

Séquençage de Vérification Sanger sur produits PCR et Plasmides

Guide de Trouble shooting

TROUBLE SHOOTING	1
I. RESULTATS ET INTERPRETATION	2
1. <i>Visualisation de la séquence</i>	2
2. <i>Contrôle qualité de la réaction de séquence</i>	4
3. <i>Problèmes rencontrés et solutions</i>	5
a. Liés à la réaction de séquence	5
a.1 Séquence négative : aucune séquence détectable	5
a.2 Séquence très faible, souvent inutilisable, bruit de fond élevé	7
a.3 Séquence avec mélange de plusieurs bases sur la même position	9
a.4 Séquence dont le signal chute rapidement	12
a.5 Séquences présentant des Pics de fluorescence parasites	13
b. Au niveau de la structure de l'ADN	15
a.1 Structures riche en G	15
a.2 Séquences riches en GC	17
a.3 Séquence avec une longue région à homopolymer (A ; T) ou STRETCH	18
a.4 Séquences riches en GT	19
a.5 Comparaison des différents kits utilisés	20
c. Au niveau technique (machine)	22
a.1 Problèmes d'électro-injection	22

I. RESULTATS ET INTERPRETATION

1. Visualisation de la séquence

Pour chaque échantillon, deux types de fichiers sont créés par le programme d'analyse:

Un fichier image couleur.ab1 contenant les données de fluorescence brutes et les données analysées.

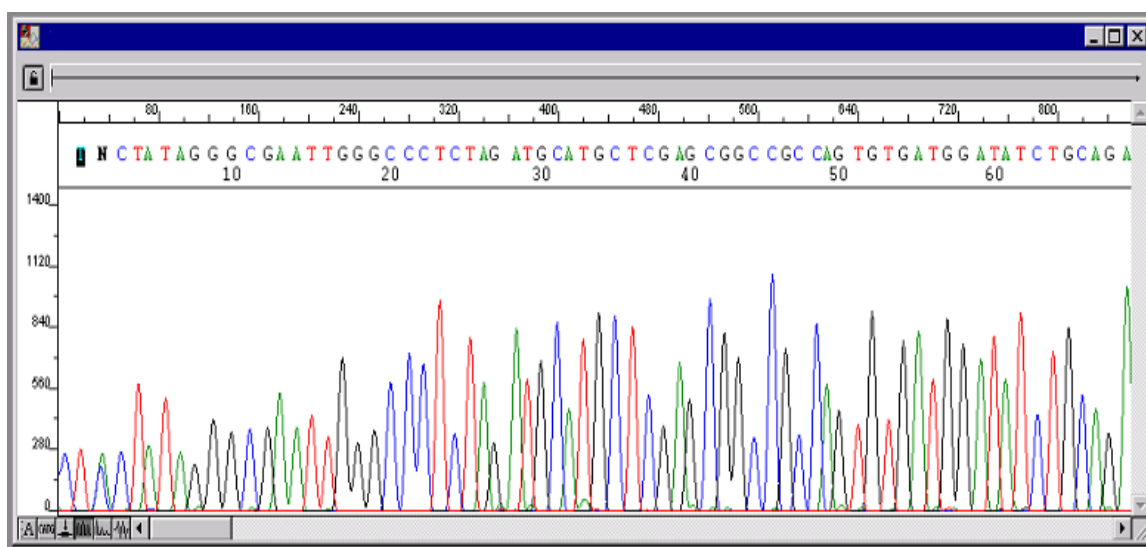
Un fichier.txt contenant le fichier texte de la séquence analysée.

Pour chaque séquence 5 types d'informations sont disponibles sur le fichier source. elles sont accessibles grâce aux boutons situés au bas de la fenêtre de séquence.

On dispose de plusieurs champs d'évaluation des données par échantillon qui permettent de déterminer la qualité de la séquence :

L'ensemble de ces données est accessible avec le logiciel Sequence Scanner, pas forcément avec les autres logiciels de visualisation des électrophorégrammes

CHROMATOGRAMME DES DONNEES ANALYSEES (Electrophorégramme)



1 2 3 4 5

→ Boutons d'accès aux différentes fenêtres :

1. Annotations
2. Séquence texte
3. Données analysées
4. Données brutes
5. Données EPT : Electrophoresis puissance tampon

L' électrophorégramme de la séquence présente les données analysées. Chaque couleur de pic correspond un des quatre fluorochromes différents (cf ci dessus).

L'électrophorégramme donne des informations sur :

- La capacité résolutive du gel, c'est à dire la capacité à distinguer 2 base successives, grâce à la taille et la forme des pics de fluorescence. La capacité résolutive d'un capillaire de 50cm neuf avec un polymère neuf est en moyenne de 600-900 bases.

La lecture de la séquence pourra être plus longue (800-1000 bases) mais la fidélité du résultat diminue avec la longueur de lecture.

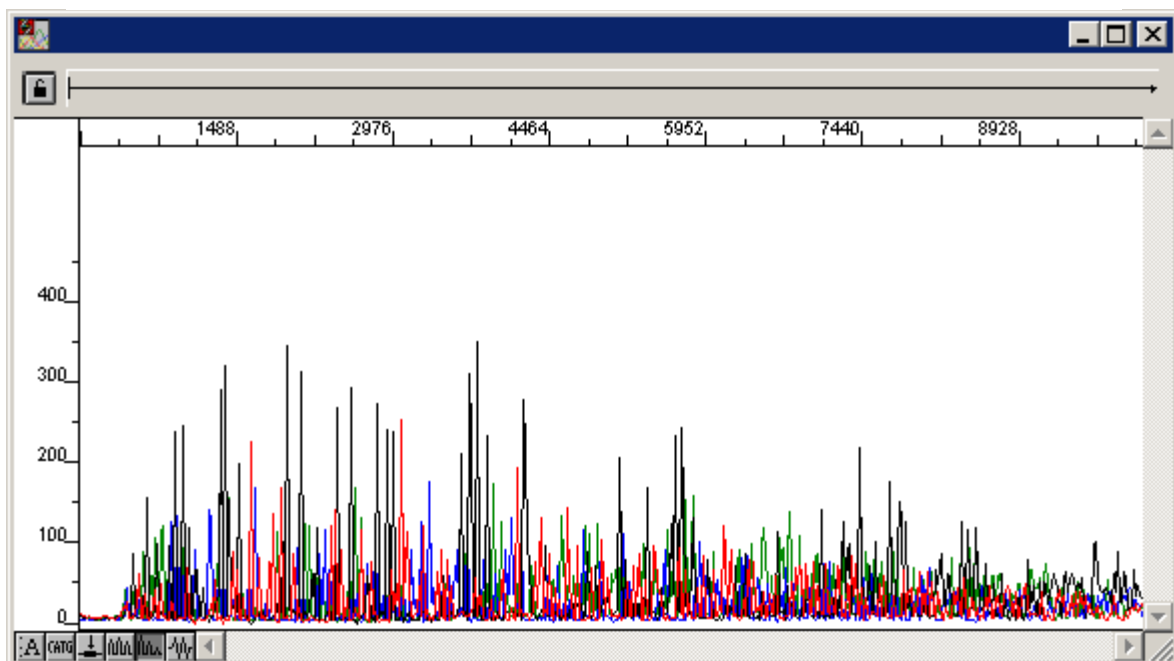
- La présence éventuelle de ddNTP non incorporés parasitant la lecture (cf.II3c1).

- La localisation du premier point lecture correspondant au premier point de fluorescence analysé. Parfois il est nécessaire de modifier la position du premier point de lecture déterminé automatiquement par le logiciel. Un positionnement éloigné du 1^{er} point de lecture entraîne la disparition du début de la séquence (quelques bases situées en 5' de l'amorce).

Le Raw Data nous permet de visualiser les données brutes qui ont tout de même subi la calibration spectrale au moment de la collecte.

