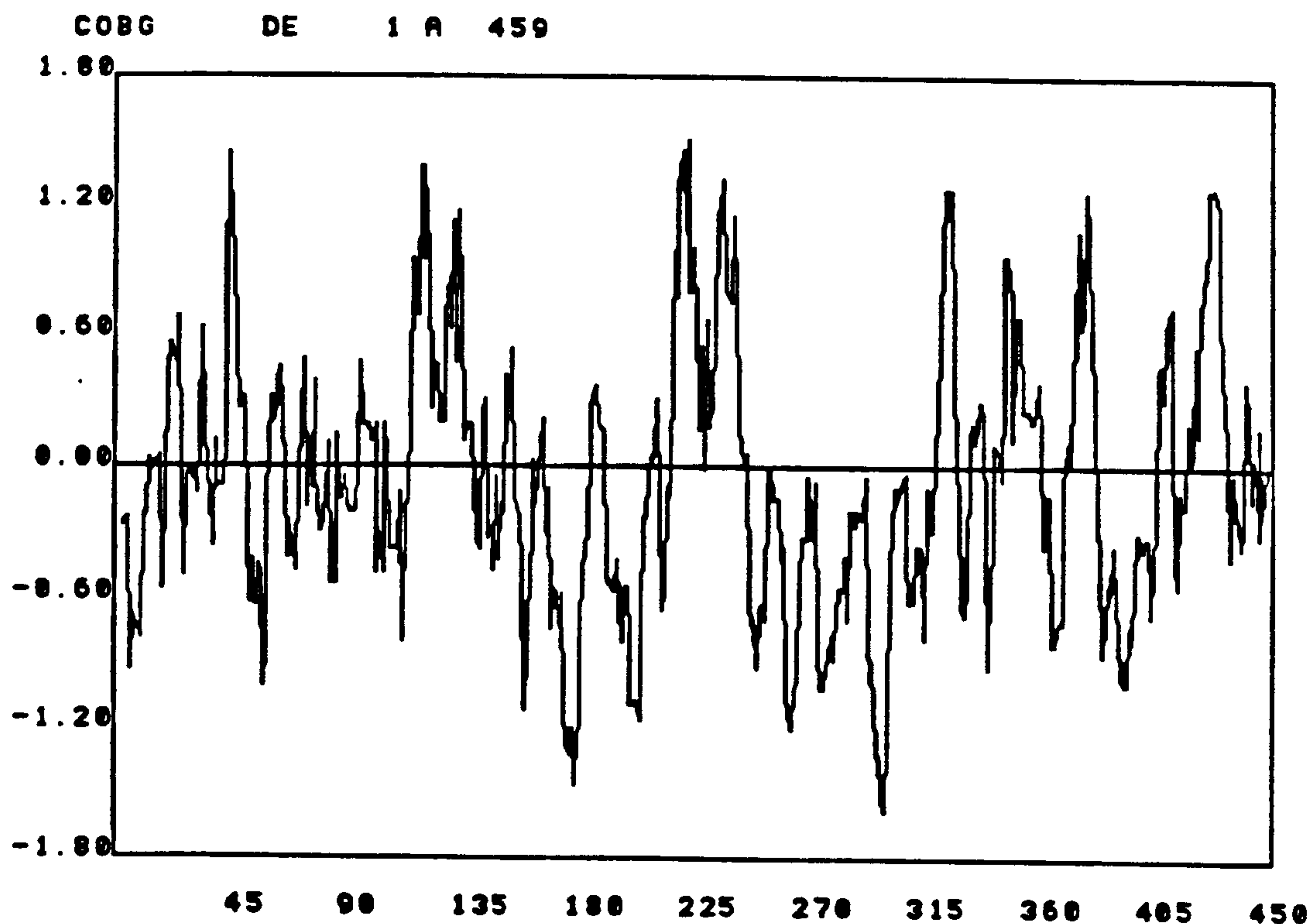


FIGURE 16.4

54/151

GENE cobH ET PROTEINE COBH

SEQUENCE DU FRAGMENT DE 8753 BP, DE 3002 A 3634

```

MetProGluTyrAspTyrIleArgAspGlyAsnAlaIleTyrGluArgSerPheAlaIleIleArgAlaGluAla
ATGCCTGAGTATGATTACATTCGCGATGGCAACGCCATCTACGAGCGTTCCTTCGCCATCATCCGCGCCGAGGCC
3002      3012      3022      3032      3042      3052      3062
AspLeuSerArgPheSerGluGluGluAlaAspLeuAlaValArgMetValHisAlaCysGlySerValGluAla
GATCTGTCGCGCTTCTCCGAAGAGGAAGCGGATCTGGCTGTGCGCATGGTGCACGCCTGCGGTTCCGTCGAGGCG
3077      3087      3097      3107      3117      3127      3137
ThrArgGlnPheValPheSerProAspPheValSerSerAlaArgAlaAlaLeuLysAlaGlyAlaProIleLeu
ACCAGGCAGTTCGTGTTTTCTCCCGATTTCGTAAGCTCGGCCCGTGC GGCGCTGAAAGCCGGTGCGCCGATCCTC
3152      3162      3172      3182      3192      3202      3212
CysAspAlaGluMetValAlaHisGlyValThrArgAlaArgLeuProAlaGlyAsnGluValIleCysThrLeu
TGCGATGCCGAGATGGTTGCGCACGGTGTACCCGCGCCCGTCTGCCGGCCGGCAACGAGGTGATCTGCACGCTG
3227      3237      3247      3257      3267      3277      3287
ArgAspProArgThrProAlaLeuAlaAlaGluIleGlyAsnThrArgSerAlaAlaAlaLeuLysLeuTrpSer
CGCGATCCTCGCACGCCCCGCACTTGCGGCCGAGATCGGCAACACCCGCTCCGCCGAGCCCTGAAGCTCTGGAGC
3302      3312      3322      3332      3342      3352      3362
GluArgLeuAlaGlySerValValAlaIleGlyAsnAlaProThrAlaLeuPhePheLeuLeuGluMetLeuArg
GAGCGGCTGGCCGGTTCGGTGGTTCGCGATCGGCAACGCGCCGACGGCGTTGTTCTTCTTCTTGAAATGCTGCGC
3377      3387      3397      3407      3417      3427      3437
AspGlyAlaProLysProAlaAlaIleLeuGlyMetProValGlyPheValGlyAlaAlaGluSerLysAspAla
GACGGCGCGCCGAAGCCGGCGGCAATCCTCGGCATGCCCGTCGGTTTCGTTCGGTTCGGCGGAATCGAAGGATGCG
3452      3462      3472      3482      3492      3502      3512
LeuAlaGluAsnSerTyrGlyValProPheAlaIleValArgGlyArgLeuGlyGlySerAlaMetThrAlaAla
CTGGCCGAGAACTCCTATGGCGTTCCCTTCGCCATCGTGCGCGGCCGCTCGGCGGGAGTGCCATGACGGCGGCA
3527      3537      3547      3557      3567      3577      3587
AlaLeuAsnSerLeuAlaArgProGlyLeu***
GCGCTTAAC TCGCTCGCGAGGCCGGGCCTGTGA
3602      3612      3622      3632

```

FIGURE 16.5

55/151

PROTEINE COBH

PREMIER RESIDU = 1
DERNIER RESIDU = 210

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	9	4.29	1323.63	6.00
2	LEU	L	20	9.52	2261.60	10.26
3	ILE	I	10	4.76	1130.80	5.13
4	MET	M	6	2.86	786.24	3.57
5	VAL	V	14	6.67	1386.98	6.29
6	SER	S	14	6.67	1218.42	5.53
7	PRO	P	12	5.71	1164.60	5.28
8	THR	T	7	3.33	707.35	3.21
9	ALA	A	40	19.05	2841.60	12.89
10	TYR	Y	4	1.90	652.24	2.96
11	*	*	0	0.00	0.00	0.00
12	HIS	H	2	0.95	274.12	1.24
13	GLN	Q	1	0.48	128.06	0.58
14	ASN	N	6	2.86	684.24	3.10
15	LYS	K	4	1.90	512.36	2.32
16	ASP	D	9	4.29	1035.27	4.70
17	GLU	E	14	6.67	1806.56	8.19
18	CYS	C	3	1.43	309.03	1.40
19	TRP	W	1	0.48	186.08	0.84
20	ARG	R	17	8.10	2653.70	12.03
21	GLY	G	17	8.10	969.34	4.40
22	-	-	0	0.00	0.00	0.00

RESIDUS = 210
 POIDS MOLECULAIRE = 22050.
 INDEX DE POLARITE (%) = 35.
 POINT ISOELECTRIQUE = 6.22
 DO 260 (1mg/ml) = 0.291 DO 280 (1mg/ml) = 0.467

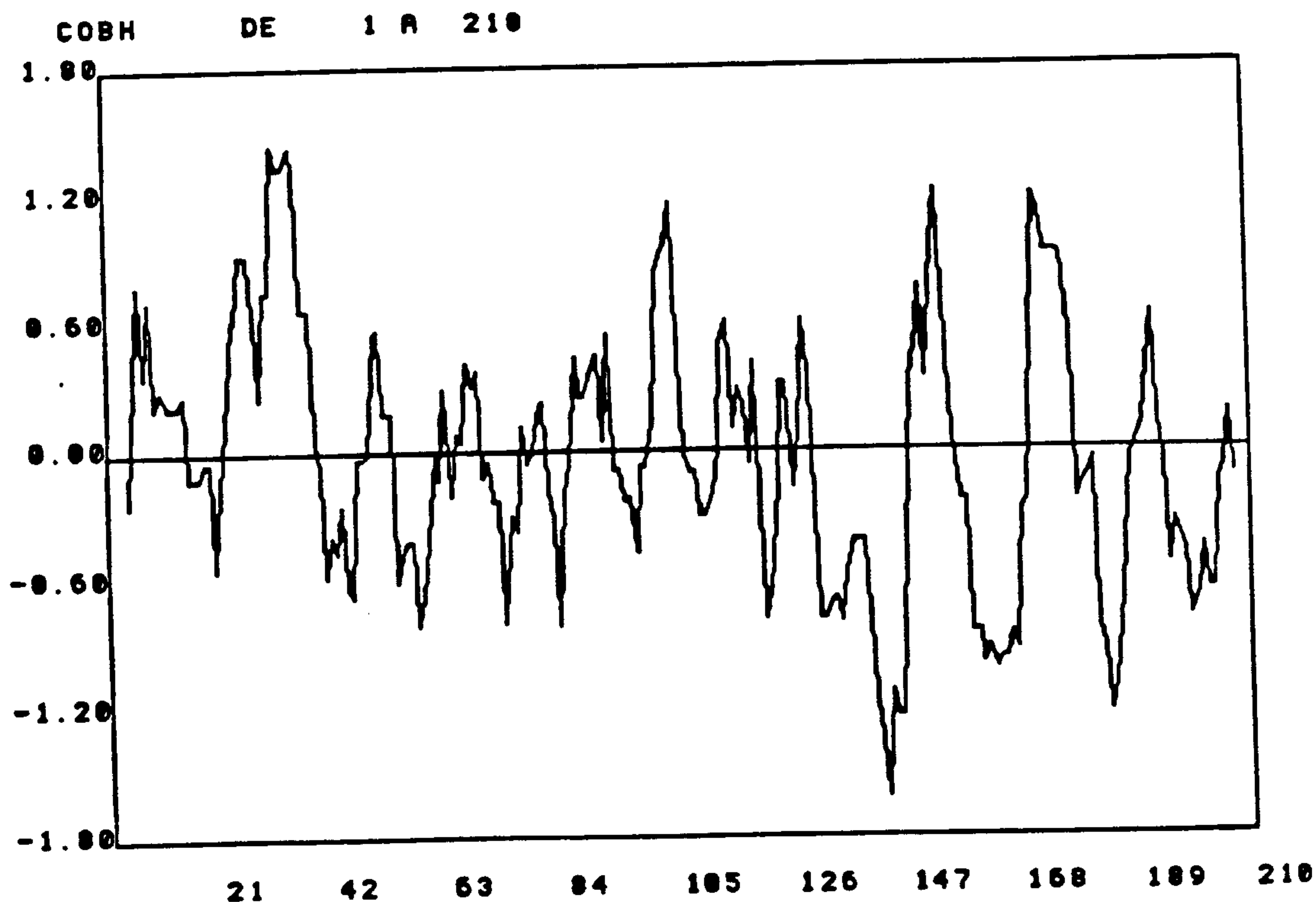


FIGURE 16.6

56 / 151

GENE cobI ET PROTEINE COBI

SEQUENCE DU FRAGMENT DE 8753 BP, DE 3631 A 4368

```

MetSerGlyValGlyValGlyArgLeuIleGlyValGlyThrGlyProGlyAspProGluLeuLeuThrValLys
GTGAGCGGCGTCGGCGTGGGGCGCCTGATCGGTGTTGGGACCGGCCCGGTGATCCGGAACCTTTGACGGTCAAG
3631      3641      3651      3661      3671      3681      3691
AlaValLysAlaLeuGlyGlnAlaAspValLeuAlaTyrPheAlaLysAlaGlyArgSerGlyAsnGlyArgAla
GCGGTGAAGGCGCTCGGGCAAGCCGATGTGCTTGCCTATTTCCGCAAGGCCGGCGAAGCGGTAACGGCCGCGCG
3706      3716      3726      3736      3746      3756      3766
ValValGluGlyLeuLeuLysProAspLeuValGluLeuProLeuTyrTyrProValThrThrGluIleAspLys
GTGGTCGAGGGTCTGCTGAAGCCCGATCTTGTCGAGCTGCCGCTATACTATCCGGTGACGACCGAAATCGACAAG
3781      3791      3801      3811      3821      3831      3841
AspAspGlyAlaTyrLysThrGlnIleThrAspPheTyrAsnAlaSerAlaGluAlaValAlaAlaHisLeuAla
GACGATGGCGCCTACAAGACCCAGATCACCGACTTCTACAATGCGTCGGCCGAAGCGGTAGCGGCGCATCTTGCC
3856      3866      3876      3886      3896      3906      3916
AlaGlyArgThrValAlaValLeuSerGluGlyAspProLeuPheTyrGlySerTyrMetHisLeuHisValArg
GCCGGGCGCACGGTCGCCGTGCTCAGTGAAGGCGACCCGCTGTTCTATGGTTCCTACATGCATCTGCATGTGCGG
3931      3941      3951      3961      3971      3981      3991
LeuAlaAsnArgPheProValGluValIleProGlyIleThrAlaMetSerGlyCysTrpSerLeuAlaGlyLeu
CTCGCCAATCGTTTCCCGGTGAGGTGATCCCCGGCATTACCGCCATGTCCGGCTGTTGGTCGCTTGCCGGCCTG
4006      4016      4026      4036      4046      4056      4066
ProLeuValGlnGlyAspAspValLeuSerValLeuProGlyThrMetAlaGluAlaGluLeuGlyArgArgLeu
CCGCTGGTGCAGGGCGACGACGTGCTCTCGGTGCTTCCGGGCACCATGGCCGAGGCCGAGCTCGGCCGCAGGCTT
4081      4091      4101      4111      4121      4131      4141
AlaAspThrGluAlaAlaValIleMetLysValGlyArgAsnLeuProLysIleArgArgAlaLeuAlaAlaSer
GCGGATACCGAAGCCGCCGTGATCATGAAGGTGCGGCGCAATTTGCCGAAGATCCGTCGGGCGCTCGCTGCCTCC
4156      4166      4176      4186      4196      4206      4216
GlyArgLeuAspGlnAlaValTyrValGluArgGlyThrMetLysAsnAlaAlaMetThrAlaLeuAlaGluLys
GGCCGTCTCGACCAGGCCGTCTATGTGCAACGCGGCACGATGAAGAACGCGCGATGACGGCTCTTGCGGAAAAG
4231      4241      4251      4261      4271      4281      4291
AlaAspAspGluAlaProTyrPheSerLeuValLeuValProGlyTrpLysAspArgPro***
GCCGACGACGAGGCGCCCTATTTCTCGCTGGTGCTCGTTCCCGGCTGGAAGGACCGACCATGA
4306      4316      4326      4336      4346      4356      4366

```

FIGURE 16.7

57/151

PROTEINE COBI			PREMIER RESIDU = 1		DERNIER RESIDU = 245	
			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	5	2.04	735.35	2.84
2	LEU	L	28	11.43	3166.24	12.24
3	ILE	I	7	2.86	791.56	3.06
4	MET	M	7	2.86	917.28	3.54
5	VAL	V	25	10.20	2476.75	9.57
6	SER	S	10	4.08	870.30	3.36
7	PRO	P	14	5.71	1358.70	5.25
8	THR	T	12	4.90	1212.60	4.69
9	ALA	A	34	13.88	2415.36	9.33
10	TYR	Y	9	3.67	1467.54	5.67
11	*	*	0	0.00	0.00	0.00
12	HIS	H	3	1.22	411.18	1.59
13	GLN	Q	4	1.63	512.24	1.98
14	ASN	N	5	2.04	570.20	2.20
15	LYS	K	11	4.49	1408.99	5.44
16	ASP	D	15	6.12	1725.45	6.67
17	GLU	E	13	5.31	1677.52	6.48
18	CYS	C	1	0.41	103.01	0.40
19	TRP	W	2	0.82	372.16	1.44
20	ARG	R	14	5.71	2185.40	8.44
21	GLY	G	26	10.61	1482.52	5.73
22	-	-	0	0.00	0.00	0.00

RESIDUS = 245
POIDS MOLECULAIRE = 25878.
INDEX DE POLARITE (%) = 36.
POINT ISOELECTRIQUE = 6.17
DO 260 (1mg/ml) = 0.512 DO 280 (1mg/ml) = 0.843

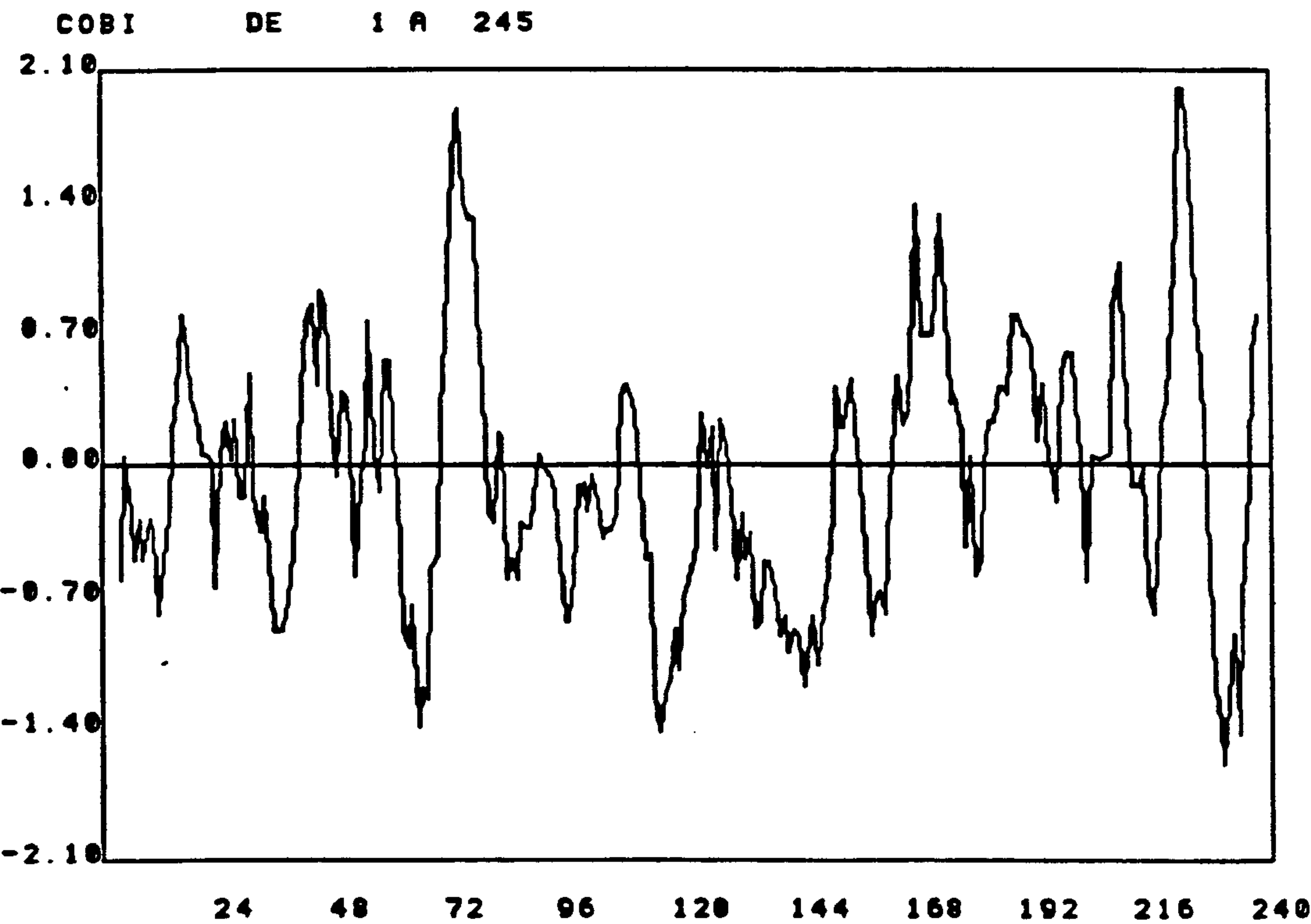


FIGURE 16.8

58 / 151

GENE cobJ ET PROTEINE COBJ

SEQUENCE DU FRAGMENT DE 8753 BP, DE 4365 A 5129

```

MetThrGlyThrLeuTyrValValGlyThrGlyProGlySerAlaLysGlnMetThrProGluThrAlaGluAla
ATGACCGGTACGCTCTATGTCGTCGGTACCGGACCGGGCAGCGCCAAGCAGATGACGCCGGAACGGCGGAAGCC
4365      4375      4385      4395      4405      4415      4425
ValAlaAlaAlaGlnGluPheTyrGlyTyrPheProTyrLeuAspArgLeuAsnLeuArgProAspGlnIleArg
GTTGCGGCCGCTCAGGAGTTTTACGGCTACTTTCCCTATCTCGACCGGCTGAACCTCAGACCGGATCAGATCCGT
4440      4450      4460      4470      4480      4490      4500
ValAlaSerAspAsnArgGluGluLeuAspArgAlaGlnValAlaLeuThrArgAlaAlaAlaGlyValLysVal
GTCGCCTCGGACAACCGCGAGGAGCTCGATCGGGCACAGGTCGCGCTGACGCGGGCTGCGGCAGGCGTGAAGGTC
4515      4525      4535      4545      4555      4565      4575
CysMetValSerGlyGlyAspProGlyValPheAlaMetAlaAlaAlaValCysGluAlaIleAspLysGlyPro
TGCATGGTCTCCGGTGGCGATCCCGGTGTCTTTGCCATGGCGGCCCGCTCTGCGAGGCGATCGACAAGGGACCG
4590      4600      4610      4620      4630      4640      4650
AlaGluTrpLysSerValGluLeuValIleThrProGlyValThrAlaMetLeuAlaValAlaAlaArgIleGly
GCGGAATGGAAGTCGGTTGAACTGGTGATCACGCCCGGCGTGACCGCGATGCTCGCCGTTGCCGCCCGCATCGGC
4665      4675      4685      4695      4705      4715      4725
AlaProLeuGlyHisAspPheCysAlaIleSerLeuSerAspAsnLeuLysProTrpGluValIleThrArgArg
GCGCCGCTCGGTCATGATTTCTGTGCGATCTCGCTTTCCGACAATCTGAAGCCCTGGGAAGTCATCACCCGGCGT
4740      4750      4760      4770      4780      4790      4800
LeuArgLeuAlaAlaGluAlaGlyPheValIleAlaLeuTyrAsnProIleSerLysAlaArgProTrpGlnLeu
CTCAGGCTGGCGGCGGAAGCGGGCTTCGTCATTGCCCTCTACAATCCGATCAGCAAGGCGCGGCCCTGGCAGCTC
4815      4825      4835      4845      4855      4865      4875
GlyGluAlaPheGluLeuLeuArgSerValLeuProAlaSerValProValIlePheGlyArgAlaAlaGlyArg
GGTGAGGCCTTCGAGCTTCTGCGCAGCGTTCTGCCGGCAAGCGTTCCGGTCATCTTCGGCCGTGCGGCCGGGCGG
4890      4900      4910      4920      4930      4940      4950
ProAspGluArgIleAlaValMetProLeuGlyGluAlaAspAlaAsnArgAlaAspMetAlaThrCysValIle
CCGGACGAACGGATCGCGGTGATGCCGCTCGGCGAGGCCGATGCCAACCGCGCCGACATGGCGACCTGCGTCATC
4965      4975      4985      4995      5005      5015      5025
IleGlySerProGluThrArgIleValGluArgAspGlyGlnProAspLeuValTyrThrProArgPheTyrAla
ATCGGCTCGCCGGAGACGCGCATCGTCGAGCGCGACGGCCAACCCGATCTCGTCTACACACCGCGCTTCTATGCA
5040      5050      5060      5070      5080      5090      5100
GlyAlaSerGln***
GGGGCGAGCCAGTGA
5115      5125

```

FIGURE 16.9

PROTEINE COBJ

59/151

PREMIER RESIDU = 1
DERNIER RESIDU = 254

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	8	3.15	1176.56	4.34
2	LEU	L	20	7.87	2261.60	8.35
3	ILE	I	13	5.12	1470.04	5.43
4	MET	M	7	2.76	917.28	3.39
5	VAL	V	23	9.06	2278.61	8.41
6	SER	S	11	4.33	957.33	3.53
7	PRO	P	18	7.09	1746.90	6.45
8	THR	T	12	4.72	1212.60	4.48
9	ALA	A	40	15.75	2841.60	10.49
10	TYR	Y	7	2.76	1141.42	4.21
11	*	*	0	0.00	0.00	0.00
12	HIS	H	1	0.39	137.06	0.51
13	GLN	Q	7	2.76	896.42	3.31
14	ASN	N	5	1.97	570.20	2.11
15	LYS	K	6	2.36	768.54	2.84
16	ASP	D	13	5.12	1495.39	5.52
17	GLU	E	16	6.30	2064.64	7.62
18	CYS	C	4	1.57	412.04	1.52
19	TRP	W	3	1.18	558.24	2.06
20	ARG	R	19	7.48	2965.90	10.95
21	GLY	G	21	8.27	1197.42	4.42
22	-	-	0	0.00	0.00	0.00

RESIDUS = 254
 POIDS MOLECULAIRE = 27088.
 INDEX DE POLARITE (%) = 35.
 POINT ISOELECTRIQUE = 5.43
 DO 260 (1mg/ml) = 0.575 DO 280 (1mg/ml) = 0.922

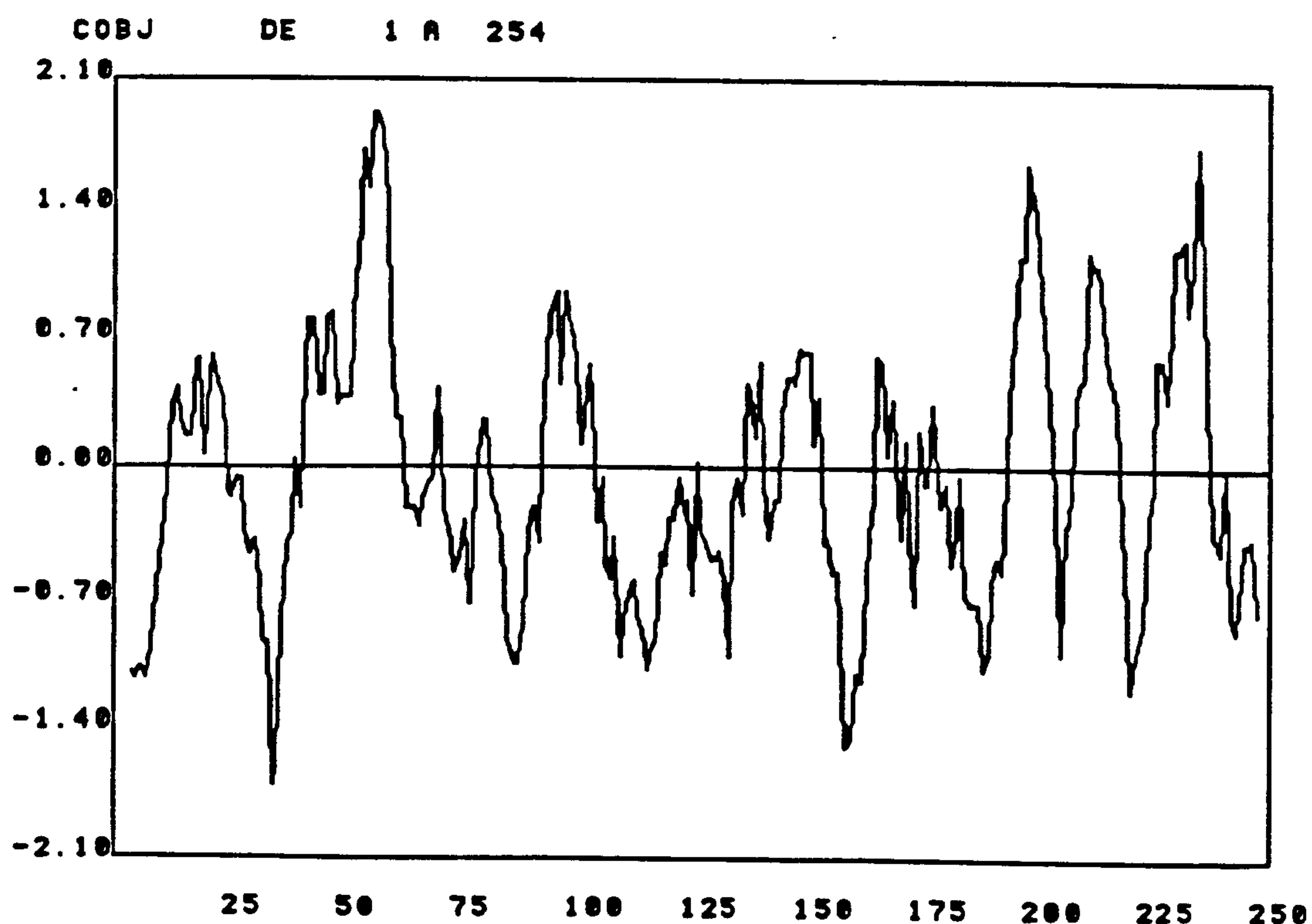


FIGURE 16.10

ENE cobK ET PROTEINE COBK

60/151

SEQUENCE DU FRAGMENT ECORI-ECORI DE 8753 PB DE 2861 A 3646 SUR LE BRIN
COMPLEMENTAIRE

```

      M A G S L F D T S A M E K P R I L I L G G T T E A
      ATGGCGGGTTCGCTGTTTCGACACGTCAGCCATGGAAAAACCTCGTATTCTGATTCTGGGTGGCACCACCGAGGCA
2861      2871      2881      2891      2901      2911      2921      2931
      R E L A R R L A E D V R Y D T A I S L A G R T A D
      CGCGAACTCGCGCGCCGCTTGGCCGAAGATGTCCGCTACGACACCGCCATCTCGCTGGCCGGCCGCACCGCGGAC
2936      2946      2956      2966      2976      2986      2996      3006
      P R P Q P V K T R I G G F G G A D G L A H F V H D
      CCGCGGCCGCGAGCCGGTCAAGACGCGCATCGGCGGCTTTGGCGGCGCCGATGGGCTGGCGCATTTCGTGCATGAC
3011      3021      3031      3041      3051      3061      3071      3081
      E N I A L L V D A T H P F A A R I S H N A A D A A
      GAAAACATCGCGCTGCTGGTCGATGCGACGCACCCCTTTGCCGCACGCATTTCGCACAACGCCGCGGACGCAGCG
3086      3096      3106      3116      3126      3136      3146      3156
      Q R T G V A L I A L R R P E W V P L P G D R W T A
      CAAAGAACCGGCGTTGCGCTTATCGCCCTCCGCCGACCGGAATGGGTGCCCTGCCTGGCGACCGCTGGACTGCT
3161      3171      3181      3191      3201      3211      3221      3231
      V D S V V E A V S A L G D R R R R R V F L A I G R Q
      GTCGATAGCGTTGTCGAGGCCGTCAGCGCGCTCGGTGATCGGCGACGCCGCGTCTTCCTGGCGATAGGTGACAG
3236      3246      3256      3266      3276      3286      3296      3306
      E A F H F E V A P Q H S Y V I R S V D P V T P P L
      GAAGCTTTCCACTTCGAGGTCGCGCCGCGACACAGCTACGTCATCCGCAGCGTCGATCCGGTGACGCCGCCGCTT
3311      3321      3331      3341      3351      3361      3371      3381
      N L P D Q E A I L A T G P F A E A D E A A L L R S
      AATCTGCCCCGACCAGGAGGCGATCCTGGCGACCGGTCCCTTTGCGGAAGCCGACGAAGCCGCGTTGCTCAGGAGT
3386      3396      3406      3416      3426      3436      3446      3456
      R Q I D V I V A K N S G G S A T Y G K I A A A R R
      CGGCAGATCGATGTGATCGTCGCCAAGAACAGCGGTGGCAGCGCCACCTACGGCAAGATTGCCGCGAGCGCGCCG
3461      3471      3481      3491      3501      3511      3521      3531
      L G I E V I M V E R R K P A D V P T V G S C D E A
      CTCGGCATCGAGGTGATCATGGTCGAGCGGCGCAAGCCCGCGGACGTGCCGACGGTCGGCAGTTGCGACGAGGCA
3536      3546      3556      3566      3576      3586      3596      3606
      L N R I A H W L A P A
      CTCAACCGCATCGCTCACTGGCTCGCCCCTGCATGA
3611      3621      3631      3641

```

NOM = COBK

61/151

PREMIER RESIDU = 1
DERNIER RESIDU = 261

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	8	3.07	1176.55	4.19
2	LEU	L	22	8.43	2487.85	8.86
3	ILE	I	16	6.13	1809.34	6.44
4	MET	M	3	1.15	393.12	1.40
5	VAL	V	21	8.05	2080.44	7.41
6	SER	S	12	4.60	1044.38	3.72
7	PRO	P	17	6.51	1649.90	5.88
8	THR	T	13	4.98	1313.62	4.68
9	ALA	A	42	16.09	2983.56	10.63
10	TYR	Y	3	1.15	489.19	1.74
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	2.68	959.41	3.42
13	GLN	Q	6	2.30	768.35	2.74
14	ASN	N	5	1.92	570.21	2.03
15	LYS	K	5	1.92	640.47	2.28
16	ASP	D	17	6.51	1955.46	6.96
17	GLU	E	15	5.75	1935.64	6.89
18	CYS	C	1	0.38	103.01	0.37
19	TRP	W	3	1.15	558.24	1.99
20	ARG	R	26	9.96	4058.63	14.45
21	GLY	G	19	7.28	1083.41	3.86
22	-	-	0	0.00	0.00	0.00

RESIDUS = 261
 Masse moleculaire (monoisotopique) = 28078.7988
 Masse moleculaire (moyenne) = 28096.0195
 INDEX DE POLARITE (%) = 40.61
 POINT ISOELECTRIQUE = 7.54
 DO 260 (1mg/ml) = 0.509 DO 280 (1mg/ml) = 0.721

PROFIL D'HYDROPHILICITE DE LA PROTEINE
COBK DE 1 A 261

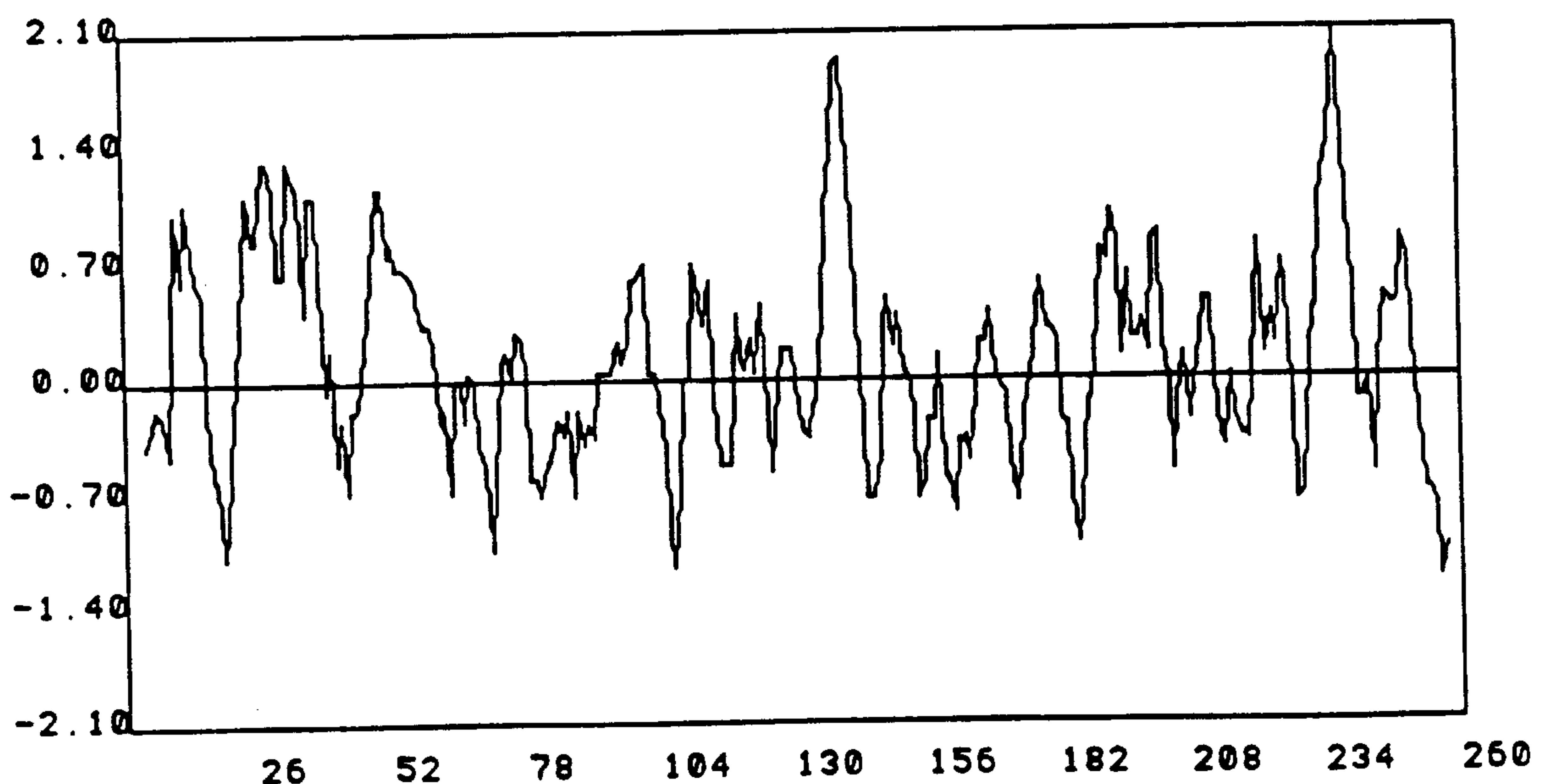


FIGURE 16.12

62/151

MetAlaAspValSerAsnSerGluProAlaIleValSerProTrpLeuThrValIleGlyIleGlyGluAspGly	ATGGCTGACGTGTCGAACAGCGAACCCGCCATAGTCTCCCCCTGGCTGACCGTCATCGGTATCGGTGAGGATGGT					
5862	5872	5882	5892	5902	5912	5922
ValAlaGlyLeuGlyAspGluAlaLysArgLeuIleAlaGluAlaProValValTyrGlyGlyHisArgHisLeu	GTAGCGGGTCTCGGCGACGAGGCCAAGCGGGCTGATCGCCGAAGCGCCGGTCTCTACGGCGGCCATCGTCATCTG					
5937	5947	5957	5967	5977	5987	5997
GluLeuAlaAlaSerLeuIleThrGlyGluAlaHisAsnTrpLeuSerProLeuGluArgSerValValGluIle	GAGCTCGCCGCTCCCTCATCACCGGCGAAGCGCACAAATTGGCTAAGCCCCCTCGAACGCTCGGTCTCGAGATC					
6012	6022	6032	6042	6052	6062	6072
ValAlaArgArgGlySerProValValValLeuAlaSerGlyAspProPhePhePheGlyValGlyValThrLeu	GTCGCGCGTCGCGGCAGCCCGGTGGTGGTGGCTTGCCTCGGGCGACCCGTTCTTCTTCGGCGTCGGCGTGACGCTG					
6087	6097	6107	6117	6127	6137	6147
AlaArgArgIleAlaSerAlaGluIleArgThrLeuProAlaProSerSerIleSerLeuAlaAlaSerArgLeu	GCGCGCCGCATCGCCTCGGCCGAAATACGCACGCTTCCGGCGCCGTCGTCGATCAGTCTTGCCGCCTCGCGCCTC					
6162	6172	6182	6192	6202	6212	6222
GlyTrpAlaLeuGlnAspAlaThrLeuValSerValHisGlyArgProLeuAspLeuValArgProHisLeuHis	GGCTGGGCGCTGCAGGATGCGACGCTCGTCTCCGTACATGGGCGGCCGCTGGATCTGGTGCACCGCATTTGCAT					
6237	6247	6257	6267	6277	6287	6297
ProGlyAlaArgValLeuThrLeuThrSerAspGlyAlaGlyProArgAspLeuAlaGluLeuLeuValSerSer	CCGGGGGCGCGTGTGCTTACGCTCACGTCGGACGGTTCGGGTCCGCGAGACCTTGCCGAGCTTCTGGTTTCAAGC					
6312	6322	6332	6342	6352	6362	6372
GlyPheGlyGlnSerArgLeuThrValLeuGluAlaLeuGlyGlyAlaGlyGluArgValThrThrGlnIleAla	GGCTTCGGTCAGTCGCGACTGACCGTGCTCGAAGCGCTGGGCGGCGCCGGCGAACGGGTGACGACGCAGATCGCC					
6387	6397	6407	6417	6427	6437	6447
AlaArgPheMetLeuGlyLeuValHisProLeuAsnValCysAlaIleGluValAlaAlaAspGluGlyAlaArg	GCGCGCTTCATGCTCGGCCTCGTGCATCCTTTGAACGTCTGCGCCATTGAGGTGGCGGCCGACGAGGGCGCGCGC					
6462	6472	6482	6492	6502	6512	6522
IleLeuProLeuAlaAlaGlyArgAspAspAlaLeuPheGluHisAspGlyGlnIleThrLysArgGluValArg	ATCCTGCCGCTTGCCGCCGGCCGCGACGATGCGCTGTTTCGAACATGACGGGCAGATACCAAGCGCGAGGTGCGG					
6537	6547	6557	6567	6577	6587	6597
AlaLeuThrLeuSerAlaLeuAlaProArgLysGlyGluLeuLeuTrpAspIleGlyGlyGlySerGlySerIle	GCGCTGACGCTGTCGGCACTCGCACCGCGCAAGGGCGAACTGCTATGGGACATCGGCGGCGGCTCCGGCTCGATC					
6612	6622	6632	6642	6652	6662	6672
GlyIleGluTrpMetLeuAlaAspProThrMetGlnAlaIleThrIleGluValGluProGluArgAlaAlaArg	GGCATCGAATGGATGCTCGCCGATCCGACCATGCAGGCGATCACCATCGAGGTTGAGCCGGAGCGGGCAGCGCGC					
6687	6697	6707	6717	6727	6737	6747
IleGlyArgAsnAlaThrMetPheGlyValProGlyLeuThrValValGluGlyGluAlaProAlaAlaLeuAla	ATCGGCCGCAACGCGACGATGTTTCGGCGTGCCCGGGCTGACGGTTGTTCGAAGGCGAGGCGCCGGCGGCGCTTGCC					
6762	6772	6782	6792	6802	6812	6822
GlyLeuProGlnProAspAlaIlePheIleGlyGlyGlyGlySerGluAspGlyValMetGluAlaAlaIleGlu	GGCCTGCCACAACCGGACGCGATCTTCATCGGCGGCGGCGGCGAGCGAAGACGGCGTCATGGAAGCAGCGATCGAG					
6837	6847	6857	6867	6877	6887	6897
AlaLeuLysSerGlyGlyArgLeuValAlaAsnAlaValThrThrAspMetGluAlaValLeuLeuAspHisHis	GCGCTCAAGTCAGGCGGACGGCTGGTTGCCAACGCGGTGACGACGGACATGGAAGCGGTGCTGCTCGATCATCAC					
6912	6922	6932	6942	6952	6962	6972
AlaArgLeuGlyGlySerLeuIleArgIleAspIleAlaArgAlaGlyProIleGlyGlyMetThrGlyTrpLys	GCGCGGCTCGGCGGTTCGCTGATCCGCATCGATATCGCGCGTGCAGGACCCATCGGCGGCATGACCGGCTGGAAG					
6987	6997	7007	7017	7027	7037	7047
ProAlaMetProValThrGlnTrpSerTrpThrLysGly***	CCGGCCATGCCGGTCACCCAATGGTCGTGGACGAAGGGCTAA					
7062	7072	7082	7092	7102		

FIGURE 16.13

63/151

PROTEINE COBL

PREMIER RESIDU = 1
DERNIER RESIDU = 413

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	8	1.94	1176.56	2.74
2	LEU	L	47	11.38	5314.76	12.39
3	ILE	I	26	6.30	2940.08	6.85
4	MET	M	9	2.18	1179.36	2.75
5	VAL	V	34	8.23	3368.38	7.85
6	SER	S	25	6.05	2175.75	5.07
7	PRO	P	24	5.81	2329.20	5.43
8	THR	T	21	5.08	2122.05	4.95
9	ALA	A	56	13.56	3978.24	9.27
10	TYR	Y	1	0.24	163.06	0.38
11	*	*	0	0.00	0.00	0.00
12	HIS	H	10	2.42	1370.60	3.19
13	GLN	Q	7	1.69	896.42	2.09
14	ASN	N	5	1.21	570.20	1.33
15	LYS	K	6	1.45	768.54	1.79
16	ASP	D	19	4.60	2185.57	5.09
17	GLU	E	27	6.54	3484.08	8.12
18	CYS	C	1	0.24	103.01	0.24
19	TRP	W	8	1.94	1488.64	3.47
20	ARG	R	28	6.78	4370.80	10.19
21	GLY	G	51	12.35	2908.02	6.78
22	-	-	0	0.00	0.00	0.00

RESIDUS = 413
 POIDS MOLECULAIRE = 42911.
 INDEX DE POLARITE (%) = 36.
 POINT ISOELECTRIQUE = 5.70
 DO 260 (1mg/ml) = 0.754 DO 280 (1mg/ml) = 1.064

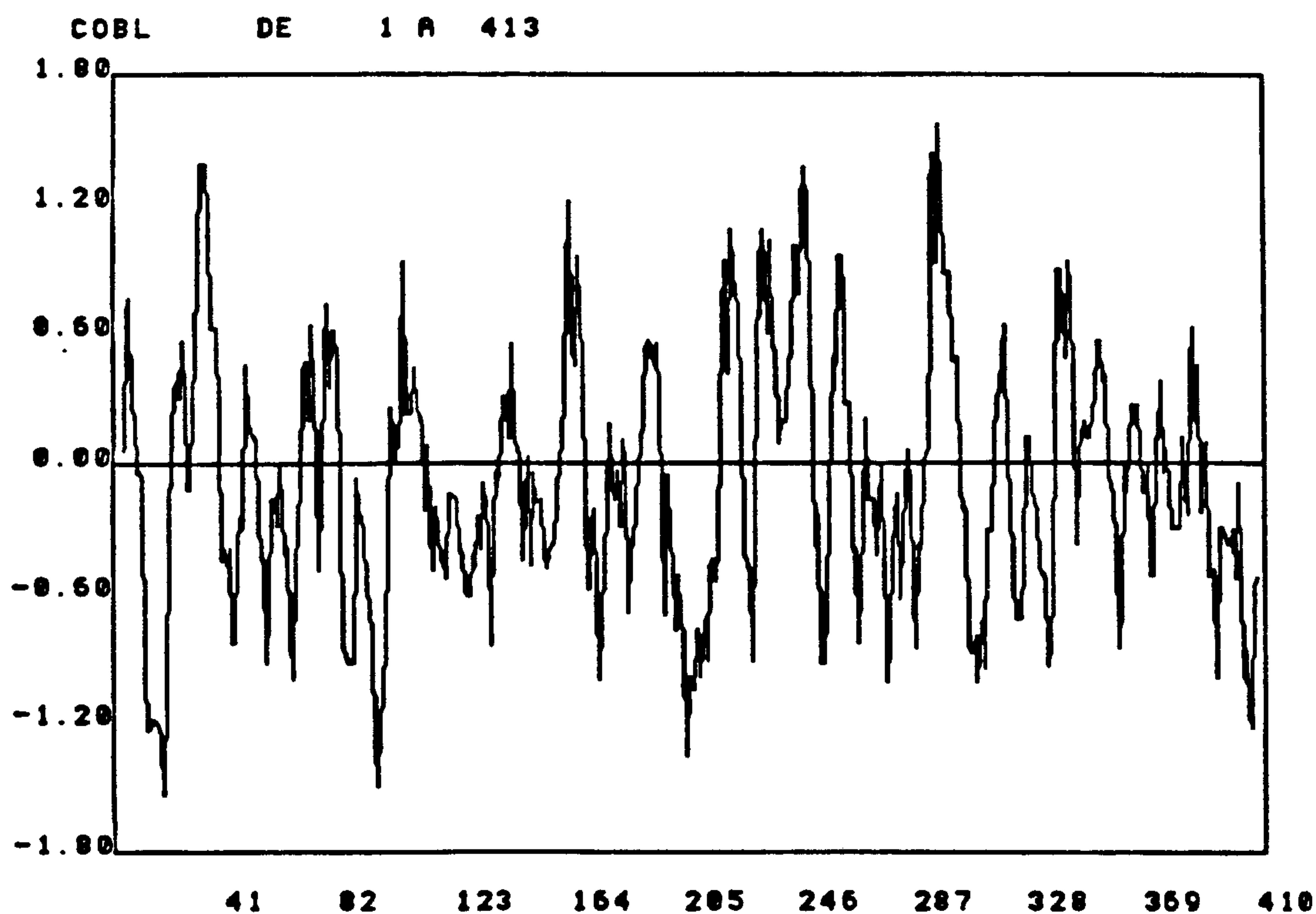


FIGURE 16.14

64 / 151

GENE cobM ET PROTEINE COBM

SEQUENCE DU FRAGMENT DE 8753 BP, DE 7172 A 7930

```

MetThrValHisPheIleGlyAlaGlyProGlyAlaAlaAspLeuIleThrValArgGlyArgAspLeuIleGly
ATGACGGTACATTTTCATCGGCGCCGGCCCGGGAGCCGACACCTGATCACGGTGCGTGGTCGCGACCTGATCGGG
7172      7182      7192      7202      7212      7222      7232
ArgCysProValCysLeuTyrAlaGlySerIleValSerProGluLeuLeuArgTyrCysProProGlyAlaArg
CGCTGCCCCGGTCTGCCTTTACGCCGGCTCGATCGTCTCGCCGGAGCTGCTGCGATATTGCCCGCCGGGCGCCCGC
7247      7257      7267      7277      7287      7297      7307
IleValAspThrAlaProMetSerLeuAspGluIleGluAlaGluTyrValLysAlaGluAlaGluGlyLeuAsp
ATTGTCGATACGGCGCCGATGTCCCTCGACGAGATCGAGGCGGAGTATGTGAAGGCCGAAGCCGAAGGGCTCGAC
7322      7332      7342      7352      7362      7372      7382
ValAlaArgLeuHisSerGlyAspLeuSerValTrpSerAlaValAlaGluGlnIleArgArgLeuGluLysHis
GTGGCGCGGCTTCATTCGGGCGACCTTTCGGTCTGGAGTGCTGTGGCCGAACAGATCCGCCGGCTCGAGAAGCAT
7397      7407      7417      7427      7437      7447      7457
GlyIleAlaTyrThrMetThrProGlyValProSerPheAlaAlaAlaAlaSerAlaLeuGlyArgGluLeuThr
GGCATCGCCTATACGATGACGCCGGGCGTTCCTTCCTTTGCGGCGGCGGCTTCAGCGCTCGGTTCGCGAATTGACC
7472      7482      7492      7502      7512      7522      7532
IleProAlaValAlaGlnSerLeuValLeuThrArgValSerGlyArgAlaSerProMetProAsnSerGluThr
ATTCCGGCCCGTGGCCCAGAGCCTGGTGCTGACCCGCGTTCGGGCCGCGCCTCGCCGATGCCGAACCTCAGAAACG
7547      7557      7567      7577      7587      7597      7607
LeuSerAlaPheGlyAlaThrGlySerThrLeuAlaIleHisLeuAlaIleHisAlaLeuGlnGlnValValGlu
CTTTCCGCTTTCGGCGCTACGGGATCGACGCTGGCAATCCACCTTGCGATCCATGCGCTTCAGCAGGTGGTCGAG
7622      7632      7642      7652      7662      7672      7682
GluLeuThrProLeuTyrGlyAlaAspCysProValAlaIleValValLysAlaSerTrpProAspGluArgVal
GAACTGACGCCGCTCTACGGTGCCGACTGCCCGGTCGCCATCGTCGTCAAGGCCTCCTGGCCGGACGAACGCGTG
7697      7707      7717      7727      7737      7747      7757
ValArgGlyThrLeuGlyAspIleAlaAlaLysValAlaGluGluProIleGluArgThrAlaLeuIlePheVal
GTGCGCGGCACGCTCGGTGACATCGCCGCCAAGGTGGCGGAAGAGCCGATCGAGCGCACGGCGCTGATCTTCGTC
7772      7782      7792      7802      7812      7822      7832
GlyProGlyLeuGluAlaSerAspPheArgGluSerSerLeuTyrAspProAlaTyrGlnArgArgPheArgGly
GGTCCGGGGCTCGAAGCCTCCGATTTCCGTGAAAGCTCGCTCTACGATCCCGCCTATCAGCGGCGCTTCAGAGGG
7847      7857      7867      7877      7887      7897      7907
ArgGlyGlu
CGCGGCGAA
7922      7932      7942      7952      7962      7972      7982

```

FIGURE 16.15

PROTEINE COBM

65/451

PREMIER RESIDU = 1
DERNIER RESIDU = 253

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	6	2.37	882.42	3.29
2	LEU	L	24	9.49	2713.92	10.11
3	ILE	I	15	5.93	1696.20	6.32
4	MET	M	4	1.58	524.16	1.95
5	VAL	V	22	8.70	2179.54	8.12
6	SER	S	18	7.11	1566.54	5.84
7	PRO	P	17	6.72	1649.85	6.15
8	THR	T	13	5.14	1313.65	4.89
9	ALA	A	35	13.83	2486.40	9.26
10	TYR	Y	7	2.77	1141.42	4.25
11	*	*	0	0.00	0.00	0.00
12	HIS	H	5	1.98	685.30	2.55
13	GLN	Q	5	1.98	640.30	2.39
14	ASN	N	1	0.40	114.04	0.42
15	LYS	K	4	1.58	512.36	1.91
16	ASP	D	11	4.35	1265.33	4.71
17	GLU	E	19	7.51	2451.76	9.13
18	CYS	C	4	1.58	412.04	1.53
19	TRP	W	2	0.79	372.16	1.39
20	ARG	R	19	7.51	2965.90	11.05
21	GLY	G	22	8.70	1254.44	4.67
22	-	-	0	0.00	0.00	0.00