Cette fenêtre nous permet d'avoir une appréciation générale de la séquence, par la visualisation de l'évolution générale de la fluorescence et de l'intensité du signal. L'intensité du signal est étroitement liée a la quantité de matrice initiale présente; par exemple, en cas d'une présence trop faible de matériel, le signal sera très faible, le bruit de fond deviendra prépondérant et l'assignation des bases sera difficile(cf.II3c1).

FENETRE DE PRESENTATION DES DONNEES BRUTES : RAW DATA



Une page d'annotation récapitule toutes les informations concernant la séquence et son analyse. Elle exprime notamment, de manière numérique, l'intensité moyenne du signal de chaque fluorochrome. Pour qu'une séquence soit lisible, le signal doit être supérieur à 10-20 pour chaque ddNTP.

La séquence texte est lisible sur le chromatogramme des données analysées; une base est assignée à chaque pic de fluorescence. Il est aussi possible d'avoir uniquement la séquence texte (pour les recherches d'homologies dans les bases de données) sur le deuxième fichier disponible (fichier.txt)

2. Contrôle qualité de la plaque de réaction de séquence

La grande hétérogénéité des échantillons fournis par les utilisateurs et la diversité des méthodes de travail des équipes introduisent d'importantes différences de résultats entre les séquences d'une même plaque. Il est donc nécessaire de vérifier la qualité générale de chaque plaque de séquence.

Afin de contrôler la qualité de notre travail, nous effectuons pour chaque plaque la réaction de séquence, la purification, le dépôt et la migration d'échantillons standards dont la séquence et la qualité sont connues. La lecture des résultats de ces « témoins » nous permet de valider la qualité de la plaque effectuée.

Plusieurs critères nous permettent de valider la qualité des résultats, il est nécessaire de regarder :

- l'intensité générale de la fluorescence des séquences de la plaque.

- La résolution de la séquence, dépend principalement de l'âge du capillaire et de la qualité du polymère utilisé.

Un set de capillaire est certifié pour 100 runs. Lorsque les capillaires sont entretenus dans de bonnes conditions, la résolution diminue doucement en fonction du nombre de runs. Aux environs de 300 runs, nos capillaires peuvent résoudre jusqu'à 400-450 bases. Chaque bouteille de polymère est testée à son arrivée afin de s'assurer de sa qualité.

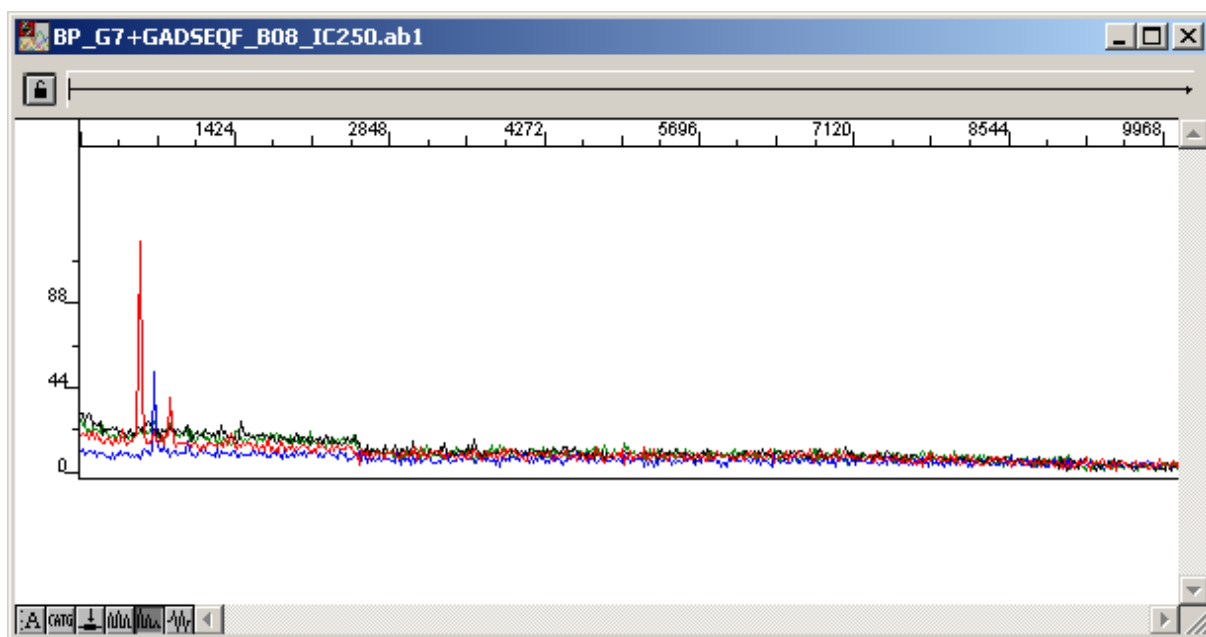
3. Problèmes rencontrés et solutions

On observe pour les utilisateurs une palette de résultats possibles liés à différentes causes qui vont de l'erreur de manipulation à des problèmes de structure du fragment à séquencer. Nous nous assurons dans un premier temps de la qualité de la plaque en observant nos séquences de référence. Leur bonne qualité permet de valider la plaque. Tous les problèmes présentés ci-dessus supposent que la plaque de séquence a été validée.

a. Liés à la réaction de séquence

a.1 *Séquence négative : aucune séquence détectable*

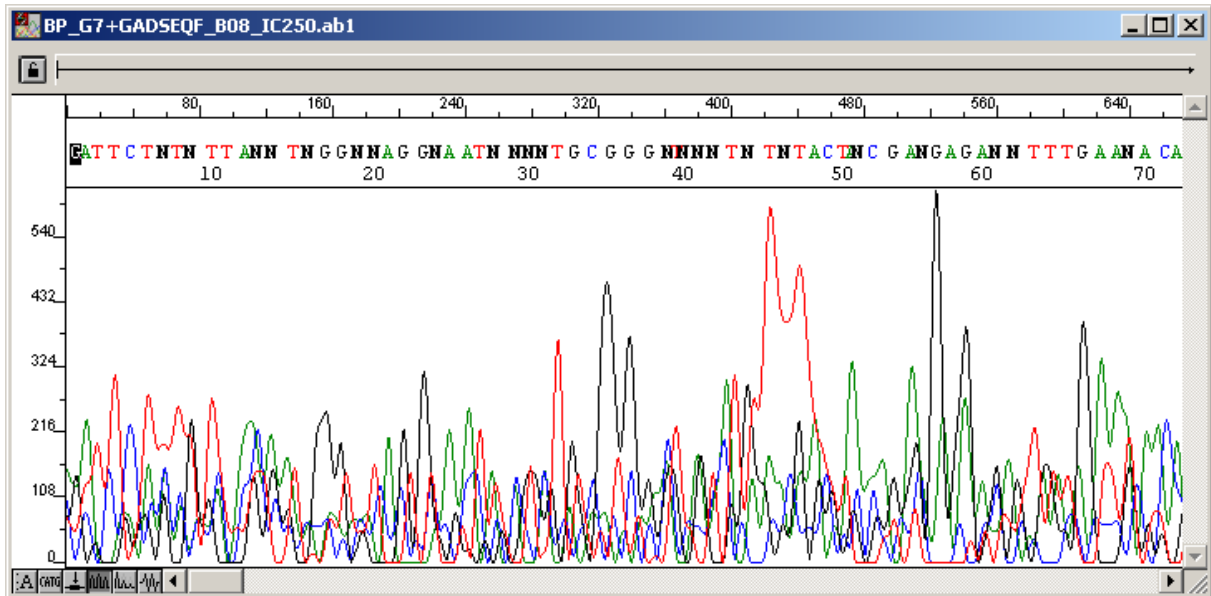
Pour ce type de résultats on n'enregistre aucune fluorescence spécifique, le seul signal présent correspond à la fluorescence basale appelée bruit de fond, cela s'observe de manière nette sur les données brutes.



Fenêtre de présentation des données brutes d'une séquence négative

Ce type de profil peut avoir plusieurs causes :

- Problème lié à l'amorce. Soit l'amorce utilisée ne possède pas de séquence complémentaire sur la matrice d'ADN, soit elle n'est pas présente dans l'échantillon (oublié), soit elle est dégradée.
- Problème lié à la matrice d'ADN présente soit en trop faible quantité (inférieure au seuil de détection de la technique), soit inexistante lors de la migration (ADN dégradée ou oublié).



Fenêtre de présentation des données analysées d'une séquence négative

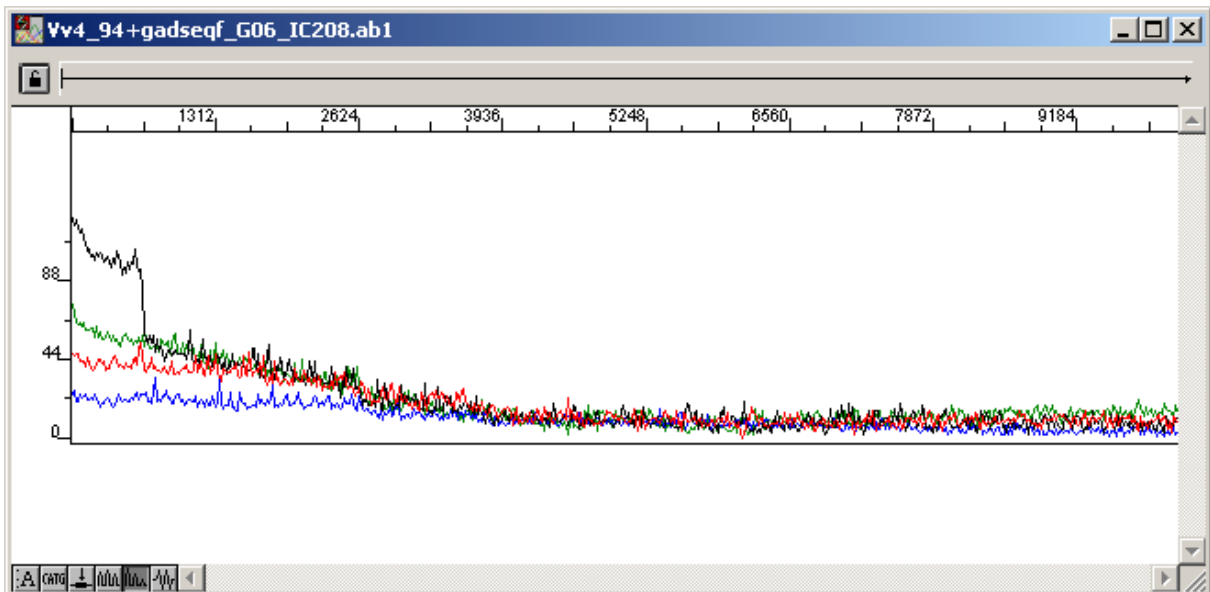
Le chromatogramme des données analysées présente des assignations de bases mais les bases identifiées n'ont aucune signification biologique.

Après l'obtention d'une séquence négative telle que celle-ci, il est nécessaire de recommencer la réaction de séquence. Il faut préparer une nouvelle dilution d'amorce et vérifier que l'amorce est bien complémentaire à la matrice. Dans le cas contraire, il faut changer d'amorce ou vérifier la qualité et la quantité de la matrice et éventuellement la re préparer.

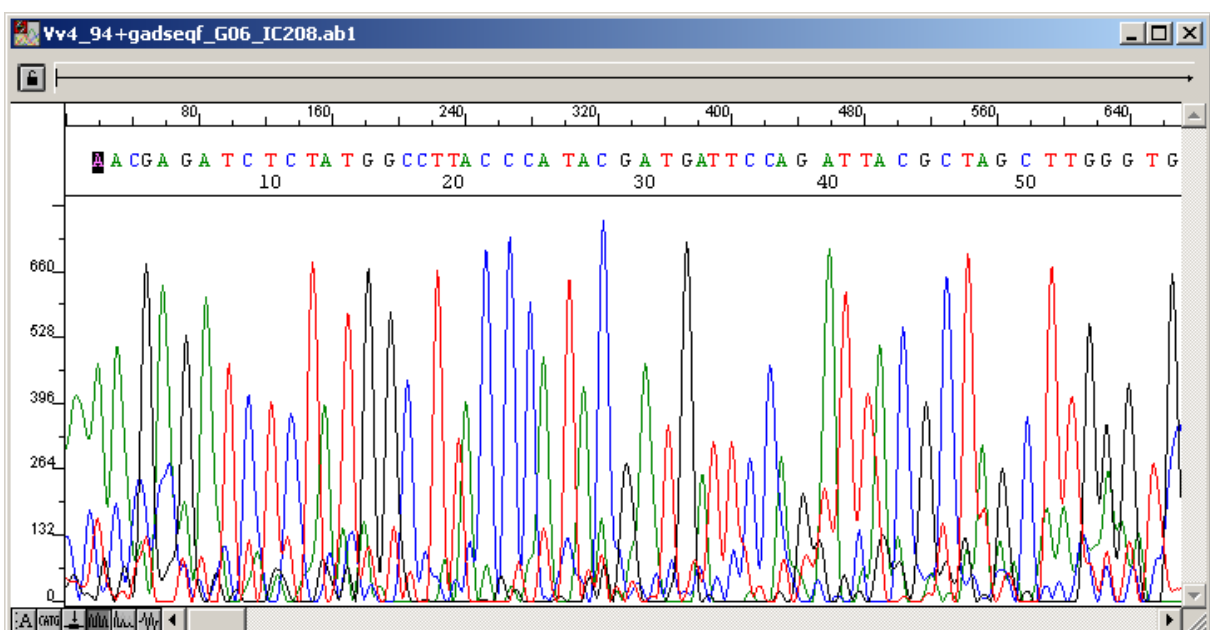
a.2 *Séquence très faible, souvent inutilisable, présence de bruit de fond élevé.*

Malgré la grande sensibilité de l'appareil, ce type de profil est la plupart du temps inutilisable. On observe une séquence où les signaux d'intensité de fluorescence des ddNTP sont très faibles (souvent inférieurs à 10-20), et vont venir se confondre avec le bruit de fond.

La visualisation des données brutes nous donne une bonne appréciation rapide de l'intensité du signal



Fenêtre de présentation des données brutes d'une séquence de très faible intensité



Fenêtre de présentation des données analysées d'une séquence de très faible intensité

Ce type de profil est en général dû à un problème dans la qualité et la quantité de la matrice d'ADN. Ainsi des préparations de plasmides, si elles sont mal extraites, présentent souvent des sels ou des impuretés (protéines, débris cellulaires, ARN résiduel). Les impuretés perturbent et parfois inhibent le fonctionnement de la taq polymérase au cours de la réaction de séquence. L'intensité du signal obtenu est donc très faible.

La purification sur chromatographie d'exclusion permet d'éliminer en partie les sels mais n'élimine pas les autres contaminants (Protéines, ARN).

Ces derniers, parfois chargés, seront électroinjectés comme les molécules d'ADN. Cela aura plusieurs conséquences :

- Leur encombrement important entraînera une saturation des capillaires (50µm) et par conséquent une perturbation de la migration électrophorétique.
- L'électroinjection des contaminants au dépend des molécules d'ADN diminuera d'autant l'intensité du signal fluorescent.
- De plus une forte présence d'ARN résiduels lors de la préparation de l'échantillon entraînera une surestimation de la détermination de la quantité lorsqu'elle est mesurée au spectrophotomètre. Cette quantité d'ADN présente est alors trop faible, il en sera de même pour l'intensité du signal.

Pour remédier à ce type de problèmes il est nécessaire de refaire la préparation des plasmides, de doser à la fois au spectrophotomètre et sur gel d'agarose coloré au Bromure d'éthidium.