RESIDUS = 253
 POIDS MOLECULAIRE = 26846.
 INDEX DE POLARITE (%) = 38.
 POINT ISOELECTRIQUE = 5.58
 DO 260 (1mg/ml) = 0.461 DO 280 (1mg/ml) = 0.724

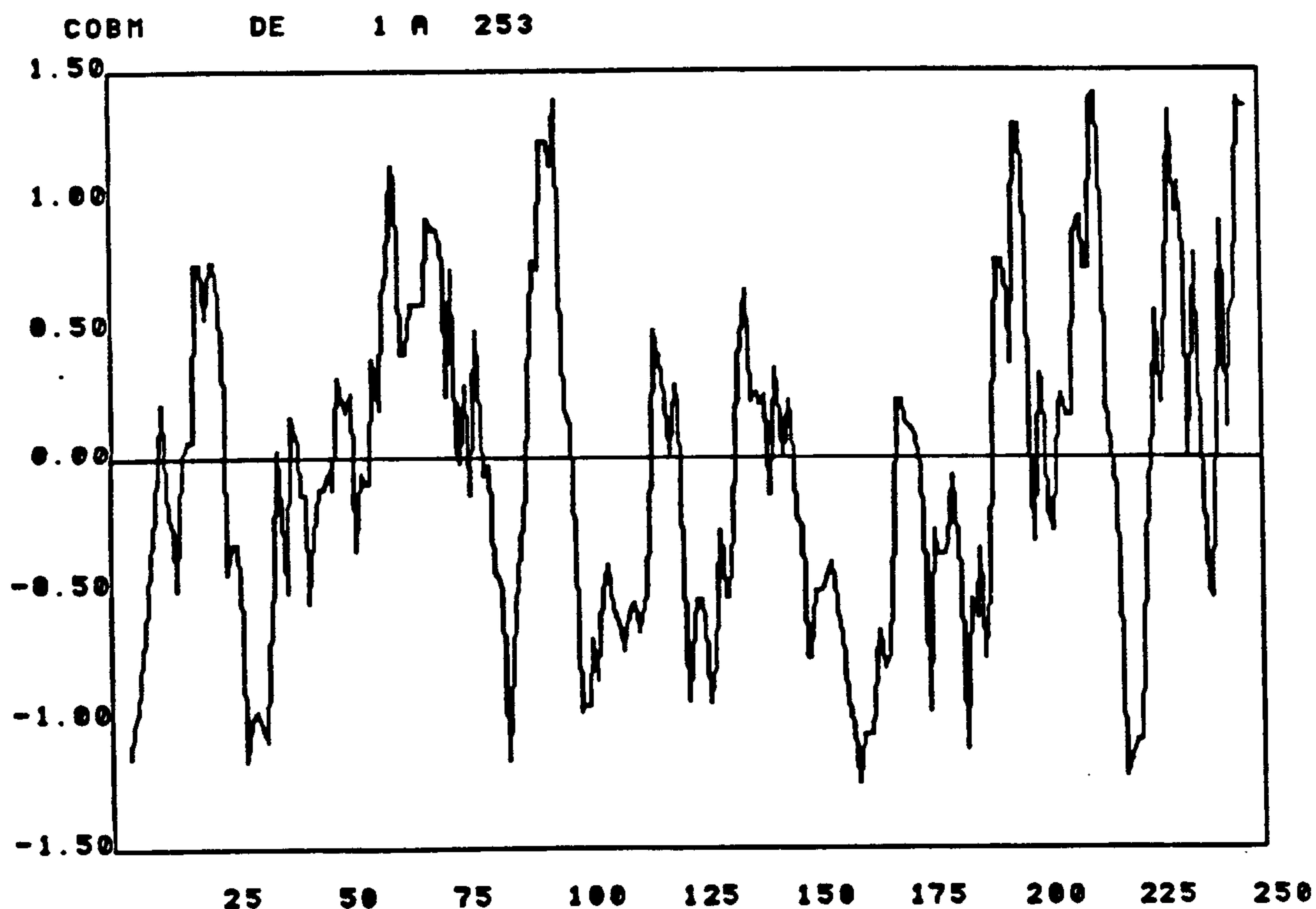


FIGURE 16.16

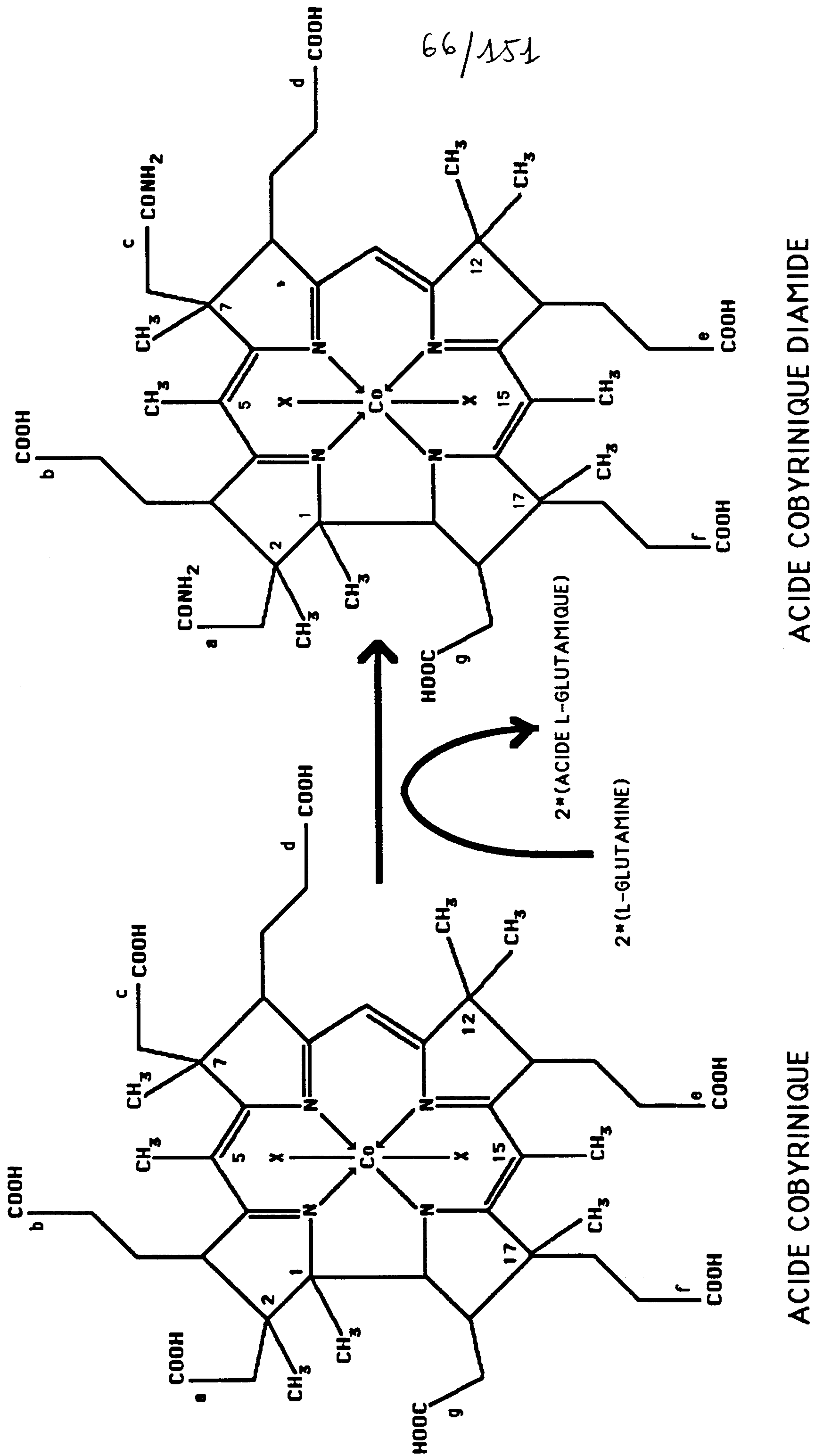


FIGURE 17

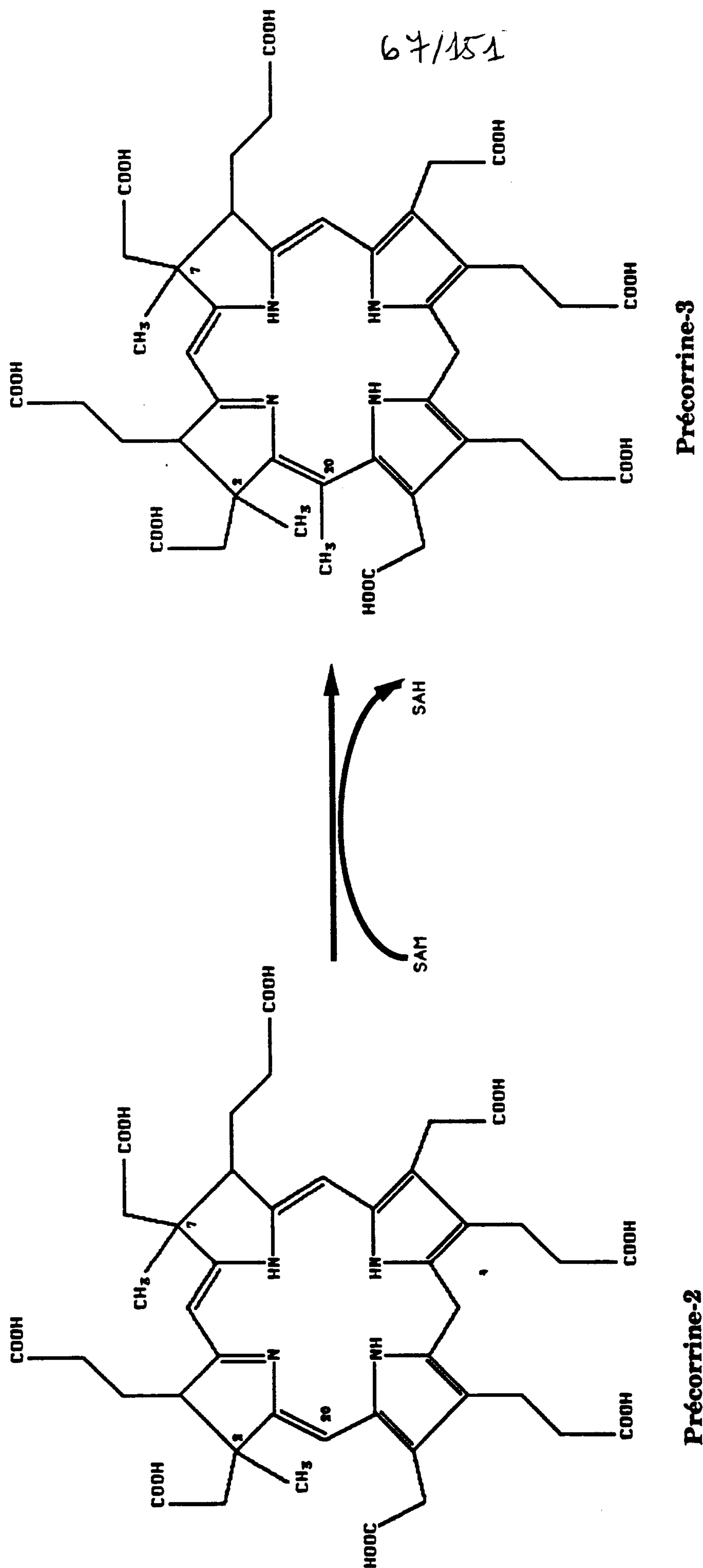


FIGURE 18

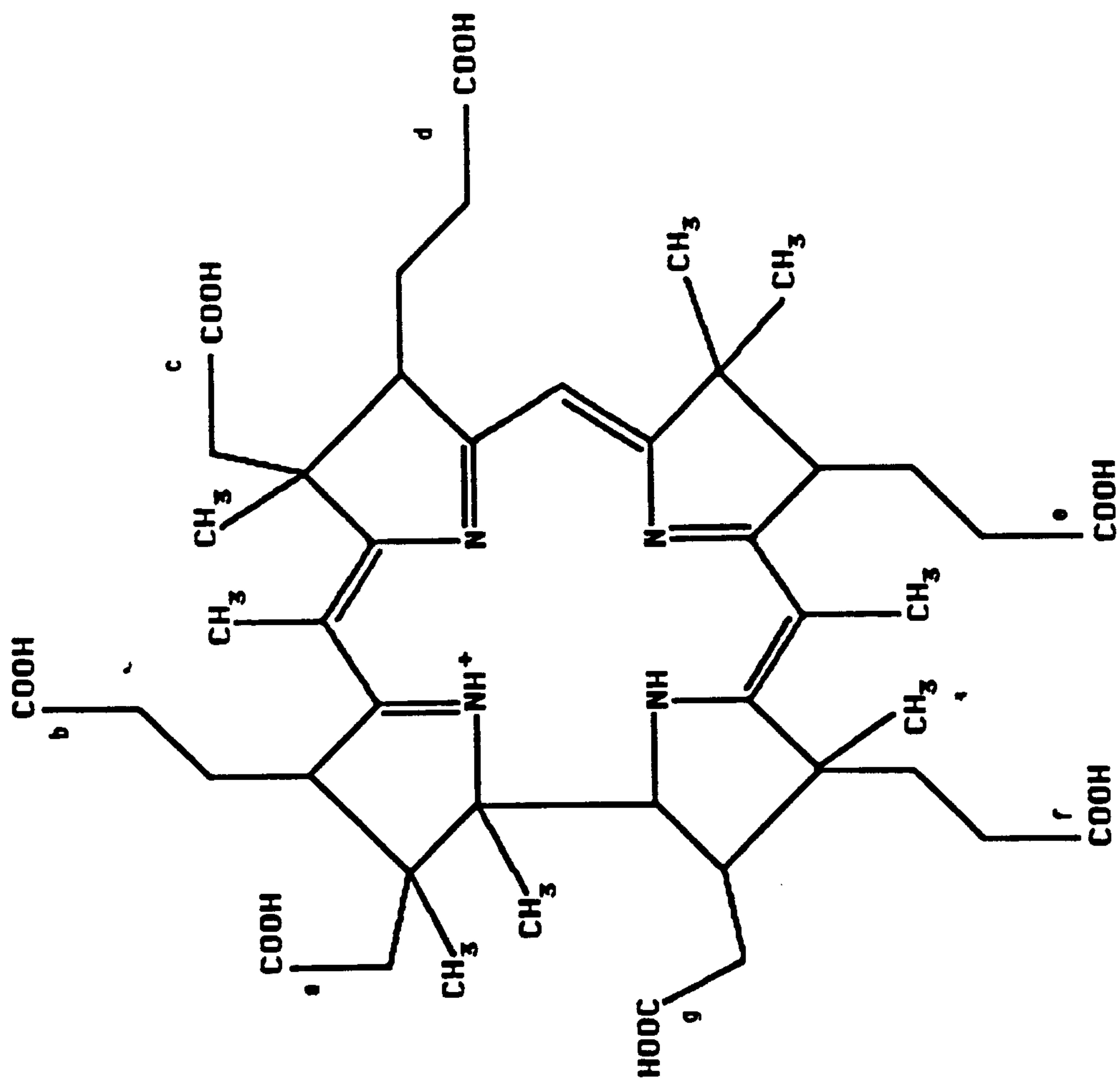
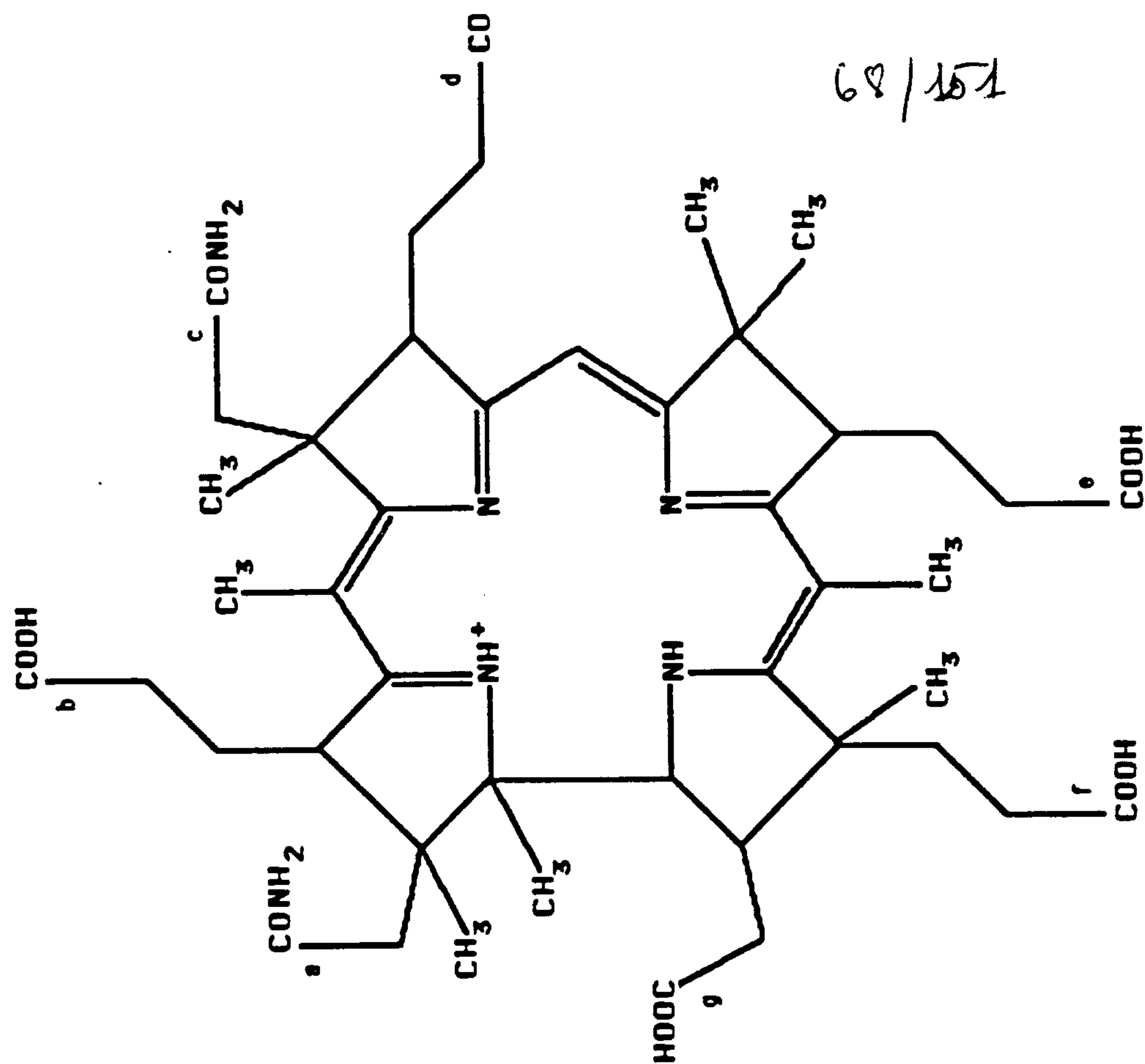


FIGURE 19

69/151

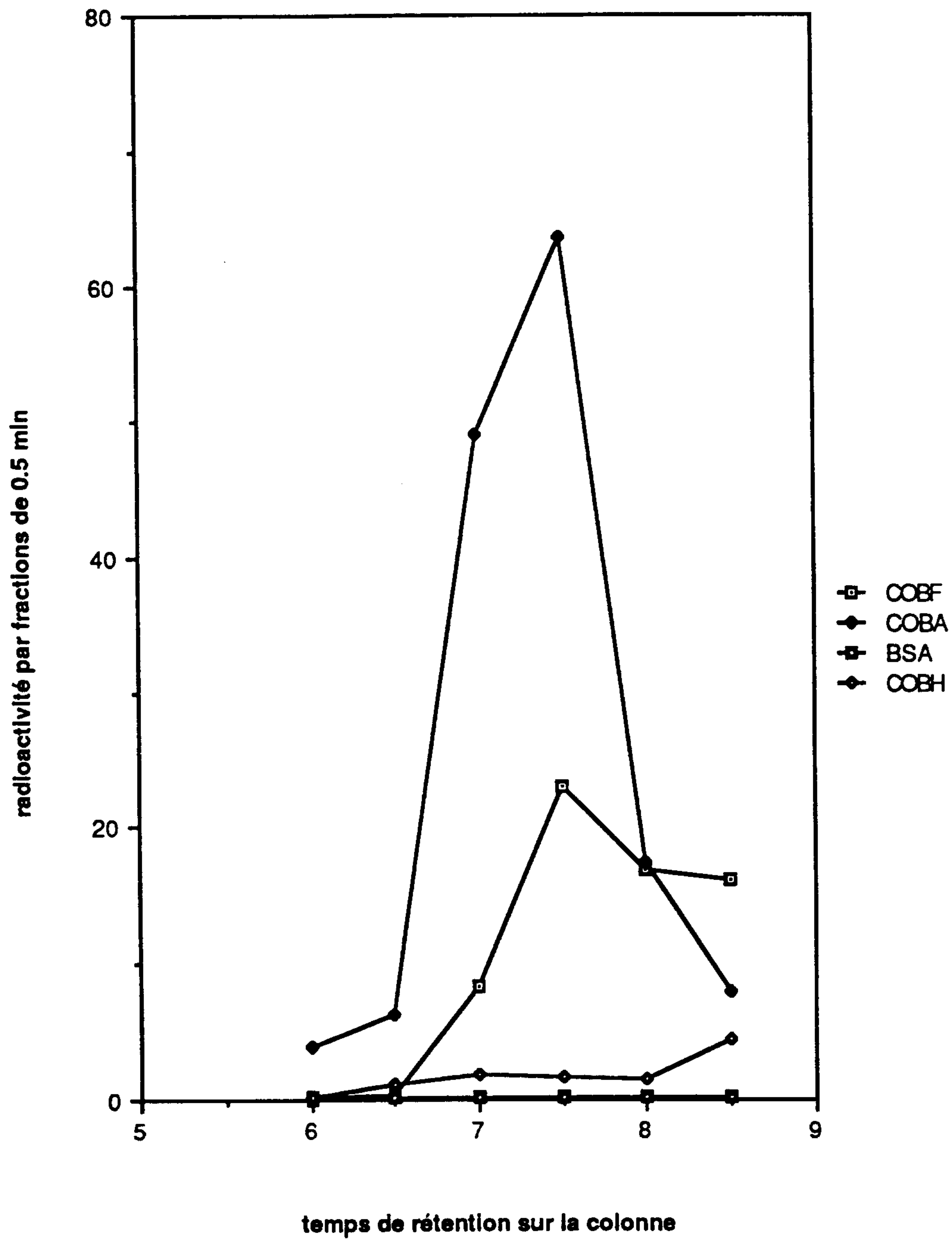


FIGURE 20

71 / 151

```

SEQUENCE COBA
SEQUENCE CYSG
1
10 20 30 40 50 60 70
DDLFAGLPALEKGSVWLVGAGPGDPGLLTLHAANALRQADVIVHDALVNEDCLKLARPGAVLEFAGKRGG
-- = -- = -- = -- = -- = -- = -- = -- = -- = -- = -- = -- =
EQLINE-PLDHRGEVVLVGAGPGDAGLLTLKGLQOIQOADVVDRLVSDDIMNLVRRDADR VFGKRAG
210 220 230 240 250 260 270
80 90 100 110 120 130 140
2
KPSPKQORDISLRLVELARAGNRVRLRLKGGDPFVFGRGGEELTIVEHQVPFRIVPGITAGIGGLAYAGIP
= = = = =
YHCVPOEEINQILLREAOKGKRVVRLKGGDPFIFGRGGEELTICNAGIPFSVVPGITAASGCSAYSGIP
280 290 300 310 320 330 340
150 160 170 180 190 200 210
3
VTHREVNHAVTFLTGHDSGLVPDRINWQGIASGSPVIVMYMAMKHIGAITANLIAGGRSPDEPVAFVCN
-- = -- = -- = -- = -- = -- = -- = -- = -- = -- = -- = -- =
LTHRDYAQSVRLITGHLKTG-G-E-LDWENLAAEKQTLVFYMG LNQAATIQOKLIEHGMPGEMPVAIVEN
350 360 370 380 390 400
220 230 240 250
AATPQQAVLETTLARA EADVAAAGLEPPAIVVVGEVVRLRAALDWIG
-- = -- = -- = -- = -- = -- = -- = -- = -- = -- = -- = -- =
GTA VTQRVIDGTLTQL-GELAAQQ-MNSPSLIIIGRVVGLRDKLNWFS
420 430 440 450

```

SEQUENCE COBA DE 3 A 259
SEQUENCE CYSG DE 204 A 460

HOMOLOGIE STRICTE EN ACIDE AMINES: 41.6%

FIGURE 22

72/151

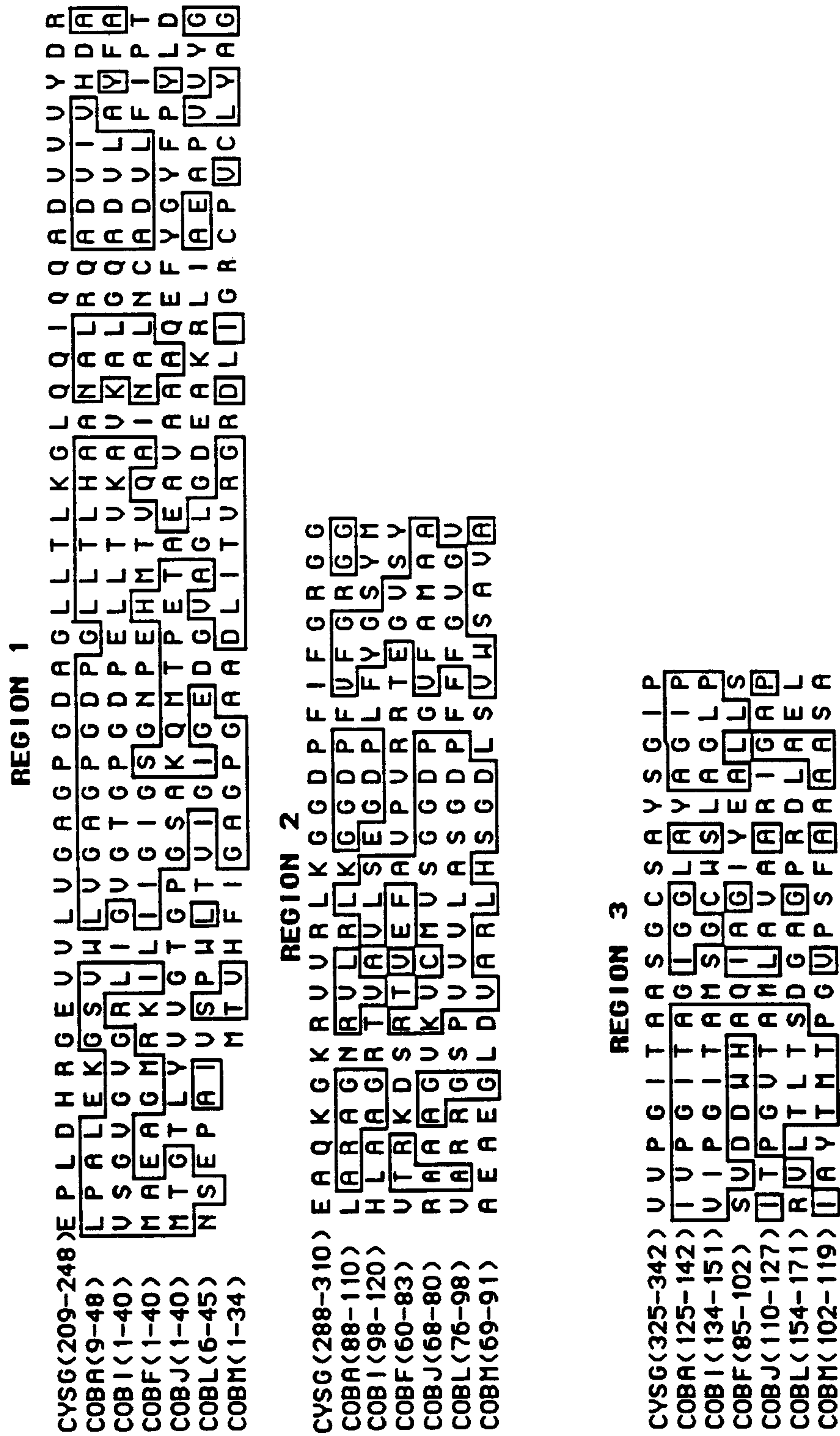


FIGURE 23

74 / 151

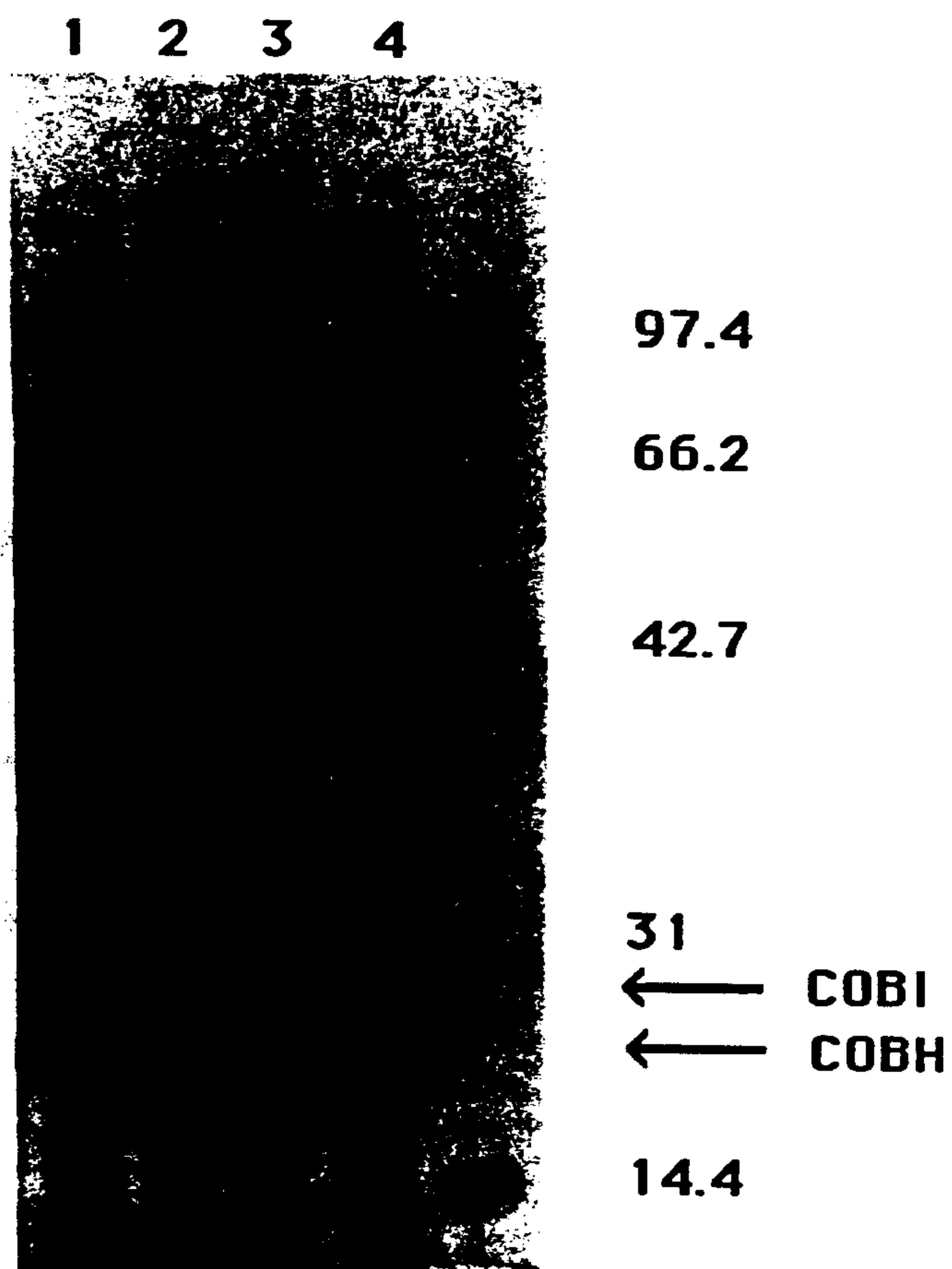


FIGURE 25

FEUILLE DE REMPLACEMENT

75 / 151

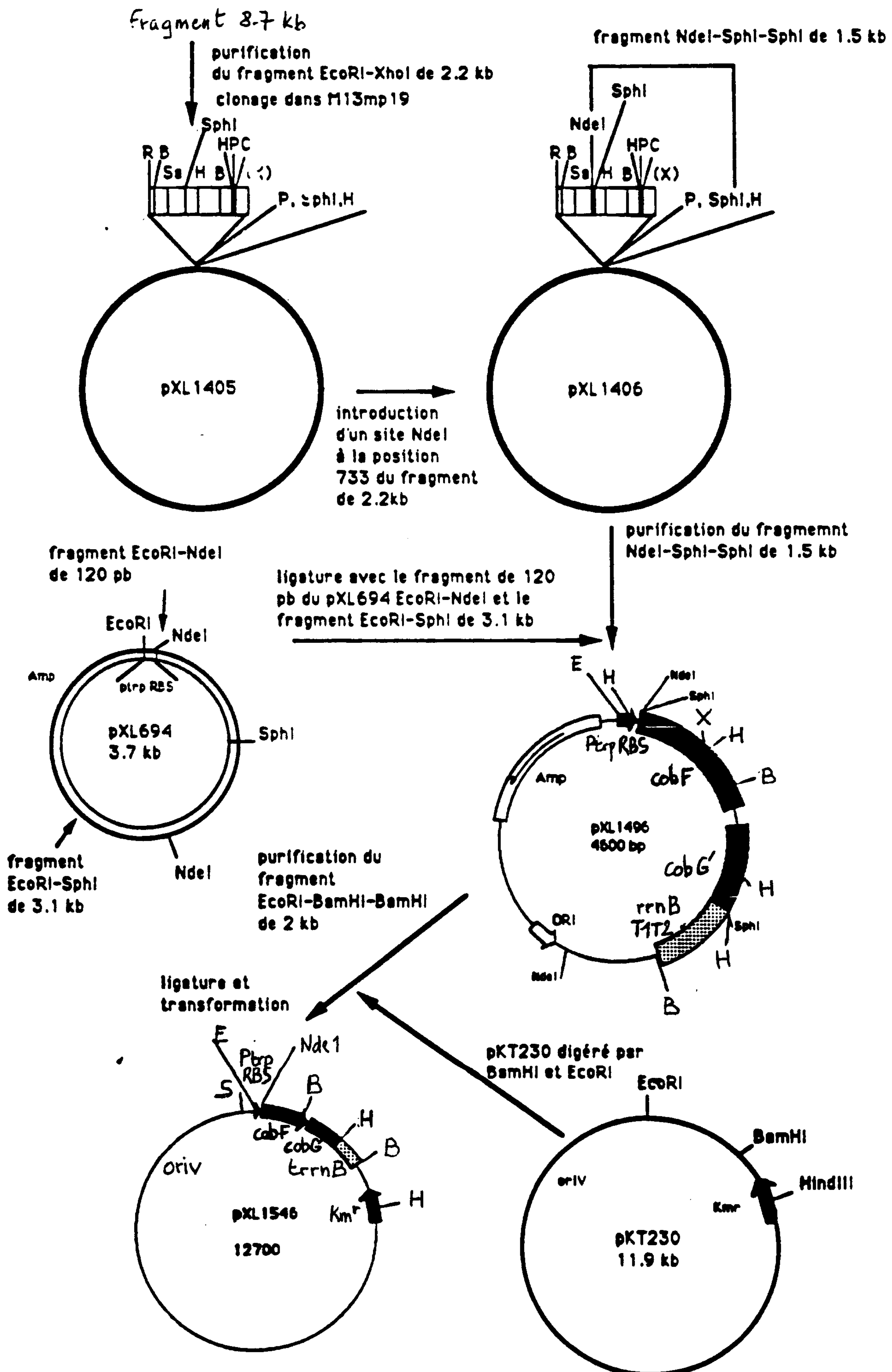


FIGURE 26

FEUILLE DE REMPLACEMENT

PL. 76/151

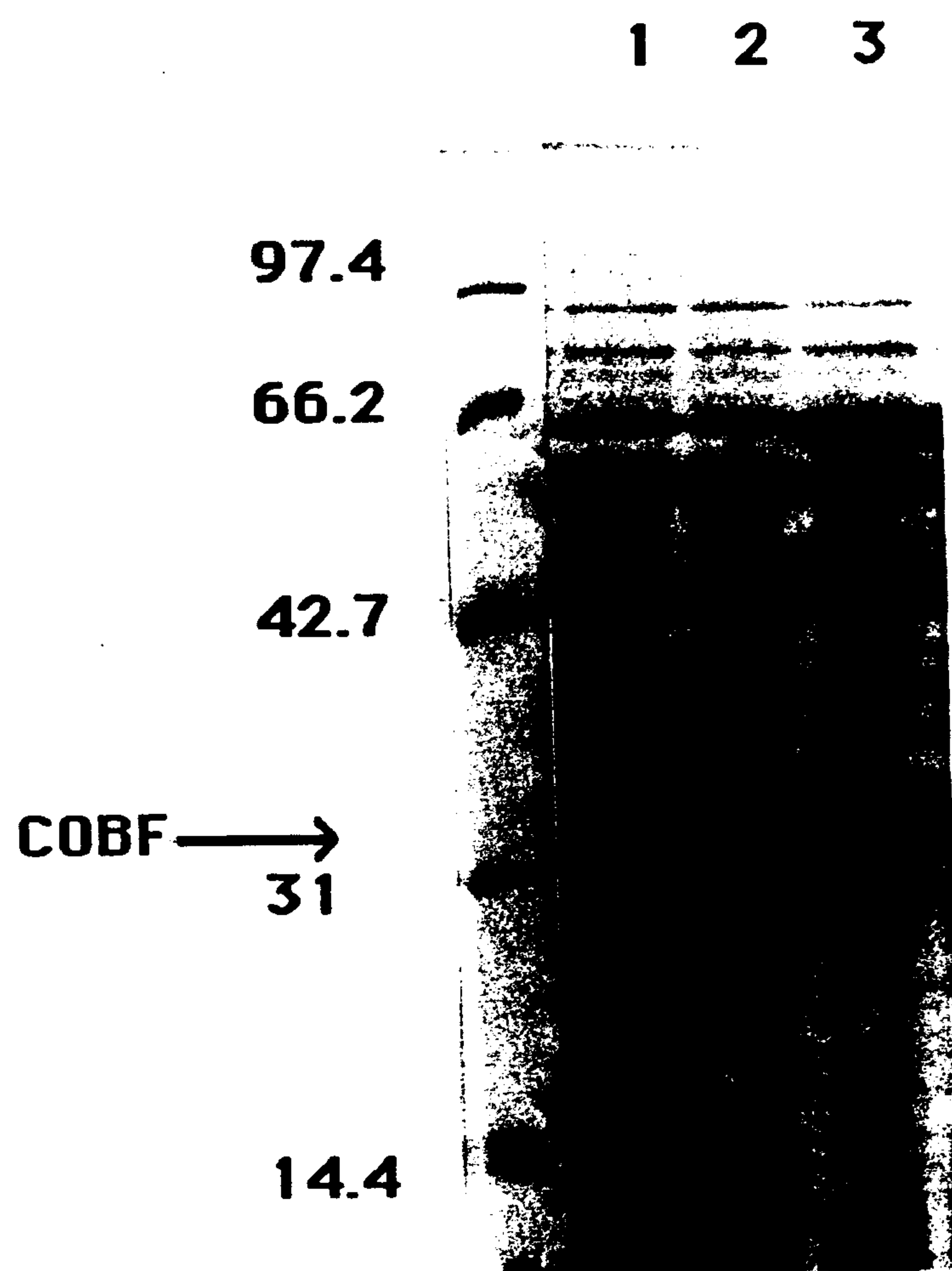


FIGURE 27

PL. 77 / 151

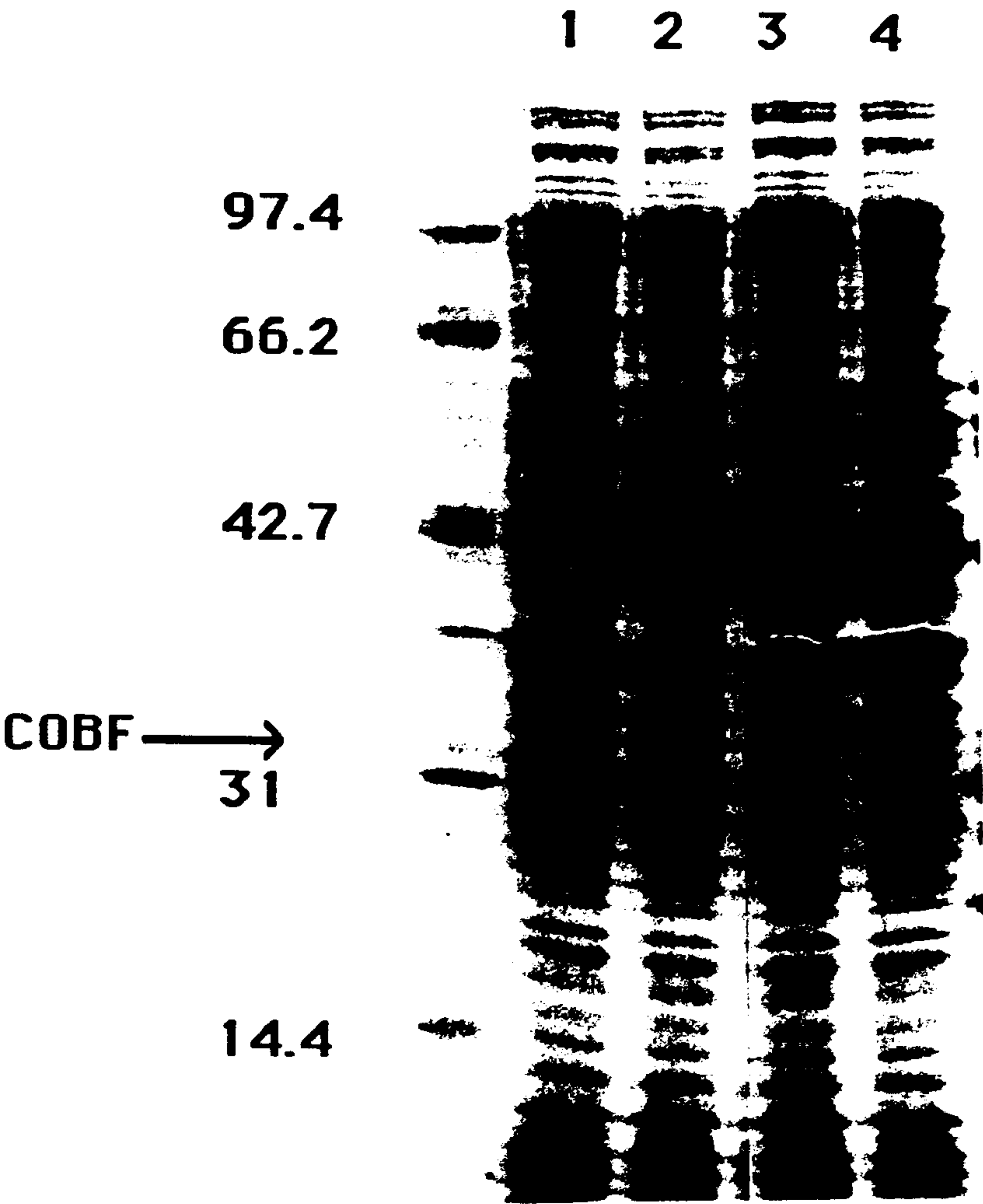


FIGURE 28

78 / 151

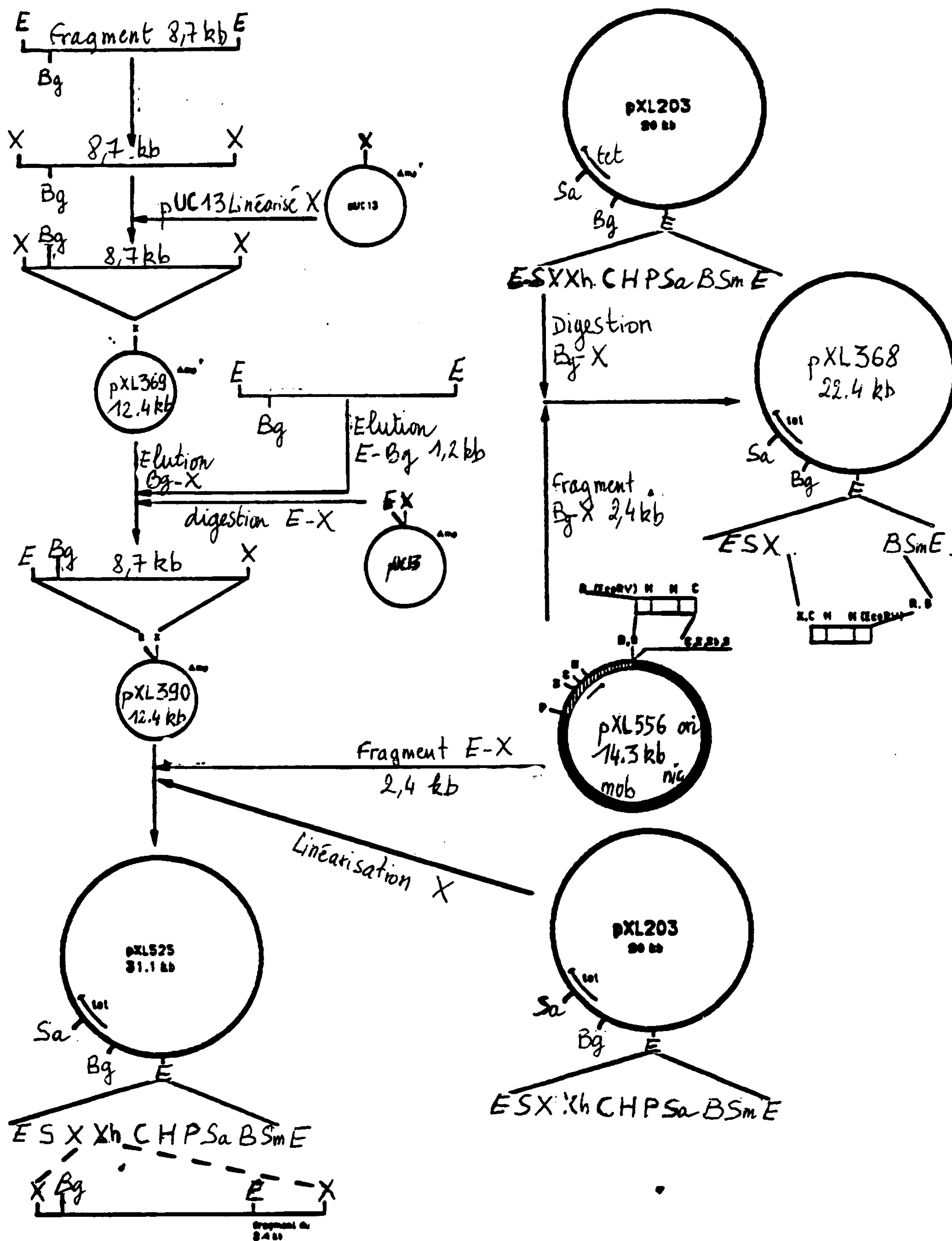


FIGURE 29

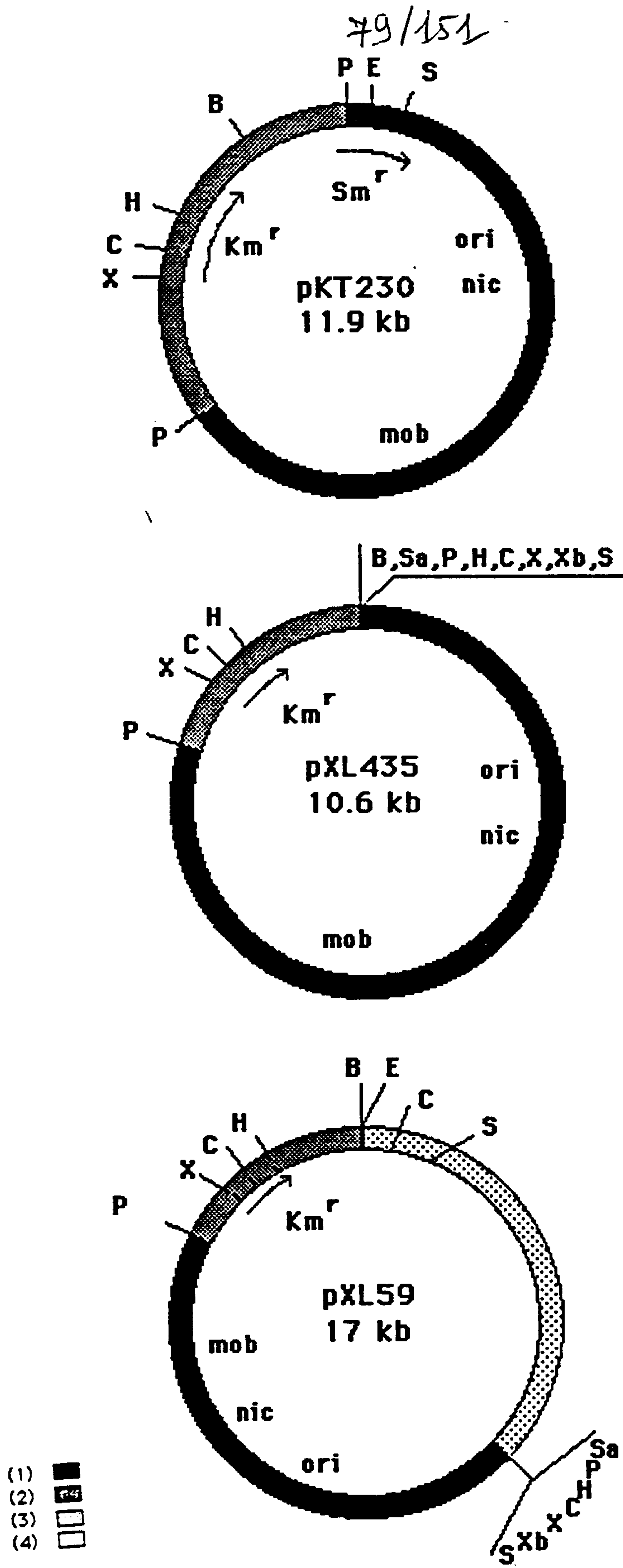


FIGURE 30

80/151

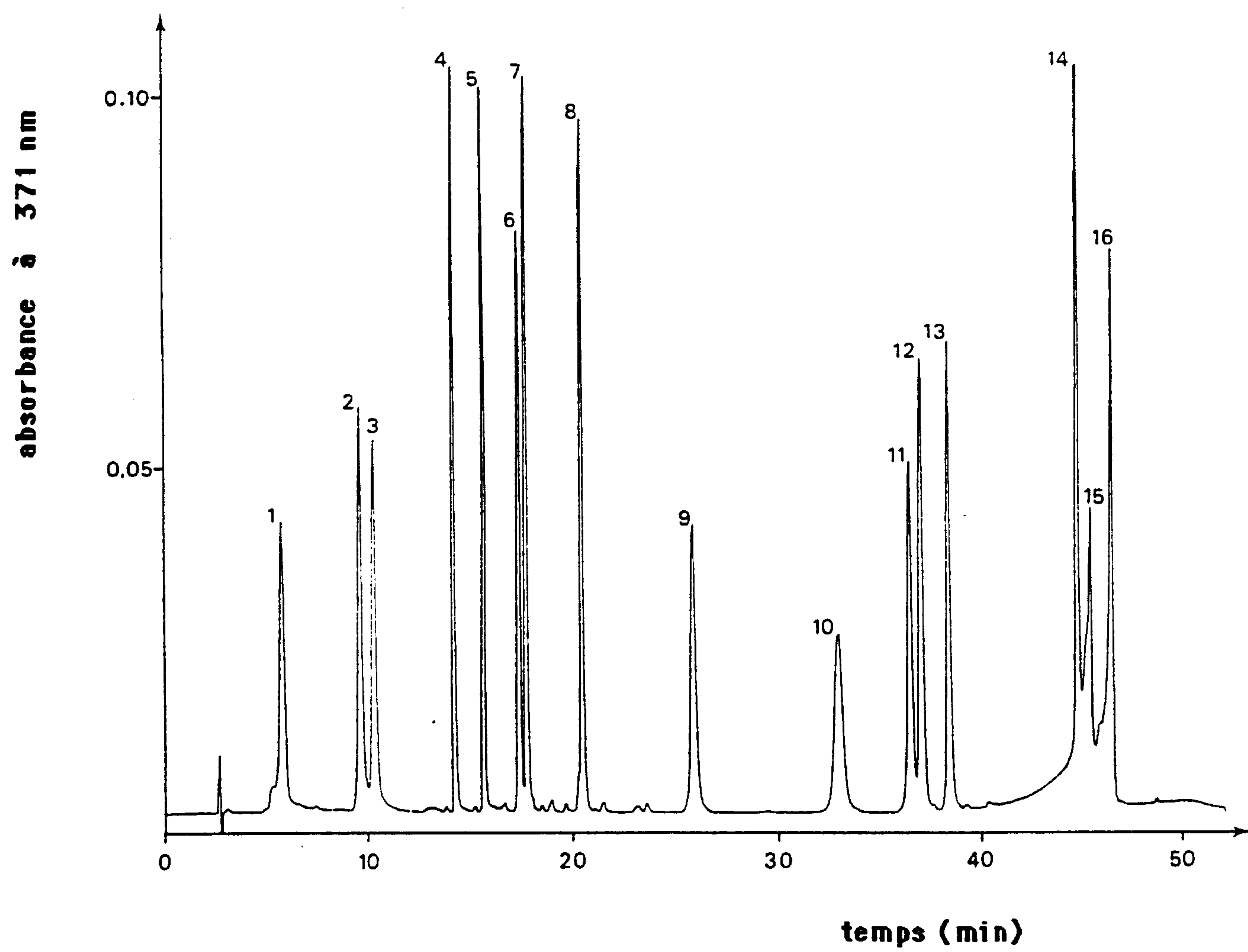


FIGURE 31

PL. 81/151

10	20	30	40	50	60
GTCGACGAGT	ATGGTCAGGT	TCAGGGTCTG	GTGACGCTGG	AGGACATTCT	GGAGGAGATC
CAGCTGCTCA	TACCAGTCCA	AGTCCCAGAC	CACTGCGACC	TCCTGTAAGA	CCTCCTCTAG
70	80	90	100	110	120
GTCGGCGATA	TCGCCGATGA	GCACGACCTC	GACATTCAGG	GCGTGCGCCA	GGAAGCCGAT
CAGCCGCTAT	AGCGGCTACT	CGTGCTGGAG	CTGTAAGTCC	CGCACGCGGT	CCTTCGGCTA
130	140	150	160	170	180
GGCTCGATCG	TCGTCTGATGG	CTCGGTGCCG	ATCCGCGATC	TCAACCGCGC	GCTCGACTGG
CCGAGCTAGC	AGCAGCTACC	GAGCCACGGC	TAGGCGCTAG	AGTTGGCGCG	CGAGCTGACC
190	200	210	220	230	240
TCGCTGCCGG	ACGAGGAGGC	GACGACGGTG	GCCGGTCTGG	TCATCCACGA	GTCCAAGAGC
AGCGACGGCC	TGCTCCTCCG	CTGCTGCCAC	CGGCCAGACC	AGTAGGTGCT	CAGGTTCTCG
250	260	270	280	290	300
ATTCCGGAGG	AGCGCCAGGC	CTTCACCTTC	CACGGCAAAC	GCTTCATCGT	GATGAAGCGG
TAAGGCCTCC	TCGCGGTCCG	GAAGTGGAAG	GTGCCGTTTG	CGAAGTAGCA	CTACTTCGCC
310	320	330	340	350	360
GTGAAGAACC	GCATTACCAA	GCTGCGCATC	CGTCCGGCGG	AAGAGGGTGC	TCCGCCGGCG
CACCTCTTGG	CGTAATGGTT	CGACGCGTAG	GCAGGCCGCC	TTCTCCCACG	AGGCGGCCCG
370	380	390	400	410	420
TGATGGCCGC	GATTGCCTCT	ACCAGCGGGT	CGGCTCGCCG	GGGGCTGCCG	GCTCGACGGC
ACTACCGGCG	CTAACGGAGA	TGGTCGCCCA	GCCGAGCGGC	CCCCGACGGC	CGAGCTGCCG
430	440	450	460	470	480
GAGCGCATGC	AGGCCGGCGT	CGAGTTCTGG	CTTCAGGAGA	TCATTGATGG	CGCGGTGGCG
CTCGCGTACG	TCCGGCCGCA	GCTCAAGACC	GAAGTCCTCT	AGTAACTACC	GCGCCACCGC
490	500	510	520	530	540
GGCGACACGG	CTCATGCCGG	CAAAGGCGCT	AGAAACGATA	CGCACCCGCA	TGTGGGACTC
CCGCTGTGCC	GAGTACGGCC	GTTTCCGCGA	TCTTTGCTAT	GCGTGGGCGT	ACACCCTGAG
550	560	570	580	590	600
GCCGGTACCA	TCGAAGCCCG	GCTGATGGCC	GGTATGCTGA	TGGCTCTCGT	TGATGACCTC
CGGCCATGGT	AGCTTCGGGC	CGACTACCGG	CCATACGACT	ACCGAGAGCA	ACTACTGGAG
610	620	630	640	650	660
GAGCCGTTCA	GGGTGGAAGG	CCTCGATCAG	CTTCTTTTCG	ATGGTCTCGG	TGAGCGACAT
CTCGGCAAGT	CCCACCTTCC	GGAGCTAGTC	GAAGAAAAGC	TACCAGAGCC	ACTCGCTGTA
670	680	690	700	710	720
TCTTCCGTCC	CATTTTGCTG	TTTGCTTGGC	GCCCCCTCGC	AGTTAAGAAC	CCGGTAATCG
AGAAGGCAGG	GTAAAACGAC	AAACGAACCG	CGGGGGAGCG	TCAATTCTTG	GGCCATTAGC
730	740	750	760	770	780
CTGGCACGGC	GGCGCAAAAT	GCCCGCACAA	AGCCAGCAAC	ATTCCGCTTT	GTCAATTCTT
GACCGTGCCG	CCGCGTTTTA	CGGGCGTGTT	TCGGTCGTTG	TAAGGCGAAA	CAGTTAAGAA
790	800	810	820	830	840
GTTGTGACTC	CCGCCAAACC	CCATAATGAG	CGCCATGAGA	CTCGATTCAA	AATACTTCGA
CAACACTGAG	GGCGGTTTGG	GGTATTACTC	GCGGTACTCT	GAGCTAAGTT	TTATGAAGCT
850	860	870	880	890	900
TCGCATTCTGA	ACCCGGGCGCA	AGGTCTGAGCC	GCAGCAGAGC	CGGCGGCTCC	CGTCTGTCTAG
AGCGTAAGCT	TGGGCCGCGT	TCCAGCTCGG	CGTCGTCTCG	GCCGCCGAGG	GCAGACAGTC
910	920	930	940	950	960
TGGGACGGCT	GCGATCAGAA	GGGTGTGCAC	CGGGCGCCCG	TCGGTCGCAA	CGCCGAGGGG
ACCCTGCCGA	CGCTAGTCTT	CCCACACGTG	GCCCGCGGGC	AGCCAGCGTT	GCGGCTCCCC