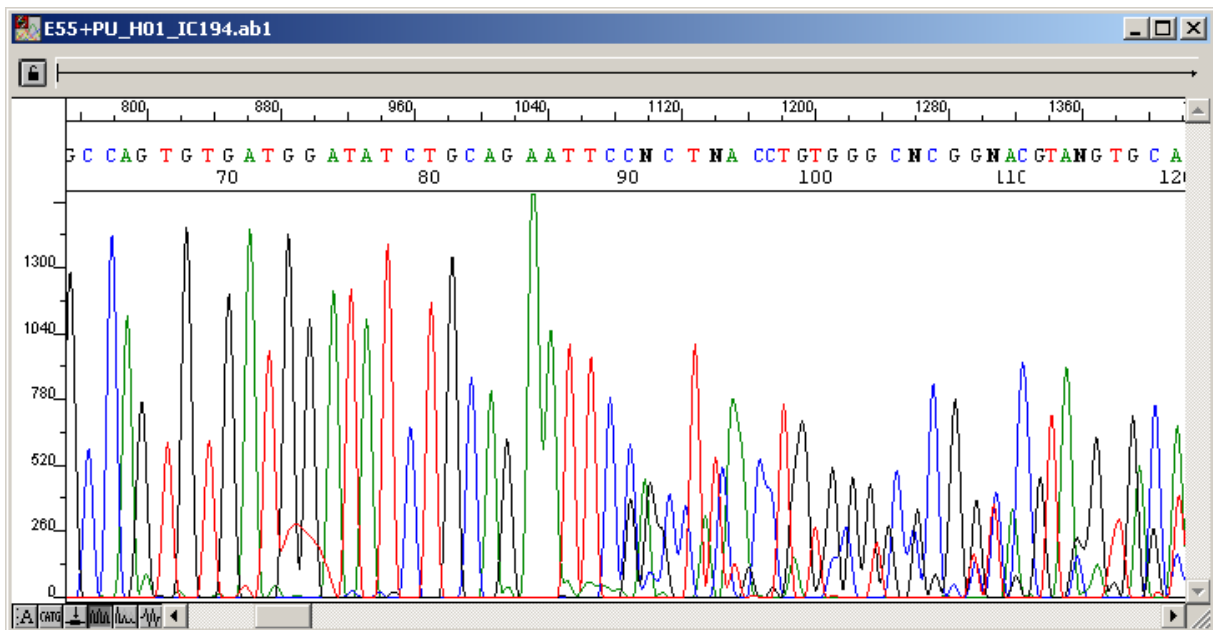
a.3 *Séquence présentant un mélange de plusieurs bases sur la même position*

De nombreuses causes peuvent-être à l'origine d'un tel profil.

- Ainsi, une mauvaise qualité de préparation de matrice entraîne la présence de contaminants susceptibles de perturber le migration électrophorétique, la résolution et donc l'assignation des bases.

- Un mélange dans l'échantillon biologique de 2 espèces moléculaires qui permettent toutes deux l'accrochage de l'amorce de séquence, va présenter 2 signaux différents sur certaines positions. Lorsque le mélange de signaux concerne une portion de séquence complète, la position de la zone du mélange par rapport à celle de la séquence normale permet de diagnostiquer le problème :

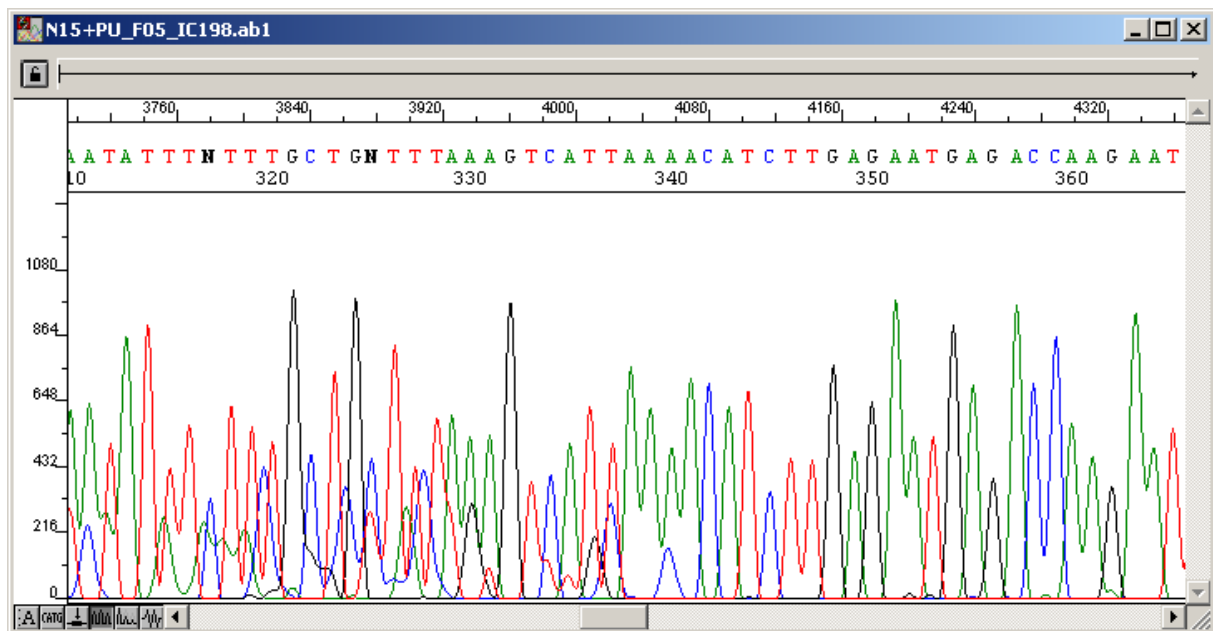
1. Lorsqu'on observe une zone de séquence normale puis une zone de mélange, cela peut-être dû au mélange d'un plasmide sauvage et d'un plasmide recombinant. Cela arrive en particulier lorsque la position à partir de laquelle le mélange apparaît correspond à un site de restriction (clonage).



Fenêtre de présentation des données analysées d'une séquence présentant un mélange de deux plasmides

Moyens d'y remédier : les plasmides proviennent de 2 clones bactériens prélevés simultanément. Le ré-étalement de ce prélèvement sur boîte de culture permettra de séparer nettement les deux clones bactériens.

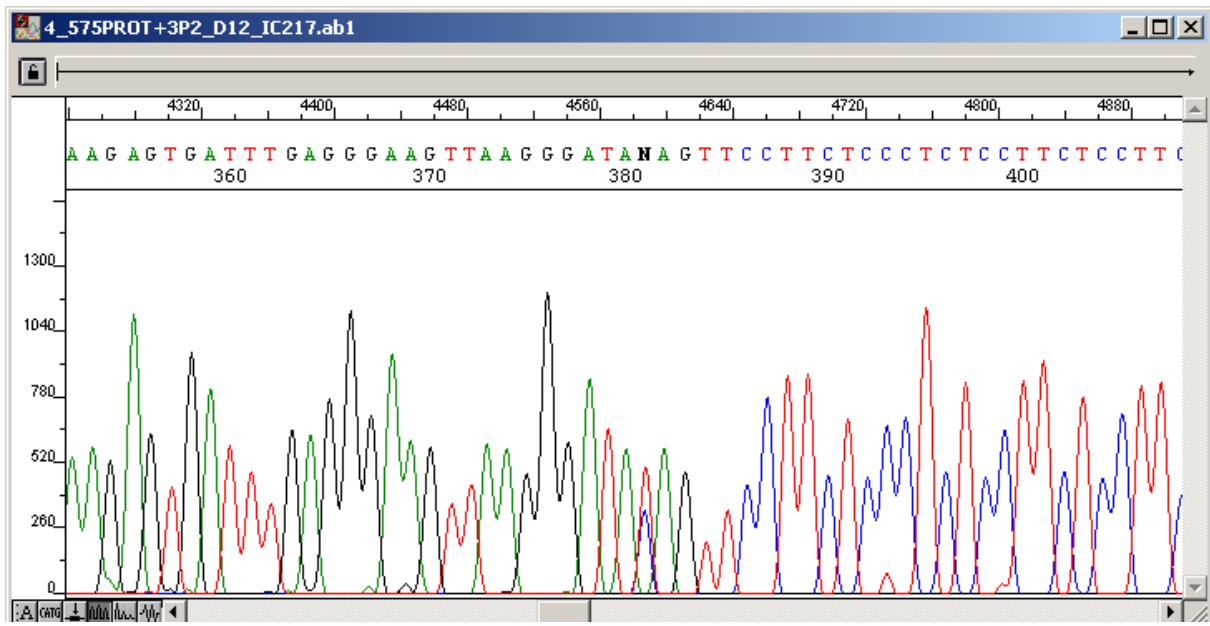
2. Lorsqu'on observe une zone de mélange puis une zone de séquence normale, cela peut provenir du mélange de 2 produits de PCR de taille différentes et peut se présenter sous forme d'un signal double jusqu'à ce que la taille du plus petit fragment soit atteinte (cas d'un produit de PCR contaminant par exemple dimère d'amorces généré lors de la PCR).



Fenêtre de présentation des données analysées d'une séquence de deux produits de PCR dont le plus petit a une taille de 340pb

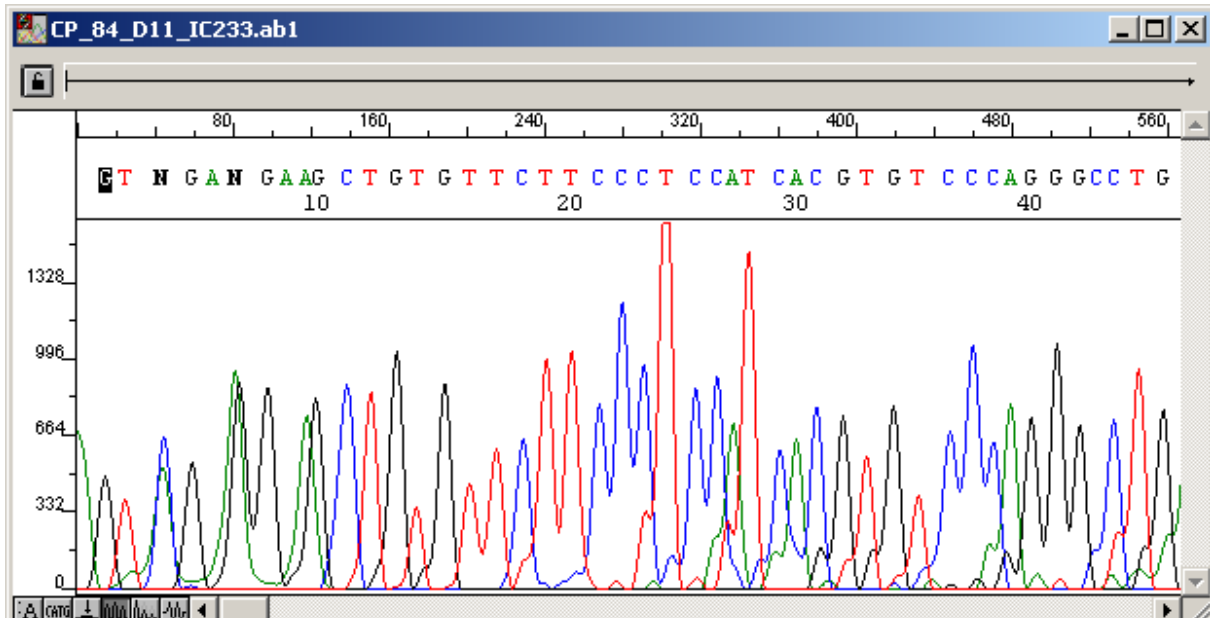
Pour y remédier, il est possible de purifier le produit de PCR sur gel ou d'utiliser une amorce de séquence interne au produit spécifique qui ne sera pas présente sur le contaminant.

3- Lorsqu'on observe un mélange de signal localisé sur une position, cela peut provenir d'un mélange de 2 produits de PCR dont la séquence diffère seulement au niveau d'une base (cas de recherche de mutations hétérozygotes par exemple). Dans ce cas au niveau de cette base seulement il y aura deux pics de fluorescence.



Fenêtre de présentation des données analysées d’une séquence présentant un mélange de signal au niveau d’une seule base

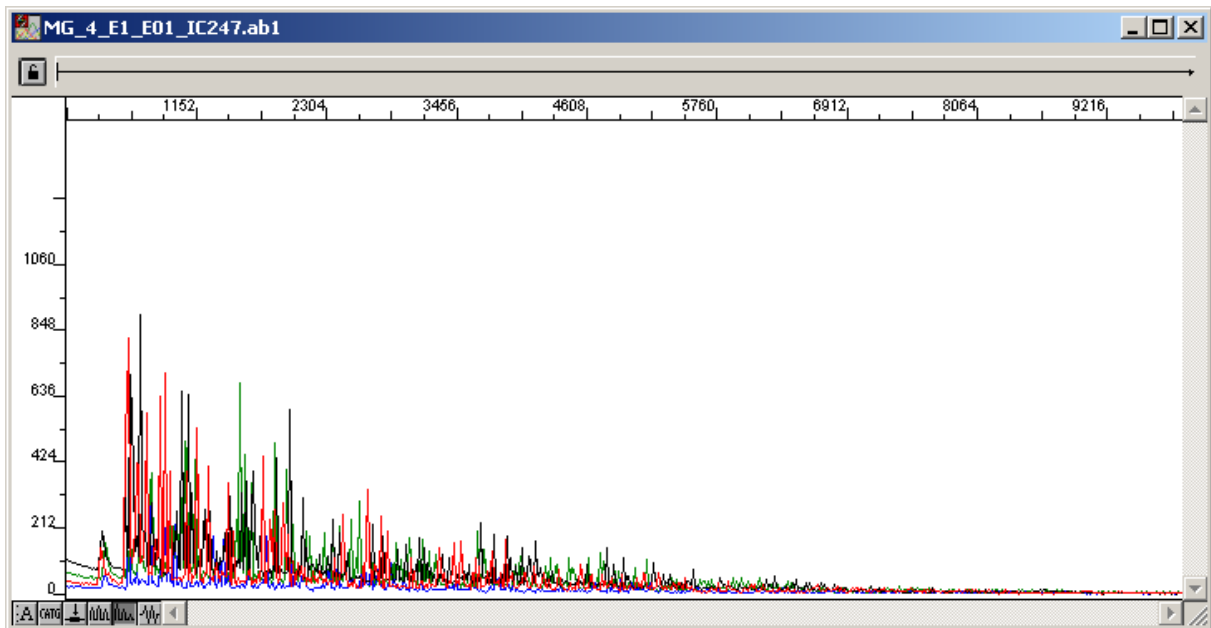
- une mauvaise qualité de l’amorce de séquence (dégradée) peut entraîner un profil double avec un décalage du cadre de lecture d’une paire de bases (profil n-1). Le moyen d’y remédier est de refaire synthétiser cette amorce.



Fenêtre de présentation des données analysées d’une séquence présentant un décalage de lecture d’une base.

a.4 Séquence dont le signal chute rapidement.

Ce type de séquence possède une bonne résolution initiale mais l'intensité du signal chute très rapidement. Les données brutes sont très significatives pour ce type de profil, le signal est très fort et diminue de manière rapide. Sur le chromatogramme la séquence devient illisible très tôt.



Fenêtre de présentation des données brutes de la séquence d'un plasmide. Cette séquence présente un excès de matrice

La plupart du temps la cause de ce résultat est la présence d'une trop grande quantité de matrice ADN. Par conséquent l'équilibre de la réaction de séquence est déplacé, il y a synthèse de nombreuses chaînes courtes.