FIGURE 32

FEUILLE DE REMPLACEMENT

PL. 82 / 151

970	980	990	1000	1010	1020
CAGTACTTCA	TGTTCTGCTT	CGAGCACGTG	AAGGAATACA	ACAAGGGCTA	CAACTTCTTC
GTCATGAAGT	ACAAGACGAA	GCTCGTGCAC	TTCCTTATGT	TGTTCCCGAT	GTTGAAGAAG
1030	1040	1050	1060	1070	1080
TCCGGCCTCT	CCGACAGCGA	GGTCGCCCCG	TACCAGAAGG	AAGCGATCAC	CGGTCATCGG
AGGCCGGAGA	GGCTGTGCTT	CCAGCGGGCG	ATGGTCTTCC	TTCGCTAGTG	GCCAGTAGCC
1090	1100	1110	1120	1130	1140
CCCACCTGGA	CCGTCGGCGT	CAACAAGAAC	GCCAAGAACG	GCCCCACCCA	GTCGCAGACG
GGGTGGACCT	GGCAGCCGCA	GTTGTTCTTG	CGGTTCTTGC	CGGGCTGGGT	CAGCGTCTGC
1150	1160	1170	1180	1190	1200
CGGTCTGGCT	CTGCCGSCGC	CCAGGCCCGC	ATGCGCGATC	CCTTCGGCTT	TGTCAGCGAG
GCCAGACCGA	GACGGCCGCG	GGTCCGGGCG	TACGCGCTAG	GGAAGCCGAA	ACAGTCGCTC
1210	1220	1230	1240	1250	1260
GCGCGGGCGC	GATCCGGTCC	TCCCGAGCCA	CGCCAGCGCA	AGCTGAAGAC	GCTCGAGGCG
CGCGCCCGCG	CTAGGCCAGC	AGGGCTCGGT	GCGGTGCGCT	TCGACTTCTG	CGAGCTCCGC
1270	1280	1290	1300	1310	1320
AAGGCCTTCG	AAACGCTTGG	TCTCGGAGCC	TCGGCGACCA	CTGCCGACAT	CAAGGCGGCC
TTCCGGAAGC	TTTGCGAACC	AGAGCCTCGG	AGCCGCTGGT	GACGGCTGTA	GTTCCGCCGG
1330	1340	1350	1360	1370	1380
TACAAGGACC	TCGTCAAGAA	GCATCACCCC	GATGCCAATG	GCGGAGATAG	AGGATCGGAA
ATGTTCTCTG	AGCAGTTCTT	CGTAGTGGGG	CTACGGTTAC	CGCCTCTATC	TCCTAGCCTT
1390	1400	1410	1420	1430	1440
GAGCGTTTTT	GCGCGGTTAT	TCAGGCCTAC	CAATTGTTAA	AACAGGCTGG	TTTCTGCTAA
CTCGCAAAAG	CGCGCCAATA	AGTCCGGATG	GTTAACAATT	TTGTCCGACC	AAAGACGATT
1450	1460	1470	1480	1490	1500
CAACCCGGAT	TAATACAGAA	GCACTTTTGC	AGGCGAATGC	GCGGGTGCCG	TCCGGTGGCC
GTTGGGCCTA	ATTATGTCTT	CGTGAAAACG	TCCGCTTACG	CGCCACCGGC	AGGCCACCGG
1510	1520	1530	1540	1550	1560
GCTCTGGAGA	CATGATGAGC	AAGATTGACC	TCGACATTTC	CAACCTCCCC	GACACCACGA
CGAGACCTCT	GTACTACTCG	TTCTAACTGG	AGCTGTAAAG	GTTGGAGGGG	CTGTGGTGCT
1570	1580	1590	1600	1610	1620
TTTCCGTCCG	GGAGGTTTTT	GGTATTGATA	CGGATTGCGG	CGTTCCTGCC	TATTCGAAGG
AAAGGCAGGC	CCTCCAAAAG	CCATAACTAT	GCCTAAACGC	GCAAGGACGG	ATAAGCTTCC
1630	1640	1650	1660	1670	1680
GCGACGCCTA	TGTCCCGGAT	CTGGATCCGG	ACTACCTCTT	CGACCGCGAA	ACGACGCTCG
CGCTGCGGAT	ACAGGGCCTA	GACCTAGGCC	TGATGGAGAA	GCTGGCGCTT	TGCTGCGAGC
1690	1700	1710	1720	1730	1740
CCATTCTCGC	AGGCTTCGCC	CACAACCGAC	GCGTGATGGT	GTCGGGCTAT	CACGGCACCG
GGTAAGAGCG	TCCGAAGCGG	GTGTTGGCTG	CGCACTACCA	CAGCCCGATA	GTGCCGTGGC
1750	1760	1770	1780	1790	1800
GCAAGTCCAC	CCATATCGAG	CAGGTCGCCG	CGCGCCTCAA	CTGGCCGTGC	GTGCGCGTCA
CGTTCAGGTG	GGTATAGCTC	GTCCAGCGGC	GCGCGGAGTT	GACCGGCACG	CACGCGCAGT
1810	1820	1830	1840	1850	1860
ACCTCGATAG	CCATGTCAGC	CGTATCGACC	TCGTGCGCAA	GGACGCGATC	GTCGTCAAGG
TGGAGCTATC	GGTACAGTCG	GCATAGCTGG	AGCAGCCGTT	CCTGCGCTAG	CAGCAGTTCC
1870	1880	1890	1900	1910	1920
ACGGCCTGCA	GGTCACCGAA	TTCAAGGACG	GCATCCTGCC	CTGGGCCTAC	CAGCACAATG
TGCCGGACGT	CCAGTGGCTT	AAGTTCCTGC	CGTAGGACGG	GACCCGGATG	GTCGTGTTAC

FIGURE 32.1

FEUILLE DE REMPLACEMENT

PL. 83 / 151

1930	1940	1950	1960	1970	1980
TCGCGCTCGT	CTTCGACGAA	TACGATGCCG	GCCGCCCGGA	CGTCATGTTT	GTCATCCAGC
AGCGCGAGCA	GAAGCTGCTT	ATGCTACGGC	CGGCGGGCCT	GCAGTACAAG	CAGTAGGTCT
1990	2000	2010	2020	2030	2040
GCGTGCTGGA	ATCCTCCGGC	CGCCTGACGC	TGCTCGACCA	GAGCCGTGTC	ATCCGTCCGC
CGCACGACCT	TAGGAGGCCG	GCGGACTGCG	ACGAGCTGGT	CTCGGCACAG	TAGGCAGGCC
2050	2060	2070	2080	2090	2100
ACCCGGCCTT	CCGCCTGTTT	GCGACCGCCA	ACACCGTCGG	CCTCGGCGAC	ACGACCGGCC
TGGGCCGGAA	GGCGGACAAA	CGCTGGCGGT	TGTGGCAGCC	GGAGCCGCTG	TGCTGGCCGG
2110	2120	2130	2140	2150	2160
TCTATCACGG	CACGCAGCAG	ATCAACCAGG	CGCAGATGGA	CCGCTGGTCG	ATCGTCACCA
AGATAGTGCC	GTGCGTCGTC	TAGTTGGTCC	GCGTCTACCT	GGCGACCAGC	TAGCAGTGGT
2170	2180	2190	2200	2210	2220
CGCTGAACTA	CCTGCCGCAC	GACAAGGAAG	TCGACATCGT	CGCCGCCAAG	GTCAAGGGCT
GCGACTTGAT	GGACGGCGTG	CTGTTCTTTC	AGCTGTAGCA	GCGGCGGTTC	CAGTTCCCGA
2230	2240	2250	2260	2270	2280
TCACCGCCGA	CAAGGGCCGC	GAGACCGTCT	CCAAGATGGT	ACGTGTCGCC	GACCTCACGC
AGTGGCGGCT	GTTCCCGGCG	CTCTGGCAGA	GGTTCTACCA	TGCACAGCGG	CTGGAGTGCG
2290	2300	2310	2320	2330	2340
GCGCAGCCTT	CATCAATGGC	GATCTCTCGA	CTGTCTATGAG	CCCGCGTACG	GTCATCACCT
CGCGTCGGAA	GTAGTTACCG	CTAGAGAGCT	GACAGTACTC	GGGCGCATGC	CAGTAGTGGA
2350	2360	2370	2380	2390	2400
GGGCCGAGAA	CGCCACATC	TTCGGCGACA	TCGCTTTTCGC	CTTCCGCGTG	ACCTTCCTCA
CCCGGCTCTT	GCGGGTGTAG	AAGCCGCTGT	AGCGAAAGCG	GAAGGCGCAC	TGGAAGGAGT
2410	2420	2430	2440	2450	2460
ACAAGTGCGA	CGAGCTGGAG	CGGGCGCTGG	TCGCCGAGCA	CTACCAGCGC	GCCTTCGGCA
TGTTACGCT	GCTCGACCTC	GCCCGCGACC	AGCGGCTCGT	GATGGTCGCG	CGGAAGCCGT
2470	2480	2490	2500	2510	2520
TCGAGCTGAA	GGAATGCGCT	GCCAACATCG	TGCTCGAAGC	CACCGCCTGA	TCCCACGGCC
AGCTCGACTT	CCTTACGCGA	CGGTTGTAGC	ACGAGCTTCG	GTGGCGGACT	AGGGTGCCGG
2530	2540	2550	2560	2570	2580
TGCCGTCCCC	TTTGGGAGGG	CGGGTCATGA	CGCTGTGGCA	AACCGGATGA	CGCCCCACTG
ACGGCAGGGG	AAACCCTCCC	GCCCAGTACT	GCGACACCGT	TTGGCCTACT	GCGGGGTGAC
2590	2600	2610	2620	2630	2640
GGGCGCCGTC	GCCTCTGGCT	GAAGAAGGAA	CTGTCTGTAG	CTCGAATTCG	AAGGCAAAGC
CCCGCGGCAG	CGGAGACCGA	CTTCTTCCTT	GACAGCACTC	GAGCTTAAGC	TTCCGTTTCG
2650	2660	2670	2680	2690	2700
CAACCACGCG	CGAGAATGCT	GCGGAACCGT	TCAAGCGGGC	GCTTTCCGGC	TGCATCCGAT
GTTGGTGCGC	GCTCTTACGA	CGCCTTGGCA	AGTTCGCCCC	CGAAAGGCCG	ACGTAGGCTA
2710	2720	2730	2740	2750	2760
CGATCGCGGG	CGATGCCGAG	GTGGAAGTCG	CCTTCGCCAA	CGAGCGGCCG	GGCATGACCG
GCTAGCGCCC	GCTACGGCTC	CACCTTCAGC	GGAAGCGGTT	GCTCGCCGGC	CCGTACTGGC
2770	2780	2790	2800	2810	2820
GCGAACGCAT	CCGTCTGCCG	GAACCTTCCA	AGCGCCCGAC	CCTGCAGGAA	CTTGCCGTGA
CGCTTGCGTA	GGCAGACGGC	CTTGAAAGGT	TCGCGGGCTG	GGACGTCCTT	GAACGGCACT
2830	2840	2850	2860	2870	2880
CCCGCGGGCT	CGGTGACAGC	ATGGCGCTGC	GCAAGGCCTG	TACGCATGCG	CGGATCCAGC
GGGCGCCCGA	GCCACTGTCT	TACCGCGACG	CGTTCCGGAC	ATGCGTACGC	GCCTAGGTCT

FIGURE 32.2

FEUILLE DE REMPLACEMENT

PL. 84 / 151

2890	2900	2910	2920	2930	2940
GCACCATGTC	GCCGCAAGGG	GCGGACGCCC	GCGCGATCTT	CGATGCGGTG	GAGCAGGCTC
CGTGGTACAG	CGGCGTTCCC	CGCCTGCGGG	CGCGCTAGAA	GCTACGCCAC	CTCGTCCGAG
2950	2960	2970	2980	2990	3000
GTGTCGAGGC	GATCGGGTCG	TTGCGCATGG	CGGGTGTCGC	CAAGAACCTC	AACGTCATGC
CACAGCTCCG	CTAGCCCAGC	AACGCGTACC	GCCCACAGCG	GTTCTTGGAG	TTGCAGTACG
3010	3020	3030	3040	3050	3060
TCGAAGAGAA	ATACGCCAAG	GCGAATTTCG	CAACGATCGA	GCGCCAGGCG	GACGCGCCGC
AGCTTCTCTT	TATGCGGTTC	CGCTTAAAGC	GTTGCTAGCT	CGCGGTCCGC	CTGCGCGGCG
3070	3080	3090	3100	3110	3120
TCGGCGAGGC	CGTAGCGCG	CTGGTGCGCG	AGAAGCTGAC	GGGCCAGAAG	CCGCCGGCGT
AGCCGCTCCG	GCATCGCGAC	GACCACGCGC	TCTTCGACTG	CCCGGTCTTC	GGCGGCCGCA
3130	3140	3150	3160	3170	3180
CTGCCGGCAA	GGTGCTCGAC	CTCTGGCGCG	AGTTCATCGA	GGGCAAGGCT	GCCGGCGACA
GACGGCCGTT	CCACGAGCTG	GAGACCGCGC	TCAAGTAGCT	CCCGTTCCGA	CGGCCGCTGT
3190	3200	3210	3220	3230	3240
TTGAGCACCT	GTCGTGACG	ATCAACAACC	AGCAGGCCTT	TGCCCCGGTC	GTTCGCGACA
AACTCGTGGA	CAGCAGCTGC	TAGTTGTTGG	TCGTCCGGAA	ACGGGCCCCAG	CAAGCGCTGT
3250	3260	3270	3280	3290	3300
TGCTGACCTC	GATGGAAGTC	GCCGAGAAAT	ACGGTGACGA	CGACAACGAG	CCGGACGAGC
ACGACTGGAG	CTACCTTCAG	CGGCTCTTTA	TGCCACTGCT	GCTGTTGCTC	GGCCTGCTCG
3310	3320	3330	3340	3350	3360
AGGAAAGCGA	GACCGACGAA	GACCAGCCGC	GCAGCCAGGA	GCAGGACGAG	AACGCCAGCG
TCCTTTCGCT	CTGGCTGCTT	CTGGTCGGCG	CGTCGGTCTT	CGTCCTGCTC	TTGCGGTGCG
3370	3380	3390	3400	3410	3420
ACGAGGAAGC	CGGCGACGAT	GCCGCACCCG	CCGACGAGAA	CCAGGCTGCC	GAAGAGCAGA
TGCTCCTTCG	GCCGCTGCTA	CGGCGTGGGC	GGCTGCTCTT	GGTCCGACGG	CTTCTCGTCT
3430	3440	3450	3460	3470	3480
TGGAAGAAGG	CGAGATGGAC	GGCGCGGAGA	TCTCCGACGA	CGATCTCCAG	GACGAAGGCG
ACCTTCTTCC	GCTCTACCTG	CCGCGCCTCT	AGAGGCTGCT	GCTAGAGGTC	CTGCTTCCGC
3490	3500	3510	3520	3530	3540
ACGAGGACAG	CGAAACGCCC	GGCGAGGTCA	AGCGTCCGAA	CCAGCCCTTC	GCCGACTTCA
TGCTCCTGTC	GCTTTGCGGG	CCGCTCCAGT	TCGCAGGCTT	GGTCGGGAAG	CGGCTGAAGT
3550	3560	3570	3580	3590	3600
ACGAGAAGGT	CGACTACGCC	GTCTTCACCC	GCGAGTTCGA	CGAGACGATT	GCCTCGGAAG
TGCTCTTCCA	GCTGATGCGG	CAGAAGTGGG	CGCTCAAGCT	GCTCTGCTAA	CGGAGCCTTC
3610	3620	3630	3640	3650	3660
AGCTTTGCGA	CGAGGCCGAG	CTCGACCGGC	TGCGCGCCTT	CCTCGACAAG	CAGCTTGCCC
TCGAAACGCT	GCTCCGGCTC	GAGCTGGCCG	ACGCGCGGAA	GGAGCTGTTC	GTCGAACGGG
3670	3680	3690	3700	3710	3720
ATCTTCAAGG	CGCGGTGCGC	CGCCTTGCCA	ACCGGCTGCA	GCGCCGCCTG	ATGGCGCAGC
TAGAAGTTCC	GCGCCAGCCG	GCGGAACGGT	TGGCCGACGT	CGCGGCGGAC	TACCGCGTCC
3730	3740	3750	3760	3770	3780
AGAACCGCTC	CTGGGAGTTC	GATCTCGAAG	AGGGGTATCT	CGATTCGGCG	CGGCTTCAGC
TCTTGCGGAG	GACCTCAAG	CTAGAGCTTC	TCCCCATAGA	GCTAAGCCGC	GCCGAAGTCG
3790	3800	3810	3820	3830	3840
GCATCATCAT	CGATCCGATG	CAGCCGCTTT	CCTTCAAGCG	CGAAAAGGAC	ACCAACTTCC
CGTAGTAGTA	GCTAGGCTAC	GTCGGCGAAA	GGAAGTTCGC	GCTTTTCCTG	TGGTTGAAGG

FIGURE 32.3

FEUILLE DE REMPLACEMENT

PL. 85 / 151

3850	3860	3870	3880	3890	3900
GCGATACCGT	CGTGACGCTG	CTGATCGACA	ATTCCGGCTC	GATGCGCGGC	CGTCCGATCA
CGCTATGGCA	GCACTGCGAC	GACTAGCTGT	TAAGGCCGAG	CTACGCGCCG	GCAGGCTAGT
3910	3920	3930	3940	3950	3960
CGGTTGCCGC	CACCTGCGCC	GATATCCTGG	CGCGCACGCT	CGAGCGCTGC	GGCGTCAAGG
GCCAACGGCG	GTGGACGCGG	CTATAGGACC	GCGCGTGCGA	GCTCGCGACG	CCGCAGTTCC
3970	3980	3990	4000	4010	4020
TCGAGATCCT	CGGTTTTACC	ACCAAGGCGT	GGAAGGGTGG	GCAGTCACGC	GAGAAGTGGC
AGCTCTAGGA	GCCAAAATGG	TGGTTCCGCA	CCTTCCCACC	CGTCAGTGCG	CTCTTCACCG
4030	4040	4050	4060	4070	4080
TGGCCGGCGG	CAAGCC CAG	GCCCCGGGTC	GCCTCAACGA	CCTGCGACAC	ATCGTCTACA
ACCGGCCGCC	GTTCGGTGTC	CGGGGCCGAG	CGGAGTTGCT	GGACGCTGTG	TAGCAGATGT
4090	4100	4110	4120	4130	4140
AGTCTGCCGA	CGCTCCGTGG	CGCCGGGCAC	GACGCAATCT	CGGCCTGATG	ATGCGGGAAG
TCAGACGGCT	GCGAGGCACC	GCGGCCCGTG	CTGCGTTAGA	GCCGGACTAC	TACGCCCTTC
4150	4160	4170	4180	4190	4200
GCCTGCTCAA	GGAAAACATC	GACGGCGAGG	CGTTGATTTG	GGCGCATGAG	CGGCTGATGG
CGGACGAGTT	CCTTTTGTAG	CTGCCGCTCC	GCAACTAAAC	CCGCGTACTC	GCCGACTACC
4210	4220	4230	4240	4250	4260
CGCGGCGCGA	ACAGCGGCGC	ATCCTGATGA	TGATTTGCGA	CGGCGCGCCG	GTCGACGACT
GCGCCGCGCT	TGTCGCGCGC	TAGGACTACT	ACTAAAGCCT	GCCGCGCGGC	CAGCTGCTGA
4270	4280	4290	4300	4310	4320
CGACGCTGTC	GGTCAATCCA	GGAAGCTATC	TGGAGCGTCA	CCTGCGCGCG	GTCATCGAGC
GCTGCGACAG	CCAGTTAGGT	CCTTTGATAG	ACCTCGCAGT	GGACGCGCGC	CAGTAGCTCG
4330	4340	4350	4360	4370	4380
AGATCGAAAC	GCGCTCGCCG	GTGGAAGTGC	TGGCGATCGG	TATCGGCCAC	GACGTGACGC
TCTAGCTTTG	CGCGAGCGGC	CACCTTGACG	ACCGCTAGCC	ATAGCCGGTG	CTGCACTGCG
4390	4400	4410	4420	4430	4440
GCTACTATCG	CCGTGCCGTC	ACCATCGTCG	ATGCCGATGA	GCTTGCCGGC	GCGATGACCG
CGATGATAGC	GGCACGGCAG	TGGTAGCAGC	TACGGCTACT	CGAACGGCCG	CGCTACTGGC
4450	4460	4470	4480	4490	4500
AACAGCTGGC	CGCACTCTTC	GAGGACGAAA	GCCAGCGCCG	CGGTTCTTCG	CGTCTTCGCC
TTGTCGACCG	GCGTGAGAAG	CTCCTGCTTT	CGGTGCGGGC	GCCAAGAAGC	GCAGAAGCGG
4510	4520	4530	4540	4550	4560
GCGCCGGGTG	ATGCTTCCCC	CTTGGGGGCG	GTGGAACATC	GCCTCCGAGC	TGCCAATCGG
CGCGGCCAC	TACGAAGGGG	GAACCCCGC	CACCTTGTAG	CGGAGGCTCG	ACGGTTAGCC
4570	4580	4590	4600	4610	4620
CACCTGCACG	CATCGCTGGC	GGCCGAAGTC	AATTTACGGA	CATAGTTTTA	CAGTCTACCA
GTGGACGTGC	GTAGCGACCG	CCGGCTTCAG	TTAAATGCCT	GTATCAAAAT	GTCAGATGGT
4630	4640	4650	4660	4670	4680
AGCTACCATG	CGTGGCGGGC	TCACTTTGAG	CGCACGCCGC	GTCATTCCCG	ATGCCCCCTG
TCGATGGTAC	GCACCGCCCG	AGTGAAACTC	GCGTGCGGCG	CAGTAAGGGC	TACGGGGGAC
4690	4700	4710	4720	4730	4740
AAGGTACTTC	TCTTGATGCT	TGGCCGCGGT	CTCCTAGCCC	TTTTCCTCCT	GGCTTCGGCC
TTCCATGAAG	AGAACTACGA	ACCGGCGCCA	GAGGATCGGG	AAAAGGAGGA	CCGAAGCCGG
4750	4760	4770	4780	4790	4800
TGCCCCGC					
ACGGGCCG					

FIGURE 32.4

PL. 86 / 151

10	20	30	40	50	60
GAGCTCATAG	AGCAGTTCCT	CGATCGACTT	CAGCAGTCGC	ATGAAATCCA	TGCCGTGCTC
CTCGAGTATC	TCGTCAAGGA	GCTAGCTGAA	GTCGTCAGCG	TACTTTAGGT	ACGGCACGAG
70	80	90	100	110	120
CCCTTGCTTC	TATGCGTGGC	ACGACCGCGC	GCCGGGGCCG	ATGCCGGTCA	GTCGCGCAGA
GGGAACGAAG	ATACGCACCG	TGCTGGCGCG	CGGCCCCGCG	TACGGCCAGT	CAGCGCGTCT
130	140	150	160	170	180
CGCAGCTCGT	CGGTACGCAT	CTGCAGCATC	TCCAGCGTCG	ACAGGAAGCT	CATGCCGAGC
GCGTCGAGCA	GCCATGCTTA	GACGTCGTAG	AGGTCGCAGC	TGTCCTTCGA	GTACGGCTCG
190	200	210	220	230	240
AGGCTCTGAT	CGAGCTTGCC	CTTGGCTGCG	ACCGTTGCGC	CGATGTTGCG	GCGGGTGATC
TCCGAGACTA	GCTCGAACGG	GAACCGACGC	TGGCAACGCG	GCTACAACGC	CGCCCACTAG
250	260	270	280	290	300
GGGCCGATCG	AGATCTCCTG	AAGCATCACG	GGGGCTGCCT	GGGCCCCGCC	ATTGGCTGTC
CCCGGCTAGC	TCTAGAGGAC	TTCGTAGTGC	CCCCGACGGA	CCCGGGCCGG	TAACCGACAG
310	320	330	340	350	360
ATGACCGTGA	CGATAAAGTT	GAGGTTGGCC	GGGTCGAGGC	CGATCTTTTC	CGCATCTTCA
TACTGGCACT	GCTATTTCAA	CTCCAACCGG	CCCAGCTCCG	GCTAGAAAAG	GCGTAGAAGT
370	380	390	400	410	420
TAGGTGAGCG	CGATGTTGCT	GGCGCCGGTA	TCGACCAGCA	TGCTGATGTC	CTTGCCGTCG
ATCCACTCGC	GCTACAACGA	CCGCGGCCAT	AGCTGGTCGT	ACGACTACAG	GAACGGCAGC
430	440	450	460	470	480
ACCGTCGCAG	TGGTCTCGAA	ATGACCGTTC	AGCATCTTCT	GCAGCACCAC	TTCCTGCTGT
TGGCAGCGTC	ACCAGAGCTT	TACTGGCAAG	TCGTAGAAGA	CGTCGTGGTG	AAGGACGACA
490	500	510	520	530	540
CCCTCGCTGT	CAGTGATGAT	GGTGGCGCGG	CCGGGGATGA	GGCCGGCGAG	CAGGCGGTTA
GGGAGCGACA	GTCACTACTA	CCACCGCGCC	GGCCCCTACT	CCGGCCGCTC	GTCCGCCAAT
550	560	570	580	590	600
CCGAAGCCCT	CCAACCTCGAA	GCGGTAGACA	TAGGCCGAGA	CCAGCGCCAG	AACGACGAAG
GGCTTCGGGA	GGTTGAGCTT	CGCCATCTGT	ATCCGGCTCT	GGTCGCGGTC	TTGCTGCTTC
610	620	630	640	650	660
AGCCAGATGG	CGATCTGACG	CAGGCCTTCG	CCGAAGCGGT	GGCGGCTCTG	CAGGATGCCG
TCGGTCTACC	GCTAGACTGC	GTCCGGAAGC	GGCTTCGCCA	CCGCCGAGAC	GTCCTACGGC
670	680	690	700	710	720
GCGCCGATCA	GCGTGGCGAT	GGCGCCGAGC	GAGACCAGTT	GCCCGAACTG	GTCATTGGCA
CGCGGCTAGT	CGCACCGCTA	CCGCGGCTCG	CTCTGGTCAA	CGGGCTTGAC	CAGTAACCGT
730	740	750	760	770	780
AGCCCCATGG	TGCGGCCGGT	GTCGTGGTTG	ATGATCAGCA	GGATGAGGCC	GATGGCCAGG
TCGGGGTACC	ACGCCGGCCA	CAGCACCAAC	TACTAGTCGT	CCTACTCCGG	CTACCGGTCC
790	800	810	820	830	840
ATCGAGAGCA	GGATGGCAAG	ACGGGTCATG	CTTCGCCGCG	TTCCCTCGCC	ATGCGCGTGC
TAGCTCTCGT	CCTACCGTTC	TGCCCAGTAC	GAAGCGGCGC	AAGGGAGCGG	TACGCGCAGC
850	860	870	880	890	900
GTCGGGTTTC	GCGCCGCGGC	TTGCGTTCGA	CGGTCTCAAG	CCGTGCAGGC	AACGCGCTCA
CAGCCCAAAG	CGCGGCGCCG	AACGCAAGCT	GCCAGAGTTC	GGCACGTCCG	TTGCGCGAGT
910	920	930	940	950	960
TGATCGCGCG	GCGTTCGGCA	TCGGTATAGA	GCGTCCAGCG	TCCGACTTCG	TCGCGGGTAC
ACTAGCGCGC	CGCAAGCCGT	AGCCATATCT	CGCAGGTCGC	AGGCTGAAGC	AGCGCCCATG

FIGURE 33

PL. 87/151

970	980	990	1000	1010	1020
GGCCGCAGCC	GAAACAGTAG	CCGGTCTTGT	CATCGATCGA	ACAGACGAGA	ATGCAGGGAG
CCGGCGTCGG	CTTTGTCATC	GGCCAGAACA	GTAGCTAGCT	TGTCTGCTCT	TACGTCCCTC
1030	1040	1050	1060	1070	1080
ATTCCATGGG	CGTGCTCAGT	TTTCCCTTGA	TATATCGATG	TTTCAAACCG	TCAGCGCAAG
TAAGGTACCC	GCACGAGTCA	AAAGGGAACT	ATATAGCTAC	AAAGTTTGGC	AGTCGCGTTC
1090	1100	1110	1120	1130	1140
GGCACCAGC	ACGGCGATTT	CGGTCAGTTG	CTGCGTCGCC	CCGATCGTGT	CGCCCGTTTG
CCGTGGCTCG	TGCCGCTAA	GCCAGTCAAC	GACGCAGCGG	GGCTAGCACA	GCGGGCAAAC
1150	1160	1170	1180	1190	1200
TCCGCCGATC	TTGCGCATCG	CCAGCCGAGC	GAAGCCCTTG	ACCGTGGCAA	GGAATGCGAC
AGGCGGCTAG	AACGCGTAGC	GGTCGGCTCG	CTTCGGGAAC	TGGCACC GTT	CCTTACGCTG
1210	1220	1230	1240	1250	1260
GAGCGCCGCG	ATGACGCCGA	GCGCCGGGAC	CTGCGCGAGA	TAGAAGAGCA	GCATTGCGAC
CTCGCGGCGC	TACTGCGGCT	GCGGGCCCTG	GACGCGCTCT	ATCTTCTCGT	CGTAACGCTG
1270	1280	1290	1300	1310	1320
AAGAAGTCCG	AAGGCAAGCG	CGAAGCGCGT	GGCCGCCGGT	TCCGGCTCGC	CAGCCGAGGC
TTCTTCAGGC	TTCCGTTCGC	GCTTCGCGCA	CCGGCGGCCA	AGGCCGAGCG	GTCGGCTCCG
1330	1340	1350	1360	1370	1380
CGCGACGCCG	CTGCTGCGCG	CCGGCGGAAG	CGACGACCAG	TGCCAGACCA	TGGCGGCGCG
GCGCTGCGGC	GACGACGCGC	GGCCGCCTTC	GCTGCTGGTC	ACGGTCTGGT	ACCGCCGCGC
1390	1400	1410	1420	1430	1440
GCTGAGGCAC	GCTGCGCCAA	GGATCGCCAT	GGCGGCGCCC	AGCGGCGAAA	AGAGCGGCAG
CGACTCCGTG	CGACGCGGTT	CCTAGCGGTA	CCGCCGCGGG	TCGCCGCTTT	TCTCGCCGTC
1450	1460	1470	1480	1490	1500
GATCGAGGCG	AACGCCGAGA	CGCGCAGGCC	GAAGGAGAGG	ATGAGGGCGA	CGGCCGCATA
CTAGCTCCGC	TTGCGGCTCT	GCGCGTCCGG	CTTCCTCTCC	TACTCCCGCT	GCCGGCGTAT
1510	1520	1530	1540	1550	1560
GGTGCCGATG	CGGCTGTCCT	TCATGATCGC	AAGCGCCGCT	TCGCGGTCGC	GACCGCCGCC
CCACGGCTAC	GCCGACAGGA	AGTACTAGCG	TTGCGGGCGA	AGCGCCAGCG	CTGGCGGCGG
1570	1580	1590	1600	1610	1620
AAAGCCATCG	GCCGTGTGCG	CAAGCCCGTC	TTCGTGCACT	GCGCCGGTGA	CAAGCGCCTG
TTTCGGTAGC	CGGCACAGCG	GTTTCGGGCG	AAGCACGTCA	CGCGGCCACT	GTTTCGGGAC
1630	1640	1650	1660	1670	1680
GATGGCGACG	ACGACAAAGG	CGGCAAAGAG	CGAGCTCACC	TGCAGCGCCA	TGAGGGCCAT
CTACCGCTGC	TGCTGTTTCC	GCCGTTTCTC	GCTCGAGTGG	ACGTCGCGGT	ACTCCCGGTA
1690	1700	1710	1720	1730	1740
GGCGACGGCC	GCCGATGGCA	GTGCGATCGC	CAGGCCGGCG	AACGGGAAGG	CGCGCACGGC
CCGCTGCCGG	CGGCTACCGT	CACGCTAGCG	GTCCGGCCGC	TTGCCCTTCC	GCGCGTGCCG
1750	1760	1770	1780	1790	1800
ACGGCTCAAG	CGCCCGTCAT	AACCTTCGAA	ATGACGCGCA	GGCATCGGGA	TGCGGCTGAG
TGCCGAGTTC	GCGGGCAGTA	TTGGAAGCTT	TACTGCGCGT	CCGTAGCCCT	ACGCCGACTC
1810	1820	1830	1840	1850	1860
AAAGCCGATC	GACCGCGCCA	CATCGTCACA	GAAATCGCCA	ACGAAGCCCA	TGGCTCCTCC
TTTCGGCTAG	CTGGCGCGGT	GTAGCAGTGT	CTTTAGCGGT	TGCTTCGGGT	ACCGAGGAGG
1870	1880	1890	1900	1910	1920
AAGGTTGCGG	CCATTGACCC	GGCCGCTGCC	AAACTCGCCG	ACTGCGGCGA	GTCTCGCAAG
TTCCAACGCC	GGTAACTGGG	CCGGCGACGG	TTTGAGCGGC	TGACGCCGCT	CAGAGCGTTC

FIGURE 33.1

FEUILLE DE REMPLACEMENT

PL. 88 / 151

1930 CCGGGCGGGC GGCCCGCCCG	1940 GCACCCGCGA CGTGGGCGCT	1950 GGGCCGCGCA CCCGGCGCGT	1960 CACTTTTCCC GTGAAAAGGG	1970 AGACCTTTCA TCTGGAAAGT	1980 TAGGCCGTCT ATCCGGCAGA
1990 GCGACCGCTC CGCTGGCGAG	2000 GCGGATCGAG CGCCTAGCTC	2010 ACGGCGACGC TGCCGCTGCG	2020 CGATTGGCGC GCTAACCGCG	2030 AAATGTCGTT TTTACAGCAA	2040 GCCCCGAATTT CGGGCTTAAA
2050 TCGGCGCCCT AGCCGCGGGA	2060 CTATGAGGGG GATACTCCCG	2070 CGTAGATAGA GCATCTATCT	2080 GCTTCACGAT CGAAGTGCTA	2090 GATGCAAGGA CTACGTTCTT	2100 TTCCTCCCAT AAGGAGGGTA
2110 GAGTGCCAGC CTCACGGTCG	2120 GGCCTGCCGT CCGGACGGCA	2130 TTGATGATTT AACTACTAAA	2140 TCGCGAATTG AGCGCTTAAC	2150 TTGCGCAACC AACGCGTTGG	2160 TGCCGGGCCC ACGGCCCGGG
2170 GGATGCGGCA CCTACGCCGT	2180 GCCCTCGTTG CGGGAGCAAC	2190 CCGCGCGGGA GGCGCGCCCT	2200 GCGGGACGCC CGCCCTGCGG	2210 CAGCTGACGA GTCGACTGCT	2220 AGCCGCCGGG TCGGCGGCCC
2230 CGCGCTCGGC GCGCGAGCCG	2240 CGCCTCGAGG GCGGAGCTCC	2250 AAATCGCCTT TTTAGCGGAA	2260 CTGGCTCGCC GACCGAGCGG	2270 GCCTGGACGG CGGACCTGCC	2280 GCAAGGCGCC CGTTCCGCGG
2290 GGTGGTCAAC CCACCAGTTG	2300 CGGCCGCTGG GCCGGCGACC	2310 TGGCGATCTT ACCGCTAGAA	2320 TGCCCGCAAC ACGGCCGTTG	2330 CACGGCGTCA GTGCCGCACT	2340 CCCGCCAGGG GGGCGGTCCC
2350 GGTGACCCCG CCACTGGGGC	2360 TTCCCGTCAT AAGGGCAGTA	2370 CCGTCACCGC GGCAGTGGCG	2380 ACAGATGGTC TGTCTACCAG	2390 GAGAATTTTG CTCTTAAAC	2400 CCGCCGCTGG GGCGGCCACC
2410 CGCTGCGATC GCGACGCTAG	2420 AACCAGATCT TTGGTCTAGA	2430 GCGTCAGCCA CGCAGTCGGT	2440 CGACCTCGGG GCTGGAGCCC	2450 CTGAAGGTCT GACTTCCAGA	2460 TCGACCTCGC AGCTGGAGCG
2470 ACTCGAATAC TGAGCTTATG	2480 CCGACCGGTG GGCTGGCCAC	2490 ATATCACCAG TATAGTGGCT	2500 GGAAGCCGCG CCTTCGGCGC	2510 CTGTCCGAGC GACAGGCTCG	2520 GCGATTGCGC CGCTAACGCG
2530 CGCGACCATG GCGCTGGTAC	2540 GCCTTTGGCA CGGAAACCGT	2550 TGGAGGCGAT ACCTCCGCTA	2560 TGCCGGCGGC ACGGCCGCCG	2570 ACGGATCTTC TGCCTAGAAG	2580 TGTGCATCGG ACACGTAGCC
2590 CGAAATGGGC GCTTTACCCG	2600 ATCGGCAACA TAGCCGTTGT	2610 CCACGATCGC GGTGCTAGCG	2620 GGCCGCGATC CCGGCGCTAG	2630 AATCTCGGCC TTAGAGCCGG	2640 TTTATGGTGG AAATACCACC
2650 CACGGCCGAA GTGCCGGCTT	2660 GAATGGGTCT CTTACCCAGC	2670 GTCCGGGTAC CAGGCCCATG	2680 CGGCTCCGAG GCCGAGGCTC	2690 GGCGAGGTGC CCGCTCCACG	2700 TGAAGCGCAA ACTTCGCGTT
2710 GATCGCCGCG CTAGCGGCGC	2720 GTCGAAAAGG CAGCTTTTCC	2730 CCGTGGCGCT GGCACCAGCA	2740 GCATCGCGAT CGTAGCGCTA	2750 CACCTGTCCG GTGGACAGGC	2760 ATCCGCTCGA TAGGCGAGCT
2770 ACTGATGCGT TGAATACGCA	2780 CGCCTCGGCG GCGGAGCCGC	2790 GTCGTGAGAT CAGCACTCTA	2800 CGCGGCCATG GCGCCGGTAC	2810 GCTGGCGCCA CGACCGCGGT	2820 TCCTGGCCGC AGGACCGGCG
2830 CCGCGTCCAG GGCGCAGGTC	2840 AAGGTACCTG TTCCATGGAC	2850 TCATCATCGA AGTAGTAGCT	2860 CGGCTACGTG GCCGATGCAC	2870 GCGACCGCTG CGCTGGCGAC	2880 CGGCTTCGAT GCCGAAGCTA

FIGURE 33.2

PL. 89/151

2890	2900	2910	2920	2930	2940
CCTGAAGGCG	GCCAACCCGT	CGGCCCTCGA	CCATTGCCTG	ATCGGCCATG	TTTCGGGCGA
GGACTTCCGC	CGGTTGGGCA	GCCGGGAGCT	GGTAACGGAC	TAGCCGGTAC	AAAGCCCCGCT
2950	2960	2970	2980	2990	3000
ACCGGGGGCAT	CTGCGCGCGA	TCGAGAAGCT	CGGCAAGACG	CCGCTGCTGG	CACTCGGCAT
TGGCCCCGTA	GACGCGCGCT	AGCTCTTCGA	GCCGTTCTGC	GGCGACGACC	GTGAGCCGTA
3010	3020	3030	3040	3050	3060
GCGGCTTGGC	GAAGGCACGG	GCGCGGCCCT	TGCCGCGCGT	ATCGTCAAGG	CGGCGGCCGC
CGCCGAACCG	CTTCCGTCGC	CGCGCCGGGA	ACGGCGGCCA	TAGCAGTTCC	GCCGCGGGCG
3070	3080	3090	3100	3110	3120
TTGCCACAGC	GGCATGGCGA	CCTTTGCCCA	GGCCGGCGTC	AGCAACAAGG	AATAGTGAAG
AACGGTGTCG	CCGTACCGCT	GGAAACGGGT	CCGGCCGCAG	TCGTTGTTCC	TTATCACTTC
3130	3140	3150	3160	3170	3180
TTCCGGCCGG	GCTTTGCAGG	AAGGCCGGCC	GGTTTCTGTC	CAAGGCCTGT	CACGGGCGCG
AAGGCCGGCC	CGAAACGTCC	TTCCGGCCGG	CCAAAGACAG	GTTCCGGACA	GTGCCCAGCG
3190	3200	3210	3220	3230	3240
AAGCTGTGCG	GTGCCGGGCC	TTGATGGATG	CGTCCTTCTC	GCCTATCCAA	AGCGCAAATG
TTGACAGCG	CACGGCCCCG	AACTACCTAC	GCAGGAAGAG	CGGATAGGTT	TCGCGTTTAC
3250	3260	3270	3280	3290	3300
CGCGCCCTAG	CTATAGTCTT	GGGTGCCTGC	AACCGAGACC	GCCTTGCAAT	CGCCTCAATC
GCGCGGGATC	GATATCAGAA	CCCACGGACG	TTGGCTCTGG	CGGAACGTAA	GCGGAGTTAG
3310	3320	3330	3340	3350	3360
ACGATGTCGA	AGCAAGCACA	GTTTCAAGCC	CTGTGCGAGC	GAAATGGACG	CCAAGAACAC
TGCTACAGCT	TCGTTCTGTG	CAAAGTTCGG	GACAGCTCTG	CTTTACCTGC	GGTTCTTGTC
3370	3380	3390	3400	3410	3420
CACGCACCGC	ATTGGACAGA	CGGGTCCTGT	CGAGAAGCAG	ACCGGCATTC	GGCATCTCTT
GTGCGTGCGG	TAACCTGTCT	GCCCAGGACA	GCTCTTCGTC	TGGCCGTAAG	CCGTAGAGAA
3430	3440	3450	3460	3470	3480
TGCCGCTGCG	AGCTATTCGC	TCGGCGGGCG	CAAGCGGCTG	ATCGGCGAGG	CTGCCTTTCG
ACGGCGACGC	TCGATAAGCG	AGCCGCGCGC	GTTGCGCCGAC	TAGCCGCTCC	GACGGAAAGC
3490	3500	3510	3520	3530	3540
CCACGAGCTG	ATCGCCTTTG	CCGCCGCGAT	GATCGCTTTC	ATCATCGTCG	GCGCAACCTT
GGTGCTCGAC	TAGCGGAAAC	GGCGGCGCTA	CTAGCGAAAG	TAGTAGCAGC	CGCGTTGGAA
3550	3560	3570	3580	3590	3600
CTTCCAATAT	GTGGCGATGG	CGATCCTGTT	CCTGCTGATG	ATGGCCTTCG	AGGCGATCAA
GAAGGTTATA	CACCGCTACC	GCTAGGACAA	GGACGACTAC	TACCGGAAGC	TCCGCTAGTT
3610	3620	3630	3640	3650	3660
CACGGCAATC	GAGGAAATTG	TCGATCGCGT	TTCTCCCGAA	ATCTCGGAAA	TGGGTAAGAA
GTGCCGTTAG	CTCCTTTAAC	AGCTAGCGCA	AAGAGGGCTT	TAGAGCCTTT	ACCCATTCTT
3670	3680	3690	3700	3710	3720
CGCCAAGGAT	CTCGGCTCCT	TCGCCTGCCT	CTGCCTGATT	GTCGCCAACG	GTGTCTATGC
GCGGTTCCCTA	GAGCCGAGGA	AGCGGACGGA	GACGGACTAA	CAGCGGTTGC	CACAGATACG
3730	3740	3750	3760	3770	3780
CGCCTATGTC	GTGATCTTCG	ACGGCTTCAT	GAACTGACCG	GCTAGCGGGC	CGGCGCCTTC
GCGGATACAG	CACTAGAAGC	TGCCGAAGTA	CTTGACTGGC	CGATCGCCCG	GCCGCGGAAG
3790	3800	3810	3820	3830	3840
ACCCGATAAA	GCACATGCGG	ACGCAGCGGG	TTGCCCCCGG	GTACCGTGAC	GTCGTGAAA
TGGGCTATTT	CGTGTACGCC	TGCGTCGCCC	AACGGGGGCC	CATGGCACTG	CAGCAGCTTT

FIGURE 33.3

3850 3860 3870 3880 3890 3900
PL. 90 / 15 /
TCATCAGCCG GATCC
AGTAGTCGGC CTAGG

FIGURE 33.4

91/151

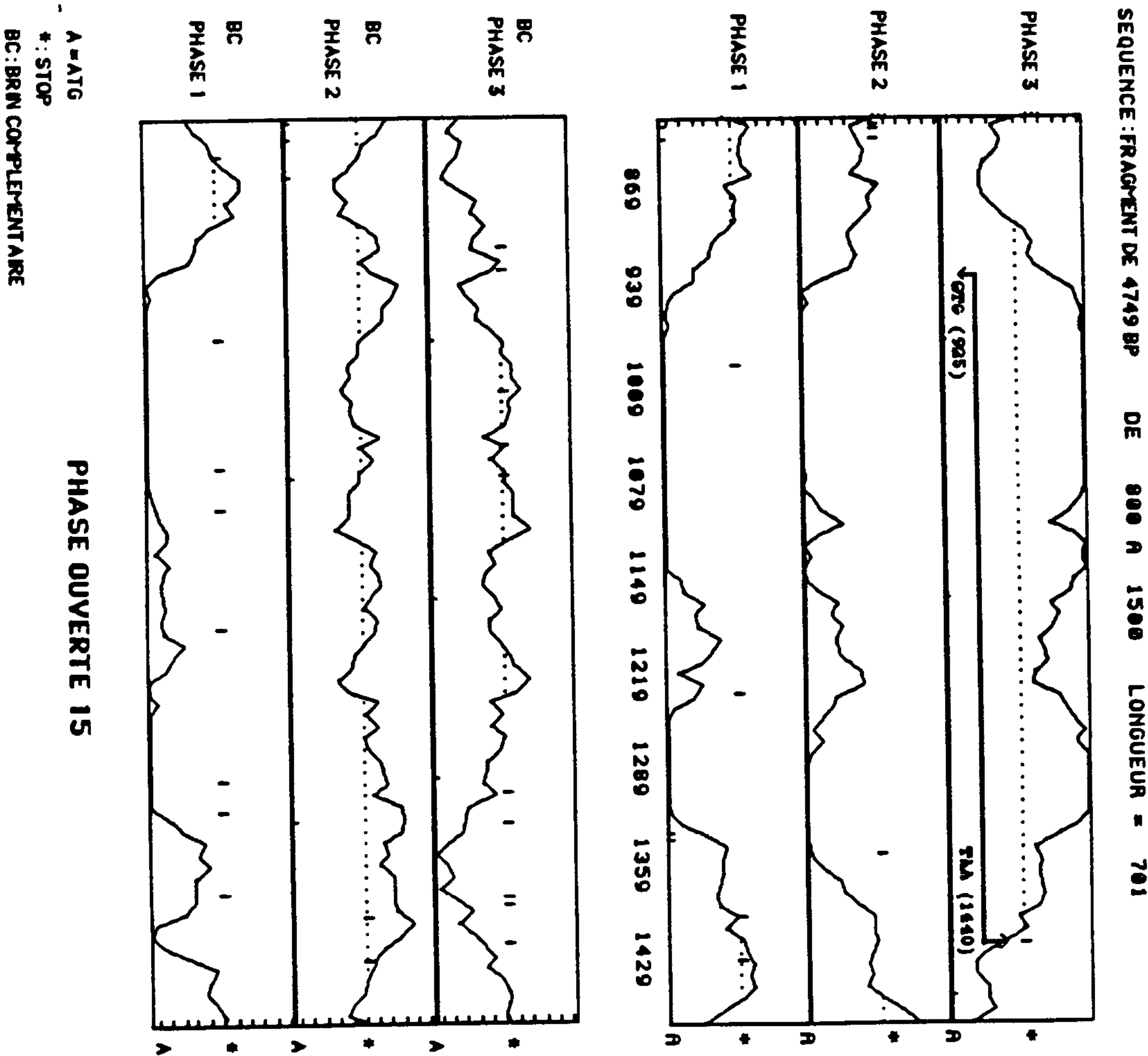
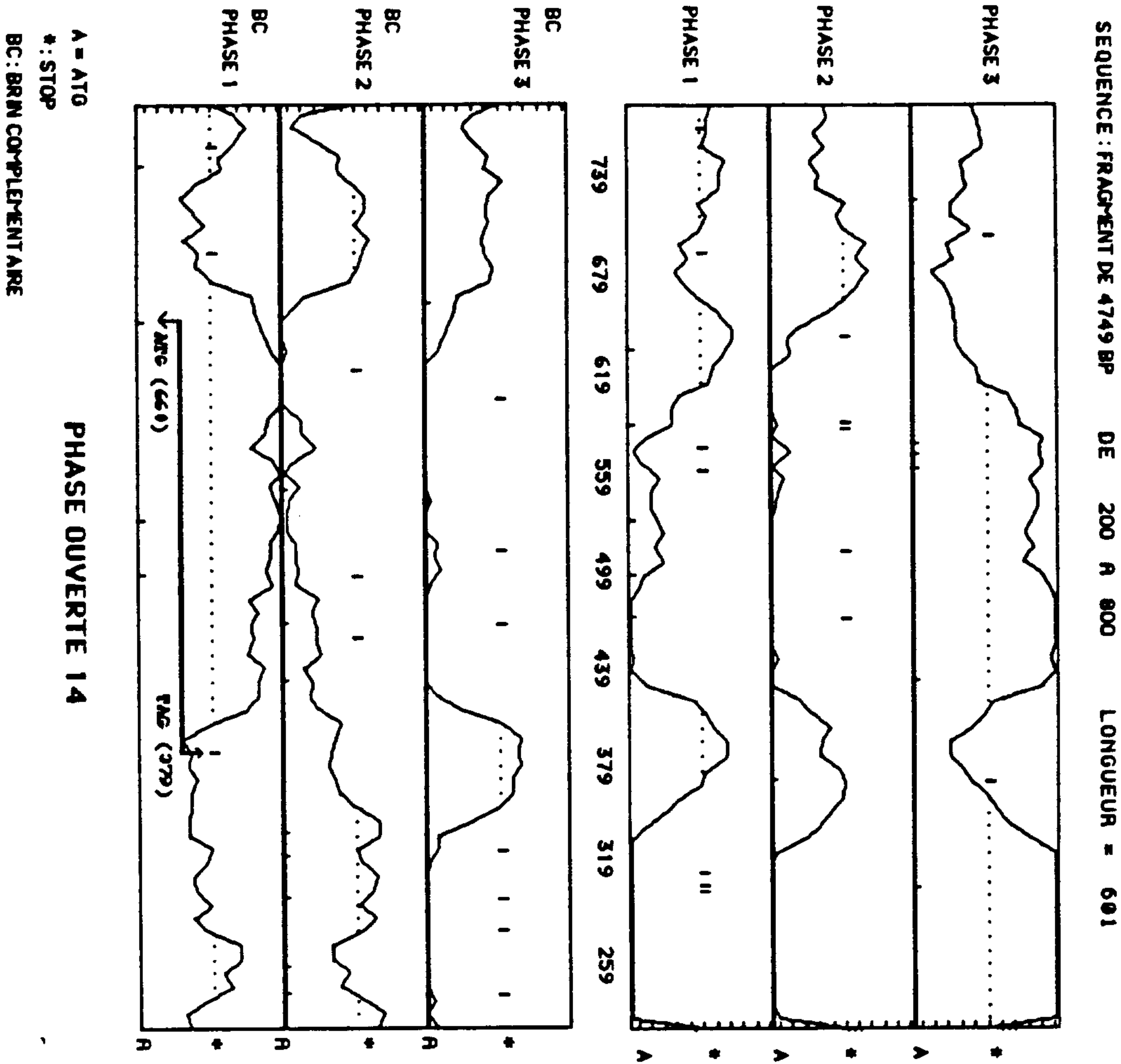