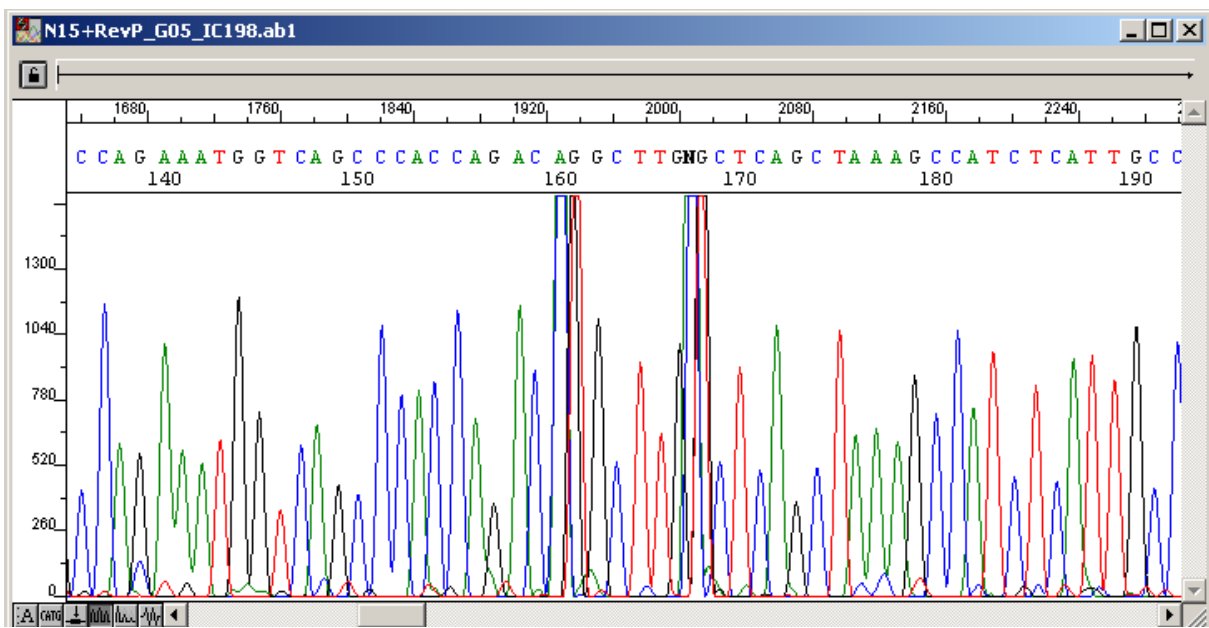
Pour y remédier il est nécessaire de refaire le séquençage en diminuant la quantité de matrice.

a.5 Séquences présentant des Pics de fluorescence parasites.

- Présence de pics parasites : pics d'urée

Ces pic apparaissent en général sur les séquences aux même positions. Ils sont dus à la présence de cristaux d'urée présents dans le polymère de migration. En effet le polymère est conservé à 4°C. Une température inférieure peut entrainer la formation des cristaux qu'il est nécessaire de faire disparaître car ils gênent la migration électrophorétique.

Sur le chromatogramme, ces pics apparaissent isolés au milieu de la séquence il y en a très peu et ils n'empêchent la lecture que d'une seule base. Ce sont des pics d'intensité très forte ou les quatre fluorescences sont présentes à la même base.

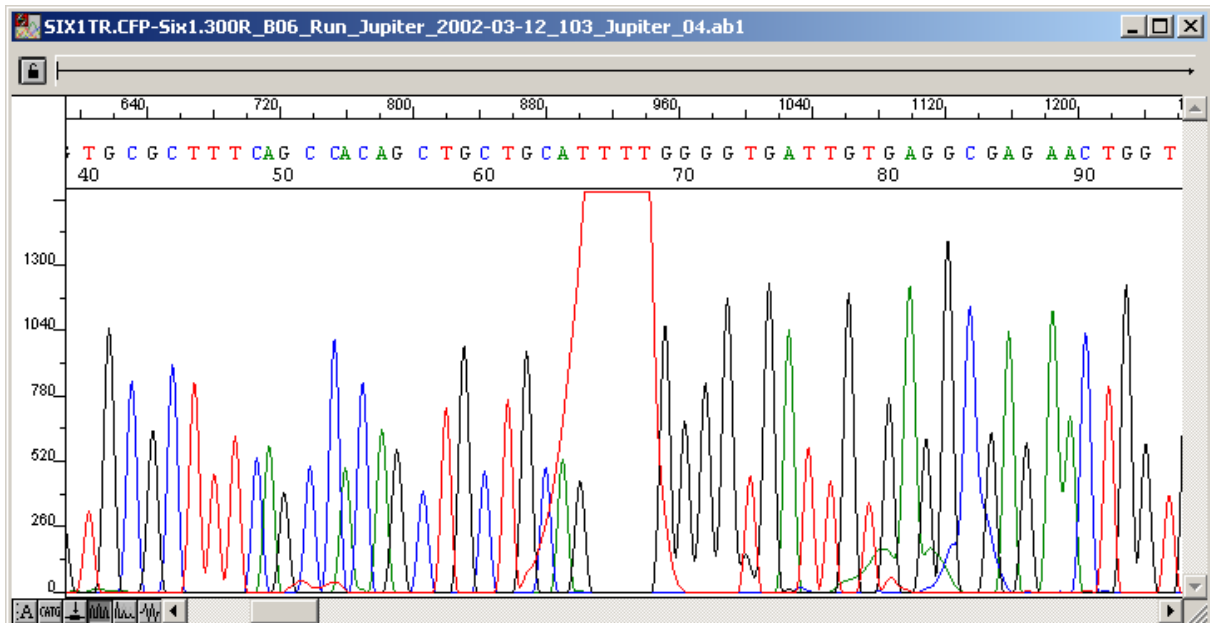


Fenêtre de présentation des données analysées d'une séquence présentant des pics d'urée

Il est possible de remédier à la formation de ces cristaux en plaçant le polymère un moment à température ambiante avant l'injection dans la machine.

- Pics non spécifiques (souvent de T)

Sur certains chromatogrammes, on observe en début de séquence et à certaines positions bien définies de larges pics de forte intensité (souvent des pics de T) qui ne sont pas spécifique à la séquence du brin matrice.



Fenêtre de présentation des données analysées d'une séquence présentant un pic de T contaminant.

Ces pics proviennent de la fluorescence parasite de ddNTP non incorporés qui ont mal été éliminés lors de l'étape de purification par chromatographie sur sephadex G50. On peut observer ce type de séquence lorsqu'on utilise un G50 qui n'est pas superfine (diamètre des pores 20-50 μ m) ou lorsque la purification à été faite trop rapidement.

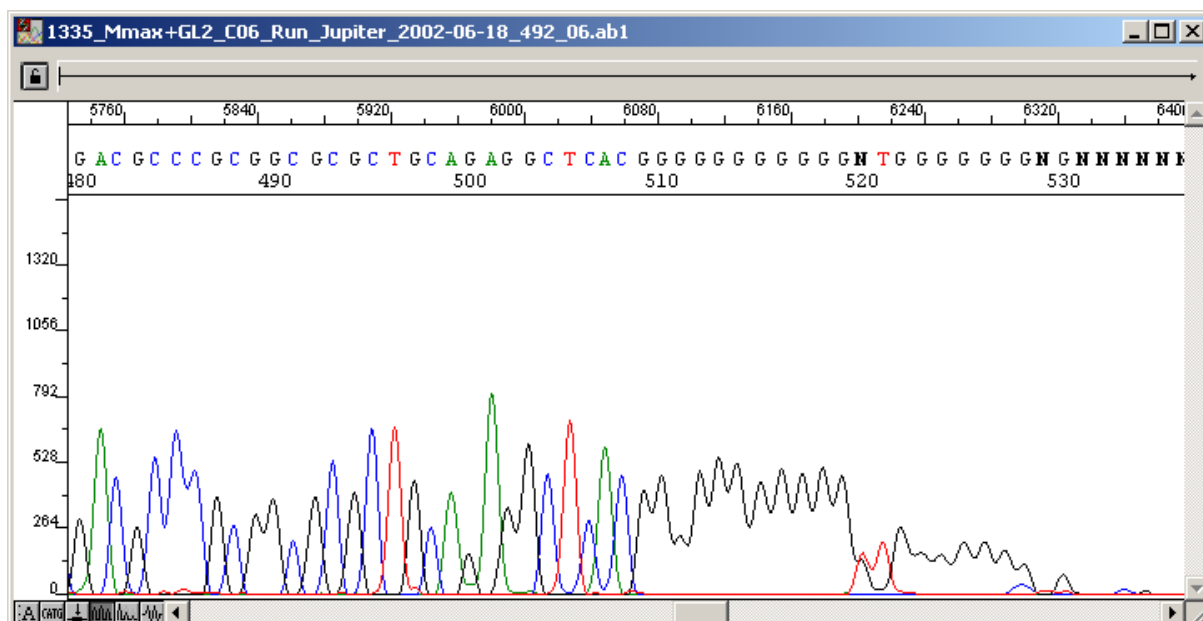
Pour résoudre ce type de problème il faut simplement repasser le produit de la réaction de séquence sur une colonne de G50 superfine.