FIGURE 34

92 / 151

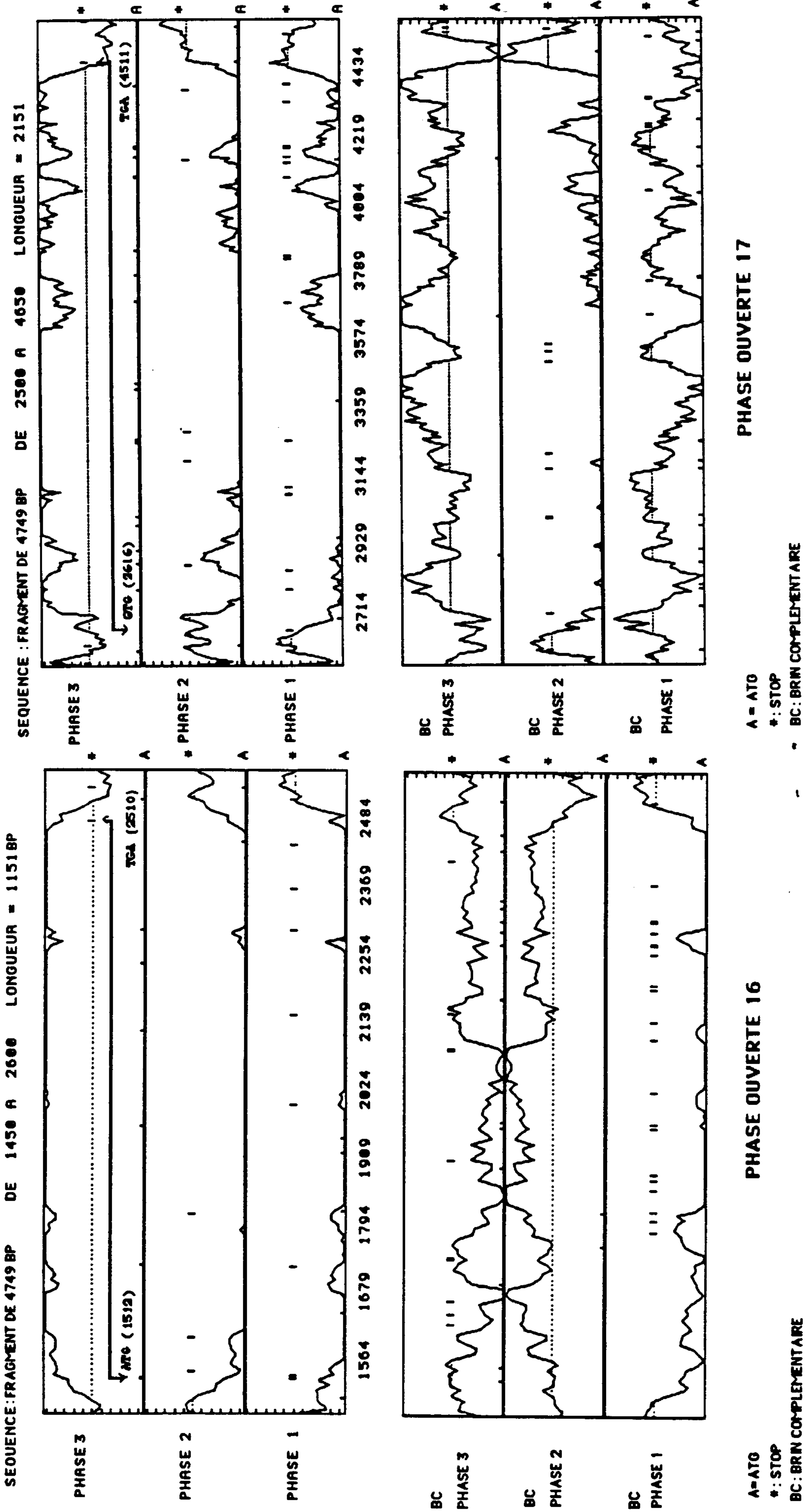


FIGURE 34.1

93 / 151

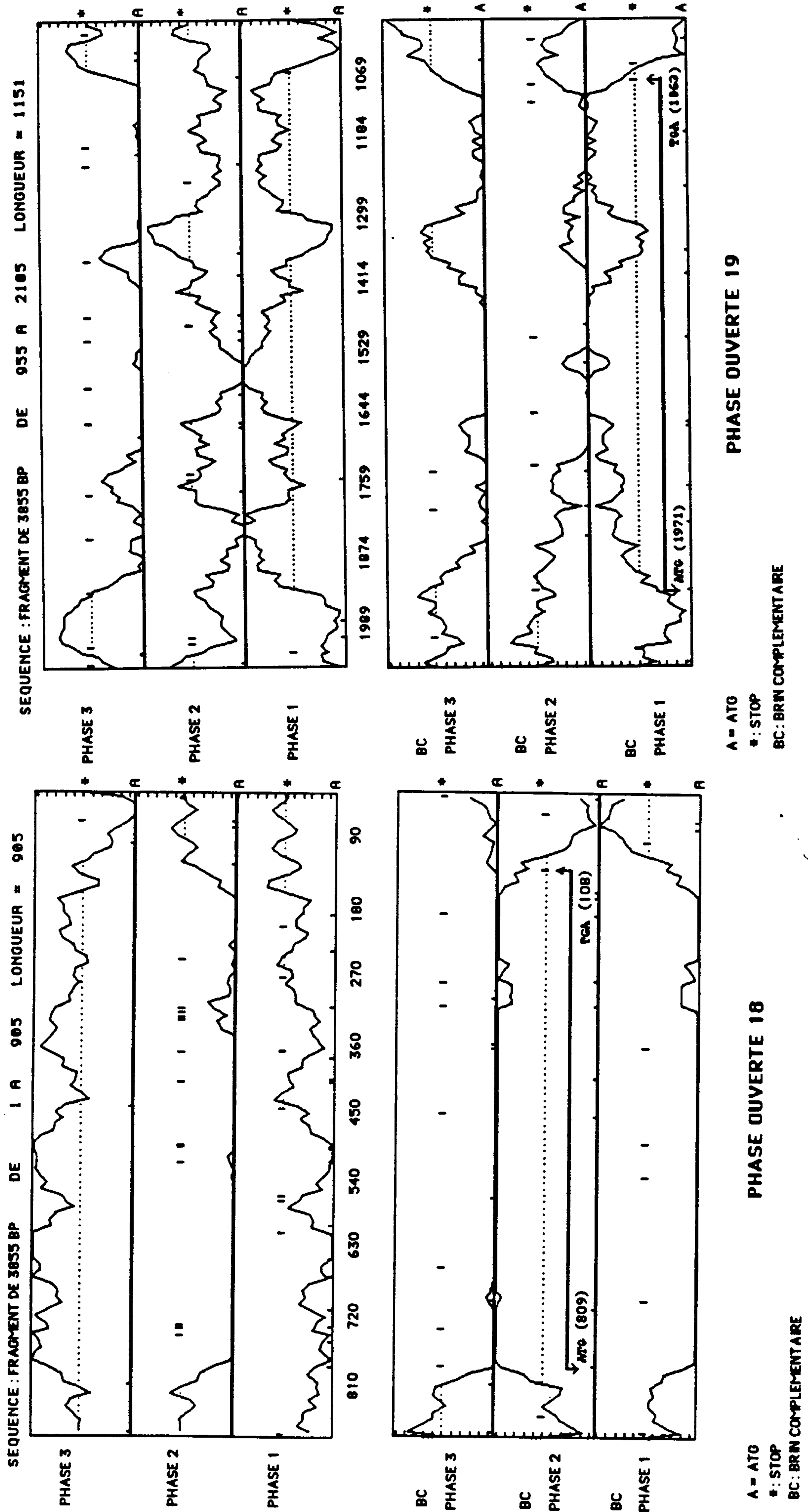


FIGURE 35

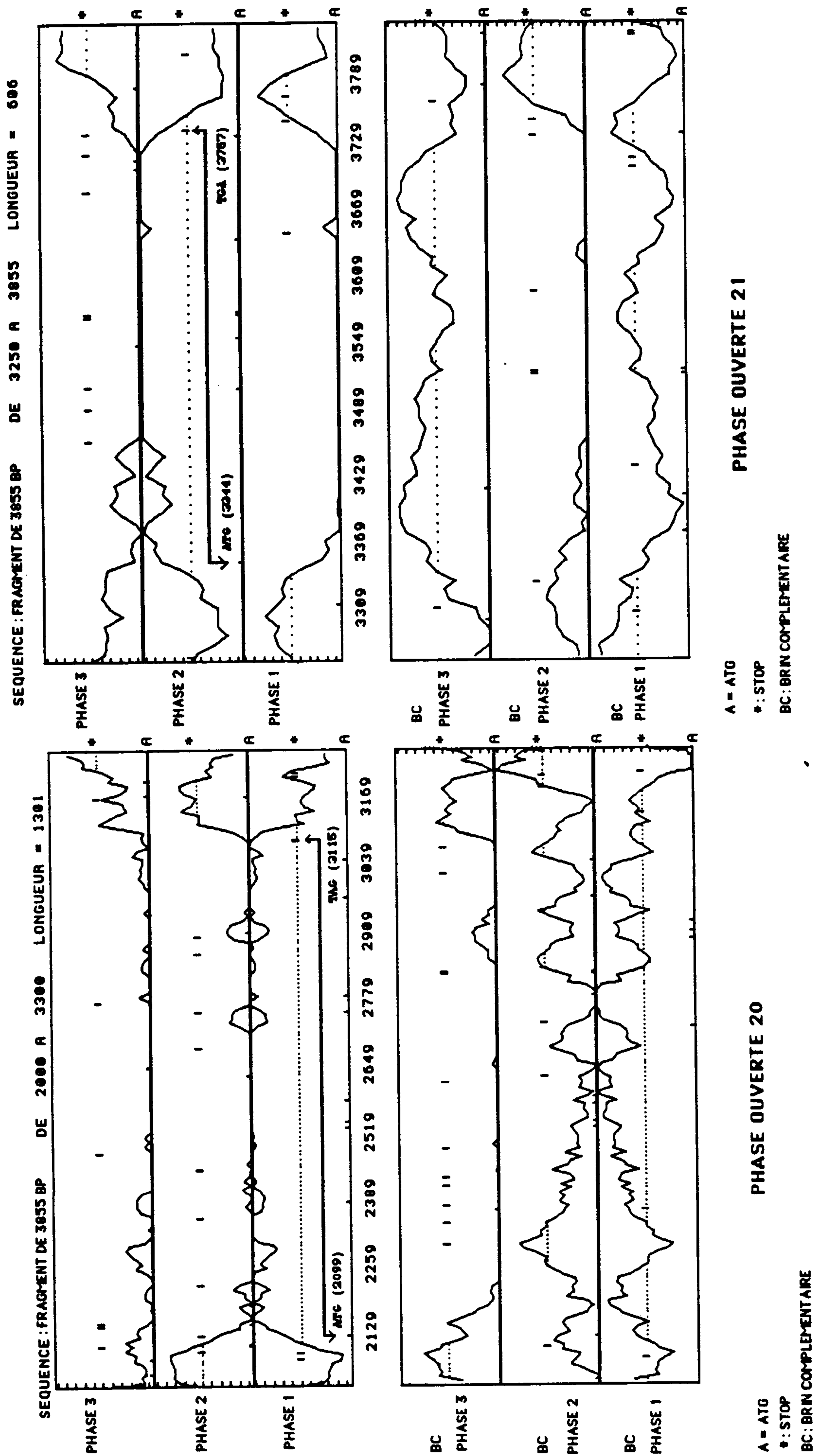


FIGURE 35.1

95 / 151

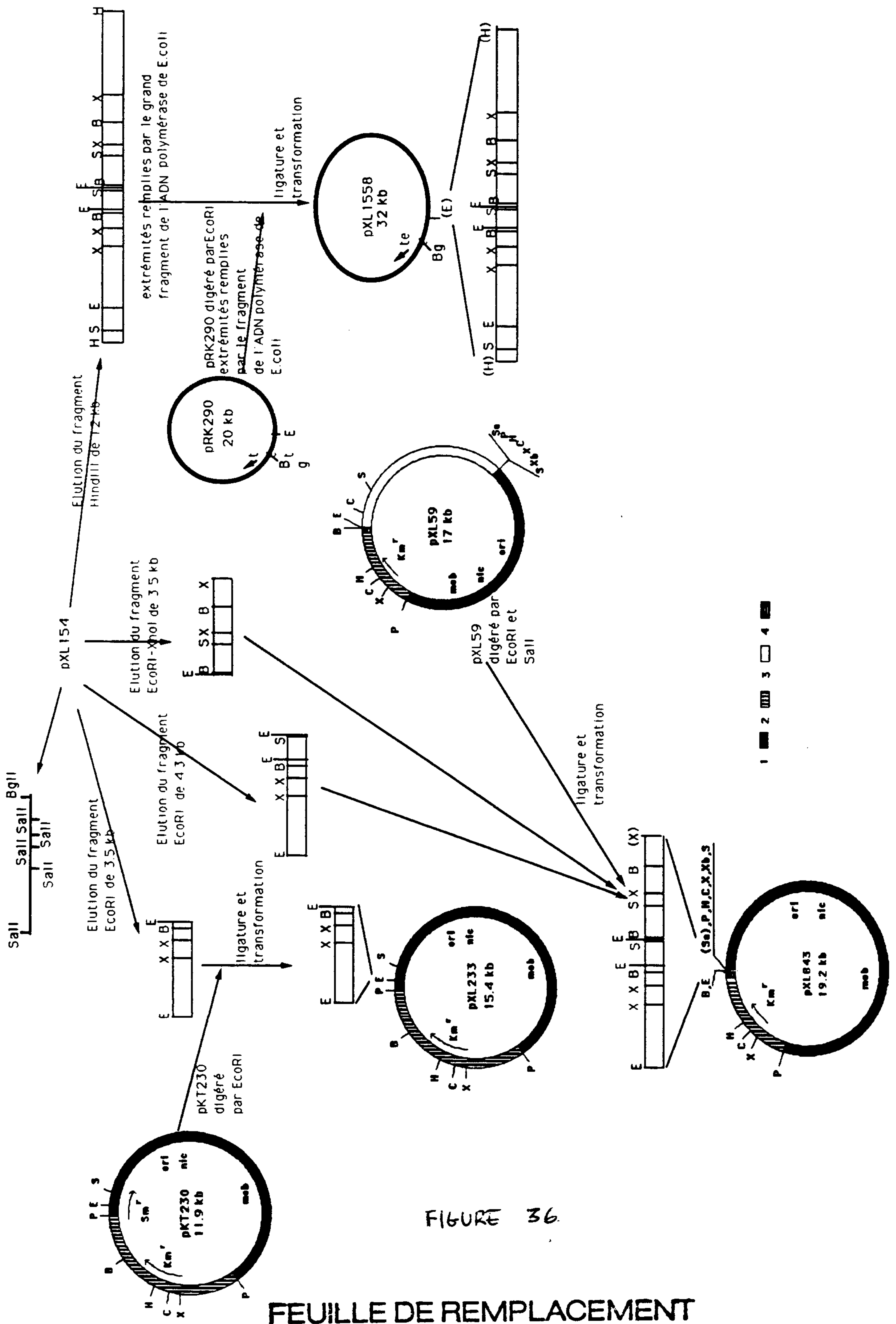


FIGURE 36

FEUILLE DE REMPLACEMENT

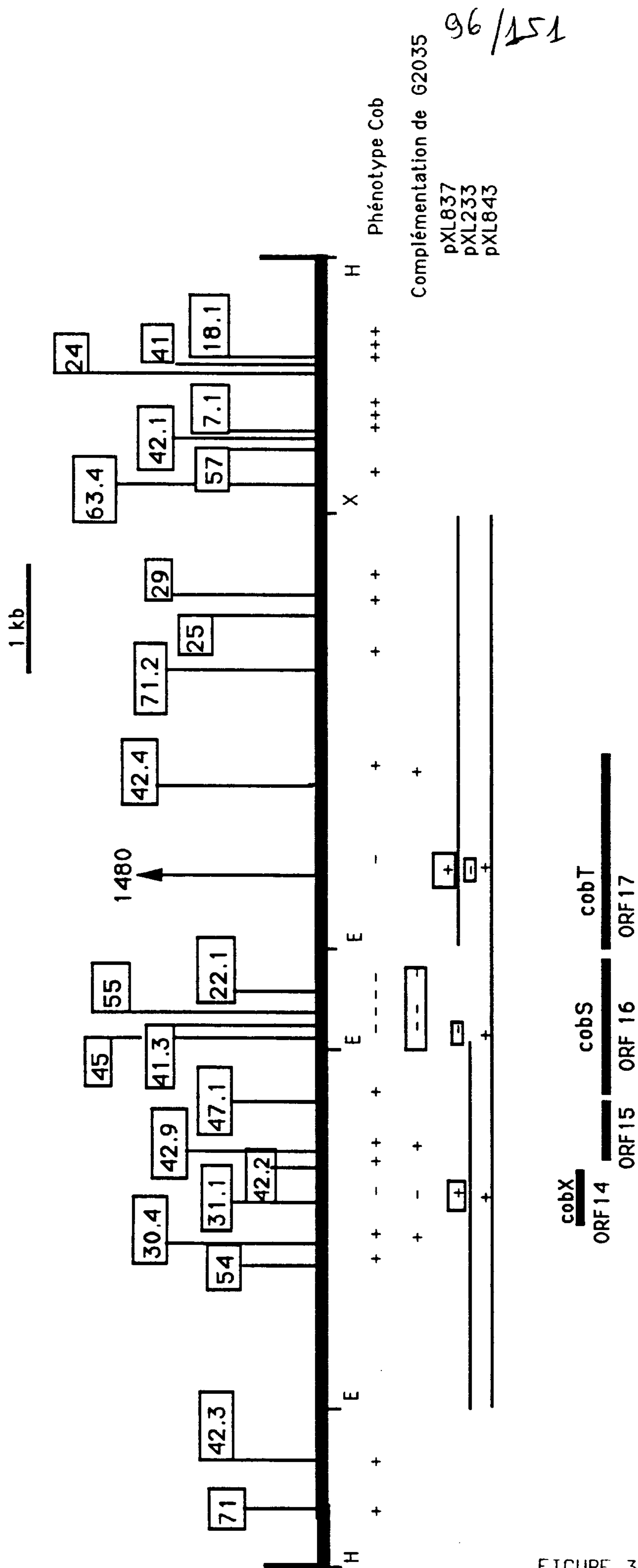


FIGURE 37

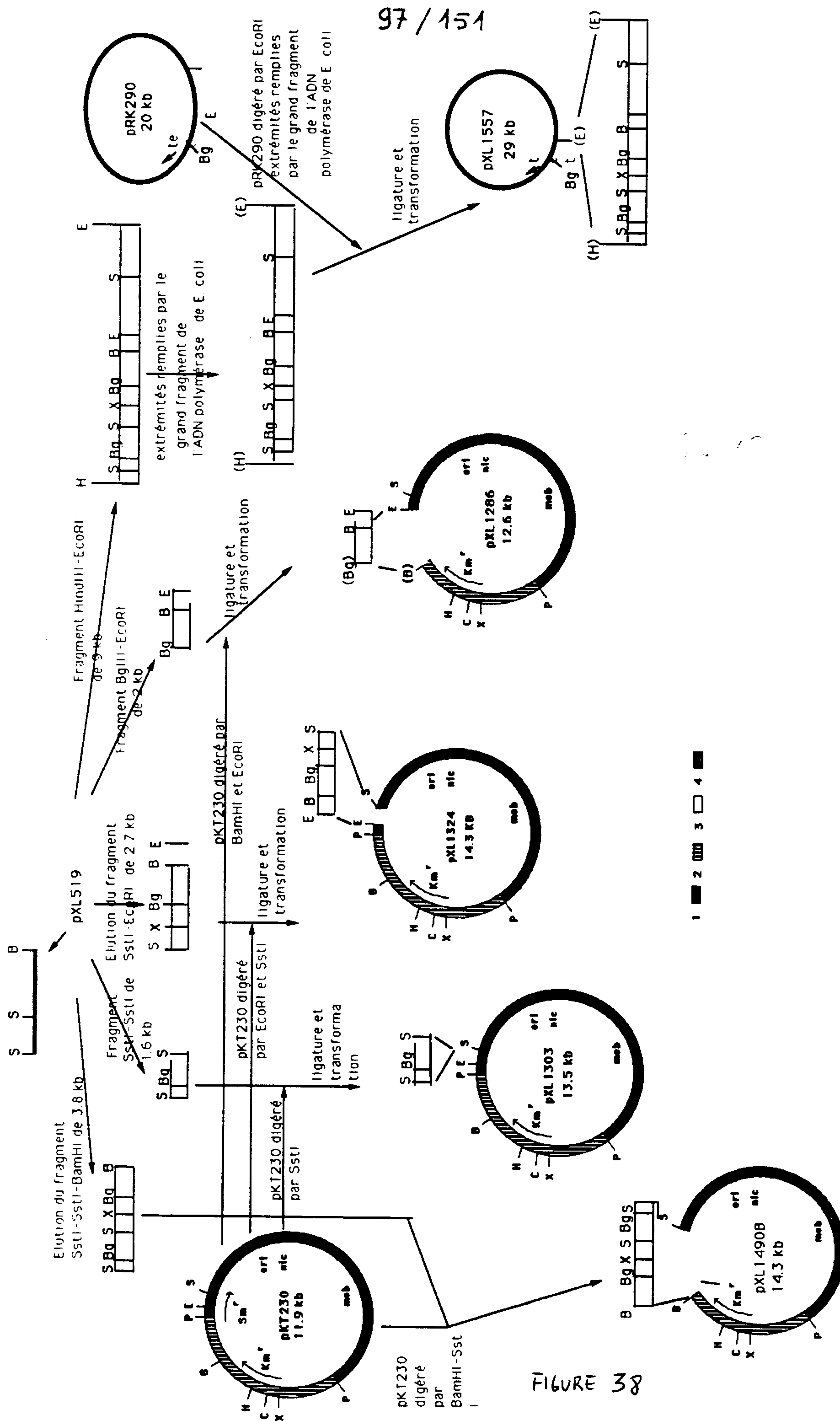
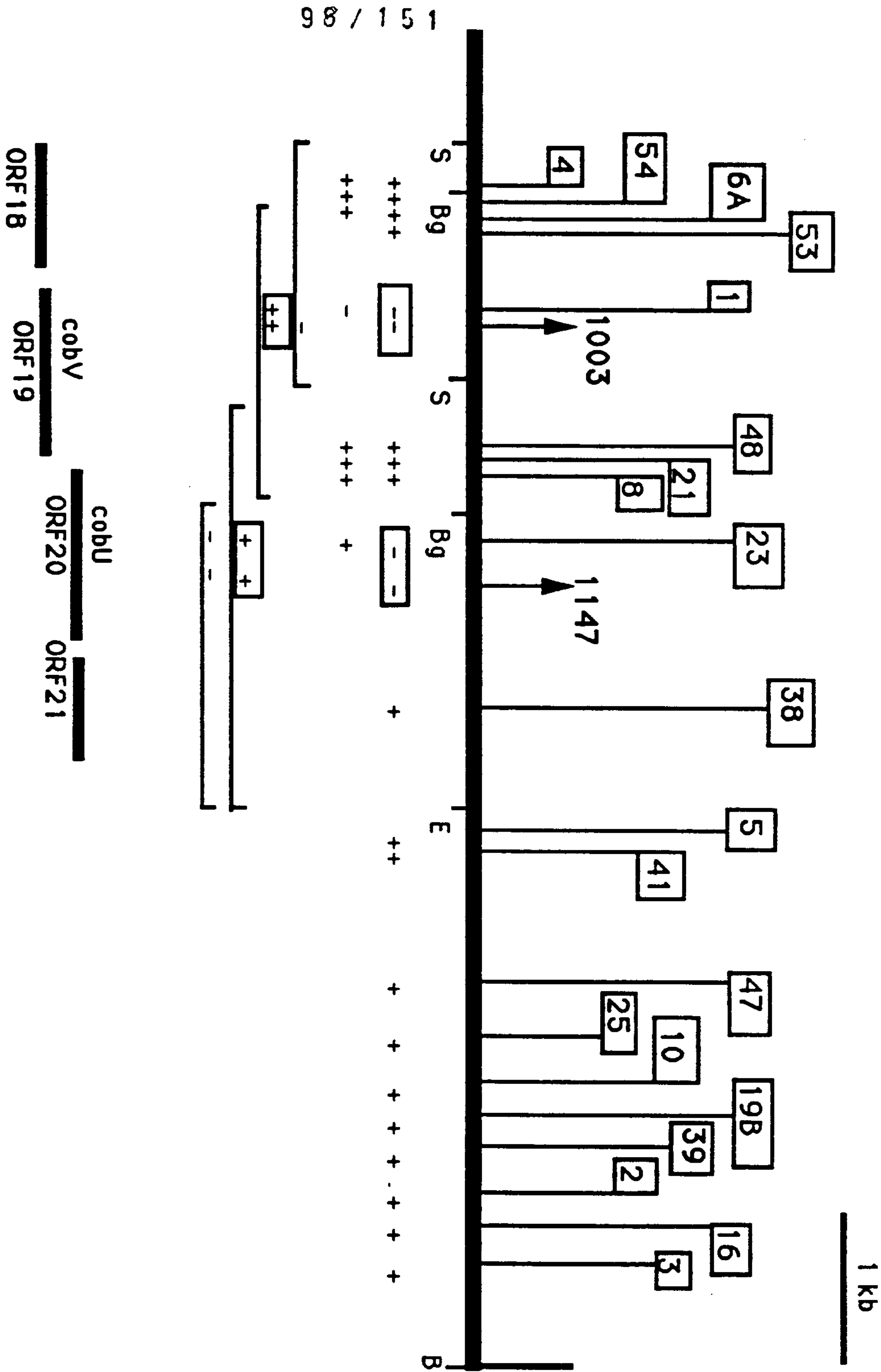


FIGURE 38



Phénotype B12
Complémentation de G2040
pXL1303
pXL699
pXL1324
pXL1286

FIGURE 39

NOM = COBS

PREMIER RESIDU = 1
DERNIER RESIDU = 332

PL. 99/151

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	15	4.52	2206.03	5.97
2	LEU	L	29	8.73	3279.44	8.87
3	ILE	I	20	6.02	2261.68	6.12
4	MET	M	7	2.11	917.28	2.48
5	VAL	V	33	9.94	3269.26	8.85
6	SER	S	15	4.52	1305.48	3.53
7	PRO	P	11	3.31	1067.58	2.89
8	THR	T	24	7.23	2425.14	6.56
9	ALA	A	32	9.64	2273.19	6.15
10	TYR	Y	9	2.71	1467.57	3.97
11	*	*	0	0.00	0.00	0.00
12	HIS	H	10	3.01	1370.59	3.71
13	GLN	Q	10	3.01	1280.59	3.46
14	ASN	N	12	3.61	1368.52	3.70
15	LYS	K	13	3.92	1665.23	4.51
16	ASP	D	28	8.43	3220.75	8.71
17	GLU	E	15	4.52	1935.64	5.24
18	CYS	C	3	0.90	309.03	0.84
19	TRP	W	4	1.20	744.32	2.01
20	ARG	R	22	6.63	3434.22	9.29
21	GLY	G	20	6.02	1140.43	3.09
22	-	-	0	0.00	0.00	0.00

RESIDUS = 332
Masse moleculaire (monoisotopique) = 36960.0000
Masse moleculaire (moyenne) = 36983.1797
INDEX DE POLARITE (%) = 44.88
POINT ISOELECTRIQUE = 6.34
DO 260 (1mg/ml) = 0.611 DO 280 (1mg/ml) = 0.891

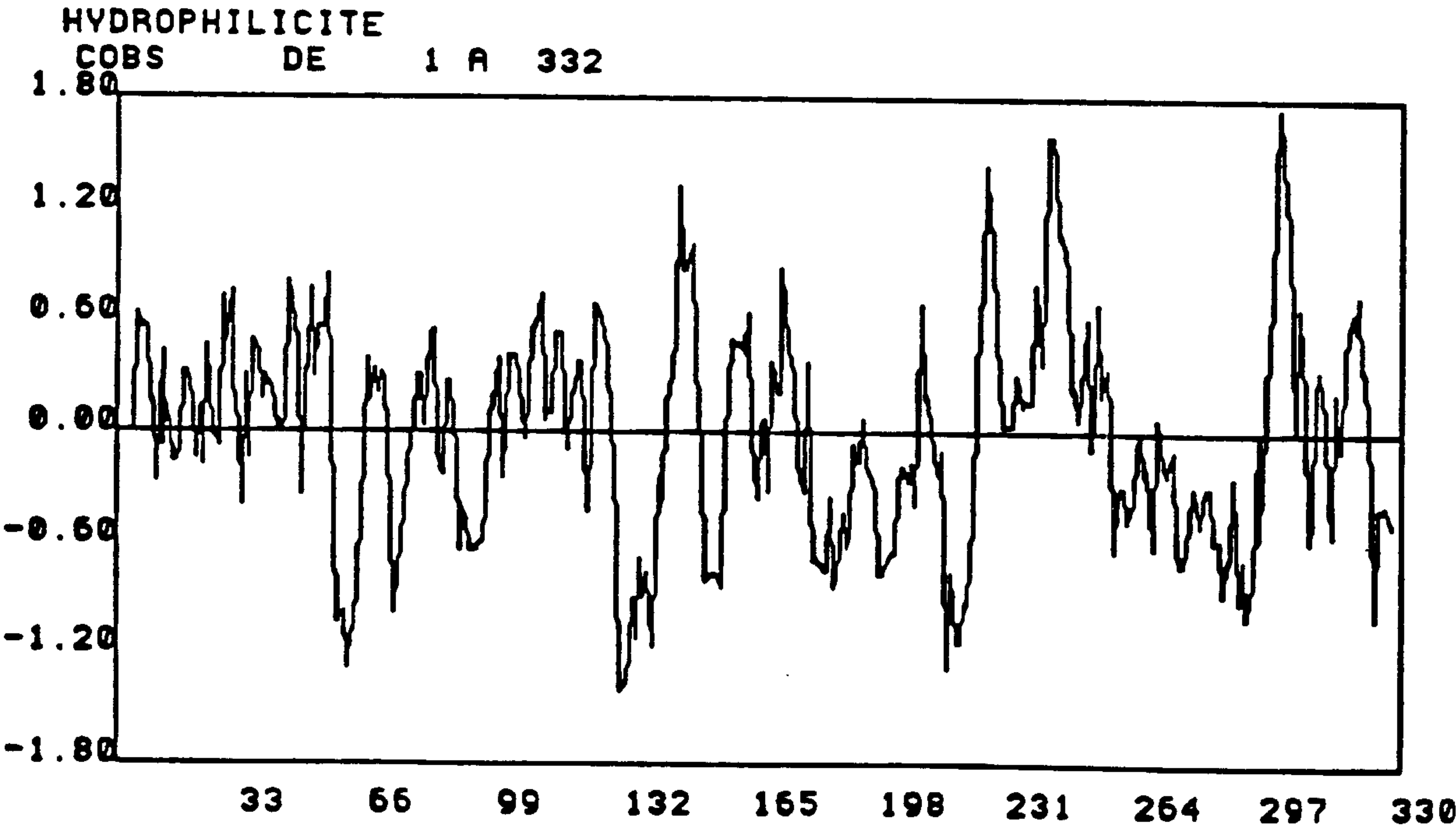


FIGURE 40

100/151

GENE cobS ET PROTEINE COBSSEQUENCE DU FRAGMENT SalI-SalI-SalI-SalI-SalI-BglI DE 4749 BP, DE 1512 A 2510

```

MetMetSerLysIleAspLeuAspIleSerAsnLeuProAspThrThrIleSerValArgGluValPheGlyIle
ATGATGAGCAAGATTGACCTCGACATTTCCAACCTCCCCGACACCACGATTTCCGTCCGGGAGGTTTTTCGGTATT
1521      1531      1541      1551      1561      1571      1581
AspThrAspLeuArgValProAlaTyrSerLysGlyAspAlaTyrValProAspLeuAspProAspTyrLeuPhe
GATACGGATTTGCGCGTTCCTGCCTATTTCGAAGGGCGACGCCTATGTCCCGGATCTGGATCCGGACTACCTCTTC
1596      1606      1616      1626      1636      1646      1656
AspArgGluThrThrLeuAlaIleLeuAlaGlyPheAlaHisAsnArgArgValMetValSerGlyTyrHisGly
GACCGCGAAACGACGCTCGCCATTCTCGCAGGCTTCGCCACAACCGACGCGTGATGGTGTCGGGCTATCACGGC
1671      1681      1691      1701      1711      1721      1731
ThrGlyLysSerThrHisIleGluGlnValAlaAlaArgLeuAsnTrpProCysValArgValAsnLeuAspSer
ACCGGCAAGTCCACCCATATCGAGCAGGTTCGCCGCGCGCCTCAACTGGCCGTGCGTGCGCGTCAACCTCGATAGC
1746      1756      1766      1776      1786      1796      1806
HisValSerArgIleAspLeuValGlyLysAspAlaIleValValLysAspGlyLeuGlnValThrGluPheLys
CATGTCAGCCGTATCGACCTCGTCGGCAAGGACGCGATCGTCGTCAAGGACGGCCTGCAGGTCACCGAATTCAAG
1821      1831      1841      1851      1861      1871      1881
AspGlyIleLeuProTrpAlaTyrGlnHisAsnValAlaLeuValPheAspGluTyrAspAlaGlyArgProAsp
GACGGCATCCTGCCCTGGGCTACCAGCACAATGTTCGCGCTCGTCTTCGACGAATACGATGCCGGCCGCCCCGGAC
1896      1906      1916      1926      1936      1946      1956
ValMetPheValIleGlnArgValLeuGluSerSerGlyArgLeuThrLeuLeuAspGlnSerArgValIleArg
GTCATGTTTCGTCATCCAGCGCGTGCTGGAATCCTCCGGCCGCTGACGCTGCTCGACCAGAGCCGTGTCATCCGT
1971      1981      1991      2001      2011      2021      2031
ProHisProAlaPheArgLeuPheAlaThrAlaAsnThrValGlyLeuGlyAspThrThrGlyLeuTyrHisGly
CCGCACCCGGCCTTCCGCCTGTTTTCGACCGCCAACACCGTCGGCCTCGGCGACACGACCGGCCTCTATCACGGC
2046      2056      2066      2076      2086      2096      2106
ThrGlnGlnIleAsnGlnAlaGlnMetAspArgTrpSerIleValThrThrLeuAsnTyrLeuProHisAspLys
ACGCAGCAGATCAACCAGGCGCAGATGGACCGCTGGTCGATCGTCACCACGCTGAACTACCTGCCGCACGACAAG
2121      2131      2141      2151      2161      2171      2181
GluValAspIleValAlaAlaLysValLysGlyPheThrAlaAspLysGlyArgGluThrValSerLysMetVal
GAAGTCGACATCGTCGCCGCAAGGTCAAGGGCTTCACCGCCGACAAGGGCCGCGAGACCGTCTCCAAGATGGTA
2196      2206      2216      2226      2236      2246      2256
ArgValAlaAspLeuThrArgAlaAlaPheIleAsnGlyAspLeuSerThrValMetSerProArgThrValIle
CGTGTGCGCCGACCTCACGCGCGCAGCCTTCATCAATGGCGATCTCTCGACTGTCATGAGCCCGCGTACGGTCATC
2271      2281      2291      2301      2311      2321      2331
ThrTrpAlaGluAsnAlaHisIlePheGlyAspIleAlaPheAlaPheArgValThrPheLeuAsnLysCysAsp
ACCTGGGCGGAGAACGCCCACATCTTCGGCGACATCGCTTTCGCCTTCCGCGTGACCTTCCTCAACAAGTGCGAC
2346      2356      2366      2376      2386      2396      2406
GluLeuGluArgAlaLeuValAlaGluHisTyrGlnArgAlaPheGlyIleGluLeuLysGluCysAlaAlaAsn
GAGCTGGAGCGGGCGCTGGTCGCCGAGCACTACCAGCGCGCCTTCGGCATCGAGCTGAAGGAATGCGCTGCCAAC
2421      2431      2441      2451      2461      2471      2481
IleValLeuGluAlaThrAla***
ATCGTGCTCGAAGCCACCGCCTGA
2496      2506

```

FIGURE 40.1

PL. 101 / 151

NOM = COBT
PREMIER RESIDU = 1
DERNIER RESIDU = 631

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	16	2.54	2353.09	3.35
2	LEU	L	56	8.87	6332.71	9.01
3	ILE	I	29	4.60	3279.44	4.67
4	MET	M	18	2.85	2358.73	3.36
5	VAL	V	31	4.91	3071.12	4.37
6	SER	S	33	5.23	2872.06	4.09
7	PRO	P	24	3.80	2329.27	3.31
8	THR	T	28	4.44	2829.34	4.03
9	ALA	A	75	11.89	5327.78	7.58
10	TYR	Y	8	1.27	1304.51	1.86
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	1.11	959.41	1.36
13	GLN	Q	29	4.60	3713.70	5.28
14	ASN	N	22	3.49	2508.94	3.57
15	LYS	K	25	3.96	3202.37	4.56
16	ASP	D	49	7.77	5636.32	8.02
17	GLU	E	67	10.62	8645.85	12.30
18	CYS	C	5	0.79	515.05	0.73
19	TRP	W	6	0.95	1116.48	1.59
20	ARG	R	61	9.67	9522.17	13.55
21	GLY	G	42	6.66	2394.90	3.41
22	-	-	0	0.00	0.00	0.00

RESIDUS = 631
Masse moleculaire (monoisotopique) = 70291.3984
Masse moleculaire (moyenne) = 70334.7656
INDEX DE POLARITE (%) = 50.87
POINT ISOELECTRIQUE = 5.10
DO 260 (1mg/ml) = 0.423 DO 280 (1mg/ml) = 0.610

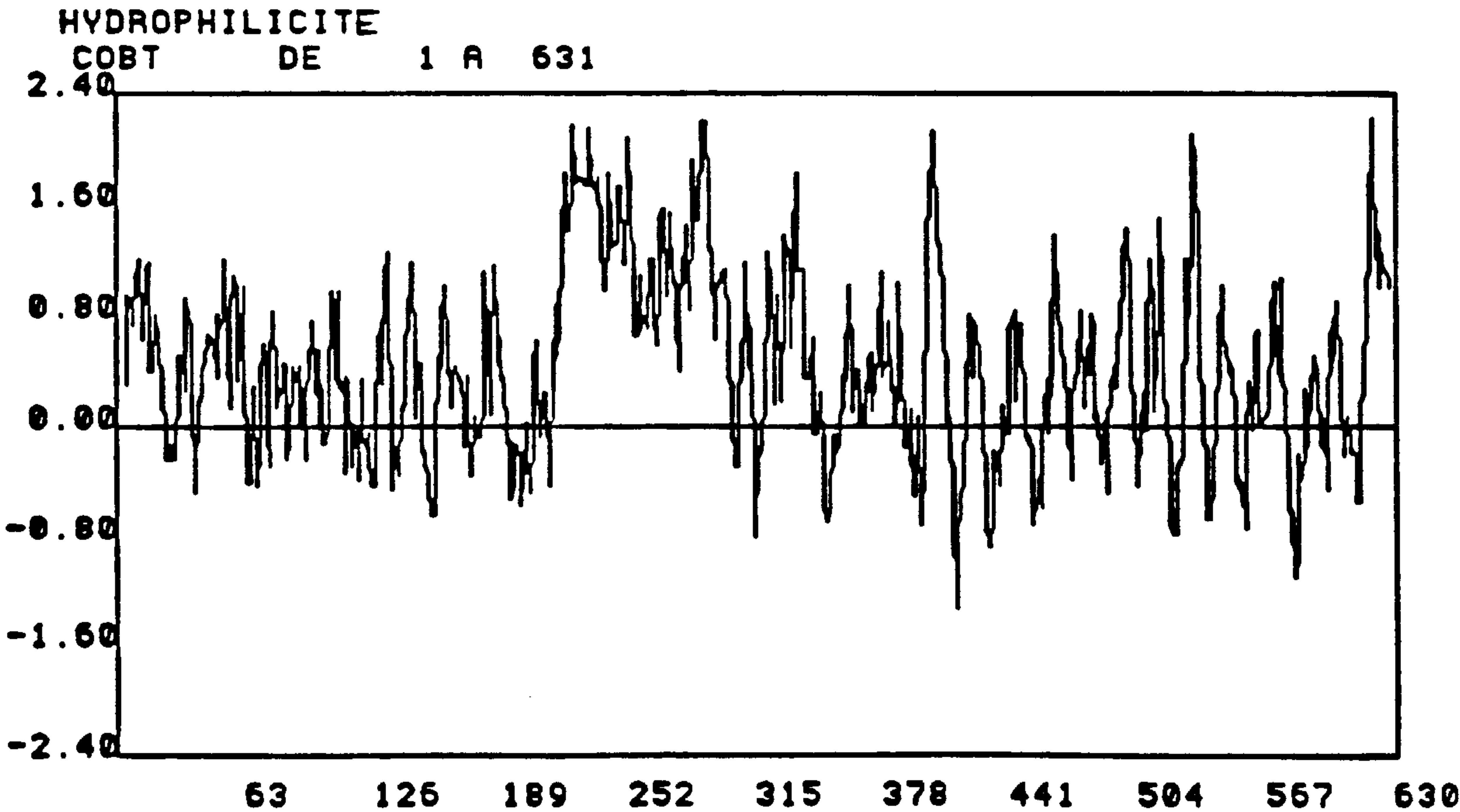


FIGURE 40.2

102 / 151

GENE cobT ET PROTEINE COBTSEQUENCE DU FRAGMENT SalI-SalI-SalI-SalI-SalI-BglI DE 4749 BP, DE 2616 A 4511

```

ValSerSerAsnSerLysAlaLysProThrThrArgGluAsnAlaAlaGluProPheLysArgAlaLeuSerGly
GTGAGCTCGAATTCGAAGGCAAAGCCAACCACGCGCGAGAATGCTGCGGAACCGTTCAAGCGGGCGCTTTCCGGC
  2625      2635      2645      2655      2665      2675      2685
CysIleArgSerIleAlaGlyAspAlaGluValGluValAlaPheAlaAsnGluArgProGlyMetThrGlyGlu
TGCATCCGATCGATCGCGGGCGATGCCGAGGTGGAAGTCGCCTTCGCCAACGAGCGGGCCGGCATGACCGGCGAA
  2700      2710      2720      2730      2740      2750      2760
ArgIleArgLeuProGluLeuSerLysArgProThrLeuGlnGluLeuAlaValThrArgGlyLeuGlyAspSer
CGCATCCGTCTGCCGGAACCTTTCCAAGCGCCCGACCCTGCAGGAAGTTGCCGTGACCCGCGGGCTCGGTGACAGC
  2775      2785      2795      2805      2815      2825      2835
MetAlaLeuArgLysAlaCysThrHisAlaArgIleGlnArgThrMetSerProGlnGlyAlaAspAlaArgAla
ATGGCGCTGCGCAAGGCCTGTACGCATGCGCGGATCCAGCGCACCATGTCGCCGCAAGGGGCGGACGCCCGCGCG
  2850      2860      2870      2880      2890      2900      2910
IlePheAspAlaValGluGlnAlaArgValGluAlaIleGlySerLeuArgMetAlaGlyValAlaLysAsnLeu
ATCTTCGATGCGGTGGAGCAGGCTCGTGTGCGAGGCGATCGGGTCGTTGCGCATGGCGGGTGTGCCCAAGAACCTC
  2925      2935      2945      2955      2965      2975      2985
AsnValMetLeuGluGluLysTyrAlaLysAlaAsnPheAlaThrIleGluArgGlnAlaAspAlaProLeuGly
AACGTCATGCTCGAAGAGAAATACGCCAAGGCGAATTTTCGCAACGATCGAGCGCCAGGCGGACGCGCCGCTCGGC
  3000      3010      3020      3030      3040      3050      3060
GluAlaValAlaLeuLeuValArgGluLysLeuThrGlyGlnLysProProAlaSerAlaGlyLysValLeuAsp
GAGGCCGTAGCGCTGCTGGTGC GCGAGAAGCTGACGGGCCAGAAGCCGCCGGCGTCTGCCGCAAGGTGCTCGAC
  3075      3085      3095      3105      3115      3125      3135
LeuTrpArgGluPheIleGluGlyLysAlaAlaGlyAspIleGluHisLeuSerSerThrIleAsnAsnGlnGln
CTCTGGCGCGAGTTTCATCGAGGGCAAGGCTGCCGGCGACATTGAGCACCTGTCGTCGACGATCAACAACAGCAG
  3150      3160      3170      3180      3190      3200      3210
AlaPheAlaArgValValArgAspMetLeuThrSerMetGluValAlaGluLysTyrGlyAspAspAspAsnGlu
GCCTTTGCCCGGGTCGTTTCGCGACATGCTGACCTCGATGGAAGTCGCCGAGAAATACGGTGACGACGACAACGAG
  3225      3235      3245      3255      3265      3275      3285
ProAspGluGlnGluSerGluThrAspGluAspGlnProArgSerGlnGluGlnAspGluAsnAlaSerAspGlu
CCGGACGAGCAGGAAAGCGAGACCGACGAAGACCAGCCGCGCAGCCAGGAGCAGGACGAGAACGCCAGCGACGAG
  3300      3310      3320      3330      3340      3350      3360
GluAlaGlyAspAspAlaAlaProAlaAspGluAsnGlnAlaAlaGluGluGlnMetGluGluGlyGluMetAsp
GAAGCCGGCGACGATGCCGCACCCGCCGACGAGAACCAGGCTGCCGAAGAGCAGATGGAAGAAGGCGAGATGGAC
  3375      3385      3395      3405      3415      3425      3435
GlyAlaGluIleSerAspAspAspLeuGlnAspGluGlyAspGluAspSerGluThrProGlyGluValLysArg
GGCGCGGAGATCTCCGACGACGATCTCCAGGACGAAGGCGACGAGGACAGCGAAACGCCCGGCGAGGTCAAGCGT
  3450      3460      3470      3480      3490      3500      3510
ProAsnGlnProPheAlaAspPheAsnGluLysValAspTyrAlaValPheThrArgGluPheAspGluThrIle
CCGAACCAGCCCTTCGCCGACTTCAACGAGAAGGTCGACTACGCCGTCTTCACCCGCGAGTTCGACGAGACGATT
  3525      3535      3545      3555      3565      3575      3585
AlaSerGluGluLeuCysAspGluAlaGluLeuAspArgLeuArgAlaPheLeuAspLysGlnLeuAlaHisLeu
GCCTCGGAAGAGCTTTGCGACGAGGCCGAGCTCGACCGGCTGCGCGCCTTCCTCGACAAGCAGCTTGCCCATCTT
  3600      3610      3620      3630      3640      3650      3660
GlnGlyAlaValGlyArgLeuAlaAsnArgLeuGlnArgArgLeuMetAlaGlnGlnAsnArgSerTrpGluPhe
CAAGGCGCGGTGCGCCGCTTGCCAACCGGCTGCAGCGCCGCTGATGGCGCAGCAGAACCGCTCCTGGGAGTTC
  3675      3685      3695      3705      3715      3725      3735
AspLeuGluGluGlyTyrLeuAspSerAlaArgLeuGlnArgIleIleIleAspProMetGlnProLeuSerPhe
GATCTCGAAGAGGGGTATCTCGATTGCGCGCGGCTTCAGCGCATCATCATCGATCCGATGCAGCCGCTTTCTCTC
  3750      3760      3770      3780      3790      3800      3810
LysArgGluLysAspThrAsnPheArgAspThrValValThrLeuLeuIleAspAsnSerGlySerMetArgGly
AAGCGCGAAAAGGACACCAACTTCCGCGATACCGTCGTGACGCTGCTGATCGACAATTCCGGCTCGATGCGCGGC
  3825      3835      3845      3855      3865      3875      3885
ArgProIleThrValAlaAlaThrCysAlaAspIleLeuAlaArgThrLeuGluArgCysGlyValLysValGlu
CGTCCGATCACGGTTGCCGCCACCTGCGCCGATATCCTGGCGCGCACGCTCGAGCGCTGCGGCGTCAAGGTGAG
  3900      3910      3920      3930      3940      3950      3960
IleLeuGlyPheThrThrLysAlaTrpLysGlyGlyGlnSerArgGluLysTrpLeuAlaGlyGlyLysProGln
ATCCTCGGTTTTTACCACCAAGGCGTGGAAGGGTGGGCAGTCACGCGAGAAGTGGCTGGCCGGCGGCAAGCCACAG
  3975      3985      3995      4005      4015      4025      4035

```

FIGURE 40.3

103/151

AlaProGlyArgLeuAsnAspLeuArgHisIleValTyrLysSerAlaAspAlaProTrpArgArgAlaArgArg
GCCCCGGGTCGCCTCAACGACCTGCGACACATCGTCTACAAGTCTGCCGACGCTCCGTGGCGCCGGGCACGACGC
4050 4060 4070 4080 4090 4100 4110
AsnLeuGlyLeuMetMetArgGluGlyLeuLeuLysGluAsnIleAspGlyGluAlaLeuIleTrpAlaHisGlu
AATCTCGGCCTGATGATGCGGGAAGGCCTGCTCAAGGAAAACATCGACGGCGAGGCGTTGATTTGGGCGCATGAG
4125 4135 4145 4155 4165 4175 4185
ArgLeuMetAlaArgArgGluGlnArgArgIleLeuMetMetIleSerAspGlyAlaProValAspAspSerThr
CGGCTGATGGCGCGGCGCGAACAGCGGCGCATCCTGATGATGATTTTCGGACGGCGCGCCGGTCGACGACTCGACG
4200 4210 4220 4230 4240 4250 4260
LeuSerValAsnProGlyAsnTyrLeuGluArgHisLeuArgAlaValIleGluGlnIleGluThrArgSerPro
CTGTCCGTCAATCCAGGAAACTATCTGGAGCGTCACCTGCGCGCGGTCATCGAGCAGATCGAAACGCGCTCGCCG
4275 4285 4295 4305 4315 4325 4335
ValGluLeuLeuAlaIleGlyIleGlyHisAspValThrArgTyrTyrArgArgAlaValThrIleValAspAla
GTGGAAGTCTGGCGATCGGTATCGGCCACGACGTGACGCGCTACTATCGCCGTGCCGTACCATCGTCGATGCC
4350 4360 4370 4380 4390 4400 4410
AspGluLeuAlaGlyAlaMetThrGluGlnLeuAlaAlaLeuPheGluAspGluSerGlnArgArgGlySerSer
GATGAGCTTGCCGGCGCGATGACCGAACAGCTGGCCGCACTCTTCGAGGACGAAAGCCAGCGCCGCGGTTCTTCG
4425 4435 4445 4455 4465 4475 4485
ArgLeuArgArgAlaGly***
CGTCTTCGCCGCGCCGGGTGA
4500 4510

FIGURE 40.4

PL. 104/151

NOM = COBX
PREMIER RESIDU = 1
DERNIER RESIDU = 93

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	3	3.23	441.21	4.29
2	LEU	L	8	8.60	904.67	8.80
3	ILE	I	5	5.38	565.42	5.50
4	MET	M	3	3.23	393.12	3.82
5	VAL	V	5	5.38	495.34	4.82
6	SER	S	6	6.45	522.19	5.08
7	PRO	P	6	6.45	582.32	5.66
8	THR	T	5	5.38	505.24	4.92
9	ALA	A	10	10.75	710.37	6.91
10	TYR	Y	0	0.00	0.00	0.00
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	7.53	959.41	9.33
13	GLN	Q	2	2.15	256.12	2.49
14	ASN	N	2	2.15	228.09	2.22
15	LYS	K	3	3.23	384.28	3.74
16	ASP	D	3	3.23	345.08	3.36
17	GLU	E	10	10.75	1290.43	12.55
18	CYS	C	0	0.00	0.00	0.00
19	TRP	W	1	1.08	186.08	1.81
20	ARG	R	7	7.53	1092.71	10.63
21	GLY	G	7	7.53	399.15	3.88
22	-	-	0	0.00	0.00	0.00

RESIDUS = 93
Masse moleculaire (monoisotopique) = 10279.2354
Masse moleculaire (moyenne) = 10285.6309
INDEX DE POLARITE (%) = 48.39
POINT ISOELECTRIQUE = 6.94
DO 260 (1mg/ml) = 0.411 DO 280 (1mg/ml) = 0.541

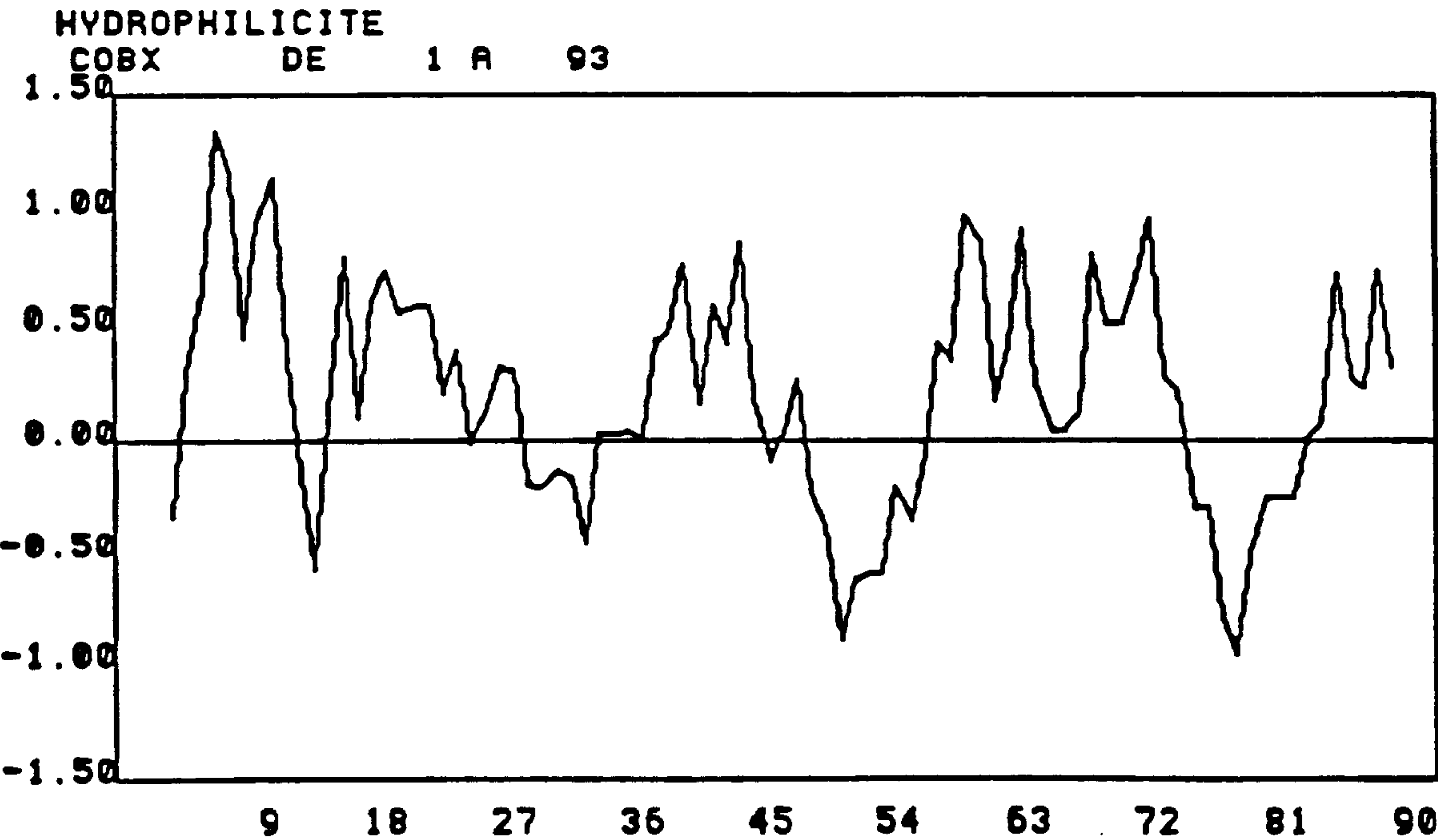


FIGURE 40.5

105/151

GENE cobX ET PROTEINE COBX
SEQUENCE DU FRAGMENT BglI-SalI-SalI-SalI-SalI-SalI DE 4749 BP, DE 4089 A 4370

MetSerLeuThrGluThrIleGluLysLysLeuIleGluAlaPheHisProGluArgLeuGluValIleAsnGlu
ATGTCGCTCACCGAGACCATCGAAAAGAAGCTGATCGAGGCCTTCCACCCTGAACGGCTCGAGGTCATCAACGAG
4098 4108 4118 4128 4138 4148 4158
SerHisGlnHisThrGlyHisGlnProGlyPheAspGlyThrGlyGluSerHisMetArgValArgIleValSer
AGCCATCAGCATACCGGCCATCAGCCGGGCTTCGATGGTACCGGCGAGTCCCACATGCGGGTGCGTATCGTTTCT
4173 4183 4193 4203 4213 4223 4233
SerAlaPheAlaGlyMetSerArgValAlaArgHisArgAlaIleAsnAspLeuLeuLysProGluLeuAspAla
AGCGCCTTTGCCGGCATGAGCCGTGTCGCCCCGCCACCGCGCCATCAATGATCTCCTGAAGCCAGAACTCGACGCC
4248 4258 4268 4278 4288 4398 4308
GlyLeuHisAlaLeuAlaValGluProAlaAlaProGlyGluProThrArgTrp**
GGCCTGCATGCGCTCGCCGTCGAGCCGGCAGCCCCGGCGAGCCGACCCGCTGGTAG
4323 4333 4343 4353 4363

FIGURE 40.6

PL. 106/151

NOM = COBU
PREMIER RESIDU = 1
DERNIER RESIDU = 338

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	9	2.66	1323.62	3.82
2	LEU	L	36	10.65	4071.03	11.75
3	ILE	I	19	5.62	2148.60	6.20
4	MET	M	9	2.66	1179.36	3.40
5	VAL	V	21	6.21	2080.44	6.00
6	SER	S	13	3.85	1131.42	3.26
7	PRO	P	16	4.73	1552.84	4.48
8	THR	T	17	5.03	1717.81	4.96
9	ALA	A	61	18.05	4333.26	12.50
10	TYR	Y	3	0.89	489.19	1.41
11	*	*	0	0.00	0.00	0.00
12	HIS	H	8	2.37	1096.47	3.16
13	GLN	Q	6	1.78	768.35	2.22
14	ASN	N	9	2.66	1026.39	2.96
15	LYS	K	12	3.55	1537.14	4.43
16	ASP	D	13	3.85	1495.35	4.31
17	GLU	E	22	6.51	2838.94	8.19
18	CYS	C	5	1.48	515.05	1.49
19	TRP	W	3	0.89	558.24	1.61
20	ARG	R	16	4.73	2497.62	7.21
21	GLY	G	40	11.83	2280.86	6.58
22	-	-	0	0.00	0.00	0.00

RESIDUS = 338
Masse moleculaire (monoisotopique) = 34659.9844
Masse moleculaire (moyenne) = 34681.9609
INDEX DE POLARITE (%) = 34.32
POINT ISOELECTRIQUE = 6.21
DO 260 (1mg/ml) = 0.416 DO 280 (1mg/ml) = 0.584

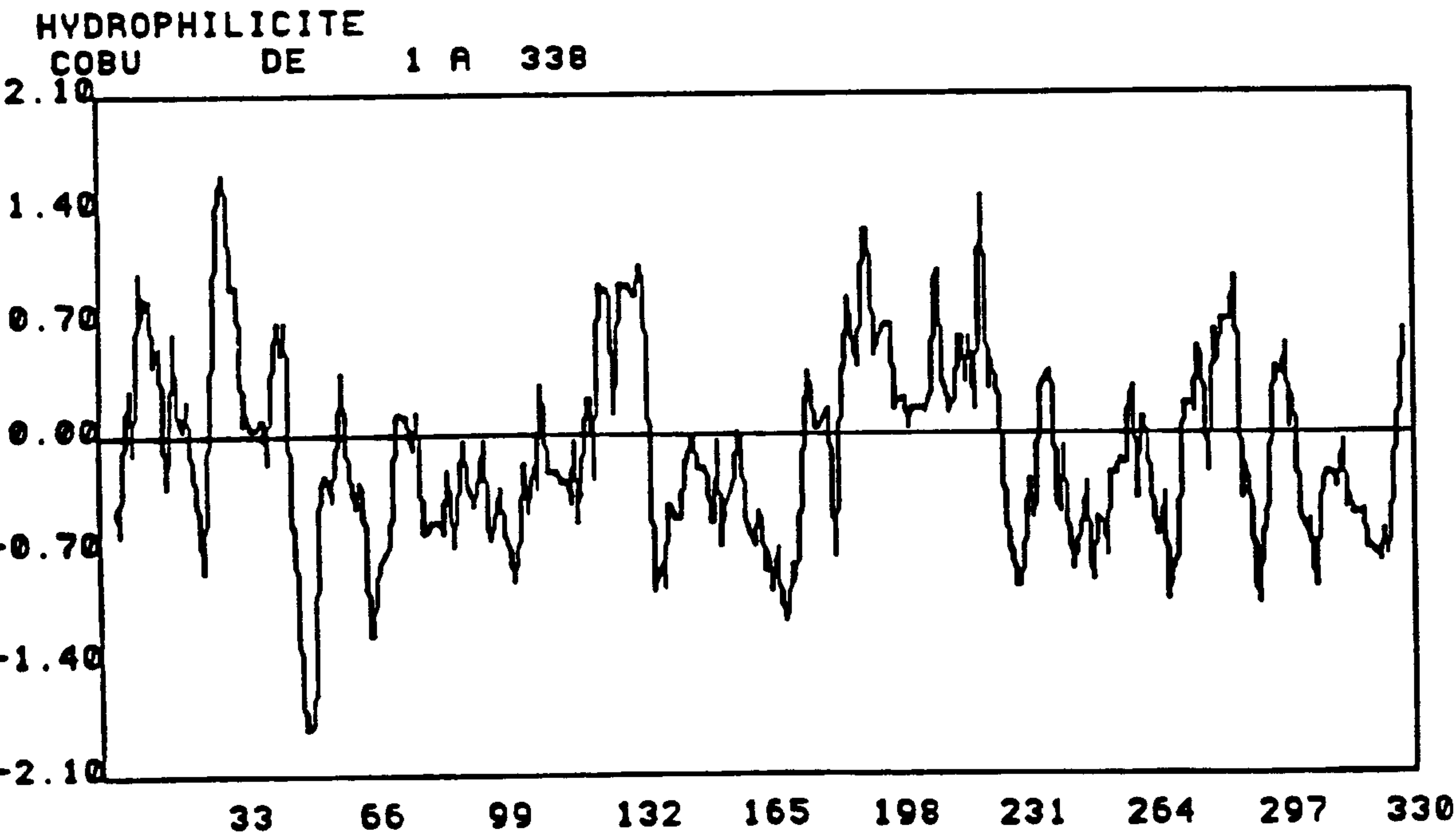


FIGURE 41

107/151

GENE cobU ET PROTEINE COBUSEQUENCE DU FRAGMENT SstI-SstI-BamHI DE 3855 BP, DE 2099 A 3115

```

MetSerAlaSerGlyLeuProPheAspAspPheArgGluLeuLeuArgAsnLeuProGlyProAspAlaAlaAla
ATGAGTGCCAGCGGCCTGCCGTTTGATGATTTTCGCGAATTGTTGCGCAACCTGCCGGGCCCCGGATGCGGCAGCC
    2108      2118      2128      2138      2148      2158      2168
LeuValAlaAlaArgGluArgAspAlaGlnLeuThrLysProProGlyAlaLeuGlyArgLeuGluGluIleAla
CTCGTTGCCGCGCGGGAGCGGGACGCCCAGCTGACGAAGCCGCGGGCGCGCTCGGCCGCTCGAGGAAATCGCC
    2183      2193      2203      2213      2223      2233      2243
PheTrpLeuAlaAlaTrpThrGlyLysAlaProValValAsnArgProLeuValAlaIlePheAlaGlyAsnHis
TTCTGGCTCGCCGCCTGGACGGGCAAGGCGCCGGTGGTCAACCGGCCGCTGGTGGCGATCTTTGCCGGCAACCAC
    2258      2268      2278      2288      2298      2308      2318
GlyValThrArgGlnGlyValThrProPheProSerSerValThrAlaGlnMetValGluAsnPheAlaAlaGly
GGCGTCACCCGCCAGGGGGTGACCCCGTTCCCGTCATCCGTCACCGCACAGATGGTTCGAGAATTTTGCCGCCGGT
    2333      2343      2353      2363      2373      2383      2393
GlyAlaAlaIleAsnGlnIleCysValSerHisAspLeuGlyLeuLysValPheAspLeuAlaLeuGluTyrPro
GGCGCTGCGATCAACCAGATCTGCGTCAGCCACGACCTCGGGCTGAAGGTCTTCGACCTCGCACTCGAATACCCG
    2408      2418      2428      2438      2448      2458      2468
ThrGlyAspIleThrGluGluAlaAlaLeuSerGluArgAspCysAlaAlaThrMetAlaPheGlyMetGluAla
ACCGGTGATATCACCGAGGAAGCCGCGCTGTCCGAGCGCGATTGCGCCGCGACCATGGCCTTTGGCATGGAGGCG
    2483      2493      2503      2513      2523      2533      2543
IleAlaGlyGlyThrAspLeuLeuCysIleGlyGluMetGlyIleGlyAsnThrThrIleAlaAlaAlaIleAsn
ATTGCCGGCGGCACGGATCTTCTGTGCATCGGCGAAATGGGCATCGGCAACACCACGATCGCGGCCGCGATCAAT
    2558      2568      2578      2588      2598      2608      2618
LeuGlyLeuTyrGlyGlyThrAlaGluGluTrpValGlyProGlyThrGlySerGluGlyGluValLeuLysArg
CTCGGCCTTTATGGTGGCACGGCCGAAGAATGGGTCTGGTCCGGGTACCGGCTCCGAGGGCGAGGTGCTGAAGCGC
    2633      2643      2653      2663      2673      2683      2693
LysIleAlaAlaValGluLysAlaValAlaLeuHisArgAspHisLeuSerAspProLeuGluLeuMetArgArg
AAGATCGCCGCGGTGCGAAAAGGCCGTGGCGCTGCATCGCGATCACCTGTCCGATCCGCTCGAACTGATGCGTCGC
    2708      2718      2728      2738      2748      2758      2768
LeuGlyGlyArgGluIleAlaAlaMetAlaGlyAlaIleLeuAlaAlaArgValGlnLysValProValIleIle
CTCGGCGGTCTGTGAGATCGCGGCCATGGCTGGCGCCATCCTGGCCGCCCCGCGTCCAGAAGGTACCTGTCATCATC
    2783      2793      2803      2813      2823      2833      2843
AspGlyTyrValAlaThrAlaAlaAlaSerIleLeuLysAlaAlaAsnProSerAlaLeuAspHisCysLeuIle
GACGGCTACGTGGCGACCGCTGCGGCTTCGATCCTGAAGGCGGCCAACCCGTCGGCCCTCGACCATTGCCTGATC
    2858      2868      2878      2888      2898      2908      2918
GlyHisValSerGlyGluProGlyHisLeuArgAlaIleGluLysLeuGlyLysThrProLeuLeuAlaLeuGly
GGCCATGTTTCGGGCGAACCAGGGGCATCTGCGCGCGATCGAGAAGCTCGGCAAGACGCCGCTGCTGGCACTCGGC
    2933      2943      2953      2963      2973      2983      2993
MetArgLeuGlyGluGlyThrGlyAlaAlaLeuAlaAlaGlyIleValLysAlaAlaAlaAlaCysHisSerGly
ATGCGGCTTGCGGAAGGCACGGGCGCGGCCCTTGCCGCCGGTATCGTCAAGGCGGCGGCCGCTTGCCACAGCGGC
    3008      3018      3028      3038      3048      3058      3068
MetAlaThrPheAlaGlnAlaGlyValSerAsnLysGlu***
ATGGCGACCTTTGCCCAGGCCGCGCTCAGCAACAAGGAATAG
    3083      3093      3103      3113

```

FIGURE 41.2

NOM = COBV

PREMIER RESIDU = 1
DERNIER RESIDU = 302

PL. 108 / 151

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	18	5.96	2647.23	8.64
2	LEU	L	39	12.91	4410.28	14.39
3	ILE	I	13	4.30	1470.09	4.80
4	MET	M	10	3.31	1310.41	4.28
5	VAL	V	23	7.62	2278.57	7.44
6	SER	S	18	5.96	1566.58	5.11
7	PRO	P	12	3.97	1164.63	3.80
8	THR	T	10	3.31	1010.48	3.30
9	ALA	A	63	20.86	4475.34	14.61
10	TYR	Y	3	0.99	489.19	1.60
11	*	*	0	0.00	0.00	0.00
12	HIS	H	3	0.99	411.18	1.34
13	GLN	Q	6	1.99	768.35	2.51
14	ASN	N	2	0.66	228.09	0.74
15	LYS	K	5	1.66	640.47	2.09
16	ASP	D	10	3.31	1150.27	3.75
17	GLU	E	7	2.32	903.30	2.95
18	CYS	C	3	0.99	309.03	1.01
19	TRP	W	2	0.66	372.16	1.21
20	ARG	R	19	6.29	2965.92	9.68
21	GLY	G	36	11.92	2052.77	6.70
22	-	-	0	0.00	0.00	0.00

RESIDUS = 302

Masse moleculaire (monoisotopique) = 30642.3359

Masse moleculaire (moyenne) = 30662.0820

INDEX DE POLARITE (%) = 26.49

POINT ISOELECTRIQUE = 9.97

DO 260 (1mg/ml) = 0.391 DO 280 (1mg/ml) = 0.479

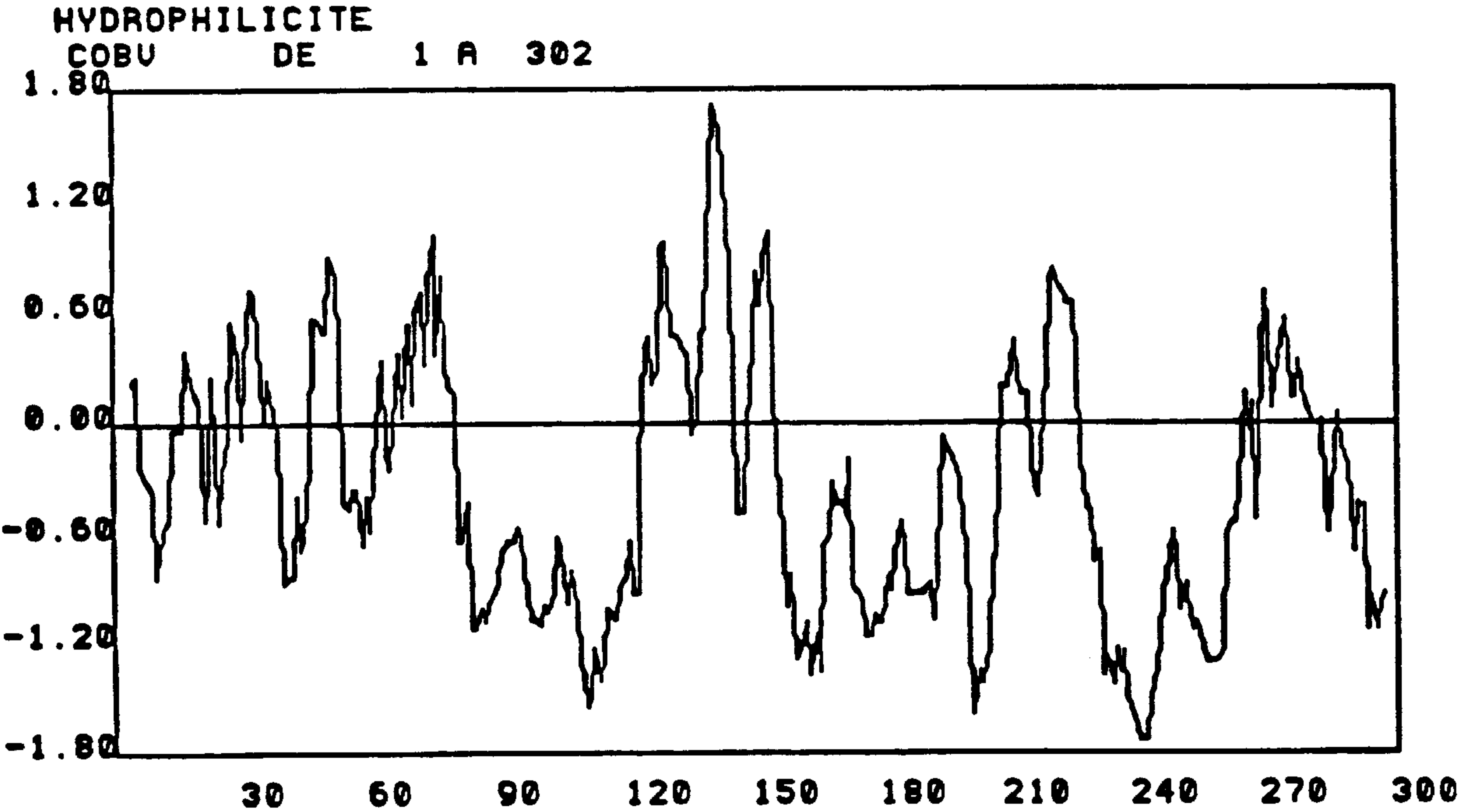


FIGURE 41.3

GENE cobV ET PROTEINE COBV

109/151

SEQUENCE DU FRAGMENT BamHI-SstI-SstI DE 3855 BP, DE 1885 A 2793

MetLysGlyLeuGlyLysValCysAlaAlaLeuAlaGlyAlaProAlaArgLeuAlaArgLeuAlaAlaValGly
 ATGAAAGGTCTGGGAAAAGTGTGCGCGGCCCTCGCGGGTGCGCCCGCCCGGCTTGCGAGACTCGCCGCAGTCGGC
 1894 1904 1914 1924 1934 1944 1954
 GluPheGlySerGlyArgValAsnGlyArgAsnLeuGlyGlyAlaMetGlyPheValGlyAspPheCysAspAsp
 GAGTTTGGCAGCGGCCGGGTCAATGGCCGCAACCTTGGAGGAGCCATGGGCTTCGTTGGCGATTTCTGTGACGAT
 1969 1979 1989 1999 2009 2019 2029
 ValAlaArgSerIleGlyPheLeuSerArgIleProMetProAlaArgHisPheGluGlyTyrAspGlyArgLeu
 GTGGCGCGGTTCGATCGGCTTTCTCAGCCGCATCCCGATGCCTGCGCGTCATTTCGAAGGTTATGACGGGCGCTTG
 2044 2054 2064 2074 2084 2094 2104
 SerArgAlaValArgAlaPheProPheAlaGlyLeuAlaIleAlaLeuProSerAlaAlaValAlaMetAlaLeu
 AGCCGTGCCGTGCGCGCCTTCCCGTTCGCGGGCCTGGCGATCGCACTGCCATCGGCGGGCGTCGCCATGGCCCTC
 2119 2129 2139 2149 2159 2169 2179
 MetAlaLeuGlnValSerSerLeuPheAlaAlaPheValValValAlaIleGlnAlaLeuValThrGlyAlaLeu
 ATGGCGCTGCAGGTGAGCTCGCTCTTTGCCGCCTTTGTCTGTCGTCGCCATCCAGGCGCTTGTCACCGGCGCACTG
 2194 2204 2214 2224 2234 2244 2254
 HisGluAspGlyLeuGlyAspThrAlaAspGlyPheGlyGlyGlyArgAspArgGluAlaAlaLeuAlaIleMet
 CACGAAGACGGGCTTGGCGACACGGCCGATGGCTTTGGCGGGCGGTCGCGACCGCGAAGCGGCGCTTGCGATCATG
 2269 2279 2289 2299 2309 2319 2329
 LysAspSerArgIleGlyThrTyrAlaAlaValAlaLeuIleLeuSerPheGlyLeuArgValSerAlaPheAla
 AAGGACAGCCGCATCGGCACCTATGCGGGCGTCGCCCTCATCCTCTCCTTCGGCCTGCGCGTCTCGGCGTTCGCC
 2344 2354 2364 2374 2384 2394 2404
 SerIleLeuProLeuPheSerProLeuGlyAlaAlaMetAlaIleLeuGlyAlaAlaCysLeuSerArgAlaAla
 TCGATCCTGCCGCTCTTTTCGCCGCTGGGCGCCGCCATGGCGATCCTTGGCGCAGCGTGCCTCAGCCGCGCCGCC
 2419 2429 2439 2449 2459 2469 2479
 MetValTrpHisTrpSerSerLeuProProAlaArgSerSerGlyValAlaAlaSerAlaGlyGluProGluPro
 ATGGTCTGGCACTGGTCGTCGCTTCCGCCGGCGCGCAGCAGCGGCGTCGCGGCCTCGGCTGGCGAGCCGGAACCG
 2494 2504 2514 2524 2534 2544 2554
 AlaAlaThrArgPheAlaLeuAlaPheGlyLeuLeuValAlaMetLeuLeuPheTyrLeuAlaGlnValProAla
 GCGGCCACGCGCTTCGCGCTTGCCTTCGGACTTCTTGTCGCAATGCTGCTCTTCTATCTCGCGCAGGTCCCGGCG
 2569 2579 2589 2599 2609 2619 2629
 LeuGlyValIleAlaAlaLeuValAlaPheLeuAlaThrValLysGlyPheAlaArgLeuAlaMetArgLysIle
 CTCGGCGTCATCGCGGCGCTCGTCGCATTCCTTGCCACGGTCAAGGGCTTCGCTCGGCTGGCGATGCGCAAGATC
 2644 2654 2664 2674 2684 2694 2704
 GlyGlyGlnThrGlyAspThrIleGlyAlaThrGlnGlnLeuThrGluIleAlaValLeuGlyAlaLeuAlaLeu
 GGCGGACAAACGGGCGACACGATCGGGGCGACGACGCAACTGACCGAAATCGCCGTGCTCGGTGCCCTTGCGCTG
 2719 2729 2739 2749 2759 2769 2779
 ThrVal***
 ACGGTTTGA

FIGURE 41.4

PL. 110/151

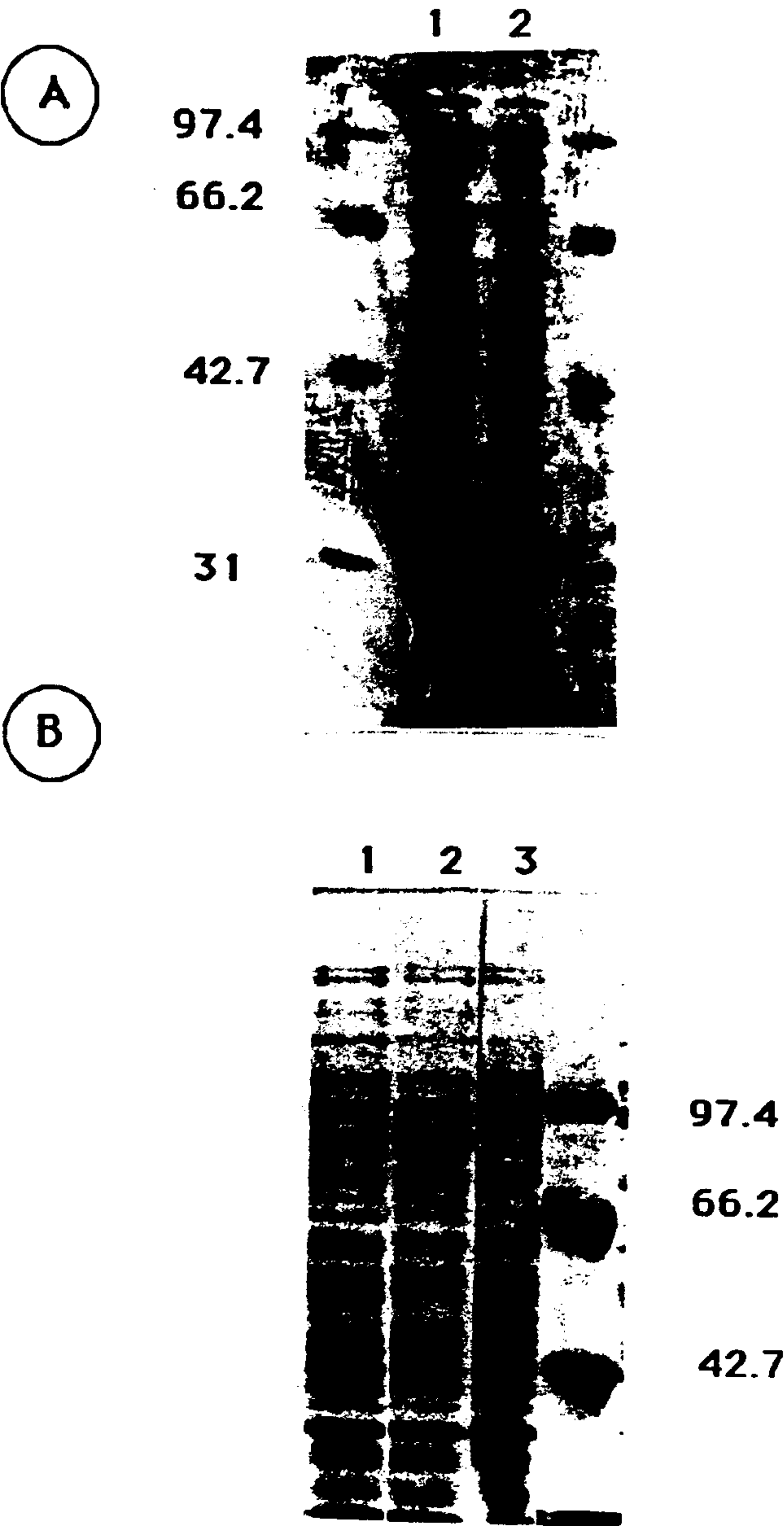


FIGURE 42

PL. 111/151

Sequence					
Longueur = 13144 de 1 a 13144					
10	20	30	40	50	60
GAGCTCGAAG	GGGCTTCCGC	CCCGATCGCT	GGCGTTAGCC	GACGTTTCGAC	GTGCGGATGA
CTCGAGCTTC	CCCGAAGGCG	GGGCTAGCGA	CCGCAATCGG	CTGCAAGCTG	CACGCCTACT
70	80	90	100	110	120
CGCCGAGCGG	GCCGAAGGGC	GCGTCGACGA	CGAGGTTGCG	TACGCGCGAC	TGGCTGGACG
GCGGCTCGCC	CGGCTTCCCG	CGCAGCTGCT	GCTCCAACGC	ATGCGCGCTG	ACCGACCTGC
130	140	150	160	170	180
GAACCTTCGA	GTTCCAGGCG	ATCTGAACGA	AATTGGGCTT	GCTGAAAATA	TACAGCATGG
CTTGAAGCT	CAAGGTCCGC	TAGACTTGCT	TTAACCCGAA	CGACTTTTAT	ATGTCGTACC
190	200	210	220	230	240
ACATGAACCT	TGAGAGGCCG	GAGGCCTATC	CTCCGGGGCG	TGTTGCTATG	CCGCTGATAT
TGTACTTGGA	ACTCTCCGGC	CTCCGGATAG	GAGGCCCCGC	ACAACGATAC	GGCGACTATA
250	260	270	280	290	300
AGGTGTGCGC	TGCAAAAAT	TGAATGCCAA	ACTCGCCACG	CCATGTCGCA	TTCTGGCTAT
TCCACACGCG	ACGTTTTTTA	ACTTACGGTT	TGAGCGGTGC	GGTACAGCGT	AAGACCGATA
310	320	330	340	350	360
CGGCCGCGAC	ATTTTCGACA	AGCCTTGCGA	AAGCGCGAAA	CAATGCGTGA	AAGGGCTTTG
GCCGGCGCTG	TAAAAGCTGT	TCGGAACGCT	TTCCGCGTTT	GTTACGCACT	TTCCCGAAAC
370	380	390	400	410	420
TCAATTGCGG	CGAAATCGTG	TCGAAACAGA	CCTTTGCCGC	TGCCCCGTTT	AGTGTTACCG
AGTTAACGCC	GCTTTAGCAC	AGCTTTGTCT	GGAAACGGCG	ACGGGCAAAG	TCACAATGGC
430	440	450	460	470	480
ATGGCCGCAT	GACACGCAGG	ATCATGTTGC	AGGGAACCGG	CTCGGATGTC	GGAAAATCGG
TACCGGCGTA	CTGTGCGTCC	TAGTACAACG	TCCCTTGGCC	GAGCCTACAG	CCTTTTAGCC
490	500	510	520	530	540
TATTGGTGGC	GGGGCTCTGC	CGGCTTGCCG	CCAATCAGGG	CCTGAAGGTC	CGGCCGTTCA
ATAACCACCG	CCCCGAGACG	GCCGAACGGC	GGTTAGTCCC	GGACTTCCAG	GCCGGCAAGT
550	560	570	580	590	600
AGCCGCAGAA	CATGTCGAAC	AACGCCGCCG	TTTCCGACGA	CGGCGGCGAG	ATCGGCCGCG
TCGGCGTCTT	GTACAGCTTG	TTGCGGCGGC	AAAGGCTGCT	GCCGCCGCTC	TAGCCGGCGC
610	620	630	640	650	660
CGCAATGGCT	GCAGGCGCTG	GCCGCGCGCG	TGCCGTCGTC	GGTGCACATG	AACCCGGTGC
GCGTTACCGA	CGTCCGCGAC	CGGCGCGCGC	ACGGCAGCAG	CCACGTGTAC	TTGGGCCACG
670	680	690	700	710	720
TCCTGAAGCC	GCAGTCGGAC	GTGGGCAGCC	AGATCGTCGT	TCAGGGCAAG	GTCGCCGGGC
AGGACTTCGG	CGTCAGCCTG	CACCCGTCGG	TCTAGCAGCA	AGTCCCGTTC	CAGCGGCCCG
730	740	750	760	770	780
AGGCCAGGGG	GCGGGAATAT	CAGGCGCTCA	AGCCCAAGCT	GCTGGGCGCC	GTCATGGAGA
TCCGGTCCCC	CGCCCTTATA	GTCCGCGAGT	TCGGGTTCGA	CGACCCGCGG	CAGTACCTCT
790	800	810	820	830	840
GTTTCGAACA	AATATCGGCC	GGTGCCGATC	TCGTGGTGGT	CGAAGGCGCC	GGCTCGCCGG
CAAAGCTTGT	TTATAGCCGG	CCACGGCTAG	AGCACCACCA	GCTTCCGCGG	CCGAGCGGCC
850	860	870	880	890	900
CCGAAATCAA	CCTCAGGCCC	GGCGACATCG	CCAATATGGG	CTTTGCGACA	CGGGCCAATG
GGCTTTAGTT	GGAGTCCGGG	CCGCTGTAGC	GGTTATACCC	GAAACGCTGT	GCCCGGTTAC
910	920	930	940	950	960

FIGURE 43

PL. 112/151

TGCCGGTCGT GCTGGTCGGC GACATCGACC GCGGGGGGGT GATCGCCTCG CTGGTCGGCA
 ACGGCCAGCA CGACCAGCCG CTGTAGCTGG CGCCCCCCA CTAGCGGAGC GACCAGCCGT

970 980 990 1000 1010 1020
 CGCATGCGAT CCTGCCCCGAG GAAGACCGGC GCATGGTGAC CGGCTATCTC ATCAACAAGT
 GCGTACGCTA GGACGGGCTC CTTCTGGCCG CGTACCACTG GCCGATAGAG TAGTTGTTCA

1030 1040 1050 1060 1070 1080
 TCCGCGGCGA CGTCACGCTG TTCGACGACG GCATTGCTGC CGTCAACCGC TACACCGGCT
 AGGCGCCGCT GCAGTGCGAC AAGCTGCTGC CGTAACGACG GCAGTTGGCG ATGTGGCCGA

1090 1100 1110 1120 1130 1140
 GGCCCTGCTT CGGCGTCGTG CCGTGGCTGA AGGCGGCGGC ACGCCTGCCG GCGGAAGATT
 CCGGGACGAA GCCGCAGCAC GGCACCGACT TCCGCCGCCG TCGGGACGGC CGCCTTCTAA

1150 1160 1170 1180 1190 1200
 CCGTCGTGCT GGAGAAGCTG ACGCGCGGCG AGGGGCGGGC GCTGAAGGTT GCCGTCCCGG
 GGCAGCACGA CCTCTTCGAC TGC GCGCCG TCCCCGCCG CGACTTCCAA CGGCAGGGCC

1210 1220 1230 1240 1250 1260
 TACTGTCGCG CATCGCCAAT TTCGACGACC TCGATCCGCT CGCCGCCGAA CCGGAGATTG
 ATGACAGCGC GTAGCGGTTA AAGCTGCTGG AGCTAGGCGA GCGGCGGCTT GGCCTCTAAC

1270 1280 1290 1300 1310 1320
 ATCTCGTCTT CGTGCGGCCT GGCAGTCCCA TTCCGGTCGA CGCTGGCCTC GTCGTCATTG
 TAGAGCAGAA GCACGCCGGA CCGTCAGGGT AAGGCCAGCT GCGACCGGAG CAGCAGTAAG

1330 1340 1350 1360 1370 1380
 CCGGGTCGAA ATCGACCATC GGCGACCTCA TCGATTTCCG TCGCAAGGG TGGGACCGTG
 GGCCAGCTT TAGCTGGTAG CCGCTGGAGT AGCTAAAGGC ACGCGTTCCC ACCCTGGCAC

1390 1400 1410 1420 1430 1440
 ACCTCGAACG TCATGTGCGC CGGGGCGGCC GGGTCATCGG CATCTGCGGC GGCTACCAGA
 TGGAGCTTGC AGTACACGCG GCCCGGCCG CCCAGTAGCC GTAGACGCCG CCGATGGTCT

1450 1460 1470 1480 1490 1500
 TGCTCGGCCG GCGCGTCACC GATCCGCTCG GCATCGAGGG CGGCGAACGT GCGGTCGAGG
 ACGAGCCGGC CGCGCAGTGG CTAGGCGAGC CGTAGCTCCC GCCGCTTGCA CGCCAGCTCC

1510 1520 1530 1540 1550 1560
 GCCTCGGGCT GCTCGAGGTC GAGACCGAGA TGGCGCCGGA AAAGACGGTG CGCAACAGCC
 CGGAGCCCGA CGAGCTCCAG CTCTGGCTCT ACCGCGGCCT TTTCTGCCAC GCGTTGTCGG

1570 1580 1590 1600 1610 1620
 GCGCCTGGTC GCTGGAGCAT GATGTGGTGC TCGAAGGCTA CGAAATCCAT CTTGGCAAGA
 CGCGGACCAG CGACCTCGTA CTACACCACG AGCTTCCGAT GCTTTAGGTA GAACCGTTCT

1630 1640 1650 1660 1670 1680
 CGCAAGGTGC GGA CTGTGGC CGGCCGTGCG TGC GCATCGA CAATCGCGCC GACGGCGCCC
 GCGTTCCACG CCTGACACCG GCCGGCAGCC ACGCGTAGCT GTTAGCGCGG CTGCCGCGGG

1690 1700 1710 1720 1730 1740
 TTTCGGCCGA TGGCCGCGTG ATGGGCACCT ACCTGCATGG GCTCTTCACC AGCGACGCCT
 AAAGCCGGCT ACCGGCGCAC TACCCGTGGA TGGACGTACC CGAGAAGTGG TCGCTGCGGA

1750 1760 1770 1780 1790 1800
 ATCGCGGCGC GCTGCTCAAG AGTTTCGGCA TCGAAGGCGG CGCCAACAAC TACCGCCAAT
 TAGCGCCGCG CGACGAGTTC TCAAAGCCGT AGCTTCCGCC GCGGTTGTTG ATGGCGGTTA

1810 1820 1830 1840 1850 1860
 CCGTCGATGC GCGGCTCGAC GATGTCGCGA ACGAACTGGA GGCTGTGCTC GATCGTCGCT
 GCCAGCTACG CCGCGAGCTG CTACAGCGCT TGCTTGACCT CCGACACGAG CTAGCAGCGA

1870 1880 1890 1900 1910 1920
 GGCTGGACGA GTTGCTCAGG CACTAGGGAC GCGGCAACGG TCAGCCAGCA GGTCCGGTAC

FIGURE 43.1

PL. 113/151

CCGACCTGCT CAACGAGTCC GTGATCCCTG CGCCGTTGCC AGTCGGTCGT CCAGGCCATG

1930	1940	1950	1960	1970	1980
GTCGGGCCCCA	ACAGGAGCAA	CGAGCTTATC	CGACGGAAC	ACGCTGCGAC	ATCGTGCTCC
CAGCCCGGGT	TGTCCTCGTT	GCTCGAATAG	GCTGCCTTGA	TGCGACGCTG	TAGCACGAGG
1990	2000	2010	2020	2030	2040
TCGCTTGCGG	CTTCCCAGAC	TTCCCGCGCG	GCATCCAGGT	TCATCAGGGC	AATCCCCAGG
AGCGAACGCC	GAAGGGTCTG	AAGGGCGCGC	CGTAGGTCCA	AGTAGTCCCG	TTAGGGGTCC
2050	2060	2070	2080	2090	2100
CCGACGATCA	GGTCCGGCCA	GGCCGACTGC	CACAGATAGG	CTGTGCGCCAG	ACCCGCGGCG
GGCTGCTAGT	CCAGGCCGGT	CCGGCTGACG	GTGTCTATCC	GACAGCGGTC	TGGGCGCCCG
2110	2120	2130	2140	2150	2160
ATGATGGCCA	CATTGGCGAA	GGCATCGTTG	CGGGCCGAGA	GAAATGCTGC	CCGCGTGAGC
TACTACCGGT	GTAACCGCTT	CCGTAGCAAC	GCCC GGCTCT	CTTTACGACG	GGCGCACTCG
2170	2180	2190	2200	2210	2220
GTGCCGCTCG	TGTGACGGTA	GGCGACGAGC	AGATAGGCGC	AGAAGAGGTT	GACCACCAGC
CACGGCGAGC	ACACTGCCAT	CCGCTGCTCG	TCTATCCGCG	TCTTCTCCAA	CTGGTGGTCC
2230	2240	2250	2260	2270	2280
GCCCCAAGTC	CGGTCAGGGA	CAGGGCAAAG	GGCTCTGGCG	GGACCGGATC	CATGAACTTC
CGGGGTTCAG	GCCAGTCCCT	GTCCCGTTTC	CCGAGACCGC	CCTGGCCTAG	GTACTTGAAG
2290	2300	2310	2320	2330	2340
GCCCAGGCCG	TCCAAAGGAA	GGCCAGCGCC	GGTACCAGTA	GAATGAAAGC	CATCGCCATG
CGGGTCCGCG	AGGTTTCCTT	CCGGTCGCGG	CCATGGTCAT	CTTACTTGCG	GTAGCGGTAC
2350	2360	2370	2380	2390	2400
CCGACCCGCG	CGCGGGTTCG	CGCCGTCCAG	GCCAGAGCAA	AGAAAATCAG	CATGTTGACG
GGCTGGGCGC	GCGCCCAAGC	GCGGCAGGTC	CGGTCTCGTT	TCTTTTAGTC	GTACAACTGC
2410	2420	2430	2440	2450	2460
GAGGCGTCTT	CGAGGAAGTC	GACGCTGTCT	GCCATGAGGG	ACACCGAGCC	GATCGAAAGC
CTCCGCAGAA	GCTCCTTCAG	CTGCGACAGC	CGGTACTCCC	TGTGGCTCGG	CTAGCTTTCC
2470	2480	2490	2500	2510	2520
GCGACAAGGA	GTTTCGACCCC	GAAATAGCCA	AGGTTCAACA	GGGAGACGAT	GAGGACGACG
CGCTGTTTCT	CAAGCTGGGG	CTTTATCGGT	TCCAAGTTGT	CCCTCTGCTA	CTCCTGCTGC
2530	2540	2550	2560	2570	2580
CGGCGCAGGT	CGGTATCCAC	TCGAAAGGTT	CCCTTTCTGG	CGAGATTTCG	CCTCGGCACT
GCCGCGTCCA	GCCATAGGTG	AGCTTTCCAA	GGGAAAGACC	GCTCTAAGCG	GGAGCCGTGA
2590	2600	2610	2620	2630	2640
TTTTTTGGCG	AGATTTCGCC	TCGGCACTTT	GGCACAGGTG	TTAGCAGCAG	TTTGCTATCC
AAAAAACCGC	TCTAAGCGGG	AGCCGTGAAA	CCGTGTCCAC	AATCGTCGTC	AAACGATAGG
2650	2660	2670	2680	2690	2700
ATAGCACTAG	GTTTCGACAT	CGGTTCGGTT	CACACTGCCG	TCGTGCCTGA	CGCCCGACAA
TATCGTGATC	CAAAGCTGTA	GCCAAGGCAA	GTGTGACGGC	AGCACGGACT	GCGGGCTGTT
2710	2720	2730	2740	2750	2760
ATCGTCGCGT	GGCGCAACTC	GGCCGGGGAG	GCGTCGCAATG	CGTCGATTGA	CTTTGGGCTG
TAGCAGCGCA	CCGCGTTGAG	CCGGCCCCCTC	CGCAGCGTAC	GCAGCTAACT	GAAACCCGAC
2770	2780	2790	2800	2810	2820
CCCGCTTCCT	AATCATCAGG	TGTTGGATGG	TTCCCCCTTG	TCGTGGCGAT	CTGGGGGAAT
GGGCGAAGGA	TTAGTAGTCC	ACAACCTACC	AAGGGGGAAC	AGCACCGCTA	GACCCCTTA
2830	2840	2850	2860	2870	2880
AATTGGGAAT	GTGACGGATG	GACCCAAATC	GGGCATCCTT	ATCGCAGCCG	ACCCCGCGAC
TTAACCTTA	CACTGCCTAC	CTGGGTTTAG	CCCGTAGGAA	TAGCGTCGGC	TGGGGCGCTG

FIGURE 43.2

FEUILLE DE REMPLACEMENT

PL. 114/151

2890	2900	2910	2920	2930	2940
TGTAGAACGG	TCAGGGTTTCG	CCATCGGGAT	TGGTGCCGGG	CTGTCGGCCG	GTTGCATGGG
ACATCTTGCC	AGTCCCAAGC	GGTAGCCCTA	ACCACGGCCC	GACAGCCGGC	CAACGTACCC
2950	2960	2970	2980	2990	3000
CAATCGGGGC	AGGTCGGGGA	TCAAGCCGGA	AAAGCCACTG	GCGTGGCATC	GTGATCAGCC
GTTAGCCCCG	TCCAGCCCCT	AGTTCGGCCT	TTTCGGTGAC	CGCACCGTAG	CACTAGTCGG
3010	3020	3030	3040	3050	3060
GGGTTTGGAC	GCCTCTTCTT	CTACGAATCG	TCCGCCTTTC	ACGATGTCCC	TCACAGCGCC
CCCAAACCTG	CGGAGAAGAA	GATGCTTAGC	AGGCGGAAAG	TGCTACAGGG	AGTGTCGCGG
3070	3080	3090	3100	3110	3120
CATGCGTCGG	AGACGACGCG	CAAAGGTTTCG	CTGTGGCACC	GGAAAGACGC	CGGGAAGGTG
GTACGCAGCC	TCTGCTGCGC	GTTTCCAAGC	GACACCGTGG	CCTTTCTGCG	GCCCTTCCAC
3130	3140	3150	3160	3170	3180
AGGCGGGCCG	CTCGGGCCCT	GACATCGGAA	CCTTGCCGTT	TAAGGGCGAG	GCGATGTTTCG
TCCGCCCGGC	GAGCCCGGGA	CTGTAGCCTT	GGAACGGCAA	ATTCCCGCTC	CGCTACAAGC
3190	3200	3210	3220	3230	3240
GCCCGTGACG	CCGTGAGCCA	GGAGACCTGC	CATCCGGCAT	GGGCATTCCG	CCCGAGGGGA
CGGGCACTGC	GGCACTCGGT	CCTCTGGACG	GTAGGCCGTA	CCCGTAAGGC	GGGCTCCCCT
3250	3260	3270	3280	3290	3300
CTTTTGTCTC	CAACGCCATC	ACGGAGGTTG	TTTTGGCTCG	CAGATGTTTT	CAAGAACGCG
GAAAACAGAG	GTTGCGGTAG	TGCCTCCAAC	AAAACCGAGC	GTCTACAAAA	GTTCTTGCGC
3310	3320	3330	3340	3350	3360
CCCGTGCGCG	GTCCGATGGC	TTTTGCCACC	GACGGCTGAT	TTGGGAATGT	TGAGGCAGCC
GGGCACCGCG	CAGGCTACCG	AAAACGGTGG	CTGCCGACTA	AACCCTTACA	ACTCCGTCGG
3370	3380	3390	3400	3410	3420
ACGATGAGCA	GTCTCAGCGC	CGGGCCCGTG	CTGGTCCTTG	GCGGCGCCCG	TTCCGGCAAG
TGCTACTCGT	CAGAGTCGCG	GCCCCGGCAC	GACCAGGAAC	CGCCGCGGGC	AAGGCCGTTC
3430	3440	3450	3460	3470	3480
TCCAGCTTTT	CCGAGAGGCT	CGTCGAAGCG	TCCGGCTTCA	CCATGCATTA	TGTCGCCACG
AGGTCGAAAA	GGCTCTCCGA	GCAGCTTCGC	AGGCCGAAGT	GGTACGTAAT	ACAGCGGTGC
3490	3500	3510	3520	3530	3540
GGCCGCGCCT	GGGACGACGA	AATGCGCGAG	CGCATCGACC	ATCACC GGAC	GCGCCGCGGC
CCGGCGCGGA	CCCTGCTGCT	TTACGCGCTC	GCGTAGCTGG	TAGTGGCCTG	CGCGGCGCCG
3550	3560	3570	3580	3590	3600
GAGGGCTGGA	CGACGCATGA	GGAGCCGCTC	GATCTCGTCG	GCATCCTCAG	ACGCATCGAT
CTCCCGACCT	GCTGCGTACT	CCTCGGCGAG	CTAGAGCAGC	CGTAGGAGTC	TGCGTAGCTA
3610	3620	3630	3640	3650	3660
GATCCCAGCC	ATGTGGTCCT	GATCGACTGC	CTGACGCTAT	GGGTCACCAA	TCTCATGCTG
CTAGGGTCGG	TACACCAGGA	CTAGCTGACG	GACTGCGATA	CCCAGTG GTT	AGAGTACGAC
3670	3680	3690	3700	3710	3720
GAAGAGCGCG	ACATGACGGC	GGAGTTCGCC	GCCCTTGTTG	CGTATCTGCC	CGAGGCGCGG
CTTCTCGCGC	TGTACTGCCG	CCTCAAGCGG	CGGGAACAAC	GCATAGACGG	GCTCCGCGCC
3730	3740	3750	3760	3770	3780
GCGCGCCTCG	TCTTTGTTTC	CAATGAGGTC	GGCCTCGGCA	TCGTGCCCGA	GAACCGCATG
CGCGCGGAGC	AGAAACAAAG	GTTACTCCAG	CCGGAGCCGT	AGCACGGGCT	CTTGCGGTAC
3790	3800	3810	3820	3830	3840
GCCCGCGAGT	TTCGCGACCA	TGCCGGCCGG	CTTCACCAGA	TCGTTGCGGA	GAAATCCGCT
CGGGCGCTCA	AAGCGCTGGT	ACGGCCGGCC	GAAGTGGTCT	AGCAACGCCT	CTTTAGGCGA

FIGURE 43.3

FEUILLE DE REMPLACEMENT

PL. 115/151

3850	3860	3870	3880	3890	3900
GAAGTTTACT	TTGTCGCGGC	CGGTTTGCCG	CTGAAAATGA	AGGGTTGATC	CATGACCACT
CTTCAAATGA	AACAGCGCCG	GCCAAACGGC	GACTTTTACT	TCCCAACTAG	GTACTGGTGA
3910	3920	3930	3940	3950	3960
GCGAGAGCCA	ACCAGGGCAA	GATCCCGGCG	ACCGTCATCA	CCGGCTTCCT	CGGCGCCGGC
CGCTCTCGGT	TGGTCCCGTT	CTAGGGCCGC	TGGCAGTAGT	GGCCGAAGGA	GCCGCGGCCG
3970	3980	3990	4000	4010	4020
AAGACGACGA	TGATCCGCAA	CCTGCTGCAG	AACGCCGACG	GCAAGCGCAT	CGGCCTGATC
TTCTGCTGCT	ACTAGGCGTT	GGACGACGTC	TTGCGGCTGC	CGTTCGCGTA	GCCGGACTAG
4030	4040	4050	4060	4070	4080
ATCAACGAGT	TCGGCGATCT	TGGCGTCGAC	GGCGATGTCT	TGAAGGGCTG	CGGTGCCGAG
TAGTTGCTCA	AGCCGCTAGA	ACCGCAGCTG	CCGCTACAGA	ACTTCCCGAC	GCCACGGCTC
4090	4100	4110	4120	4130	4140
GCCTGCACCG	AGGACGACAT	CATCGAGCTC	ACCAATGGCT	GCATCTGCTG	CACCGTGGCT
CGGACGTGGC	TCCTGCTGTA	GTAGCTCGAG	TGGTTACCGA	CGTAGACGAC	GTGGCACCGA
4150	4160	4170	4180	4190	4200
GACGATTTCA	TCCCGACCAT	GACGAAGCTG	CTCGAGCGTG	AAAACCGTCC	TGACCACATC
CTGCTAAAGT	AGGGCTGGTA	CTGCTTCGAC	GAGCTCGCAC	TTTTGGCAGG	ACTGGTGTAG
4210	4220	4230	4240	4250	4260
ATCATCGAAA	CCTCGGGCCT	TGCCCTGCCG	CAGCCGCTGA	TCGCCGCTTT	CAACTGGCCG
TAGTAGCTTT	GGAGCCCGGA	ACGGGACGGC	GTCGGCGACT	AGCGGCGAAA	GTTGACCGGC
4270	4280	4290	4300	4310	4320
GATATCCGCA	GCGAAGTGAC	CGTCGATGGC	GTCGTCACCG	TGGTCGACAG	CGCCGCCGTT
CTATAGGCGT	CGCTTCACTG	GCAGCTACCG	CAGCAGTGGC	ACCAGCTGTC	GCGGCGGCAA
4330	4340	4350	4360	4370	4380
GCCGCTGGCC	GCTTTGCCGA	CGACCACGAC	AAGGTCGATG	CGCTGCGCGT	CGAGGACGAC
CGGCGACCGG	CGAAACGGCT	GCTGGTGCTG	TTCCAGCTAC	GCGACGCGCA	GCTCCTGCTG
4390	4400	4410	4420	4430	4440
AATCTCGATC	ACGAAAGCCC	GATCGAGGAG	CTGTTTCGAGG	ATCAACTGAC	GGCTGCCGAT
TTAGAGCTAG	TGCTTTCGGG	CTAGCTCCTC	GACAAGCTCC	TAGTTGACTG	CCGACGGCTA
4450	4460	4470	4480	4490	4500
CTCATCGTTC	TCAACAAGAC	CGATCTGATC	GATGCCTCCG	GCCTCAAGGC	CGTGCGCGAC
GAGTAGCAAG	AGTTGTTCTG	GCTAGACTAG	CTACGGAGGC	CGGAGTTCCG	GCACGCGCTG
4510	4520	4530	4540	4550	4560
GAGGTGTCTT	CGCGCACCAG	CCGCAAGCCC	ACGATGATCG	AGGCGAAAAA	CGGCGAAGTC
CTCCACAGAA	GCGCGTGGTC	GGCGTTCCGG	TGCTACTAGC	TCCGCTTTTT	GCCGCTTCAG
4570	4580	4590	4600	4610	4620
GCCGCTGCCA	TCCTGCTTGG	CCTCGGTGTC	GGCACGGAAA	GCGATATCGC	CAACCGCAAG
CGGCGACGGT	AGGACGAACC	GGAGCCACAG	CCGTGCCTTT	CGCTATAGCG	GTTGGCGTTC
4630	4640	4650	4660	4670	4680
TCGCATCACG	AGATGGAGCA	CGAGGCAGGT	GAGGAGCACG	ATCACGACGA	GTTTCGACAGC
AGCGTAGTGC	TCTACCTCGT	GCTCCGTCCA	CTCCTCGTGC	TAGTGCTGCT	CAAGCTGTGC
4690	4700	4710	4720	4730	4740
TTCGTCGTCTG	AGCTCGGTTC	GATCGCCGAT	CCGGCCGCCT	TCATCGATCG	CCTGAAGGGC
AAGCAGCAGC	TCGAGCCAAG	CTAGCGGCTA	GGCCGGCGGA	AGTAGCTAGC	GGACTTCCCC
4750	4760	4770	4780	4790	4800
GTAATCGCGG	AGCACGACGT	TCTGCGCCTC	AAGGGTTTTG	CAGACGTGCC	CGGCAAGCCG
CATTAGCGCC	TCGTGCTGCA	AGACGCGGAG	TTCCCAAAC	GTCTGCACGG	GCCGTTCCGGC
4810	4820	4830	4840	4850	4860