b. Au niveau de la structure de l'ADN

Parfois; même s'il n'y a aucun problème technique au niveau de la réaction de séquence (matrice propre, bien dosée et amorce en bon état), la structure de l'ADN peut influencer sur la qualité, de la séquence.

a.1 Structures riche en G

Le kit Big Dye terminator version2(ABI) utilisé en routine qui utilise l'inosine (dITP) à la place de la guanine (dGTP), est bloqué par certaines structures (on se réfère ci-dessous à la séquence du brin allongé c'est à dire la séquence qu'on lit sur le chromatogramme). Il s'agit en particulier des séquences répétées de G. En effet le dITP se lie de manière moins forte avec le dCTP que le dGTP. Lors de l'allongement, l'hybridation entre le brin lu et le brin allongé est moins stable et provoque un décrochage de l'enzyme au niveau des zones riches en G. Ce phénomène se visualise par un arrêt de la lecture de la séquence qui suit la structure G.

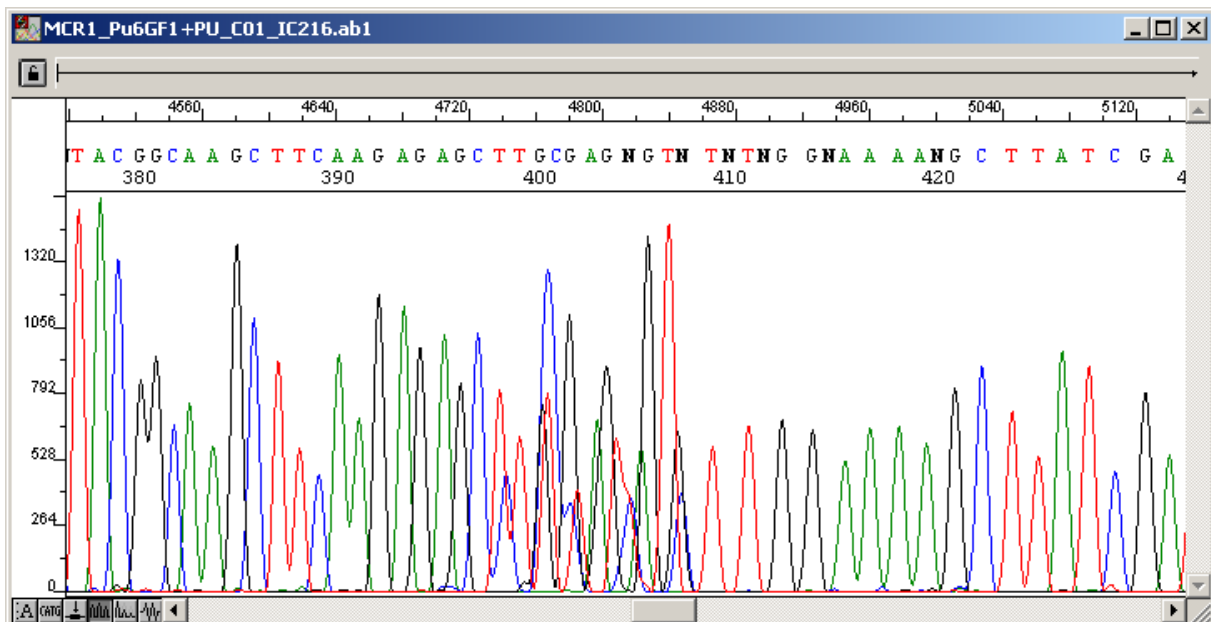


Fenêtre de présentation des données analysées d'une séquence G riche

Lorsque l'on utilise le kit dGTP (version 2 ABI), il est possible de passer ces structures avec plus de succès. Avec ce kit, les brins synthétisés au cours de la réaction de séquence sont riche en G. Les liaisons GC sont fortes et l'enzyme ne se décroche plus et passe la structure.

Le seul inconvénient de ce kit spécial est que le dGTP provoque des bandes de compressions. Ces dernières résultent de la formation de structures secondaires a l'intérieur du fragment d'ADN allongé(cf.II1c).

Sur le chromatogramme ce profil est caractérisé par la présence d'une part de pics très rapprochés voir confondus et d'autre part de pics très éloignés.



Fenêtre de présentation des données analysées d'une séquence présentant des bandes de compressions

Même si ce type de profil dû le plus souvent aux structures GC n'empêche en général pas la lecture de la séquence, il est préférable de ne pas utiliser le dGTP en routine.

a.2 Séquences riches en GC

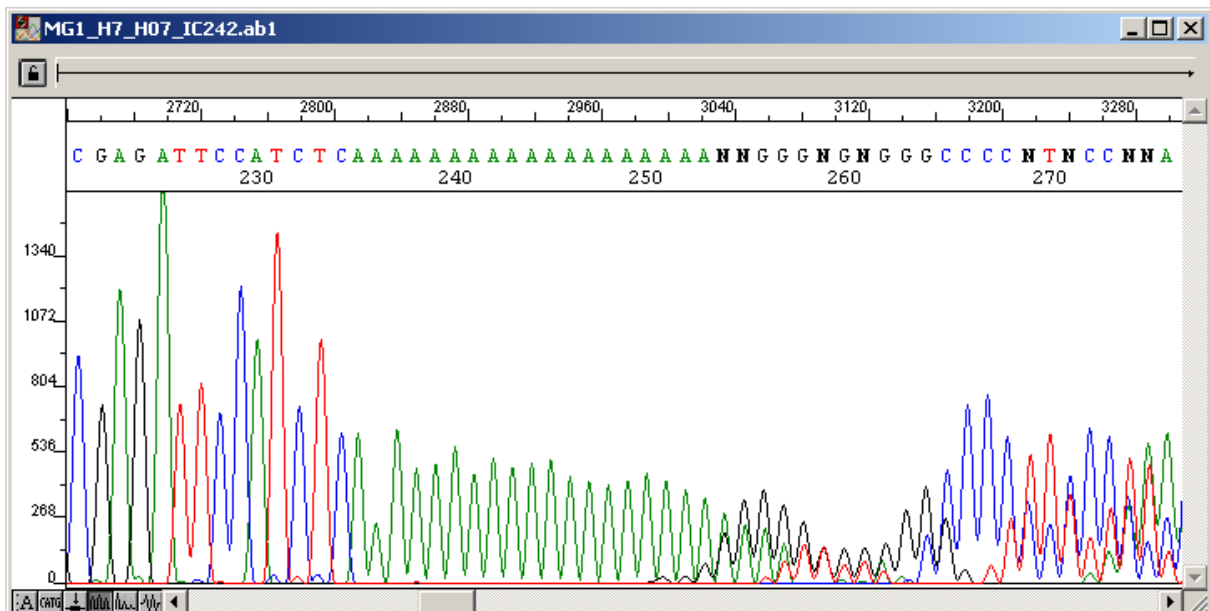
Les séquences où le pourcentage en GC% est élevé peuvent poser problème pour effectuer la réaction de séquence.

En effet, la liaison GC étant plus stable que la liaison AT, plus la proportion de GC dans le fragment est élevée plus la température de fusion du fragment augmente. Dans ce cas, les conditions dénaturantes standard (principalement la température) utilisées au cours de la réaction de séquence ne sont pas suffisantes à la dénaturation du brin matrice et ne donnera pas un séquence lisible au niveau et au delà de la région GC riche.

Le profil observé dans ce cas montre un faible signal ou un arrêt brusque du signal.