FIGURE 43.4

FEUILLE DE REMPLACEMENT

PL. 116 / 151

ATGCGCCTCC	TGATCCAGGC	GGTCGGCGCC	CGCATCGACC	AATATTACGA	CCGCGCCTGG
TACGCGGAGG	ACTAGGTCCG	CCAGCCGCGG	GCGTAGCTGG	TTATAATGCT	GGCGCGGACC
4870	4880	4890	4900	4910	4920
GGCGCTGGCG	AAAAGCGCGG	TACGCGCCTC	GTCGTCATCG	GCCTGCACGA	CATGGACGAG
CCGCGACCGC	TTTTCGCGCC	ATGCGCGGAG	CAGCAGTAGC	CGGACGTGCT	GTACCTGCTC
4930	4940	4950	4960	4970	4980
GCGGCGGTGC	GCGCCGCGAT	CACCGCGCTC	GTGTAGATCG	TTCTTTGAAT	GAAATGATCT
CGCCGCCACG	CGCGGCGCTA	GTGGCGCGAG	CACATCTAGC	AAGAACTTA	CTTTACTAGA
4990	5000	5010	5020	5030	5040
AACGCATTGA	AATGATGCAG	TTCCGGATGG	AGAACGCTTT	TAGCGTTTTTC	GTTTCGGAATT
TTGCGTAACT	TTACTACGTC	AAGGCCTACC	TCTTGCGAAA	ATCGCAAAAG	CAAGCCTTAA
5050	5060	5070	5080	5090	5100
GCCCCAACGG	ACAAGACGAA	TGCATCTGCT	TCTCGCCCAG	AAAGGAACGA	TCGCCGACGG
CGGGGTGTC	TGTTCTGCTT	ACGTAGACGA	AGAGCGGGTC	TTTCCTTGCT	AGCGGCTGCC
5110	5120	5130	5140	5150	5160
CAACGAGGCG	ATCGACCTTG	GGCAAACGCC	GGCCGATATC	CTTTTCCTAT	CGGCCGCCGA
GTTGCTCCGC	TAGCTGGAAC	CCGTTTGCGG	CCGGCTATAG	GAAAAGGATA	GCCGGCGGCT
5170	5180	5190	5200	5210	5220
CACCGAGCTC	TCCTCGATCG	CCGCGGCTCA	CGGCCGACGC	GACGGAGGCT	TGAGCCTGCG
GTGGCTCGAG	AGGAGCTAGC	GGCGCCGAGT	GCCGGCTGCG	CTGCCTCCGA	ACTCGGACGC
5230	5240	5250	5260	5270	5280
CATCGCCAGC	CTGATGAGCC	TGATGCACCC	GATGTCGGTC	GACACTTACG	TCGAGCGCAC
GTAGCGGTCG	GACTACTCGG	ACTACGTGGG	CTACAGCCAG	CTGTGAATGC	AGCTCGCGTG
5290	5300	5310	5320	5330	5340
GGCGCGTCAC	GCCAAGCTGA	TCGTCTGTCG	GCCGCTCGGT	GGCGCCAGCT	ATTTCCGTTA
CCGCGCAGTG	CGGTTCGACT	AGCAGCAGGC	CGGCGAGCCA	CCGCGGTCGA	TAAAGGCAAT
5350	5360	5370	5380	5390	5400
TCTGCTGGAA	GCCCTGCATG	CGGCTGCCGT	CACCCATCGT	TTGAGATCG	CGGTTCTGCC
AGACGACCTT	CGGGACGTAC	GCCGACGGCA	GTGGGTAGCA	AAGCTCTAGC	GCCAAGACGG
5410	5420	5430	5440	5450	5460
GGGTGACGAC	AAGCCGGATC	CGGGGCTGGA	GCCTTTCTCC	ACCGTCGCAG	CCGACGACCG
CCCACTGCTG	TTCGGCCTAG	GCCCCGACCT	CGGAAAGAGG	TGGCAGCGTC	GGCTGCTGGC
5470	5480	5490	5500	5510	5520
CCAGCGCCTT	TGGGCTTACT	TCACCGAAGG	CGGCTCGGAC	AATGCCGGGC	TGTTTCTCGA
GGTCGCGGAA	ACCCGAATGA	AGTGGCTTCC	GCCGAGCCTG	TTACGGCCCG	ACAAAGAGCT
5530	5540	5550	5560	5570	5580
CTATGCGGCC	GCACTGGTCA	CAGGTGCGGA	GAAGCCGCAG	CCGGCAAAGC	CCCTGTTGAA
GATACGCCGG	CGTGACCAGT	GTCCACGCCT	CTTCGGCGTC	GGCCGTTTCG	GGGACAACTT
5590	5600	5610	5620	5630	5640
GGCCGGCATC	TGGTGGCCGG	GTGCTGGTGT	GATCGGCGTC	AGCGAATGGC	AGTCCCTTGT
CCGGCCGTAG	ACCACCGGCC	CACGACCACA	CTAGCCGCAG	TCGCTTACCG	TCAGGGAACA
5650	5660	5670	5680	5690	5700
TCAGGGACGG	ATGGTAGCGA	GGGAGGGATT	CGAACCCCCG	ACGGTCGGGA	TCTGCTTTTA
AGTCCCTGCC	TACCATCGCT	CCCTCCCTAA	GCTTGGGGGC	TGCCAGCCCT	AGACGAAAAT
5710	5720	5730	5740	5750	5760
CCGCGCGCTC	GTGCAGAGTG	GCGAGACACG	GCCTGTGGAG	GCGCTGATCG	ATGCGCTGGA
GGCGCGCGAG	CACGTCTCAC	CGCTCTGTGC	CGGACACCTC	CGCGACTAGC	TACGCGACCT
5770	5780	5790	5800	5810	5820
GGCTGAAGGT	GTGCGGGCAC	TGCCGGTGTT	TGTCTCAAGC	CTCAAGGATG	CCGTTTCCGT

FIGURE 43.5

PL. 117/151

CCGACTTCCA	CACGCCCGTG	ACGGCCACAA	ACAGAGTTCTG	GAGTTCCTAC	GGCAAAGGCA
5830	5840	5850	5860	5870	5880
CGGCACGCTG	CAGGCGATTT	TTTCCGAGGC	CGCACCCGAC	GTGGTGATGA	ACGCCACTGG
GCCGTGCGAC	GTCCGCTAAA	AAAGGCTCCG	GCGTGGGCTG	CACCACTACT	TGCGGTGACC
5890	5900	5910	5920	5930	5940
CTTTGCGGTC	TCGTGCGCCG	GTGCCGACCG	TCAGCCGACG	GTGCTGGAAT	CGACCGGTGC
GAAACGCCAG	AGCAGCGGGC	CACGGCTGGC	AGTCGGCTGC	CACGACCTTA	GCTGGCCACG
5950	5960	5970	5980	5990	6000
GCCGGTGCTG	CAGGTGATTT	TCTCCGGCTC	GTCGCGGGCG	CAATGGGAAA	CGTCGCCGCA
CGGCCACGAC	GTCCACTAAA	AGAGGCCGAG	CAGCGCCCGC	GTTACCCTTT	GCAGCGGCGT
6010	6020	6030	6040	6050	6060
GGGGCTGATG	GCGCGCGACC	TCGCCATGAA	CGTGGCACTC	CCCGAAGTCG	ATGGCCGCAT
CCCCGACTAC	CGCGCGCTGG	AGCGGTACTT	GCACCGTGAG	GGGCTTCAGC	TACCGGCGTA
6070	6080	6090	6100	6110	6120
CCTTGCGCGC	GCCGTCTCCT	TCAAGGCGGC	GTCGATCTAT	GACGCCAAGG	TGGAGGCCAA
GGAACGCGCG	CGGCAGAGGA	AGTTCCGCCG	CAGCTAGATA	CTGCGGTTCC	ACCTCCGGTT
6130	6140	6150	6160	6170	6180
TATCGTCGGC	CATGAGCCGC	TCGAAGGCCG	GGTGCGCTTT	GCCGCTGATC	TTGCCGTCAA
ATAGCAGCCG	GTACTCGGCG	AGCTTCCGGC	CCACGCGAAA	CGGCGACTAG	AACGGCAGTT
6190	6200	6210	6220	6230	6240
CTGGGCGAAC	GTGCGCCGGG	CAGAGCCCGC	CGAGCGCCGT	ATTGCCATCG	TCATGGCCAA
GACCCGCTTG	CACGCGGCCC	GTCTCGGGCG	GCTCGCGGCA	TAACGGTAGC	AGTACCGGTT
6250	6260	6270	6280	6290	6300
CTATCCGAAC	CGCGACGGTC	GCCTCGGCAA	CGGTGTCGGG	CTCGACACGC	CGGCCGGTAC
GATAGGCTTG	GCGCTGCCAG	CGGAGCCGTT	GCCACAGCCC	GAGCTGTGCG	GCCGGCCATG
6310	6320	6330	6340	6350	6360
CGTCGAGGTG	CTTAGCGCCA	TGGCGCGGGA	AGGCTATGCG	GTCGGTGAGG	TTCCCGCCGA
GCAGCTCCAC	GAATCGCGGT	ACCGCGCCCT	TCCGATACGC	CAGCCACTCC	AAGGGCGGCT
6370	6380	6390	6400	6410	6420
TGGCGACGCG	CTGATCCGCT	TTCTGATGGC	CGGGCCGACC	AATGCGGCGA	GCCATGACCG
ACCGCTGCGC	GACTAGGCGA	AAGACTACCG	GCCC GGCTGG	TTACGCCGCT	CGGTACTGGC
6430	6440	6450	6460	6470	6480
TGAAATCCGC	GAGCGTATTT	CGCTGAACGA	TTACAAAACG	TTCTTCGATT	CGCTTCCGAA
ACTTTAGGCG	CTCGCATAAA	GCGACTTGCT	AATGTTTTGC	AAGAAGCTAA	GCGAAGGCTT
6490	6500	6510	6520	6530	6540
ACAGATAAAG	GATGAAGTTG	CCGGTCGCTG	GGGCGTGCCG	GAGGCCGATC	CCTTTTTCCT
TGTCTATTTT	CTACTTCAAC	GGCCAGCGAC	CCCGCACGGC	CTCCGGCTAG	GGAAAAAGGA
6550	6560	6570	6580	6590	6600
CGATGGCGCC	TTCGCGCTGC	CGCTCGCCCG	CTTCGGCGAG	GTGATCGTCG	GCATCCAACC
GCTACCGCGG	AAGCGCGACG	GCGAGCGGGC	GAAGCCGCTC	CACTAGCAGC	CGTAGGTTGG
6610	6620	6630	6640	6650	6660
GGCGCGCGGC	TACAACATCG	ATCCGAAGGA	AAGCTACCAT	TCCCCGGACC	TCGTGCCGCC
CCGCGCGCCG	ATGTTGTAGC	TAGGCTTCCT	TTCGATGGTA	AGGGGCCTGG	AGCACGGCGG
6670	6680	6690	6700	6710	6720
GCATGGCTAT	CTCGCCTTCT	ACGCCTTCCT	GCGCCAGCAG	TTCGGAGCGC	AGGCGATCGT
CGTACCGATA	GAGCGGAAGA	TGCGGAAGGA	CGCGGTCGTC	AAGCCTCGCG	TCCGCTAGCA
6730	6740	6750	6760	6770	6780
CCACATGGGC	AAGCACGGCA	ATCTCGAATG	GCTGCCGGGC	AAGGCGCTGG	CGCTGTCGGA
GGTGTACCCG	TTCGTGCCGT	TAGAGCTTAC	CGACGGCCCG	TTCCGCGACC	GCGACAGCCT

FIGURE 43.6

FEUILLE DE REMPLACEMENT

PL. 118 / 151

6790 AACCTGCTAT TTGGACGATA	6800 CCCGAAGCGA GGGCTTCGCT	6810 TCTTCGGGCC AGAAGCCCGG	6820 GCTGCCGCAC CGACGGCGTG	6830 ATCTATCCCT TAGATAGGGA	6840 TCATCGTCAA AGTAGCAGTT
6850 CGATCCGGGC GCTAGGCCCG	6860 GAAGGTACGC CTTCCATCGC	6870 AGGCCAAGCG TCCGGTTCGC	6880 CCGCACCAGC GGCGTGGTCG	6890 GCCGTCATCA CGGCAGTAGT	6900 TCGACCACCT AGCTGGTGGA
6910 GACCCCGCCC CTGGGGCGGG	6920 TTGACGCGCG AACTGCGCGC	6930 CCGAATCCTA GGCTTAGGAT	6940 CGGCCCCGCTC GCCGGGCGAG	6950 AAGGATCTGG TTCCTAGACC	6960 AAGCGCTCGT TTCGCGAGCA
6970 CGACGAATAT GCTGCTTATA	6980 TACGACGCCG ATGCTGCGGC	6990 CCGGCGGTGA GGCCGCCACT	7000 TCCGCGCCGC AGGCGCGGCG	7010 CTCAGGCTGC GAGTCCGACG	7020 TCAGCCGCCA AGTCGGCGGT
7030 GATCCTCGAT CTAGGAGCTA	7040 CTCGTGCGCG GAGCACGCGC	7050 ACATCGGCCT TGTAGCCGGA	7060 CGACAGCGAC GCTGTCGCTG	7070 GCAGGCATCG CGTCCGTAGC	7080 ACAGGGGCGA TGTCCCCGCT
7090 CAGCGACGAC GTCGCTGCTG	7100 AAGGCGCTGG TTCCGCGACC	7110 AAAAGCTCGA TTTTCGAGCT	7120 CGCCTATCTC GCGGATAGAG	7130 TGCGACCTCA ACGCTGGAGT	7140 AGGAAATGCA TCCTTTACGT
7150 GATCCGCGAC CTAGGCGCTG	7160 GGCCTGCACA CCGGACGTGT	7170 TCTTCGGCGT AGAAGCCGCA	7180 TGCGCCGGAA ACGCGGCCTT	7190 GGGCGGTTGT CCCGCCAACA	7200 TGACGGACCT ACTGCCTGGA
7210 CACCGTAGCG GTGGCATCGC	7220 CTGGCGCGCG GACCGCGCGC	7230 TGCCCCGAGG ACGGGGCTCC	7240 TCTCGGCGAG AGAGCCGCTC	7250 GGCGGCGACC CCGCCGCTGG	7260 AGAGCCTGCA TCTCGGACGT
7270 GCGGGCGATC CGCCCCGCTAG	7280 GCAGCGGATG CGTCGCCTAC	7290 CGGGGCTGCG GCCCCGACGC	7300 TGGGTTTGCT ACCCAAACGA	7310 ATTCCCACCT TAAGGGTGGA	7320 CGGCGGGGGG GCCGCCCCCC
7330 CAACCCCGCA GTTGGGGCGT	7340 CGCGACGCCC GCGCTGCGGG	7350 AACCCTTCGA TTGGGAAGCT	7360 CCCGCTCGAC GGGCGAGCTG	7370 TGCGTCATGT ACGCAGTACA	7380 CCGACACCTG GGCTGTGGAC
7390 GACAGGCCCG CTGTCCGGGC	7400 AAACCGTCCA TTTGCCAGGT	7410 TCCTCGCTGA AGGAGCGACT	7420 CCTCTCGGAC GGAGAGCCTG	7430 GCCCCCTGGC CGGGGGACCG	7440 GCACCGCCGG CGTGGCGGCC
7450 CGATACGGTC GCTATGCCAG	7460 GAGCGCATCG CTCGCGTAGC	7470 AGTTGCTTGC TCAACGAACG	7480 CGCAAATCTC GCGTTTAGAG	7490 GTGTCGGGTG CACAGCCCAC	7500 AACTGGCTTG TTGACCGAAC
7510 CCCGGACCAC GGGCCTGGTG	7520 TGGGCCAACA ACCCGGTTGT	7530 CCCGCGCCGT GGGCGCGGCA	7540 GCTCGGCGAA CGAGCCGCTT	7550 ATCGAAACGC TAGCTTTGCG	7560 GCCTGAAGCC CGGACTTCGG
7570 GTCGATTTC CAGCTAAAGT	7580 AACTCGGGTG TTGAGCCCAC	7590 CCGCCGAGAT GGCGGCTCTA	7600 GACCGGCTTC CTGGCCGAAG	7610 CTCACC GGTC GAGTGGCCAG	7620 TCAGCGGCCG AGTCGCCGGC
7630 CTTCGTCGCC GAAGCAGCGG	7640 CCCGGTCCAT GGGCCAGGTA	7650 CGGGCGCGCC GCCCGCGCGG	7660 GACGCGCGGC CTGCGCGCCG	7670 CGGCCGGATG GCCGGCCTAC	7680 TGTTGCCGAC ACAACGGCTG
7690 GGGGCGCAAT CCCCGCGTTA	7700 TTCTACTCGG AAGATGAGCC	7710 TCGACAGCCG AGCTGTCGGC	7720 CGCCGTGCCG GCGGCACGGC	7730 ACGCCGGCGG TGCGGCCGCC	7740 CTTACGAGCT GAATGCTCGA

FIGURE 43.7

FEUILLE DE REMPLACEMENT

PL. 119/151

7750	7760	7770	7780	7790	7800
TGGCAAGAAA	TCGGCCGAGC	TTCTGATCCG	CCGCTACCTG	CAGGACCATG	GCGAATGGCC
ACCGTTCTTT	AGCCGGCTCG	AAGACTAGGC	GGCGATGGAC	GTCCTGGTAC	CGCTTACCGG
7810	7820	7830	7840	7850	7860
GTCCTCCTTT	GGCCTGACCG	CCTGGGGCAC	GGCGAACATG	CGCACC GGCG	GCGACGACAT
CAGGAGGAAA	CCGGACTGGC	GGACCCCGTG	CCGCTTGTAC	GCGTGGCCGC	CGCTGCTGTA
7870	7880	7890	7900	7910	7920
CGCCCAGGCC	CTGGCGCTGA	TCGGCGCCAA	GCCCACCTGG	GACATGGTCT	CTCGCCGGGT
GCGGGTCCGG	GACCGCGACT	AGCCGCGGTT	CGGGTGGACC	CTGTACCAGA	GAGCGGCCCA
7930	7940	7950	7960	7970	7980
GATGGGCTAC	GAGATCGTGC	CGCTCGCAGT	CCTCGGCCGC	CCACGCGTCG	ACGTGACCTT
CTACCCGATG	CTCTAGCACG	GCGAGCGTCA	GGAGCCGGCG	GGTGCGCAGC	TGCACTGGAA
7990	8000	8010	8020	8030	8040
GCGCATTTC	GGCTTCTTCC	GCGATGCCTT	CCCGGACCAG	ATCGCGCTCT	TCGACAAGGC
CGCGTAAAGG	CCGAAGAAGG	CGCTACGGAA	GGGCCTGGTC	TAGCGCGAGA	AGCTGTTCCG
8050	8060	8070	8080	8090	8100
GATCCGCGCC	GTCGCGCTGG	AGGAAGACGA	TGCCGACAAC	ATGATCGCCG	CACGCATGCG
CTAGGCGCGG	CAGCGCGACC	TCCTTCTGCT	ACGGCTGTTG	TACTAGCGGC	GTGCGTACGC
8110	8120	8130	8140	8150	8160
GGCGGAAAGC	CGGCGGCTGG	AGGCCGAAGG	CGTGGAAGCC	GCCGAGGCCG	CGCGTCGCGC
CCGCCTTTTCG	GCCGCCGACC	TCCGGCTTCC	GCACCTTCGG	CGGCTCCGGC	GCGCAGCGCG
8170	8180	8190	8200	8210	8220
CTCCTACCGC	GTCTTTGGCG	CAAAGCCCGG	TGCCTATGGC	GCCGCCCTGC	AGGCGCTGAT
GAGGATGGCG	CAGAAACCGC	GTTTCGGGCC	ACGGATACCG	CGGCGGGACG	TCCGCGACTA
8230	8240	8250	8260	8270	8280
CGACGAGAAG	GGCTGGGAAA	CCAAAGCAGA	TCTCGCCGAG	GCCTATCTTA	CCTGGGGCGC
GCTGCTCTTC	CCGACCCTTT	GGTTTCGTCT	AGAGCGGCTC	CGGATAGAAT	GGACCCCGCG
8290	8300	8310	8320	8330	8340
CTATGCCTAT	GGCGCCGGCG	AGGAGGGCAA	GGCCGAGCGC	GATCTTTTCG	AGGAGCGCCT
GATACGGATA	CCGCGGCCGC	TCCTCCCGTT	CCGGCTCGCG	CTAGAAAAGC	TCCTCGCGGA
8350	8360	8370	8380	8390	8400
GCGCACGATA	GAGGCGGTGG	TGCAGAACCA	GGACAACCGC	GAGCACGATC	TGCTCGACAG
CGCGTGCTAT	CTCCGCCACC	ACGTCTTGGT	CCTGTTGGCG	CTCGTGCTAG	ACGAGCTGTC
8410	8420	8430	8440	8450	8460
CGACGACTAC	TACCAGTTTCG	AAGGCGGCAT	GAGCGCTGCC	GCCGAACAGC	TCGGCGGTCA
GCTGCTGATG	ATGGTCAAGC	TTCCGCCGTA	CTCGCGACGG	CGGCTTGTCG	AGCCGCCAGT
8470	8480	8490	8500	8510	8520
CCGTCCGGCG	ATCTACCACA	ACGACCATTG	CCGTCCGGAA	AAGCCTGTGA	TCCGGTCGCT
GGCAGGCCGC	TAGATGGTGT	TGCTGGTAAG	GGCAGGCCTT	TTCGGACACT	AGGCCAGCGA
8530	8540	8550	8560	8570	8580
CGAAGAAGAG	ATCGGCCGCG	TGGTCCGGGC	CCGCGTCGTC	AATCCCAAGT	GGATCGATGG
GCTTCTTCTC	TAGCCGGCGC	ACCAGGCCCG	GGCGCAGCAG	TTAGGGTTCA	CCTAGCTACC
8590	8600	8610	8620	8630	8640
CGTCATGCGC	CACGGATACA	AGGGCGCCTT	CGAGATCGCT	GCCACGGTCG	ACTACATGTT
GCAGTACGCG	GTGCCTATGT	TCCCGCGGAA	GCTCTAGCGA	CGGTGCCAGC	TGATGTACAA
8650	8660	8670	8680	8690	8700
CGCCTTTGCC	GCGACCACGG	GTGCGGTGCG	CGACCATCAT	TTGAGGCCCG	CTTATCAGGC
GCGGAAACGG	CGCTGGTGCC	CACGCCACGC	GCTGGTAGTA	AAGCTCCGGC	GAATAGTCCG

FIGURE 43.8

PL. 120/151

8710 GTTTCATTGTC CAAGTAACAG	8720 GACGAGCGCG CTGCTCGCGC	8730 TGGCTGACTT ACCGACTGAA	8740 CATGCGCGAC GTACGCGCTG	8750 AAGAACCCGG TTCTTGGGCC	8760 CCGCCTTTGC GGCGGAAACG
8770 CGAGCTTGCC GCTCGAACGG	8780 GAACGCCTGC CTTGCGGACG	8790 TTGAAGCAAT AACTTCGTTA	8800 CGACCGCAAT GCTGGCGTTA	8810 CTCTGGACGC GAGACCTGCG	8820 CGCGCTCGAA GCGCGAGCTT
8830 TTCGGCGCGG AAGCCGCGCC	8840 TTTGAACCTG AAACTTGAAC	8850 CCGGCATCGG GGCCGTAGCC	8860 CACGGCAGCA GTGCCGTCGT	8870 ACCCGGCTTC TGGGCCGAAG	8880 GTGCCGGCAA CACGGCCGTT
8890 TGAATAGAGC ACTTATCTCG	8900 GGTTCCGGGC CCAAGGCCCG	8910 TGGCGGTTAT ACCGCCAATA	8920 CCGTCCGGAA GGCAGGCCTT	8930 TTGCTTGGA AACGAACCTT	8940 ACAAAGACCT TGTTTCTGGA
8950 GGTTCCGTTT CCAAGGCAAA	8960 CGCTGCTCAG GCGACGAGTC	8970 TGAAGTGCAG ACTTCACGCT	8980 AAAGGAACCG TTTCCTTGGC	8990 AAGCGGGACG TTCGCCCTGC	9000 AGGGCGTCTG TCCCGCAGAC
9010 CCCATCCCGA GGGTAGGGCT	9020 ACTTGAGAAC TGAACCTTG	9030 TGAGGGAGTG ACTCCCTCAC	9040 ATCATGAGCG TAGTACTCGC	9050 ACGAGACGAC TGCTCTGCTG	9060 AGTAGGCGGC TCATCCGCCG
9070 GAAGCCCCGG CTTCGGGGCC	9080 CCGAGAAGGA GGCTCTTCCT	9090 CGATGCCCGC GCTACGGGCG	9100 CACGCCATGA GTGCGGTACT	9110 AGATGGCGAA TCTACCGCTT	9120 GAAGAAGGCA CTTCTTCCGT
9130 GCACGCGAAA CGTGCGCTTT	9140 AGATCATGGC TCTAGTACCG	9150 GACGAAGACC CTGCTTCTGG	9160 GACGAGAAGG CTGCTCTTCC	9170 GTCTGATCAT CAGACTAGTA	9180 CGTCAACACC GCAGTTGTGG
9190 GGCAAAGGCA CCGTTTCCGT	9200 AGGGCAAGTC TCCCGTTCAG	9210 GACCGCCGGC CTGGCGGCCG	9220 TTCGGCATGA AAGCCGTACT	9230 TCTTCCGCCA AGAAGGCGGT	9240 TATCGCCCAC ATAGCGGGTG
9250 GGCATGCCCT CCGTACGGGA	9260 GCGCCGTCGT CGCGGCAGCA	9270 GCAGTTCATC CGTCAAGTAG	9280 AAGGGTGCGA TTCCCACGCT	9290 TGGCAACCGG ACCGTTGGCC	9300 CGAGCGCGAG GCTCGCGCTC
9310 TTGATCGAGA AACTAGCTCT	9320 AGCATTTCGG TCGTAAAGCC	9330 CGATGTCTGC GCTACAGACG	9340 CAGTTCTACA GTCAAGATGT	9350 CGCTCGGCCA GCGAGCCGCT	9360 GGGCTTCACC CCCGAAGTGG
9370 TGGGAAACGC ACCCTTTGCG	9380 AGGATCGCGC TCCTAGCGCG	9390 CCGCGATGTT GGCGCTACAA	9400 GCGATGGCTG CGCTACCGAC	9410 AAAAGGCCTG TTTTCCGGAC	9420 GGAGAAGGCG CCTCTTCCGC
9430 AAGGAACTGA TTCCTTGACT	9440 TCCGTGACGA AGGCACTGCT	9450 GCGCAACTCG CGCGTTGAGC	9460 ATGGTGCTGC TACCACGACG	9470 TCGACGAGAT AGCTGCTCTA	9480 CAACATTGCT GTTGTAACGA
9490 CTGCGCTACG GACGCGATGC	9500 ACTACATCGA TGATGTAGCT	9510 CGTCGCCGAA GCAGCGGCTT	9520 GTGGTGCGCT CACCACGCGA	9530 TCCTGAAGGA AGGACTTCCT	9540 AGAAAAGCCG TCTTTTCGGC
9550 CACATGACGC GTGTACTGCG	9560 ATGTGGTGCT TACACCACGA	9570 CACCGGCCGC GTGGCCGGCG	9580 AACGCGAAAG TTGCGCTTTC	9590 AAGACCTGAT TTCTGGACTA	9600 CGAAGTCGCC GCTTCAGCGG
9610 GATCTCGTCA CTAGAGCAGT	9620 CTGAGATGGA GACTCTACCT	9630 GCTGATCAAG CGACTAGTTC	9640 CATCCGTTCC GTAGGCAAGG	9650 GTTCCGGCAT CAAGGCCGTA	9660 CAAGGCGCAG GTTCCGCGTC
9670	9680	9690	9700	9710	9720

FIGURE 43.9

FEUILLE DE REMPLACEMENT

PL. 121/151

CAGGGCGTGG	AGTTCTGATG	AGCCAGAGCT	GGCAGTTCTG	GGCGCTGCTT	TCGGCCGCCT
GTCCCCGACC	TCAAGACTAC	TCGGTCTCGA	CCGTCAAGAC	CCGCGACGAA	AGCCGGCGGA
9730	9740	9750	9760	9770	9780
TCGCTGCGCT	CACGGCGGTG	TTTGCCAAGG	TCGGGGTTGC	GCAGATCAAC	TCCGACTTCG
AGCGACGCGA	GTGCCGCCAC	AAACGGTTCC	AGCCCCAACG	CGTCTAGTTG	AGGCTGAAGC
9790	9800	9810	9820	9830	9840
CAACGCTGAT	CCGCACCGTC	GTCATCCTCT	GCGTGATCGC	CGCCATCGTG	GCGGCGACAG
GTTGCGACTA	GGCGTGGCAG	CAGTAGGAGA	CGCACTAGCG	GCGGTAGCAC	CGCCGCTGTC
9850	9860	9870	9880	9890	9900
GGCAGTGGCA	GAAGCCATCG	GAAATCCCGG	GCCGCACCTG	GCTGTTCTTG	GCGCTGTCAG
CCGTCACCGT	CTTCGGTAGC	CTTTAGGGCC	CGGCGTGGAC	CGACAAGGAC	CGCGACAGTC
9910	9920	9930	9940	9950	9960
GGCTTGCGAC	TGGCGCTTCC	TGGCTTGCCCT	ATTTCCGCGC	GCTGAAGCTC	GGCGACGCCG
CCGAACGCTG	ACCGCGAAGG	ACCGAACGGA	TAAAGGCGCG	CGACTTCGAG	CCGCTGCGGC
9970	9980	9990	10000	10010	10020
CCCGCGTGGC	GCCGCTCGAC	AAGCTCTCGA	TCGTCAATGGT	CGCGATCTTC	GGCGTGCTCT
GGGCGCACCG	CGGCGAGCTG	TTCGAGAGCT	AGCAGTACCA	GCGCTAGAAG	CCGCACGAGA
10030	10040	10050	10060	10070	10080
TCCTCGGTGA	AAAGCTCAAC	CTGATGAACT	GGCTCGGCGT	CGCCTTCATT	GCCGCCGGGG
AGGAGCCACT	TTTCGAGTTG	GACTACTTGA	CCGAGCCGCA	GCGGAAGTAA	CGGCGGCCCC
10090	10100	10110	10120	10130	10140
CGCTGTTGCT	GGCGGTGTTT	TGAGCGCGCC	TGCTCTGGTG	CCTGTTCACT	GAATGCTCGC
GCGACAACGA	CCGCCACAAA	ACTCGCGCGG	ACGAGACCAC	GGACAAGTGA	CTTACGAGCG
10150	10160	10170	10180	10190	10200
CTCAATCAAT	CCGTAATCCC	GACACATGCA	GTGGTTGTGA	CGAGCGGGAG	GACGGCATGC
GAGTTAGTTA	GGCATTAGGG	CTGTGTACGT	CACCAACACT	GCTCGCCCTC	CTGCCGTACG
10210	10220	10230	10240	10250	10260
AGATTGAAGG	CAATTGGAGC	GAGCGCCTTC	CTGATCCGTC	GGGCCACGTC	GCGCAGTTTCG
TCTAACTTCC	GTTAACCTCG	CTCGCGGAAG	GACTAGGCAG	CCCGGTGCAG	CGCGTCAAGC
10270	10280	10290	10300	10310	10320
GCAGACGCTG	GAAGCGTCGC	AGCCTGAGGG	TGAGCCCTGC	TTCAGACCCA	CCGGCGGACA
CGTCTGCGAC	CTTCGCAGCG	TCGGACTCCC	ACTCGGGACG	AAGTCTGGGT	GGCCGCCTGT
10330	10340	10350	10360	10370	10380
CGCCTGCAAT	AGGCACCGTA	GGCGTCGCCG	AAGACCTTGG	CGAGGTGGGT	TTCTTCCATG
GCGGACGTTA	TCCGTGGCAT	CCGCAGCGGC	TTCTGGAACC	GCTCCACCCA	AAGGAGGTAC
10390	10400	10410	10420	10430	10440
CGGATCTGGT	AGGAAATCGA	GATCCAGGCG	GAGAGCGCCA	GCGCCACCGA	GATGACGTTG
GCCTAGACCA	TCCTTTAGCT	CTAGGTCCGC	CTCTCGCGGT	CGCGGTGGCT	CTACTGCAAC
10450	10460	10470	10480	10490	10500
GGCACCGCCA	TCACCGTGCC	GATCAGCGCG	GTCACCATGC	CGACATAGAT	CGGGTTGCGC
CCGTGGCGGT	AGTGGCACGG	CTAGTCGCGC	CAGTGGTACG	GCTGTATCTA	GCCCAACGCG
10510	10520	10530	10540	10550	10560
GAGAAGGCAT	AGAGGCCTGA	GGTCACAAGC	GGCGCGTCCT	GCTTTTCAGG	GATGCCGATC
CTCTTCCGTA	TCTCCGGAAT	CCAGTGTTCC	CCGCGCAGGA	CGAAAAGTCC	CTACGGCTAG
10570	10580	10590	10600	10610	10620
TTCCAGGAAT	GACGCATCGC	CCATTGCGAC	AGCATCGTCA	GCCCCCGGCC	GAGCGTCATC
AAGGTCCTTA	CTGCGTAGCG	GGTAACGCTG	TCGTAGCAGT	CGGGCGGCGG	CTCGCAGTAG
10630	10640	10650	10660	10670	10680
AGCGCCAGGC	CGACGGCGTG	AAGGATGGGC	GTGTGCGAGC	CCGGGATCCG	GCCGAGGGCA

FIGURE 43.10

FEUILLE DE REMPLACEMENT

PL. 122/151

TCGCGGTCCG GCTGCCGCAC TTCCTACCCG CACAGCTCGC GGCCCTAGGC CGGCTCCCGT

10690 10700 10710 10720 10730 10740
GCATCGACGG AGGCCGGGAG CATGGCGACC GCCAGCAGGT GGATCACCAG CGCTGCGACG
CGTAGCTGCC TCCGGCCCTC GTACCGCTGG CGGTCTGTCCA CCTAGTGGTC GCGACGCTGC

10750 10760 10770 10780 10790 10800
ATCAGGCGGA AAAGCCTGCC CGCAAACCTT TCCGCATCGT CGCCATAGGT TAGCACGACC
TAGTCCGCCT TTTCGGACGG GCGTTTGGA AGGCGTAGCA GCGGTATCCA ATCGTGCTGG

10810 10820 10830 10840 10850 10860
GGCGAGCGGC CGGATTGCAC GCGGCGGAGG ATCGCCAGCG CGAGCGTGGA CAATCCCACG
CCGCTCGCCG GCCTAACGTG CGCCGCCTCC TAGCGGTCTG GCTCGCACCT GTTAGGGTGC

10870 10880 10890 10900 10910 10920
ACGAGCATCA GGATGGTGGG AAGGGTGGTG GACATGGAAA CCTCTGGAGC GAGCTGACAA
TGCTCGTAGT CCTACCACCC TTCCACCAC CTGTACCTTT GGAGACCTCG CTCGACTGTT

10930 10940 10950 10960 10970 10980
GACAGGAGCG CACGACGGGT AGGCGGCCCA TATGAGCGTC TACCCGGCGA AGCATTCTGA
CTGTCTCTGC GTGCTGCCCA TCCGCCGGGT ATACTCGCAG ATGGGCCGCT TCGTAAGACT

10990 11000 11010 11020 11030 11040
TCACCTTGCA ATCTCTAGTA ACTAGAGGTT CAAGCGTCGG ACCTGTCCGA CTTTCGTCTG
AGTGGAACGT TAGAGATCAT TGATCTCCAA GTTCGCAGCC TGGACAGGCT GAAAGCAGCA

11050 11060 11070 11080 11090 11100
GGTTACCGGA TCTTATTGCC AAGCGTTGGA GGCTGTCTATC GTCGCCCCCG CCGTGTCTGGA
CCAATGGCCT AGAATAACGG TTCGCAACCT CCGACAGTAG CAGCGGGGGC GGCACAGCCT

11110 11120 11130 11140 11150 11160
AGGTCGGCAA AATTCGTCTC TTGACGGCTG CTCCTTCCGT CGAGCGATTG CATAGGCAGG
TCCAGCCGTT TTAAGCAGAG AACTGCCGAC GAGGAAGGCA GCTCGCTAAC GTATCCGTCC

11170 11180 11190 11200 11210 11220
AGGCCGCACC CATGTTAGAC CGTCGACAGG CTAAATACGG GTGAACCTTG AAGAATACTC
TCCGGCGTGG GTACAATCTG GCAGCTGTCC GATTTATGCC CACTTGGAAC TTCTTATGAG

11230 11240 11250 11260 11270 11280
TCAGAGCTGC GGTTGGTGTG GCATCGGTCT TGCTGTTCTT GTCATCAGGT GTGGCGGGGC
AGTCTCGACG CCAACCACAG CGTAGCCAGA ACGACAAGAA CAGTAGTCCA CACCGCCCCG

11290 11300 11310 11320 11330 11340
AGGCGCAAAC CGTGAAGAGC GGGGCGTCAC GAGCTCAAGA AACGACGACC ACCCAGAAGG
TCCGCGTTTG GCACTTCTCG CCCCAGAGTG CTCGAGTTCT TTGCTGCTGG TGGGTCTTCC

11350 11360 11370 11380 11390 11400
CGAAACCGAA AACTAAAACG ACGCGCAAGC AAAGGGCTGC GGATGAAGCC AAGGCCAAGG
GCTTTGGCTT TTGATTTTGC TGCGCGTTTC TTTCCCGACG CCTACTTCGG TTCCGGTTCC

11410 11420 11430 11440 11450 11460
CGCTCGCCGA AGCGCGCCGT CCACGGATTT GCAAGACGCG GGAGAGCGAA TGCAGCTATG
GCGAGCGGCT TCGCGCGGCA GGTGCCTAAA CGTTCTGCGC CCTCTCGCTT ACGTCGATAC

11470 11480 11490 11500 11510 11520
GCGCAGGTCC GGTCGGAGAG CAGTGCTCGT GCTGGTCGAA ATCCGGTGCG CCTGATCTTG
CGCGTCCAGG CCAGCCTCTC GTCACGAGCA CGACCAGCTT TAGGCCACGC GGACTAGAAC

11530 11540 11550 11560 11570 11580
GCATAACTGT CAGGCGTTGA CCGCCCGCGA CCTTCGCGCG GGCAGGCAAG CGTGCGTCTG
CGTATTGACA GTCCGCAACT GCGGGGCGCT GGAAGCGCGC CCGTCCGTTT GCACGCAGCG

11590 11600 11610 11620 11630 11640
TCGAAGCGAC GCCTGACGCG ATAGAAATCA CGGGTCTGCT GGTTCGTTCT GAAAGCTTGG
AGCTTCGCTG CGGACTGCGC TATCTTTAGT GCCCAGCGGA CCAAGCAAGA CTTTCGAACC

FIGURE 43.11

PL. 123/151

11650 GATTGGGTTT CTAACCCTAAA	11660 AGGTGATGGA TCCACTACCT	11670 AGCCGGCGTT TCGGCCGCAA	11680 GAACGCAAAA CTTGCGTTTT	11690 TAATGATCGA ATTACTAGCT	11700 TCTCGAGAAC AGAGCTCTTG
11710 AGCGCGCTCC TCGCGCGAGG	11720 AGTTTGCAAC TCAAACGTTG	11730 CCGAGCACAC GGCTCGTGTG	11740 GGCGAACAGA CCGCTTGTCT	11750 AGCGTAAGTA TCGCATTTCAT	11760 TGACGGTCGG ACTGCCAGCC
11770 CCCTATATCG GGGATATAGC	11780 TTCATCCGAT AAGTAGGCTA	11790 TGCGGTGGCG ACGCCACCGC	11800 GAGATTGTTC CTCTAACAAG	11810 GAAGCGTGCC CTTCGCACGG	11820 CCATACGCCC GGTATGCGGG
11830 GAAATGATCG CTTTACTAGC	11840 CCGCAGCGCT GGCGTCGCGA	11850 GCTTCACGAT CGAAGTGCTA	11860 ACGGTCGAAG TGCCAGCTTC	11870 ATACCGACGC TATGGCTGCG	11880 GACGCTGCTG CTGCGACGAC
11890 GAGATCAAGG CTCTAGTTCC	11900 AAGCGTTCGG TTCGCAAGCC	11910 CCCCAAGGTC GGGGTTCCAG	11920 GCAACACTGG CGTTGTGACC	11930 TTGCGTGGCT AACGCACCGA	11940 CACCGACATA GTGGCTGTAT
11950 TCCACTCCGT AGGTGAGGCA	11960 TCCACGGCAA AGGTGCCGTT	11970 CCGACAGGTG GGCTGTCCAC	11980 CGCAAGGAAC GCGTTCCTTG	11990 TGGATCGCCA ACCTAGCGGT	12000 GCACCTCGCA CGTGGAGCGT
12010 TCGGCGCCCG AGCCGCGGGC	12020 CCGCGGCGAA GGCGCCGCTT	12030 AACCGTCAAG TTGGCAGTTC	12040 CTCGCCGACC GAGCGGCTGG	12050 TGATCGACAA ACTAGCTGTT	12060 TGCGATAGCG ACGCTATCGC
12070 ATCAAAGCCG TAGTTTCGGC	12080 GCGATCCGAA CGCTAGGCTT	12090 TTTCTGGAAA AAAGACCTTT	12100 GTGTTTCGGC CACAAGCCGC	12110 CCGAGATGAA GGCTCTACTT	12120 ACGCTTGCTG TGCGAACGAC
12130 GAGGTCTTGG CTCCAGAACC	12140 GCGACGGCGA CGCTGCCGCT	12150 CGAGACCCTT GCTCTGGGAA	12160 CTCGCAAAGG GAGCGTTTCC	12170 CCCGTGCAAT GGGCACGTAA	12180 AGCGCCGGAA TCGCGGCCCT
12190 TGAGAGTGCC ACTCTCACGG	12200 GCCGTTTATC CGGCAAATAG	12210 GGCAAGCATG CCGTTTCGTAC	12220 TCTGTGCCAT AGACACGGTA	12230 GTCGACCCGG CAGCTGGGCC	12240 TCAACCGGTC AGTTGGCCAG
12250 ATCCAAGATC TAGGTTCTAG	12260 GCAGAACGGA CGTCTTGCTT	12270 CATGCATTTG GTACGTAAAC	12280 CGGTTTTGCC GCCAAAACGG	12290 CGCCGGTGTG GCGGCCACAC	12300 GCCCAGCCAC CGGGTCGGTG
12310 GCCTCACAGG CGGAGTGTCC	12320 CTGCGCGGTT GACGCGCCAA	12330 GCGGCCGTTA CGCCGGCAAT	12340 GGACAGCGCA CCTGTGCGCT	12350 GAATTTGCCG CTTAAACGGC	12360 ACCGCGCCGC TGGCGCGGCG
12370 GCCTCAATGC CGGAGTTACG	12380 CCCAGCCAGA GGGTCGGTCT	12390 TCCGCAAGGG AGGCGTTCCC	12400 ATGCGTCGGA TACGCAGCCT	12410 TCTGCGAGCA AGACGCTCGT	12420 GCCGGATCGC CGGCCTAGCG
12430 GAGCGCGATC CTCGCGCTAG	12440 GAGACGATGA CTCTGCTACT	12450 CGAGCAGCGG GCTCGTCGCC	12460 CTTGATGATC GAACTACTAG	12470 TTGGCGCCCT AACC GCGGGA	12480 TGGCCATGGC ACCGGTACCG
12490 ATAGCGCGAG TATCGCGCTC	12500 CCGACCTGGG GGCTGGACCC	12510 CGCCGAGGAA GCGGCTCCTT	12520 CTGGCCGAGG GACCGGCTCC	12530 CCCATCAACA GGGTAGTTGT	12540 GGCCGACCTT CCGGCTGGAA
12550 CCAGAGAACG GGTCTCTTGC	12560 GCGCCGAAGA CGCGGCTTCT	12570 AGAGGAAGAC TCTCCTTCTG	12580 GCCGAAGGCG CGGCTTCCGC	12590 CCGACGTTGG GGCTGCAACC	12600 AGCCAAAGTT TCGGTTTCAA

FIGURE 43.12

FEUILLE DE REMPLACEMENT

PL. 124 / 15 /

12610	12620	12630	12640	12650	12660
GAGGAACTTC	GTGTGCGCCG	TCGCCTTCAA	CACGCCGAAG	CCGGCGAGGG	TAACGAAGCC
CTCCTTGAAG	CACACGCGGC	AGCGGAAGTT	GTGCGGCTTC	GGCCGCTCCC	ATTGCTTCGG
12670	12680	12690	12700	12710	12720
GAGCATGAAG	AACGAGCCGG	TGCCGGGGCC	GAAGACGCCG	TCATAAAAGC	CGATTAGCGG
CTCGTACTTC	TTGCTCGGCC	ACGGCCCCCG	CTTCTGCGGC	AGTATTTTCG	GCTAATCGCC
12730	12740	12750	12760	12770	12780
CACCAGTGTC	AGCGTGAAGA	CGAAGGGGGT	GACGCGGCTG	TGCTGGTCGA	CGTCGCCCAT
GTGGTCACAG	TCGCACTTCT	GCTTCCCCCA	CTGCGCCGAC	ACGACCAGCT	GCAGCGGGTA
12790	12800	12810	12820	12830	12840
GTTTCGGCTTC	AGGCCGAAAT	AAAGCGCAAT	GGCGATCAGC	AGAAAGGGCA	GGATCGCCTT
CAAGCCGAAG	TCCGGCTTTA	TTTCGCGTTA	CCGCTAGTCG	TCTTTCCCGT	CCTAGCGGAA
12850	12860	12870	12880	12890	12900
CAGCACGTCG	CCGGGAACGA	TGGTTGCGAG	CAGGGCGCCG	AGCACGGCGC	CGGCGGCCGA
GTCGTGCAGC	GGCCCTTGCT	ACCAACGCTC	GTCCCGCGGC	TCGTGCCGCG	GCCGCCGGCT
12910	12920	12930	12940	12950	12960
CATCAGCGCC	ATCGGCAGCT	GCTCTTTTCA	GTTACGTGG	CCGCGCCGGG	CATAGGACAG
GTAGTCGCGG	TAGCCGTCGA	CGAGAAAGTC	CAAGTGCACC	GGCGCGGCCC	GTATCCTGTC
12970	12980	12990	13000	13010	13020
CGTGGCCGAG	CCGGAGCCGA	ACAATCCCTG	CAGCTTGTTG	GTGCCGAGCG	TCTGCAAGGG
GCACCGGCTC	GGCCTCGGCT	TGTTAGGGAC	GTCGAACAAC	CACGGCTCGC	AGACGTTCCC
13030	13040	13050	13060	13070	13080
CGGGATGCCC	GCAATGAGCA	TGGCCGGAAT	GGTGATCATG	CCACCGCCGC	CGGCGATCGA
GCCCTACGGG	CGTTACTCGT	ACCGGCCTTA	CCACTAGTAC	GGTGGCGGCG	GCCGCTAGCT
13090	13100	13110	13120	13130	13140
ATCGATGAAG	CCTGCGATGA	AGGCGGCGAC	GAACAGGAAG	GCGAGCAGGT	GGAAGGCGAG
TAGCTACTTC	GGACGCTACT	TCCGCCGCTG	CTTGTCTTTC	CGCTCGTCCA	CCTTCCGCTC

ATCT
TAGA

FIGURE 43.13

125 / 151

 Carte de restriction de la sequence de 13144 pb

ApaLI	642,							
EcoRI	8818,							
HindIII	11633,							
MluI	7963,							
NdeI	10950,							
PvuII	12918,							
SfiI	3133,							
SplI	99,							
BglII	8248,	13139,						
KpnI	2315,	6300,						
NotI	5526,	7615,						
SmaI	1322,	9868,						
SspI	4843,	6968,						
XmnI	9313,	12091,						
AatII	1033,	9503,	12773,					
AflIII	550,	7963,	8634,					
BalI	2107,	6236,	12473,					
BamHI	2266,	5416,	10664,					
BspMII	5002,	8494,	8914,					
EcoRV	4263,	4605,	5137,					
NcoI	6318,	7786,	12474,					
NsiI	3467,	5064,	12266,					
PflMI	7870,	10718,	11065,					
XhoI	1512,	4171,	11692,					
ApaI	1928,	3138,	3386,	8551,				
AsuII	784,	5670,	8418,	11799,				
FspII	784,	5670,	8418,	11799,				
MaeI	1883,	2647,	10995,	11002,				
NruI	1827,	3794,	10002,	12419,				
SauI	852,	7001,	10284,	10517,				
BstEII	995,	3642,	8456,	10470,	11041,			
Eco47III	6954,	7209,	8434,	10731,	11837,			
SacI	5,	4109,	4694,	5169,	11315,			
StuI	204,	4081,	8261,	9406,	10515,			
BstXI	761,	2982,	3612,	6031,	6232,	9102,		
SacII	932,	1025,	2096,	3537,	5184,	12014,		
SphI	966,	2740,	5360,	8098,	9246,	10199,		
BclI	2992,	4016,	9029,	9164,	9623,	10978,	13053,	
RsaI	101,	1201,	1918,	2313,	4881,	6298,	6856,	
Tth111I	1821,	2424,	4351,	7361,	7904,	12227,	12697,	
PstI	613,	3989,	5832,	5952,	7260,	7782,	8211,	12992,
Clai	1351,	3596,	4469,	4724,	5748,	6618,	8574,	11687,
	13082,							
FspI	1363,	1551,	1653,	5219,	7841,	7982,	8342,	9760,
	11971,							
HinfI	1137,	2564,	2592,	3025,	5667,	5927,	6467,	6923,
	13079,							
StyI	2488,	3396,	5116,	6105,	6318,	7786,	9745,	10355,
	11389,	11395,	11903,	12468,	12474,			
DdeI	852,	1875,	3373,	3586,	6311,	7001,	7010,	7610,
	8956,	9020,	9611,	10284,	10517,	11220,		
Nsp7524I	554,	966,	2394,	2740,	5360,	7840,	8098,	8638,
	9246,	9553,	10168,	10199,	12210,	12264,		
PvuI	26,	1853,	2453,	4403,	4703,	4728,	5091,	5112,
	5178,	6717,	7269,	9991,	12429,	13077,		
AvaI	975,	1320,	1503,	1512,	3131,	3231,	3709,	3766,
	4171,	4212,	7224,	7573,	9866,	11692,	11720,	
BanII	5,	496,	1723,	1928,	2254,	3138,	3386,	4109,
	4694,	5169,	6207,	6282,	8551,	10296,	11315,	
SalI	83,	1296,	2418,	4045,	4303,	5258,	6959,	7700,
	7967,	8627,	8708,	9198,	11182,	12221,	12766,	
XhoII	2266,	3920,	5416,	5688,	6943,	7020,	7140,	8248,
	10382,	10400,	10664,	11048,	12378,	12398,	13139,	

FIGURE 44

. 126 / 151

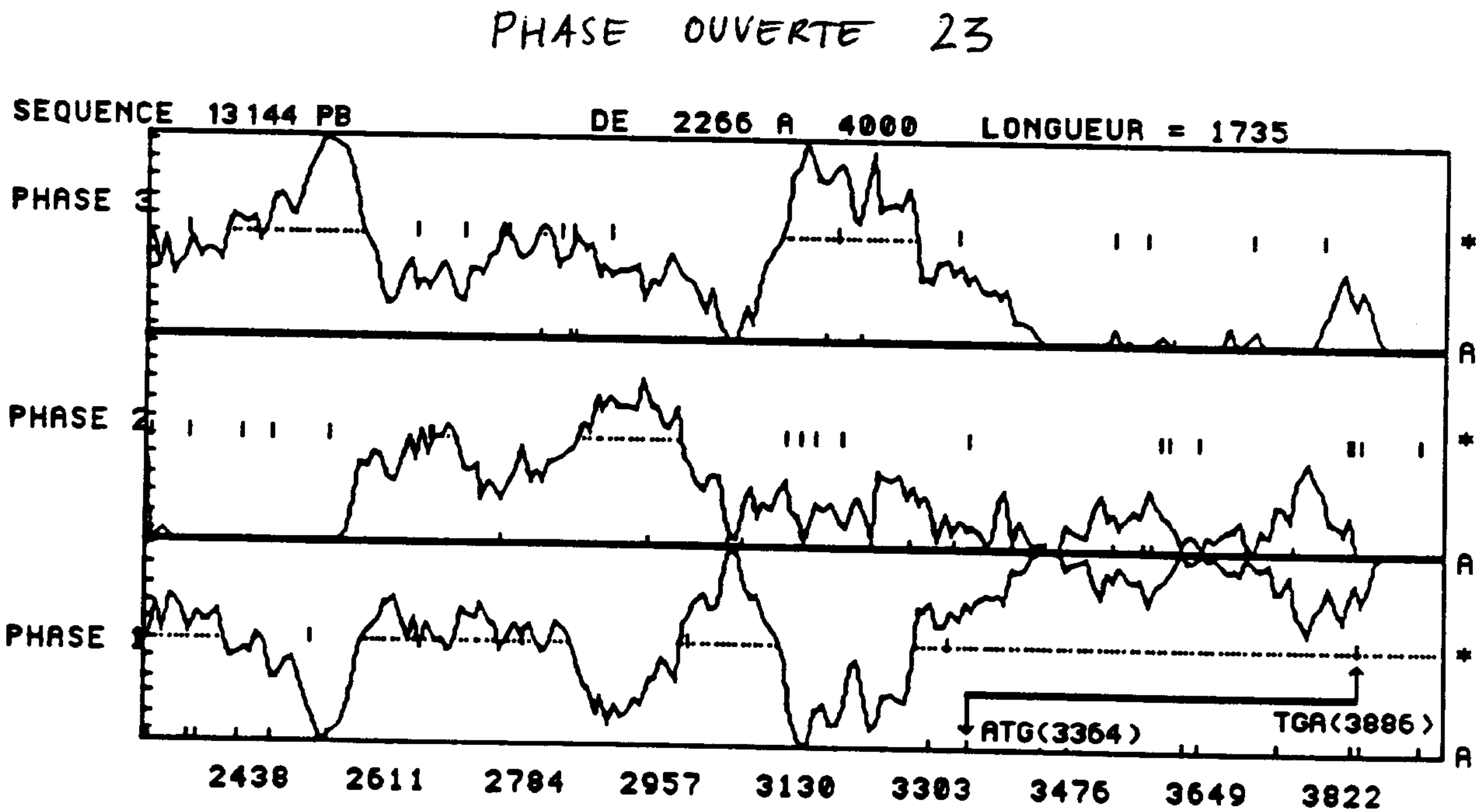
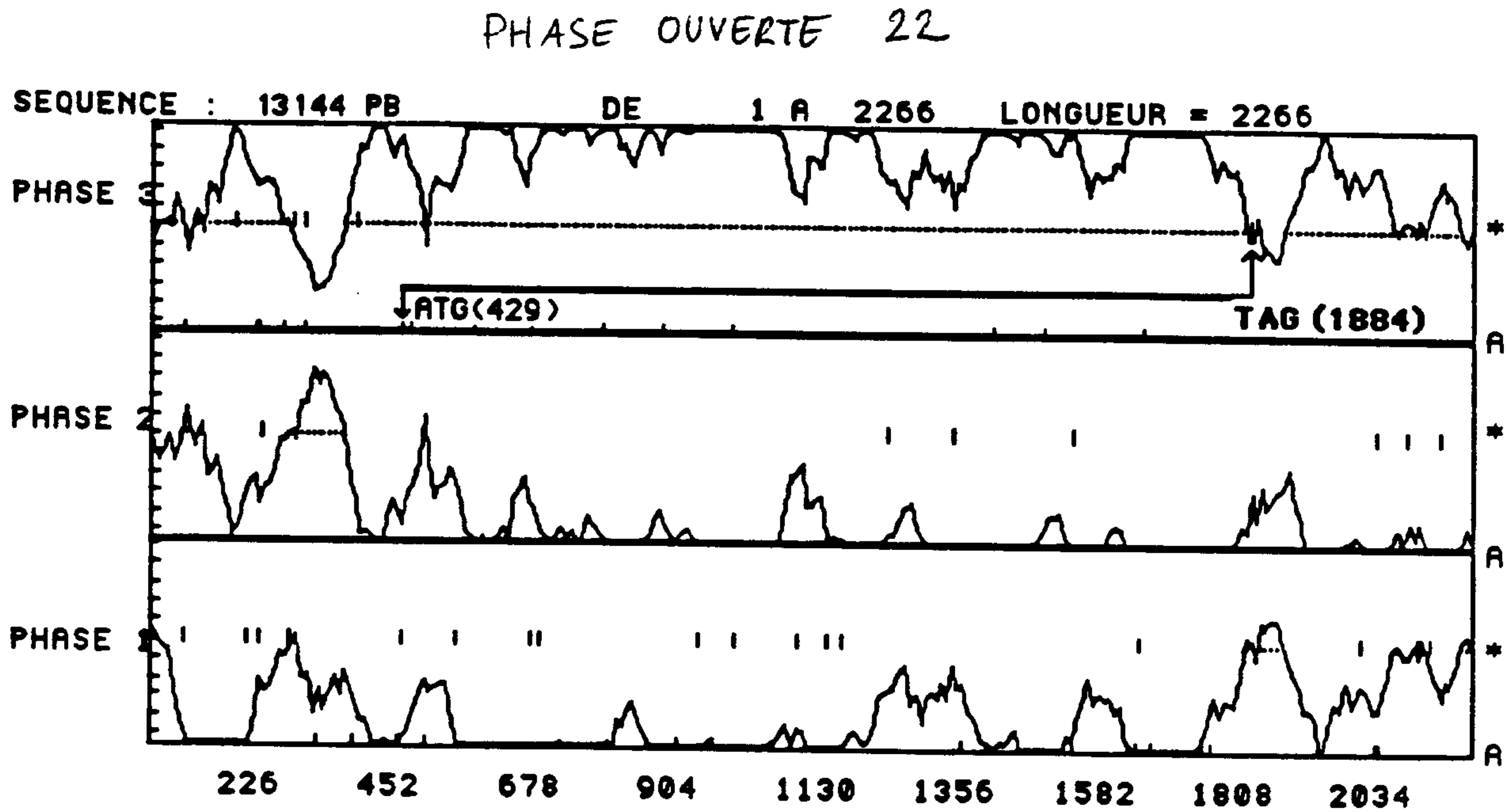
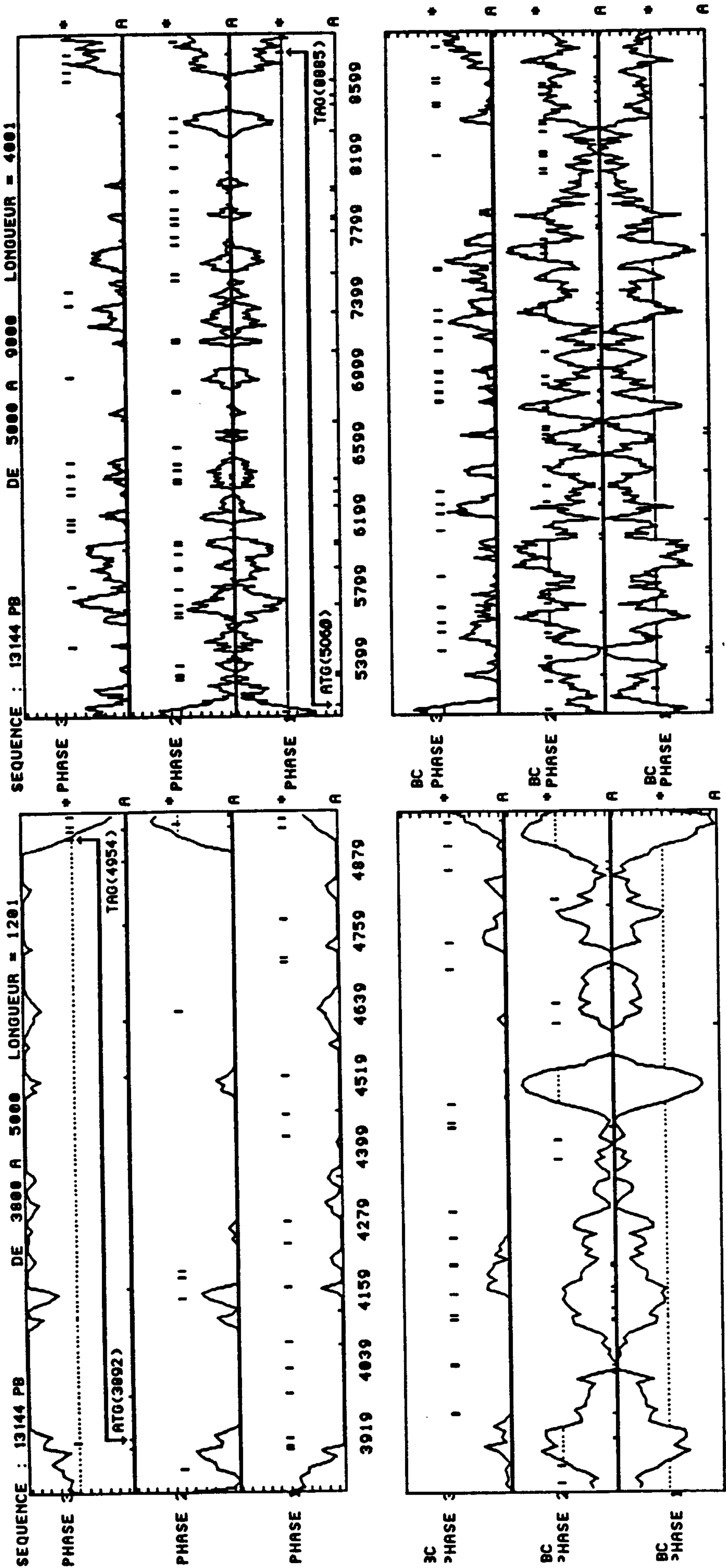


FIGURE 45

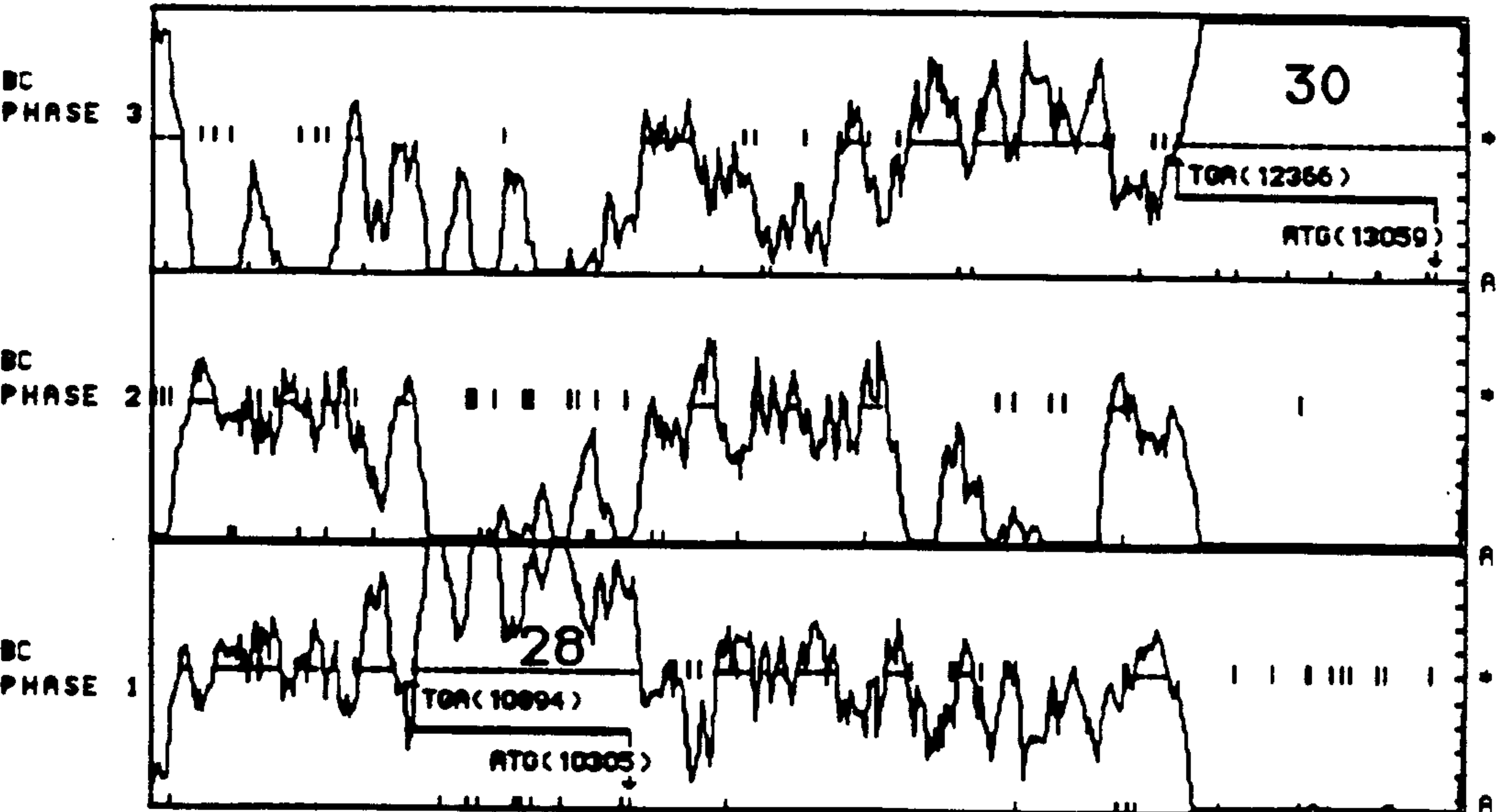
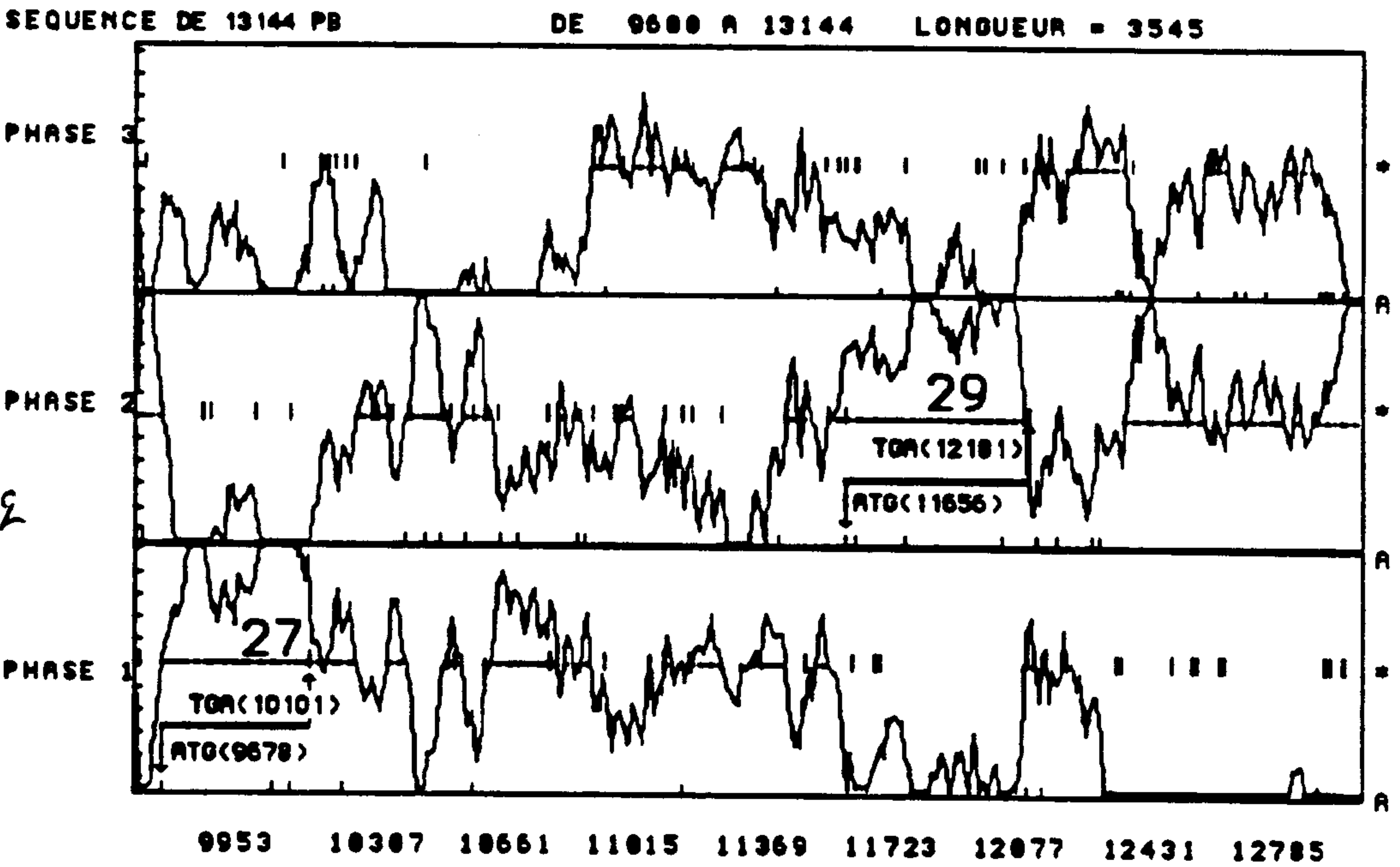
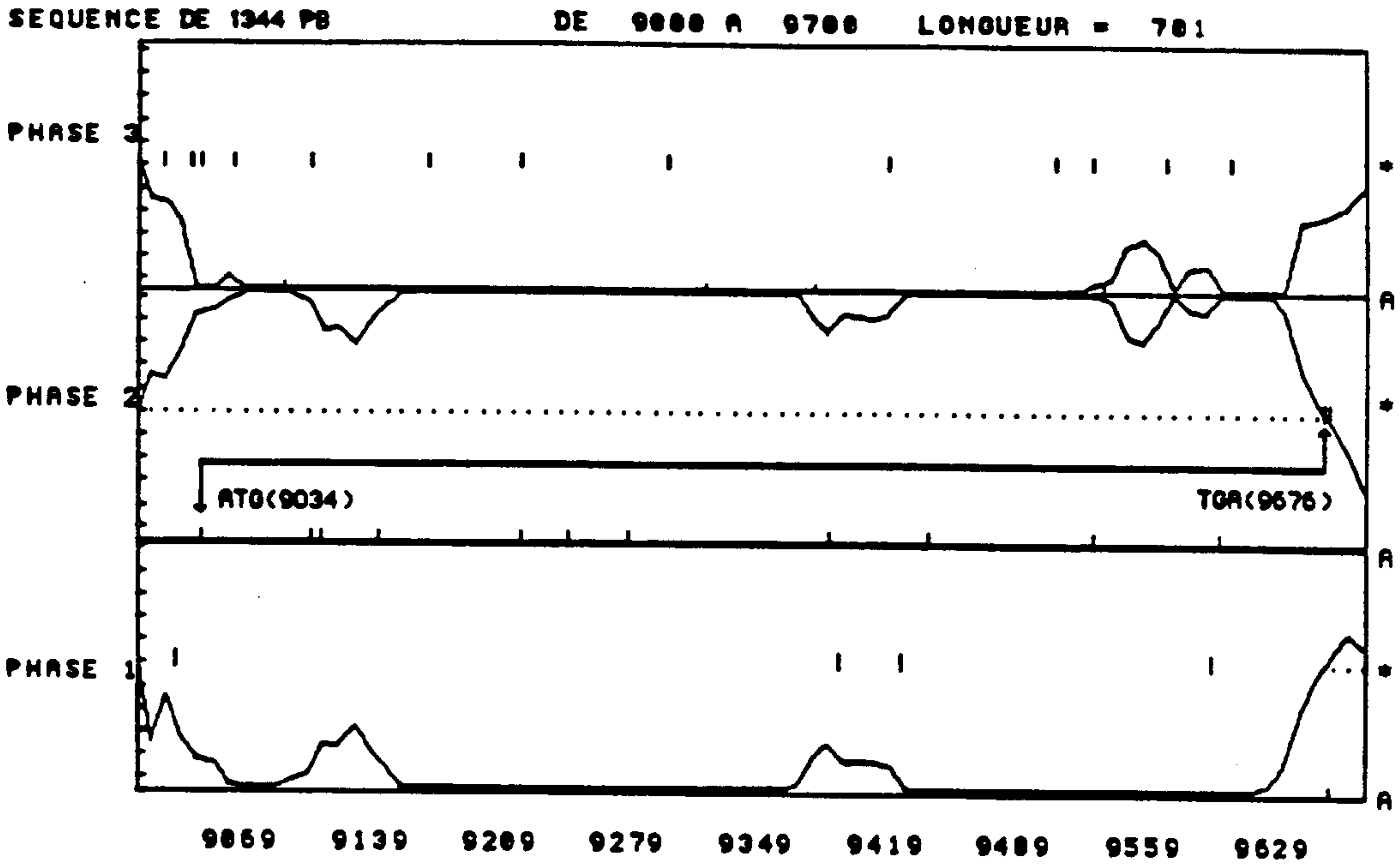


PHASE OUVERTE 25

PHASE OUVERTE 24

Fig.45.1

128 / 151



A ATO
• CODON STOP
BC BRIN COMPLEMENTAIRE

PHASES OUVERTES 26, 27, 28, 29, 30

FEUILLE DE REMPLACEMENT

FIG. 45.2

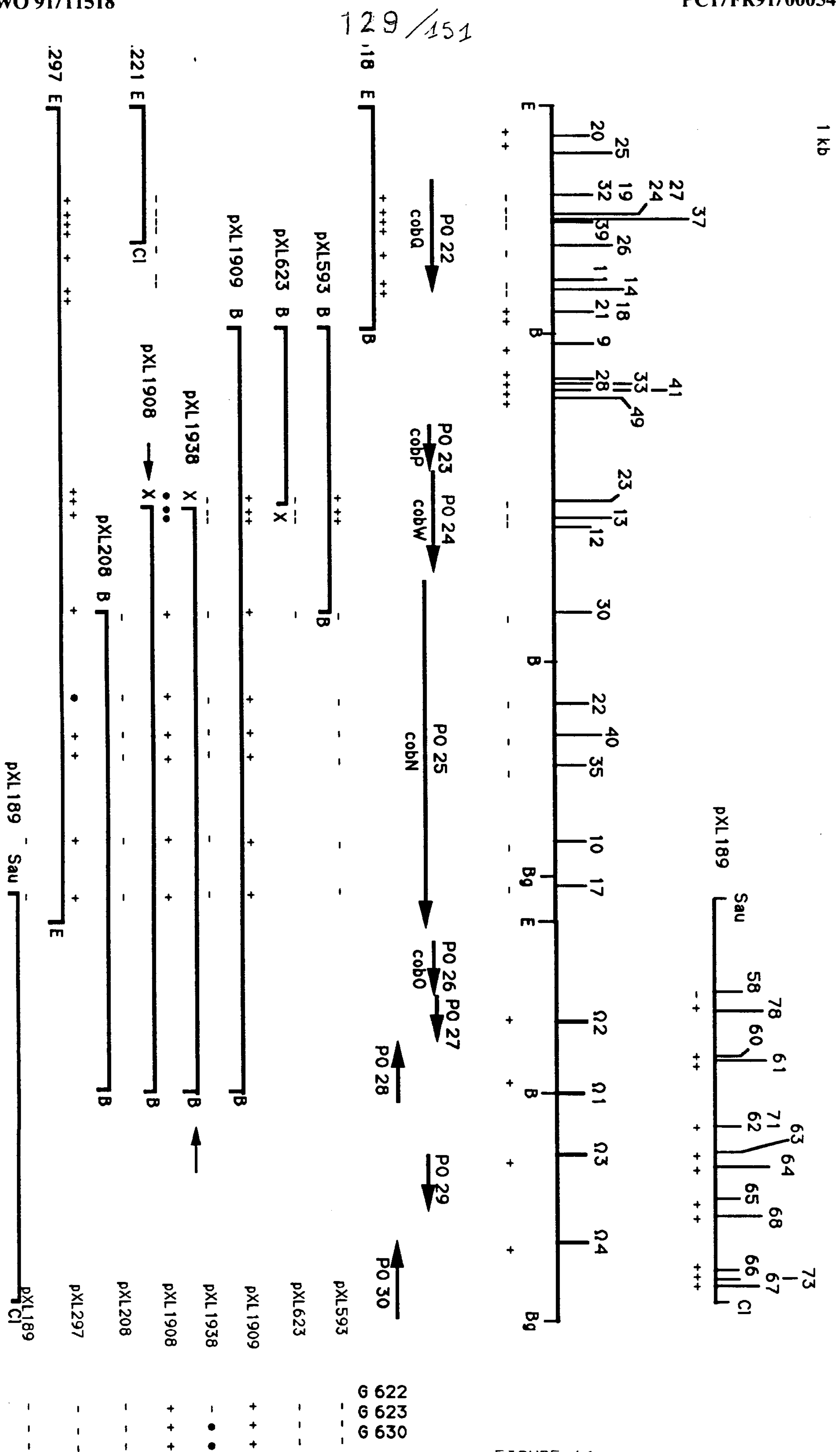


FIGURE 46

1451

Sequence de 13144 pb de 429 a 1886 gène cobQ

M T R R I M L Q G T G S D U G K S U L U A G L C R
ATGACACGCAGGATCATGTTGCAGGGAAACCGGCTCGGATGTCGGAAATCGGTATTGGTGGCGGGGCTCTGCCGG
429 439 449 459 469 479 489 499
L A A N Q G L K U R P F K P Q N M S N N A A U S D
CTTGCCGCCAATCAGGGCCTGAAGGTCCGGCCGTTCAAGCCGCAGAACATGTCGAACACGCCGCCGTTTCCGAC
504 514 524 534 544 554 564 574
D G G E I G A A Q M L Q A L A A R U P S S U H M N
GACGGCGGCGAGATCGGCCGCGCGCAATGGCTGCAGGCGCTGGCCGCGCGCGTGGCGTCTCGGTGCACATGAAC
579 589 599 609 619 629 639 649
P U L L K P Q S D U G S Q I U U Q G K U A G Q A R
CCGGTGCTCCTGAAGCCGCAGTCGGACGTGGGCAGCCAGATCGTCGTTCAAGGCAAGGTCGCCGGGCAGGCCAGG
654 664 674 684 694 704 714 724
G A E Y Q A L K P K L L G A U M E S F E Q I S A G
GGGCGGGAATATCAGGCGCTCAGCCCAAGCTGCTGGGCGCCGTCATGGAGAGTTTCAACAAATATCGGCCGGT
729 739 749 759 769 779 789 799
A D L U U U E G A G S P A E I N L R P G D I A N M
GCCGATCTCGTGGTGGTCAAGGCGCCGGCTCGCCGGCCGAATCAACCTCAGGCCCGGCGACATCGCCATATG
804 814 824 834 844 854 864 874
G F A T R A A N U P U U L U G D I D R G G U I A S L
GGCTTTGCGACACGGGCCAATGTGCCGGTCTGTGCTGGTCGGCGACATCGACCGCGGGGGGGTGATCGCCTCGCTG
879 889 899 909 919 929 939 949
U G T H A I L P E E D R R M U T G Y L I N K F R G
GTCGGCACGCATGCGATCCTGCCCGAGGAAGACCGGCGCATGGTGACCGGCTATCTCATCAACAGTTCCGCGGC
954 964 974 984 994 1004 1014 1024
D U T L F D D G I A A U N R Y T G W P C F G U U P
GACGTCACGCTGTTGACGACGGCATTGCTGCCGTCAACCGCTACACCGGCTGGCCCTGCTTCGGCGTCTGTCCG
1029 1039 1049 1059 1069 1079 1089 1099
W L K A A A R L P A E D S U U L E K L T R G E G A
TGGCTGAAGGCGGCGGCACGCCTGCCGGCGGAAGATTCCGTCTGTCTGGAGAGCTGACGCGCGGCGAGGGGCGG
1104 1114 1124 1134 1144 1154 1164 1174
A L K U A U P U L S R I A N F D D L D P L A A E P
GCGCTGAAGGTTGCCGTCCCGGTACTGTCGCGCATCGCCATTTTCGACGACCTCGATCCGCTCGCCGCCGAACCG
1179 1189 1199 1209 1219 1229 1239 1249
E I D L U F U R P G S P I P U D A G L U U I P G S
GAGATTGATCTCGTCTTCTGTCGGGCTGGCAGTCCCATTCGGTCTGACGCTGGCCTCGTCTCATTTCCGGGGTCTG
1254 1264 1274 1284 1294 1304 1314 1324
K S T I G D L I D F R A Q G W D A D L E A H U A R
AATCGACCATCGGCGACCTCATCGATTTCCGTGCGCAAGGGTGGGACCGTGACCTCGAACGTCATGTGCGCCGG
1329 1339 1349 1359 1369 1379 1389 1399
G G R U I G I C G G Y Q M L G R R U T D P L G I E
GGCGGCCGGGTCATCGGCATCTGCGGCGGCTACCAGATGCTCGGCCGGCGCGTCACCGATCCGCTCGGCATCGAG
1404 1414 1424 1434 1444 1454 1464 1474
G G E R A U E G L G L L E U E T E M A P E K T U A
GGCGGCGAACGTGCGGTGAGGGGCTCGGGCTGCTCGAGGTGAGACCGAGATGGCGCCGGAAAGACGGTGCGC
1479 1489 1499 1509 1519 1529 1539 1549
H S R A W S L E H D U U L E G Y E I H L G K T Q G
AACAGCCGCGCCTGGTCGCTGGAGCATGATGTGGTGTCTGAAGGCTACGAATCCATCTTGGCAAGACGCAAGGT
1554 1564 1574 1584 1594 1604 1614 1624
A D C G R P S U R I D N R A D G A L S A D G A U M
GCGGACTGTGGCCGGCCGTCGGTGCACATCGACAATCGCGCCGACGGCGCCCTTTCCGCCGATGGCCGCGTGATG
1629 1639 1649 1659 1669 1679 1689 1699
G T Y L H G L F T S D A Y A G A L L K S F G I E G
GGCACCTACCTGCATGGGCTCTTACCAGCGACGCCTATCGCGGCGCGCTGCTCAAGAGTTTCCGCATCGAAGGC
1704 1714 1724 1734 1744 1754 1764 1774
G A N H Y R Q S U D A A L D D U A N E L E A U L D
GGCGCCAACTACCGCCATCGGTGATGCGGCGCTCGACGATGTCGCGAACGAACCTGGAGGCTGTGCTCGAT
1779 1789 1799 1809 1819 1829 1839 1849
R A W L D E L L R H *
CGTCGCTGGCTGGACGAGTTGCTCAGGCACTAG
1854 1864 1874 1884

FIGURE 47

131 / 151

PROTEINE COBQ

PREMIER RESIDU = 1

DERNIER RESIDU = 485

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	11	2.27	1617.75	3.11
2	LEU	L	50	10.31	5654.20	10.88
3	ILE	I	23	4.74	2600.93	5.01
4	MET	M	10	2.06	1310.41	2.52
5	VAL	V	50	10.31	4953.42	9.53
6	SER	S	24	4.95	2088.77	4.02
7	PRO	P	23	4.74	2232.21	4.30
8	THR	T	15	3.09	1515.72	2.92
9	ALA	A	49	10.10	3480.82	6.70
10	TYR	Y	8	1.65	1304.51	2.51
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	1.44	959.41	1.85
13	GLN	Q	15	3.09	1920.88	3.70
14	ASN	N	16	3.30	1824.69	3.51
15	LYS	K	15	3.09	1921.42	3.70
16	ASP	D	34	7.01	3910.92	7.53
17	GLU	E	28	5.77	3613.19	6.96
18	CYS	C	4	0.82	412.04	0.79
19	TRP	W	6	1.24	1116.48	2.15
20	ARG	R	40	8.25	6244.04	12.02
21	GLY	G	57	11.75	3250.22	6.26
22	-	-	0	0.00	0.00	0.00

RESIDUS = 485

Masse moleculaire (monoisotopique) = 51950.1016

Masse moleculaire (moyenne) = 51982.3711

INDEX DE POLARITE (%) = 40.00

POINT ISOELECTRIQUE = 6.16

DO 260 (1mg/ml) = 0.558 DO 280 (1mg/ml) = 0.825

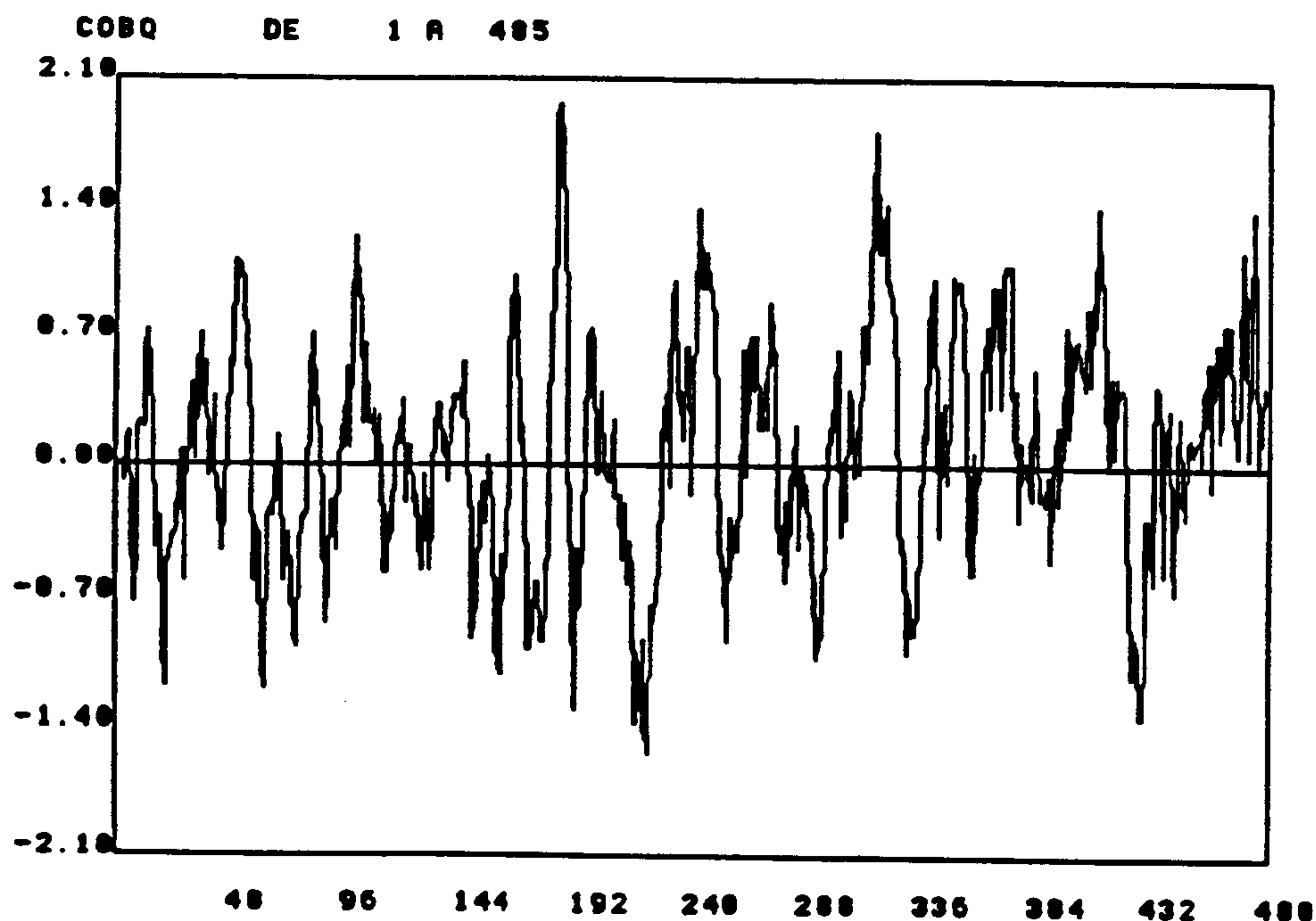


FIGURE 47.1

132/151

Sequence de 13144 pb

de

3364 a

3888

gène cobP

```

M S S L S A G P U L U L G G A R S G K S S F S E R
ATGAGCAGTCTCAGCGCCGGGCCCCGTGCTGGTCCTTGGCGGGCGCCCGTTCCGGCAAGTCCAGCTTTTCCGAGAGG
3364      3374      3384      3394      3404      3414      3424      3434
L U E A S G F T M H Y U A T G R A W D D E M R E R
CTCGTCGAAGCGTCCGGCTTCACCATGCATTATGTCGCCACGGGGCCGCGCCTGGGACGACGAAATGCGCGAGCGC
3439      3449      3459      3469      3479      3489      3499      3509
I D H H R T R R G E G W T T H E E P L D L U G I L
ATCGACCATCACCGGACGCGCCGCGGCGAGGGCTGGACGACGCATGAGGAGCCGCTCGATCTCGTCGGCATCCTC
3514      3524      3534      3544      3554      3564      3574      3584
R R I D D P S H U U L I D C L T L W U T N L M L E
AGACGCATCGATGATCCCAGCCATGTGGTCCTGATCGACTGCCTGACGCTATGGGTCACCAATCTCATGCTGGAA
3589      3599      3609      3619      3629      3639      3649      3659
E R D M T A E F A R A L U A Y L P E A R A R A L U F U
GAGCGCGACATGACGGCGGAGTTCGCCGCCCTTGTTGCGTATCTGCCCGAGGCGCGGGCGCGCCTCGTCTTTGTT
3664      3674      3684      3694      3704      3714      3724      3734
S H E U G L G I U P E N R M A R A E F A D H A G R L
TCCAATGAGGTCCGCCTCGGCATCGTGCCCGAGAACCGCATGGCCCGCGAGTTTCGCGACCATGCCGGCCGGCTT
3739      3749      3759      3769      3779      3789      3799      3809
H Q I U A E K S A E U Y F U A A G L P L K M K G *
CACCAGATCGTTGCGGAGAAATCCGCTGAAGTTTACTTTGTGCGGGCCGGTTTGCCGCTGAAAATGAAGGGTTGA
3814      3824      3834      3844      3854      3864      3874      3884

```

FIGURE 47.2

133/151

PROTEINE COBP

PREMIER RESIDU = 1
DERNIER RESIDU = 174

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	6	3.45	882.41	4.54
2	LEU	L	19	10.92	2148.60	11.06
3	ILE	I	6	3.45	678.50	3.49
4	MET	M	7	4.02	917.28	4.72
5	VAL	U	16	9.20	1585.09	8.16
6	SER	S	11	6.32	957.35	4.93
7	PRO	P	6	3.45	582.32	3.00
8	THR	T	8	4.60	808.38	4.16
9	ALA	A	17	9.77	1207.63	6.22
10	TYR	Y	3	1.72	489.19	2.52
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	4.02	959.41	4.94
13	GLN	Q	1	0.57	128.06	0.66
14	ASN	N	3	1.72	342.13	1.76
15	LYS	K	4	2.30	512.38	2.64
16	ASP	D	9	5.17	1035.24	5.33
17	GLU	E	16	9.20	2064.68	10.63
18	CYS	C	1	0.57	103.01	0.53
19	TRP	W	3	1.72	558.24	2.87
20	ARG	R	17	9.77	2653.72	13.66
21	GLY	G	14	8.05	798.30	4.11
22	-	-	0	0.00	0.00	0.00

RESIDUS = 174
Masse moleculaire (monoisotopique) = 19429.9473
Masse moleculaire (moyenne) = 19442.2637
INDEX DE POLARITE (%) = 43.68
POINT ISOELECTRIQUE = 6.71
DO 260 (1mg/ml) = 0.720 DO 280 (1mg/ml) = 1.042

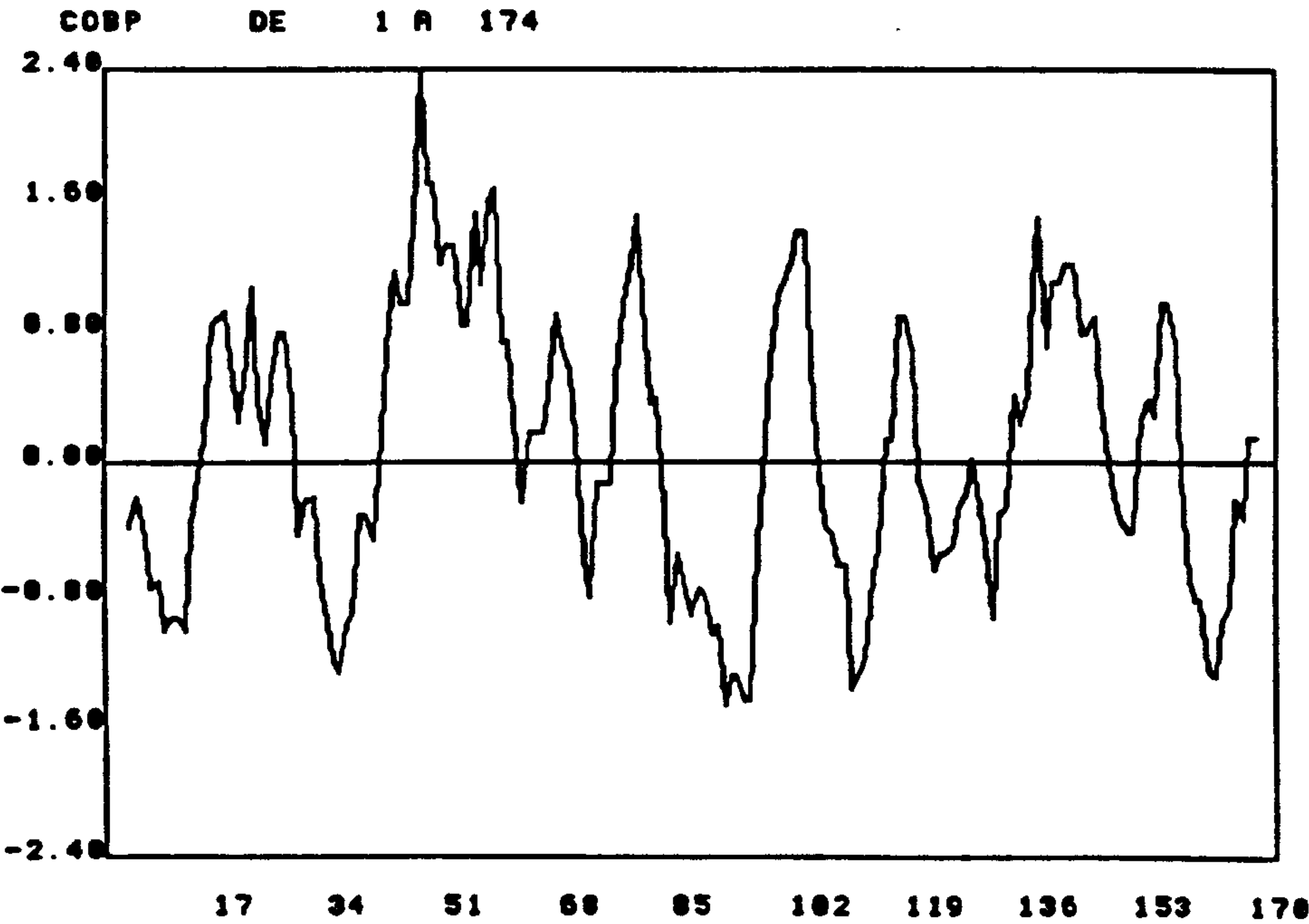


FIGURE 47.3

134/151

Sequence de 13144 pb

de 3892 a 4956 gène cobH

M T T A R A N Q G K I P A T U I T G F L G A G K T
 ATGACCACTGCGAGAGCCAACCAAGGGCAAGATCCCGGCGACCGTCATCACCGGCTTCCTCGGGCGCCGGCAAGACG
 3892 3902 3912 3922 3932 3942 3952 3962
 T M I R N L L Q N A D G K R I G L I I N E F G D L
 ACGATGATCCGCAACCTGCTGCAGAACGCCGACGGCAAGCGCATCGGCCTGATCATCAACGAGTTTCGGCGATCTT
 3967 3977 3987 3997 4007 4017 4027 4037
 G U D G D U L K G C G A E A C T E D D I I E L T N
 GCGTCGACGGCGATGTCTTGAGGGGCTGCGGTGCCGAGGCCTGCACCGAGGACGACATCATCGAGCTCACCRAAT
 4042 4052 4062 4072 4082 4092 4102 4112
 G C I C C T U A D D F I P T M T K L L E A E N R P
 GGCTGCATCTGCTGCACCGTGGCTGACGATTTTCATCCCGACCATGACGAAGCTGCTCGAGCGTGAAACCGTCTT
 4117 4127 4137 4147 4157 4167 4177 4187
 D H I I I E T S G L A L P Q P L I A A F N W P D I
 GACCACATCATCATCGAACCTCGGGCCTTGCCCTGCCGACGCCGCTGATCGCCGCTTTCAACTGGCCGGATATC
 4192 4202 4212 4222 4232 4242 4252 4262
 R S E U T U D G U U T U U D S A A U A A G R F A D
 CGCAGCGAAGTGACCGTCGATGGCGTCGTCACCGTGGTTCGACAGCGCCGCCGTTGCCGCTGGCCGCTTTGCCGAC
 4267 4277 4287 4297 4307 4317 4327 4337
 D H D K U D A L R U E D D N L D H E S P I E E L F
 GACCACGACAAGGTCGATGCGCTGCGCGTCGAGGACGACATCTCGATCACGAAGCCCGATCGAGGAGCTGTTT
 4342 4352 4362 4372 4382 4392 4402 4412
 E D Q L T A A D L I U L N K T D L I D A S G L K A
 GAGGATCAACTGACGGCTGCCGATCTCATCGTTCTCAACAAGACCGATCTGATCGATGCCTCCGGCCTCAAGGCC
 4417 4427 4437 4447 4457 4467 4477 4487
 U R D E U S S R T S R K P T M I E A K N G E U A A
 GTGCGCGACGAGGTGTCTTCGCGCACCAAGCCGCAAGCCACGATGATCGAGGGCAAAACGGCGAAGTCGCCGCT
 4492 4502 4512 4522 4532 4542 4552 4562
 A I L L G L G U G T E S D I A N R K S H H E M E H
 GCCATCCTGCTTGGCCTCGGTGTCGGCACGGAAGCGATATCGCCAACCGCAAGTCGCATCACGAGATGGAGCAC
 4567 4577 4587 4597 4607 4617 4627 4637
 E A G E E H D H D E F D S F U U E L G S I A D P A
 GAGGCAGGTGAGGAGCACGATCACGACGAGTTTCGACAGCTTCGTCGTCGAGCTCGGTTTCGATCGCCGATCCGGCC
 4642 4652 4662 4672 4682 4692 4702 4712
 A F I D R L K G U I A E H D U L R L K G F A D U P
 GCCTTCATCGATCGCCTGAGGGCGTAATCGCGGAGCACGACGTTCTGCGCCTCAAGGGTTTTGCAGACGTGCCC
 4717 4727 4737 4747 4757 4767 4777 4787
 G K P M R L L I Q A U G A R I D Q Y Y D R A W G A
 GGCAAGCCGATGCGCCTCCTGATCCAGGCGGTGCGCGCCCGCATCGACCAATATTACGACCGCGCCTGGGGCGCT
 4792 4802 4812 4822 4832 4842 4852 4862
 G E K R G T A L U U I G L H D M D E A A U A A A I
 GGCGAAAGCGCGGTACGCGCCTCGTCGTCATCGGCCTGCACGACATGGACGAGGGCGGCGGTGCGCGCCGCGATC
 4867 4877 4887 4897 4907 4917 4927 4937
 T A L U *
 ACCGCGCTCGTGTAG
 4942 4952

FIGURE 47.4

135 / 151

PROTEINE COBU
PREMIER RESIDU = 1
DERNIER RESIDU = 354

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	10	2.82	1470.68	3.86
2	LEU	L	32	9.04	3618.69	9.50
3	ILE	I	28	7.91	3166.35	8.31
4	MET	M	7	1.98	917.28	2.41
5	VAL	U	28	7.91	2773.92	7.28
6	SER	S	12	3.39	1044.38	2.74
7	PRO	P	11	3.11	1067.58	2.80
8	THR	T	21	5.93	2122.00	5.57
9	ALA	A	41	11.58	2912.52	7.64
10	TYR	Y	2	0.56	326.13	0.86
11	*	*	0	0.00	0.00	0.00
12	HIS	H	10	2.82	1370.59	3.60
13	GLN	Q	6	1.69	768.35	2.02
14	ASN	N	11	3.11	1254.47	3.29
15	LYS	K	15	4.24	1921.42	5.04
16	ASP	D	36	10.17	4140.97	10.87
17	GLU	E	27	7.63	3484.15	9.15
18	CYS	C	5	1.41	515.05	1.35
19	TRP	W	2	0.56	372.16	0.98
20	ARG	R	20	5.65	3122.02	8.19
21	GLY	G	30	8.47	1710.64	4.49
22	-	-	0	0.00	0.00	0.00

RESIDUS = 354
Masse moleculaire (monoisotopique) = 38097.4258
Masse moleculaire (moyenne) = 38121.1055
INDEX DE POLARITE (%) = 44.63
POINT ISOELECTRIQUE = 4.90
DO 260 (1mg/ml) = 0.268 DO 280 (1mg/ml) = 0.354

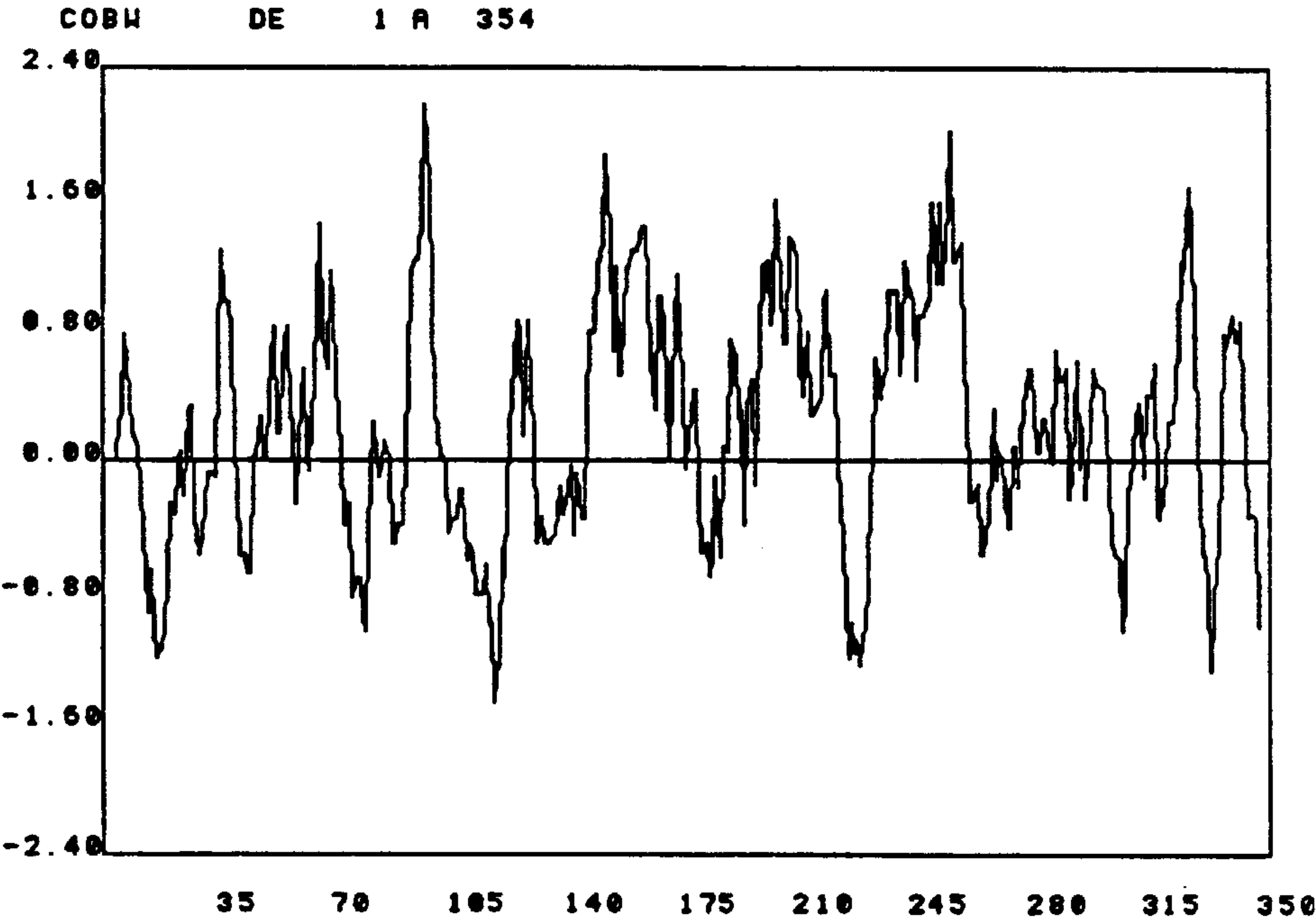


FIGURE 47.5

136/151

Sequence de 13144 pb de 5060 à 8887 gène cobH

M H L L L A Q K G T I A D G N E A I D L G Q T P A
 ATGCATCTGCTTCTCGCCCAGAAAGGAACGATCGCCGACGGCAACGAGGCGATCGACCTTGGGCAACGCGCGCC
 5060 5070 5080 5090 5100 5110 5120 5130
 D I L F L S A A D T E L S S I A A A H G R A R D G G
 GATATCCTTTTCTATCGGCCGCCGACACCGAGCTCTCCTCGATCGCCGCGGCTCACGGCCGACGCGACGGAGGC
 5135 5145 5155 5165 5175 5185 5195 5205
 L S L R I A S L M S L M H P M S U D T Y U E R T A
 TTGAGCCTGCGCATCGCCAGCCTGATGAGCCTGATGCACCCGATGTCGGTCGACACTTACGTCGAGCGCACGGCG
 5210 5220 5230 5240 5250 5260 5270 5280
 R H A K L I U U R P L G G A S Y F R Y L L E A L H
 CGTCACGCCAAGCTGATCGTCGTCCGGCCGCTCGGTGGCGCCAGCTATTTCCGTTATCTGCTGGAAGCCCTGCA
 5285 5295 5305 5315 5325 5335 5345 5355
 A A A U T H R F E I A U L P G D D K P D P G L E P
 GCGGCTGCCGTCACCCATCGTTTCGAGATCGCGGTTCTGCCGGGTGACGACAAGCCGGATCCGGGGCTGGAGCCT
 5360 5370 5380 5390 5400 5410 5420 5430
 F S T U A A D D R Q R L W A Y F T E G G S D N A G
 TTCTCCACCGTCGACGCCGACGACCGCCAGCGCCTTTGGGCTTACTTCACCGAAGGCGGCTCGGACATGCCGGG
 5435 5445 5455 5465 5475 5485 5495 5505
 L F L D Y A A A L U T G A E K P Q P A K P L L K A
 CTGTTTCTCGACTATGCGGCCGCACTGGTCACAGGTGCGGAGAGCCGACGCCGGCAAGCCCTGTTGAAGGCC
 5510 5520 5530 5540 5550 5560 5570 5580
 G I W W P G A G U I G U S E W Q S L U Q G R M U A
 GGCATCTGGTGGCCGGGTGCTGGTGTGATCGGCGTCAGCGAATGGCAGTCCCTTGTTCAGGGACGGATGGTAGCG
 5585 5595 5605 5615 5625 5635 5645 5655
 R E G F E P P T U G I C F Y R A L U Q S G E T R P
 AGGGAGGGATTGCAACCCCGACGGTCGGGATCTGCTTTTACCGCGCGCTCGTGCAGAGTGGCGAGACACGGCCT
 5660 5670 5680 5690 5700 5710 5720 5730
 U E A L I D A L E A E G U R A L P U F U S S L K D
 GTGGAGGCGCTGATCGATCGCTGGAGGCTGAAGGTGTGCGGGCACTGCCGGTGTGTTGTCTCAAGCCTCAAGGAT
 5735 5745 5755 5765 5775 5785 5795 5805
 A U S U G T L Q A I F S E A A P D U U M N A T G F
 GCCGTTTCCGTCGGCACGCTGCAGGCGATTTTTTCCGAGGCCGACCCGACGTGGTGATGAACGCCACTGGCTTT
 5810 5820 5830 5840 5850 5860 5870 5880
 A U S S P G A D R Q P T U L E S T G A P U L Q U I
 GCGGTCTCGTCGCCCGGTGCCGACCGTCAGCCGACGGTGCTGGAATCGACCGGTGCGCCGGTGCTGCAGGTGATT
 5885 5895 5905 5915 5925 5935 5945 5955
 F S G S S R A Q W E T S P Q G L M A R D L A M N U
 TTCTCCGGCTCGTCGCGGGCGCAATGGGAACGTCGCCGACGGGGCTGATGGCGCGGACCTCGCCATGAACGTG
 5960 5970 5980 5990 6000 6010 6020 6030
 A L P E U D G R I L A R A U S F K A A S I Y D A K
 GCACTCCCCGAAGTCGATGGCCGATCCTTGCGCGCGCCGTCTCCTTCAAGGCGGCGTCGATCTATGACGCCAAG
 6035 6045 6055 6065 6075 6085 6095 6105
 U E A N I U G H E P L E G R U R F A A D L A U N W
 GTGGAGGCCAATATCGTCGGCCATGAGCCGCTCGAAGGCCGGGTGCGCTTTGCCGCTGATCTTGCCGTCAACTGG
 6110 6120 6130 6140 6150 6160 6170 6180
 A N U R R A E P A E R R I A I U M A N Y P N R D G
 GCGAACGTGCGCCGGGACAGCCCGCCGAGCGCCGTATTGCCATCGTCATGGCCAACTATCCGAACCGCGACGGT
 6185 6195 6205 6215 6225 6235 6245 6255
 R L G H G U G L D T P A G T U E U L S A M A R E G
 CGCCTCGGCAACGGTGTGCGGCTCGACACGCCGGCCGGTACCGTCGAGGTGCTTAGCGCCATGGCGCGGGAGGC
 6260 6270 6280 6290 6300 6310 6320 6330
 Y A U G E U P A D G D A L I R F L M A G P T N A A
 TATGCGGTGCGTGAGGTTCGCCCGATGGCGACGCGCTGATCCGCTTTCTGATGGCCGGGCGGACCAATGCGGCG
 6335 6345 6355 6365 6375 6385 6395 6405
 S H D R E I R E R I S L N D Y K T F F D S L P K Q
 AGCCATGACCGTGAAATCCGCGAGCGTATTTTCGCTGAACGATTACAAACGTTCTTCGATTGCTTCCGAACAG
 6410 6420 6430 6440 6450 6460 6470 6480
 I K D E U A G R W G U P E A D P F F L D G A F A L
 ATAAGGATGAAGTTGCCGGTCGCTGGGGCGTGCCGGAGGCCGATCCCTTTTCTCGATGGCGCCTTCGCGCTG
 6485 6495 6505 6515 6525 6535 6545 6555

FIGURE 47.6

137/151

P L A R F G E U I U G I Q P A R G Y N I D P K E S
CCGCTCGCCCGCTTCGGCGAGGTGATCGTCGGCATCCAACCGGCGCGGGCTACACATCGATCCGAAGGAAGC
6560 6570 6580 6590 6600 6610 6620 6630
Y H S P D L U P P H G Y L A F Y A F L R Q Q F G A
TACCATTCCCCGGACCTCGTGCCGCGCATGGCTATCTCGCCTTCTACGCCTTCTGCGCCAGCAGTTTCGGAGCG
6635 6645 6655 6665 6675 6685 6695 6705
Q A I U H N G K H G N L E M L P G K A L A L S E T
CAGGCGATCGTCCACATGGGCAAGCACGGCAATCTCGAATGGCTGCCGGGCAAGGCGCTGGCGCTGTCGGAACC
6710 6720 6730 6740 6750 6760 6770 6780
C Y P E A I F G P L P H I Y P F I U N D P G E G T
TGCTATCCCGAAGCGATCTTCGGGCGCTGCCGCACATCTATCCCTTCATCGTCAACGATCCGGGCGAAGGTACG
6785 6795 6805 6815 6825 6835 6845 6855
Q A K R R T S A U I I D H L T P P L T R A E S Y G
CAGGCCAAGCGCCGCACCAGCGCCGTCATCATCGACCACCTGACCCCGCCCTTGACGCGCGCCGAATCCTACGGC
6860 6870 6880 6890 6900 6910 6920 6930
P L K D L E A L U D E Y Y D A A G G D P A R A L R L
CCGCTCAAGGATCTGGAAGCGCTCGTCGACGAATATTACGACGCCCGCGGGTGTATCCGCGCCGCTCAGGCTG
6935 6945 6955 6965 6975 6985 6995 7005
L S R Q I L D L U R D I G L D S D A G I D R G D S
CTCAGCCGCCAGATCCTCGATCTCGTGCGGACATCGGCCTCGACAGCGACGCAGGCATCGACAGGGGCGACAGC
7010 7020 7030 7040 7050 7060 7070 7080
D D K A L E K L D A Y L C D L K E M Q I R D G L H
GACGACAAGGCGCTGGAAAGCTCGACGCCTATCTCTGCGACCTCAAGGAATGCAGATCCGCGACGGCCTGCAC
7085 7095 7105 7115 7125 7135 7145 7155
I F G U A P E G R L L T D L T U A L A R U P R G L
ATCTTCGGCGTTGCGCCGGAAGGGCGGTTGTTGACGGACCTCACCGTAGCGCTGGCGCGCGTGCCCCGAGGTCTC
7160 7170 7180 7190 7200 7210 7220 7230
G E G G D Q S L Q R A I A A D A G L R G F A I P T
GGCGAGGGCGGCGACCAGAGCCTGCAGCGGGCGATCGCAGCGGATGCGGGGCTGCGTGGGTTTGCTATTCCCACC
7235 7245 7255 7265 7275 7285 7295 7305
S A G G N P A R D A Q P F D P L D C U M S D T M T
TCGGCGGGGGCAACCCCGCACGCGACGCCCAACCCCTTCGACCCGCTCGACTGCGTCATGTCCGACACCTGGACA
7310 7320 7330 7340 7350 7360 7370 7380
G P K P S I L A D L S D A P M R T A G D T U E R I
GGCCCGAACCCTCATCCTCGCTGACCTCTCGGACGCCCCCTGGCGCACCGCCGGCGATACGGTCGAGCGCATC
7385 7395 7405 7415 7425 7435 7445 7455
E L L A A N L U S G E L A C P D H M A N T R A U L
GAGTTGCTTGCCGCAATCTCGTGTCGGGTGAAGTGGCTTGCCCGGACCACTGGGCCAACACCCGCGCCGTGCTC
7460 7470 7480 7490 7500 7510 7520 7530
G E I E T R L K P S I S N S G A A E M T G F L T G
GGCGAATCGAACGCGCCTGAAGCCGTCGATTTCAACTCGGGTGCCGCGGAGATGACCGGCTTCTCACCAGGT
7535 7545 7555 7565 7575 7585 7595 7605
L S G R F U A P G P S G A P T R G R P D U L P T G
CTCAGCGGCGCTTCGTGCCCCCGGTCCATCGGGCGCGCCGACGCGCGGCGCGGCGGATGTGTTGCCGACGGGG
7610 7620 7630 7640 7650 7660 7670 7680
R H F Y S U D S R A U P T P A A Y E L G K K S A E
CGCAATTTCTACTCGGTGACAGCCGCGCCGTGCCGACGCGGGCGGCTTACGAGCTTGGAAGAATCGGCCGAG
7685 7695 7705 7715 7725 7735 7745 7755
L L I R A Y L Q D H G E M P S S F G L T A M G T A
CTTCTGATCCGCGCTACCTGCAGGACCATGGCGAATGGCCGTCCTCCTTTGGCCTGACCGCCTGGGGCACGGCG
7760 7770 7780 7790 7800 7810 7820 7830
H M R T G G D D I A Q A L A L I G A K P T M D M U
AACATGCGCACCGGCGGGCGACGACATCGCCAGGCCCTGGCGCTGATCGGCGCCAGCCACCTGGGACATGGTC
7835 7845 7855 7865 7875 7885 7895 7905
S R A U M G Y E I U P L A U L G A P R U D U T L R
TCTCGCCGGGTGATGGGCTACGAGATCGTGCCGCTCGCAGTCCTCGGCCGCCACGCGTCGACGTGACCTTGCGC
7910 7920 7930 7940 7950 7960 7970 7980
I S G F F A D A F P D Q I A L F D K A I R A U A L
ATTTCCGGCTTCTTCGCGATGCCTTCCCGGACCAGATCGCGCTCTTCGACAAGGCGATCCGCGCCGTGCGCTG
7985 7995 8005 8015 8025 8035 8045 8055
E E D D A D N M I A A R M A A E S R A R L E A E G U
GAGGAAGACGATGCCGACAACATGATCGCCGACGCGATGCGGGCGGAAGCCGGCGGCTGGAGGCCGAAGGCGTG
8060 8070 8080 8090 8100 8110 8120 8130

FIGURE 47.7

138 / 151

E A A E A A R R A S Y R U F G A K P G A Y G A A L
GAAGCCGCCGAGGCCGCGCGTCGCGCCTCCTACCGCGTCTTTGGCGCAAGCCCGGTGCCTATGGCGCCGCCCTG
8135 8145 8155 8165 8175 8185 8195 8205
Q A L I D E K G W E T K A D L A E A Y L T W G A Y
CAGGCGCTGATCGACGAGAGGGCTGGGAARCCAAAGCAGATCTCGCCGAGGCCCTATCTTACCTGGGGCGCCTAT
8210 8220 8230 8240 8250 8260 8270 8280
A Y G A G E E G K A E R D L F E E R L R T I E A U
GCCTATGGCGCCGGCGAGGAGGGCAAGGCCGAGCGCGATCTTTTCGAGGAGCGCCTGCGCACGATAGAGGCGGTG
8285 8295 8305 8315 8325 8335 8345 8355
U Q N Q D N A E H D L L D S D D Y Y Q F E G G M S
GTGCAGAACCCAGGACACCGCGAGCAGATCTGCTCGACAGCGACGACTACTACCGTTTCGAAGGCGGCATGAGC
8360 8370 8380 8390 8400 8410 8420 8430
A A A E Q L G G H R P A I Y H N D H S R P E K P U
GCTGCCGCCGAACAGCTCGGCGGTACCGTCCGGCGATCTACCACACGACCATTCCTCGTCCGGAAAGCCTGTG
8435 8445 8455 8465 8475 8485 8495 8505
I R S L E E E I G R U U R A R U U N P K W I D G U
ATCCGGTCGCTCGAAGAGAGATCGGCCGCGTGGTCCGGGCCCGCGTCGTCATCCCAAGTGGATCGATGGCGTC
8510 8520 8530 8540 8550 8560 8570 8580
M A H G Y K G A F E I A A T U D Y N F A F A A T T
ATGCGCCACGGATACAAGGGCGCCTTCGAGATCGCTGCCACGGTCGACTACATGTTGCCTTTGCCGCGACCACG
8585 8595 8605 8615 8625 8635 8645 8655
G A U R D H H F E A A Y Q A F I U D E R U A D F M
GGTGCGGTGC GCGACCATCATTTTCGAGGCCGCTTATCAGGCGTTTCATTGTCGACGAGCGCGTGGCTGACTTCATG
8660 8670 8680 8690 8700 8710 8720 8730
R D K N P A A F A E L A E R L L E A I D R N L W T
CGCGACAGAACCCGGCCGCCTTTGCCGAGCTTGCCGAACGCCTGCTTGAGCAATCGACCGCAATCTCTGGACG
8735 8745 8755 8765 8775 8785 8795 8805
P A S N S A R F E L A G I G T A A T R L R A G N E
CCGCGCTCGAATTCGGCGCGGTTTGAAGTTGCCGGCATCGGCACGGCAGCAACCCGGCTTCGTGCCGGCAATGAA
8810 8820 8830 8840 8850 8860 8870 8880
*
TAG
8885

FIGURE 47.8

139/151

PROTEINE COBN			PREMIER RESIDU = 1			
			DERNIER RESIDU = 1275			
			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	48	3.76	7059.28	5.12
2	LEU	L	121	9.49	13683.17	9.92
3	ILE	I	60	4.71	6785.04	4.92
4	MET	M	24	1.88	3144.97	2.28
5	VAL	V	82	6.43	8123.61	5.89
6	SER	S	64	5.02	5570.05	4.04
7	PRO	P	76	5.96	7376.01	5.35
8	THR	T	53	4.16	5355.53	3.88
9	ALA	A	180	14.12	12786.68	9.27
10	TYR	Y	35	2.75	5707.22	4.14
11	*	*	0	0.00	0.00	0.00
12	HIS	H	24	1.88	3289.41	2.38
13	GLN	Q	32	2.51	4097.87	2.97
14	ASN	N	30	2.35	3421.29	2.48
15	LYS	K	34	2.67	4355.23	3.16
16	ASP	D	90	7.06	10352.42	7.50
17	GLU	E	85	6.67	10968.62	7.95
18	CYS	C	5	0.39	515.05	0.37
19	TRP	W	18	1.41	3349.43	2.43
20	ARG	R	99	7.76	15454.01	11.20
21	GLY	G	115	9.02	6557.47	4.75
22	-	-	0	0.00	0.00	0.00

RESIDUS = 1275
Masse moleculaire (monoisotopique) = 137970.5000
Masse moleculaire (moyenne) = 138055.8594
INDEX DE POLARITE (%) = 40.08
POINT ISOELECTRIQUE = 5.42
DO 260 (mg/ml) = 0.693 DO 280 (mg/ml) = 1.027

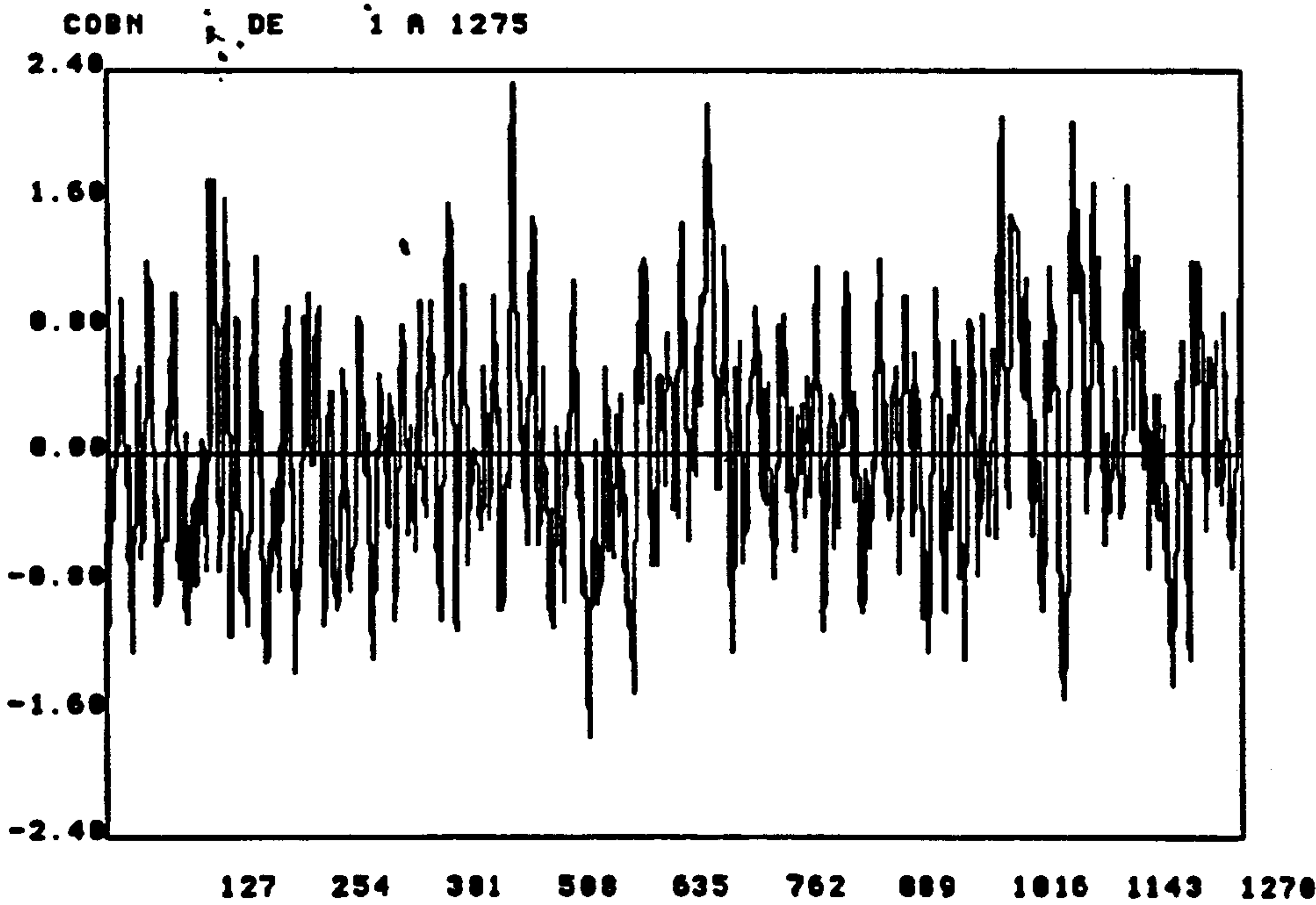


FIGURE 47.9

140/151

Sequence de 13144 pb 9034 a 9678 gène cobQ

M S D E T T U G G E A P A E K D D A R H A M K M A
 ATGAGCGACGAGACGACAGTAGGCGGCGAAGCCCCGGCCGAGAAGGACGATGCCCGCCACGCCATGAGATGGCG
 9034 9044 9054 9064 9074 9084 9094 9104
 K K K A A R E K I M A T K T D E K G L I I U N T G
 AAGAAGAAGGCGACGCGGAAAGATCATGGCGACGAAGACCGACGAGAAGGGTCTGATCATCGTCAACACCGGC
 9109 9119 9129 9139 9149 9159 9169 9179
 K G K G K S T A G F G M I F R H I A H G M P C A U
 AAAGGCAAGGGCAAGTCGACCGCCGGCTTCGGCATGATCTTCGCCCATATCGCCACGGCATGCCCTGCGCCGTC
 9184 9194 9204 9214 9224 9234 9244 9254
 U Q F I K G A M A T G E R E L I E K H F G D U C Q
 GTGCAGTTTCATCAAGGGTGGCATGGCAACCGGCGAGCGCGAGTTGATCGAGAAGCATTTCGGCGATGTCTGCCAG
 9259 9269 9279 9289 9299 9309 9319 9329
 F Y T L G E G F T W E T Q D R A A R D U A M A E K A
 TTCTACACGCTCGGCGAGGGCTTCACCTGGGAACGCGAGGATCGCGCCCGCGATGTTGCGATGGCTGAAAGGCC
 9334 9344 9354 9364 9374 9384 9394 9404
 W E K A K E L I R D E R N S M U L L D E I N I A L
 TGGGAGAAGGCGAAGGAAGTCCGTGACGAGCGCAACTCGATGGTGCTGCTCGACGAGATCAACATTGCTCTG
 9409 9419 9429 9439 9449 9459 9469 9479
 R Y D Y I D U A E U U R F L K E E K P H M T H U U
 CGCTACGACTACATCGACGTCGCCGAAGTGGTGGCTTCCTGAAGGAAGAAAGCCGCACATGACGCATGTGGTG
 9484 9494 9504 9514 9524 9534 9544 9554
 L T G R N A K E D L I E U A D L U T E M E L I K H
 CTCACCGGCCGCAACGCGAAGAGACCTGATCGAAGTCGCCGATCTCGTCACTGAGATGGAGCTGATCAAGCAT
 9559 9569 9579 9589 9599 9609 9619 9629
 P F R S G I K A Q Q G U E F *
 CCGTTCCGTTCCGGCATCAAGGCGCAGCAGGGCGTGGAGTTCTGA
 9634 9644 9654 9664 9674

FIGURE 47.10

141/151

PROTEINE COBO
PREMIER RESIDU = 1
DERNIER RESIDU = 214

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	9	4.21	1323.62	5.51
2	LEU	L	12	5.61	1357.01	5.65
3	ILE	I	14	6.54	1583.18	6.59
4	MET	M	11	5.14	1441.45	6.00
5	VAL	U	15	7.01	1486.03	6.19
6	SER	S	4	1.87	348.13	1.45
7	PRO	P	4	1.87	388.21	1.62
8	THR	T	13	6.07	1313.62	5.47
9	ALA	A	23	10.75	1633.85	6.80
10	TYR	Y	3	1.40	489.19	2.04
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	3.27	959.41	4.00
13	GLN	Q	5	2.34	640.29	2.67
14	ASN	N	4	1.87	456.17	1.90
15	LYS	K	21	9.81	2689.99	11.20
16	ASP	D	13	6.07	1495.35	6.23
17	GLU	E	23	10.75	2967.98	12.36
18	CYS	C	2	0.93	206.02	0.86
19	TRP	W	2	0.93	372.16	1.55
20	ARG	R	12	5.61	1873.21	7.80
21	GLY	G	17	7.94	969.36	4.04
22	-	-	0	0.00	0.00	0.00

RESIDUS = 214
Masse moleculaire (monoisotopique) = 24012.2500
Masse moleculaire (moyenne) = 24027.6973
INDEX DE POLARITE (%) = 47.66
POINT ISOELECTRIQUE = 6.94
DO 260 (mg/ml) = 0.443 DO 280 (mg/ml) = 0.612

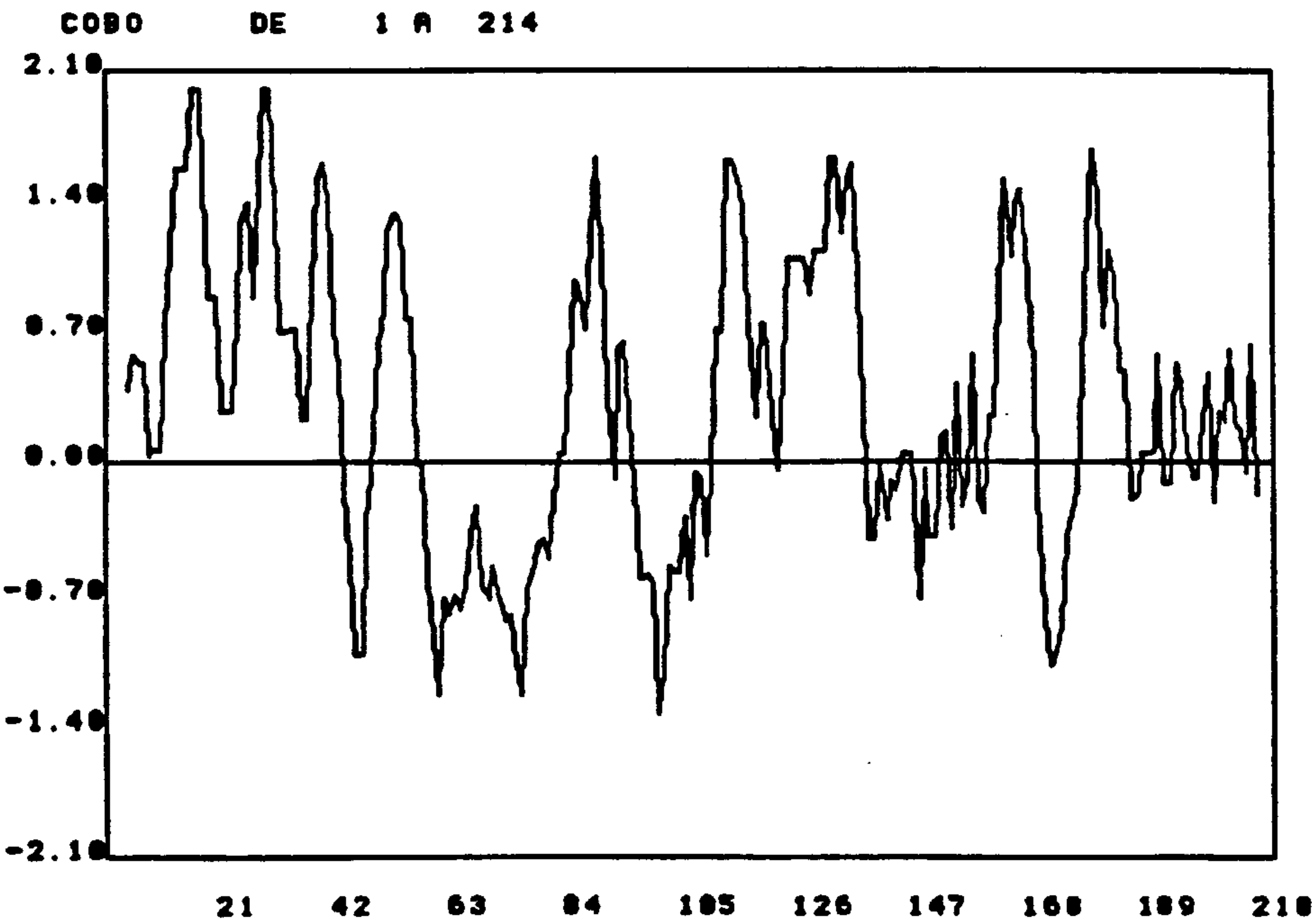


FIGURE 47.11

A

PL. 142/151

Séquence NH₂-terminale de la SUMT de M. ivanovii**VVYLVGAGPGDPELITLKAVNVLK-ADVVL****923****946**

Oligonucléotide sens 946 (27-mers)

		P	G	D	P	E	L	
5'	CGCGGAATTC	CCA	GGA	GAT	CCA	GAA	CT	3'
	EcoRI	T	T	C	T	G		
		C	C		C			
		G	G		G			

Oligonucléotide sens 923 (27-mers)

		V	Y	L	V	G	A	
5'	CGCGGAATTC	GTA	TAT	CTA	GTA	GGA	GC	3'
	EcoRI	G	C	T	T	T		
					G			

Séquence NH₂-terminale d'un fragment issu d'une digestion trypsique de la SUMT de M. ivanovii**IITGTLENIAGK****947**

Oligonucléotide antisens 947 (25-mers)

		N	E	L	T	G	
5'	CGCGAAGCTT	GTT	TTC	TAG	AGT	ACC	3'
	HindIII	A	C	A	T	T	
					G	C	
						G	

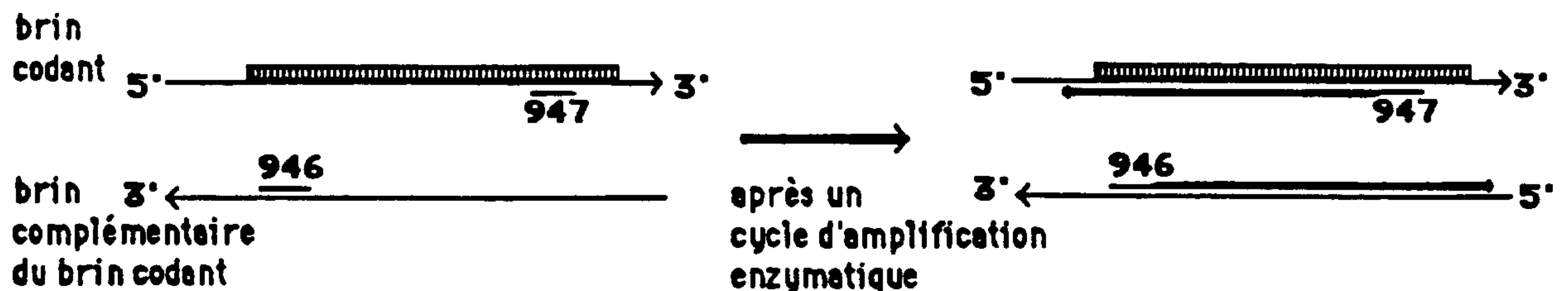
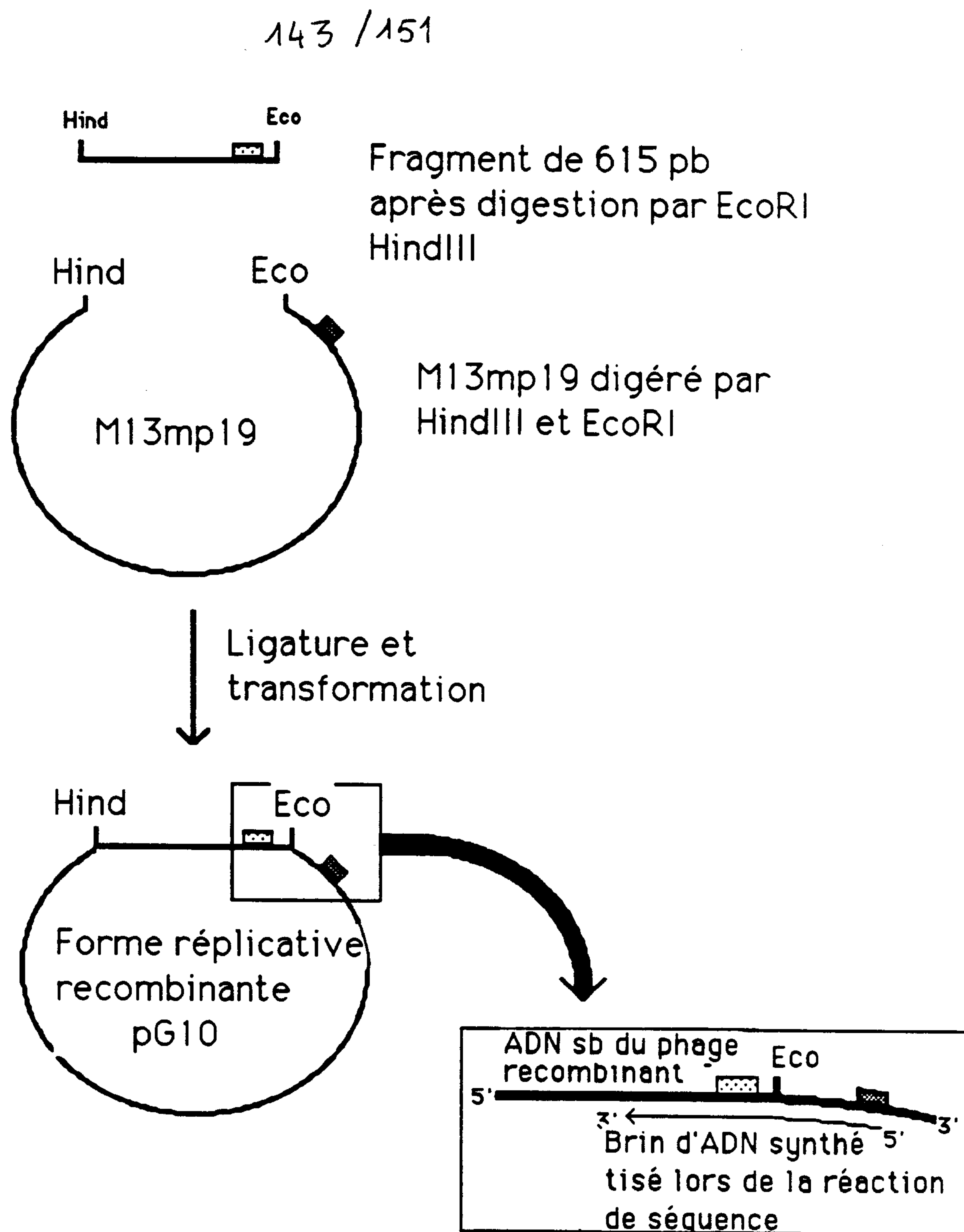
B
 gène de structure de la SUMT de M. ivanovii

FIGURE 48



- Site d'hybridation du primer -20 de
l'ADN sb du phage M13mp19
- ▤ Séquence complémentaire de
l'oligonucléotide sens 946

FIGURE 49

PL. 144 | 15 |

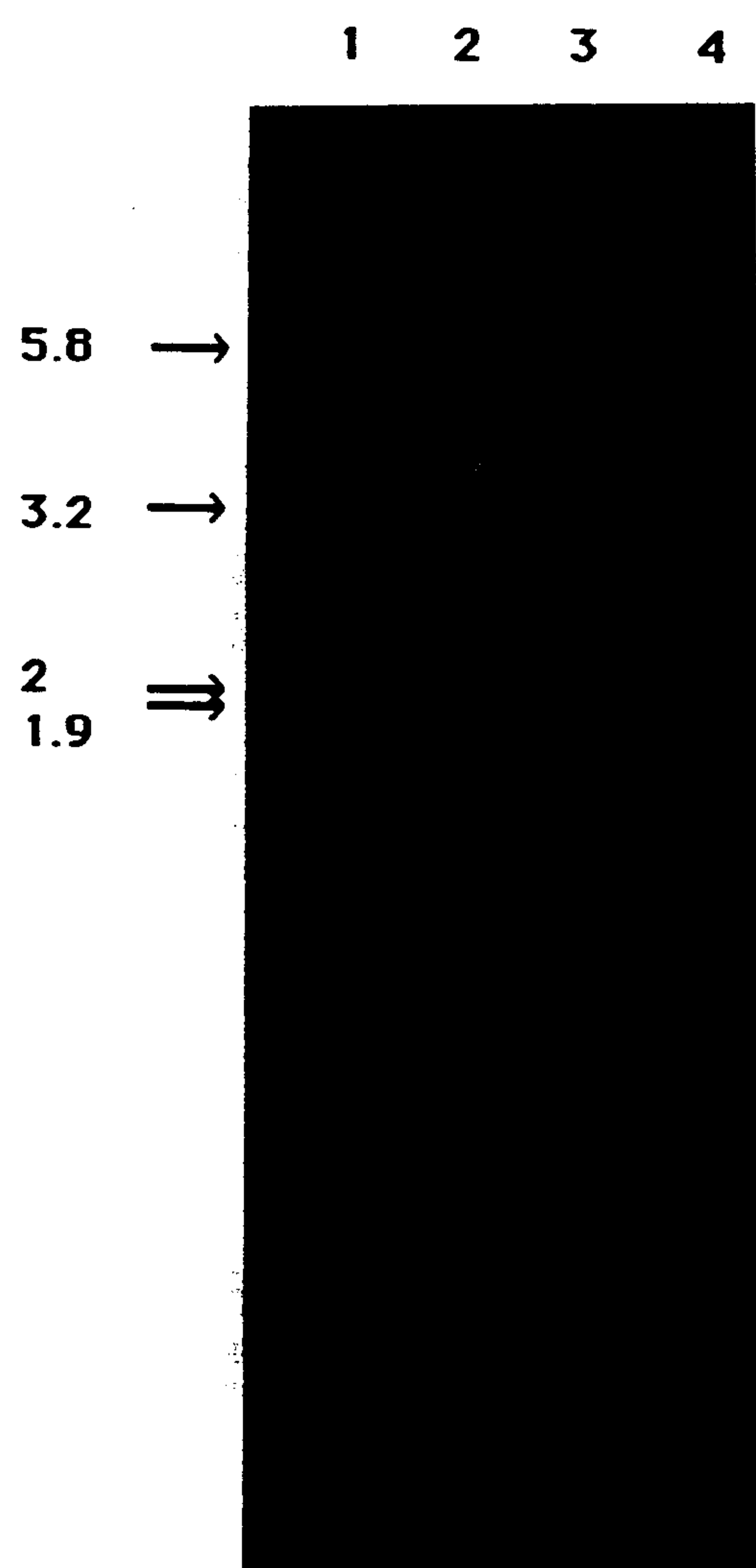


FIGURE 50

• PL: 145/151

10	20	30	40	50	60
CCATAATTCT	TTTATAATTT	AAACGGTGAA	CACATGGTAG	TTTATTTAGT	AGGTGCGGGT
GGTATTAAGA	AAATATTAAA	TTTGCCACTT	GTGTACCATC	AAATAAATCA	TCCACGCCCA
70	80	90	100	110	120
CCAGGAGATC	CCGAACCTAT	CACTCTCAAA	GCTGTAAACG	TGTTAAAAAA	AGCGGATGTT
GGTCCTCTAG	GGCTTGAATA	GTGAGAGTTT	CGACATTTGC	ACAATTTTTT	TCGCCTACAA
130	140	150	160	170	180
GTACTGTACG	ACAAACCTGC	AAATGAAGAA	ATTTTAAAGT	ATGCTGAAGG	TGCAAAACTA
CATGACATGC	TGTTTGGACG	TTTACTTCTT	TAAATTTTCA	TACGACTTCC	ACGTTTTGAT
190	200	210	220	230	240
ATATATGTCG	GAAAACAAGC	AGGACATCAT	TACAAATCTC	AAAATGAAAT	CAATACTCTT
TATATACAGC	CTTTTGTTTCG	TCCTGTAGTA	ATGTTTAGAG	TTTTACTTTA	GTTATGAGAA
250	260	270	280	290	300
CTTGTTGAAG	AAGCAAAAGA	AAATGATTTA	GTAGTACGCC	TTAAAGGTGG	AGACCCCTTT
GAACAACTTC	TTCGTTTTCT	TTTACTAAAT	CATCATGCGG	AATTTCCACC	TCTGGGGAAA
310	320	330	340	350	360
GTATTTGGAA	GAGGAGGCGA	GGAAATTCTG	GCCCTTGTAG	AAGAAGGAAT	TGATTTTGAG
CATAAACCTT	CTCCTCCGCT	CCTTTAAGAC	CGGGAACATC	TTCTTCCTTA	ACTAAAACTC
370	380	390	400	410	420
TTAGTTCCAG	GGGTAACCTC	TGCAATTGGA	GTTCCAACAA	CAATTGGGCT	TCCAGTTACT
AATCAAGGTC	CCCATTTGAAG	ACGTTAACCT	CAAGGTTGTT	GTAAACCCGA	AGGTCAATGA
430	440	450	460	470	480
CATAGAGGTG	TTGCAACATC	GTTTACAGTT	GTTACAGGTC	ATGAAGACCC	AACAAAATGC
GTATCTCCAC	AACGTTGTAG	CAAATGTCAA	CAATGTCCAG	TACTTCTGGG	TTGTTTTACG
490	500	510	520	530	540
AAGAAACAGG	TAGGATGGGA	CTTTAAAGCA	GATACTATTG	TAATACTTAT	GGGTATTGGA
TTCTTTGTCC	ATCCTACCCT	GAAATTTTCG	CTATGATAAC	ATTATGAATA	CCCATAACCT
550	560	570	580	590	600
AATTTAGCTG	AAAATACAGC	AGAAATTATG	AAACATAAAG	ATCCTGAAAC	TCCAGTTTGT
TTAAATCGAC	TTTTATGTCT	TCTTTAATAC	TTTGTATTTT	TAGGACTTTG	AGGTCAAACA
610	620	630	640	650	660
GTAATTGAAA	ATGGTACGAT	GGAAGGTCAA	AGGATAATAA	CGGGCACACT	GGAAAATATA
CATTAACCTT	TACCATGCTA	CCTTCCAGTT	TCCTATTATT	GCCCGTGTGA	CCTTTTATAT
670	680	690	700	710	720
GCTGGAAAGG	ATATTAAACC	TCCTGCTTTA	GTGGTATTGG	AAATGTTGTC	AATGTTTTTA
CGACCTTTCC	TATAATTTGG	AGGACGAAAT	CACCATAACC	TTTACAACAG	TTACAAAAAT
730	740	750	760	770	780
AAGAAATGAA	TCAAATCAGT	GGCTGATCTA	TTAAGAAGGC	AATATCATGA	ATGGATTAGA
TTCTTTACTT	AGTTTAGTCA	CCGACTAGAT	AATTCTTCCG	TTATAGTACT	TACCTAATCT
790	800	810	820	830	840
AGGTAAAAAA	ATTGTTATAA	CAAGACCTGC	TGAAAGGGCT	AAAGACTCAG	TTGAAATGGT
TCCATTTTTT	TAACAATATT	GTTCTGGACG	ACTTTCCCGA	TTTCTGAGTC	AACTTTACCA
850	860	870	880	890	900
AAAATCTTAT	GGAGCAGTTC	CAATTGTAAC	TCCTACAATT	GAAGCTCAAAG	ATTCCAAGCC
TTTTAGAATA	CCTCGTCAAG	GTTAACATTG	AGGATGTAA	CTTGAGTTTC	TAAGGTTTCG
910	920	930	940	950	960
AGAAGAAGTG	ATAAAATTAT	GTAATATGAT	AAATGAACCT	TGATTGGCCT	TATAT
TCTTCTTCAC	TATTTTAATA	CATTATACTA	TTTACTTGGA	ACTAACCGBA	ATATA

FIGURE 51

PL. 146 / 151

GENE corA ET PROTEINE CORA (SUMT) DE METHANOBACTERIUM
IVANOVII

SEQUENCE DU FRAGMENT DE 955 PB DE 34 A 729

```

      M V V Y L V G A G P G D P E L I T L K A V N V L K
      ATGGTAGTTTATTTAGTAGGTGCGGGTCCAGGAGATCCCGAACTTATCACTCTCAAAGCTGTAAACGTGTTAAAA
34      44      54      64      74      84      94      104
      K A D V V L Y D K P A N E E I L K Y A E G A K L I
      AAAGCGGATGTTGTACTGTACGACAAACCTGCAAATGAAGAAATTTTAAAGTATGCTGAAGGTGCAAACTAATA
109     119     129     139     149     159     169     179
      Y V G K Q P G H H Y K S Q N E I N T L L V E E A K
      TATGTCGGAAAACAAGCAGGACATCATTACAAATCTCAAATGAAATCAATACTCTTCTTGTTGAAGAAGCAAAA
184     194     204     214     224     234     244     254
      E N D L V V R L K G G D P F V F G R G G E E I L A
      GAAATGATTTAGTAGTACGCCTTAAAGGTGGAGACCCCTTTGTATTTGGAAGAGGAGGCGAGGAAATTCTGGCC
259     269     279     289     299     309     319     329
      L V E E G I D F E L V P G V T S A I G V P T T I G
      CTTGTAGAAGAAGGAATTGATTTTGAGTTAGTTCCAGGGGTAACCTTCTGCAATTGGAGTTCCAACAACAATTGGG
334     344     354     364     374     384     394     404
      L P V T H R G V A T S F T V V T G H E D P T K C K
      CTTCCAGTTACTCATAGAGGTGTTGCAACATCGTTTACAGTTGTTACAGGTCATGAAGACCCAACAAAATGCAAG
409     419     429     439     449     459     469     479
      K Q V G W D F K A D T I V I L M G I G N L A E N T
      AAACAGGTAGGATGGGACTTTAAAGCAGATACTATTGTAATACTTATGGGTATTGGAAATTTAGCTGAAAATACA
484     494     504     514     524     534     544     554
      A E I M K H K D P E T P V C V I E N G T M E G Q R
      GCAGAAATTATGAAACATAAAGATCCTGAAACTCCAGTTTGTGTAATTGAAATGGTACGATGGAAGGTCAAAGG
559     569     579     589     599     609     619     629
      I I T G T L E N I A G K D I K P P A L V V L E M L
      ATAATAACGGGCACACTGGAAAATATAGCTGGAAAGGATATTAAACCTCCTGCTTTAGTGGTATTGGAAATGTTG
634     644     654     664     674     684     694     704
      S M F L K K *
      TCAATGTTTTTAAAGAAATGA
709     719     729

```

FIGURE 52

FEUILLE DE REMPLACEMENT

..ROTEINE CORA

147/151

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	6	2.60	882.41	3.54
2	LEU	L	22	9.52	2487.85	9.99
3	ILE	I	17	7.36	1922.43	7.72
4	MET	M	6	2.60	786.24	3.16
5	VAL	V	26	11.26	2575.78	10.34
6	SER	S	4	1.73	348.13	1.40
7	PRO	P	12	5.19	1164.63	4.68
8	THR	T	16	6.93	1616.76	6.49
9	ALA	A	16	6.93	1136.59	4.56
10	TYR	Y	5	2.16	815.32	3.27
11	*	*	0	0.00	0.00	0.00
12	HIS	H	5	2.16	685.29	2.75
13	GLN	Q	4	1.73	512.23	2.06
14	ASN	N	9	3.90	1026.39	4.12
15	LYS	K	20	8.66	2561.90	10.29
16	ASP	D	11	4.76	1265.30	5.08
17	GLU	E	21	9.09	2709.89	10.88
18	CYS	C	2	0.87	206.02	0.83
19	TRP	W	1	0.43	186.08	0.75
20	ARG	R	4	1.73	624.40	2.51
21	GLY	G	24	10.39	1368.52	5.50
22	-	-	0	0.00	0.00	0.00

RESIDUS = 231
 Masse moleculaire (monoisotopique) = 24900.1855
 Masse moleculaire (moyenne) = 24915.9766
 INDEX DE POLARITE (%) = 40.69
 POINT ISOELECTRIQUE = 5.45
 DO 260 (1mg/ml) = 0.304 DO 280 (1mg/ml) = 0.462

FIGURE 52.1

PL. 148/151

CORA DE PREMIER RESIDU=1
DERNIER RESIDU=231

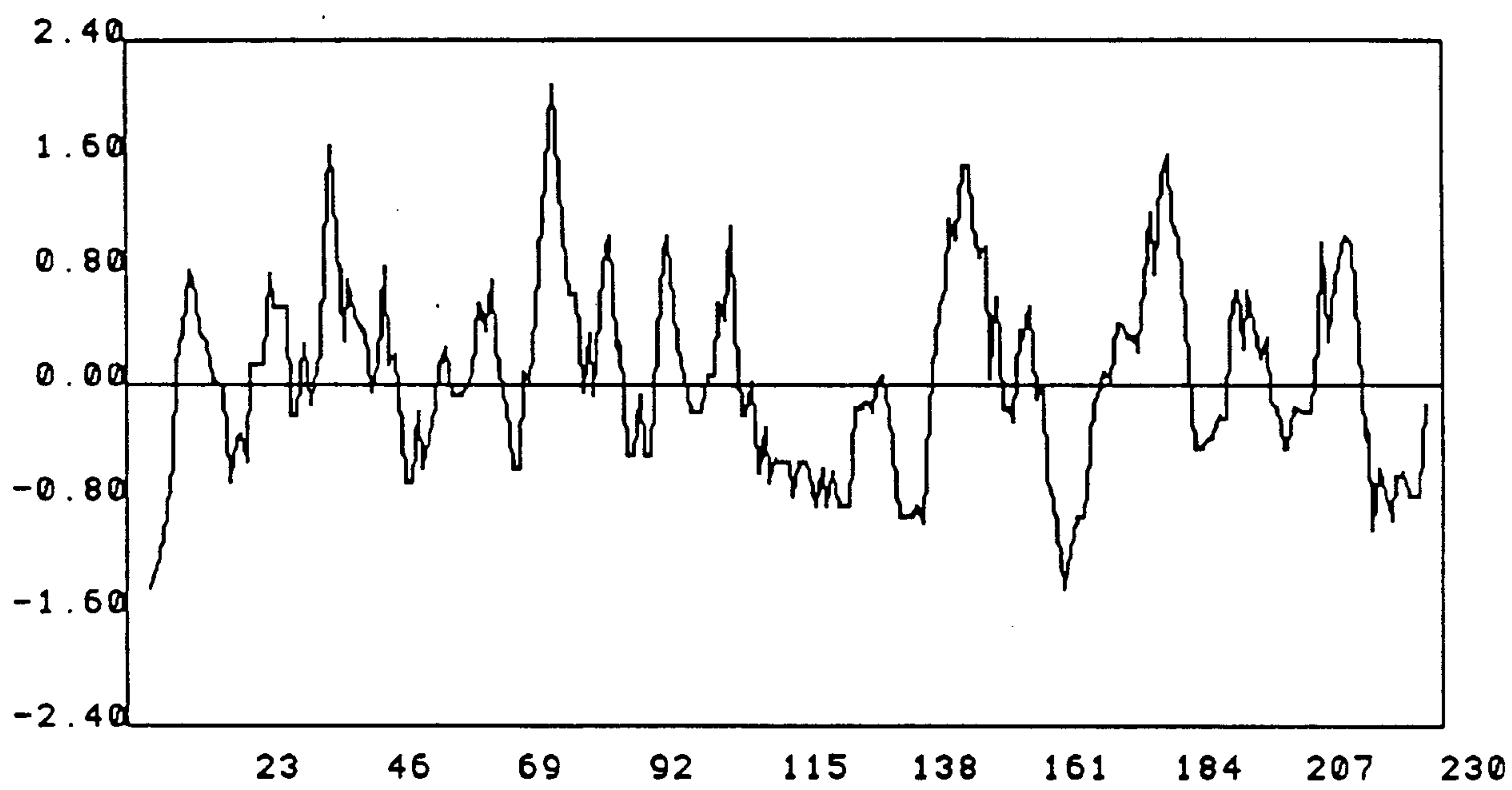


FIGURE 52.3

149/151

M. IVANOVII CORA 3 A 227
 P. DENITRIFICANS COBA 17 A 251

```

      10      20      30      40      50      60      70
VYLVGAGPGDPELITLKAVNVLLKKADVVLVDKPNEEILKYAE-GAKLIYVGKQAGHHYKSQNEINTLLV
=====
      20      30      40      50      60      70      80
VWLVGAGPGDPGLLTLHAANALRQADVIVHDALVNEDCLKLARPGAVLEFAGKRGGKPSPKQORDISLRV
      80      90      100      110      120      130      140
EEAKENDLVVRLKGGDPFVFGRGGEEILALVEEGIDFELVPGVTSAGVPTTIGLPVTHRGVATSFTVVT
=====
      90      100      110      120      130      140      150
ELARAGNRVRLRLKGGDPFVFGRGGEEALTLVEHQVPFRIVPGITAGIGGLAYAGIPVTHREVNHAVTFLT
      150      160      170      180      190      200
GHEDPTKCKKQVGWD-FKADT-IVIL-MGIGNLAENTAIEIMKH-KDPETPVCVIENGTMEGQRIITGTL-
=====
      160      170      180      190      200      210      220
GHDSSGLVPDRINWQGIASGSPVIVMYMAMKHIGAITANLIAGGRSPDEPVAFCNAATPQQAVLETTLA
      210      220
--E-NIAGKDIKPPALVVL-EMLSM
=====
      230      240      250
RAEADVAAAGLEPPAIVVVGEVVRL

```

POURCENTAGE D'HOMOLOGIE STRICTE : 40.4

FIGURE 53

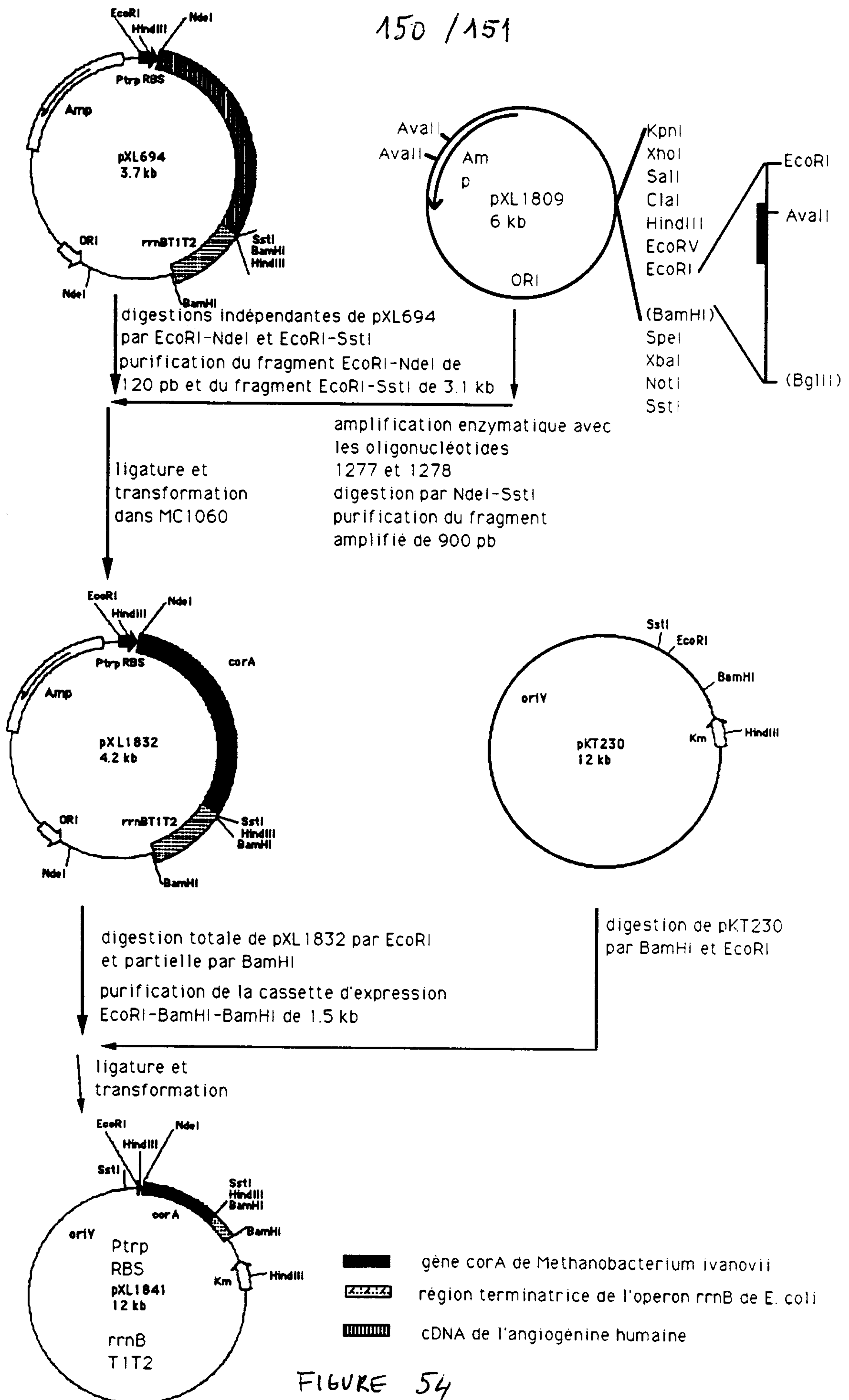


FIGURE 54

151/152

tableau: purification de la cobinamide kinase-cobinamide phosphate guanylyltransferase de *Pseudomonas denitrificans*.

Etape de purification	vol (ml)	protein (mg)	Cobinamide kinase		Cobinamide phosphate guanylyltransferase		ratio des act. spe. 2/1
			act spe 1 (U/mg de protein)	rendement (%)	facteur de purification	act. spe. 2 (U / mg de protein)	
Extrait brut ²	4.5	120	16	-	-	214	13
Eluat Mono Q HR 10/10	9.0	8.98	188	88	12	-	-
Eluat Hydroxyapatite	2.0	4.55	325	77	20	3640	11
Eluat Phenyl-Superose	2.0	1.51	560	44	35	-	-
Eluat Mono Q HR 5/5	3.0	0.90	786	37	49	11282	14

²A partir de 1 g de cellules humides de SC510 pXL622 cultivées sur milieu PS4 (Cameron et al., 1989) sans cobalt.

TABLÉAU A