Les conditions classiques n'étant pas suffisamment dénaturantes pour « ouvrir » la matrice double brin, on peut améliorer les conditions dénaturantes en utilisant du DMSO(5-10%), de la Bétaine, du glycérol ou du formamide (5-10%). Si rien n'est amélioré, on peut utiliser le kit Big dye terminator dGTP.

a.3 Séquence avec une longue région contenant un homopolymer (A ; T) ou STRETCH



Fenêtre de présentation des données analysées d'une séquence qui contient un stretch de A

Le chromatogramme de ce type de séquences est bien lisible avant la répétition (stretch), ensuite on observe un décalage des cadres de lecture.

La présence d'une base répétée de nombreuses fois sur la séquence ralentit la Taq polymérase. Cela entraîne un décrochage prématuré de cette dernière, il n'y a alors pas incorporation d'un ddNTP, on appelle ce phénomène un faux stop.

Le brin non terminé servira au cycle suivant, le stretch de A ou T présent à son extrémité 3' lui permettra de s'hybrider à des positions variables sur la répétition du brin matrice. A la fin des 30 cycles, les brins séquencés présentent des tailles de stretch variables. On a alors un mélange de séquences décalées, on dit que l'enzyme dérape.

La longueur de la répétition influence directement ce phénomène. Plus le stretch est grand et plus l'enzyme dérape ce qui rend en général la lecture impossible après la zone de répétition.

On peut résoudre ce type de problème par séquençage de la matrice dans l'autre sens en utilisant une amorce antisens qui permet de connaître la séquence qui se trouve en 3' du stretch.

Pour un homopolymère T on peut éventuellement synthétiser un mélange d'amorces polyT avec une queue 3' A, G ou C. Ce type d'amorce se fixe en 3' de l'homopolymère et permet le séquençage de la partie située en 3'. En pratique ce type d'amorce est plus compliquée à synthétiser et à maintenir stable car son faible T_m rend son utilisation difficile.

a.5 Comparaison des différents kits utilisés

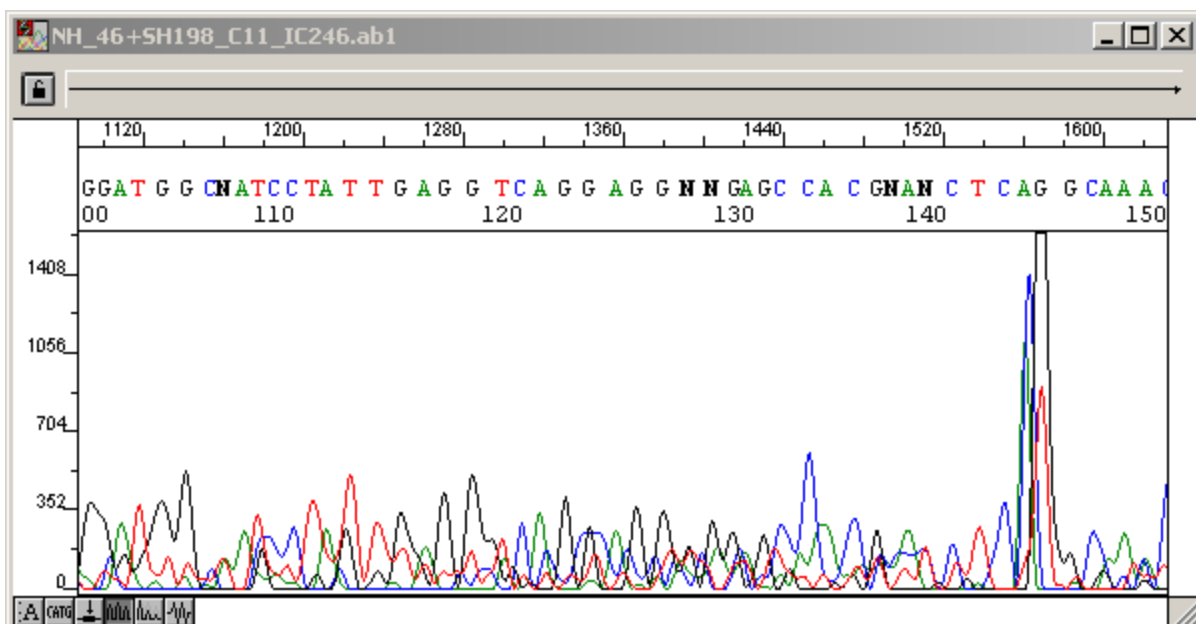
Les différents kits de réaction utilisés au laboratoire ont des propriétés différentes qui permettent de passer des structures différentes. On utilise pour le moment en routine le kit Big dye terminator version 2.

Lorsque la matrice possède des structures spécifiques qui empêchent de lire en 3' de cette structure, il est possible d'utiliser d'autres kits nous permettant de passer ces structures.

Le kit Big dye terminator version 2 spécial dGTP permet de passer certaines structures telles que les structures G riche, GC riche et GT riche.

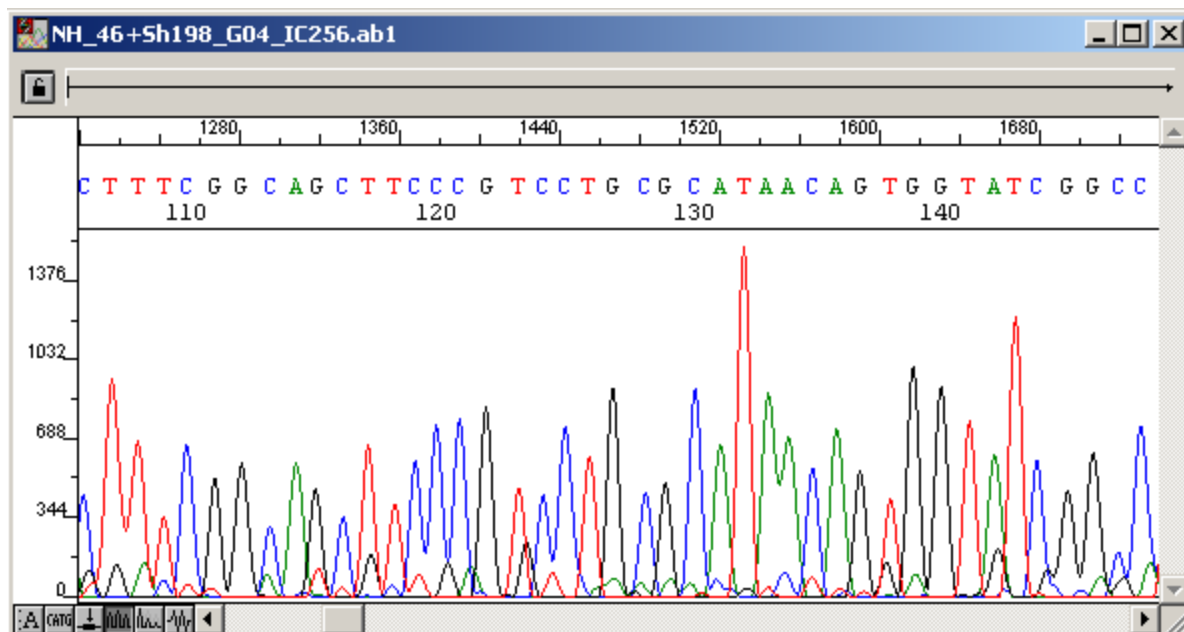
Le nouveau kit que l'on utilise DYEnamic ET Terminator Cycle sequencing a la particularité de pouvoir passer certaines structures qui ne passent pas avec le kit utilisé en routine ni avec le dGTP.

Par exemple la séquence suivante passe mal en dGTP, le signal est très faible, le bruit de fond est élevé et les bases attribuées par le logiciel d'analyse ne correspondent pas à la séquence recherchée.



Données analysées de séquences effectuées à partir du kit dGTP big dye terminator

Losque l'on utilise le kit « DYEnamic ET Terminator Cycle sequencing » (voir ci dessous), on obtient un signal nettement plus fort, un bruit de fond pratiquement nulle et une séquence fiable.



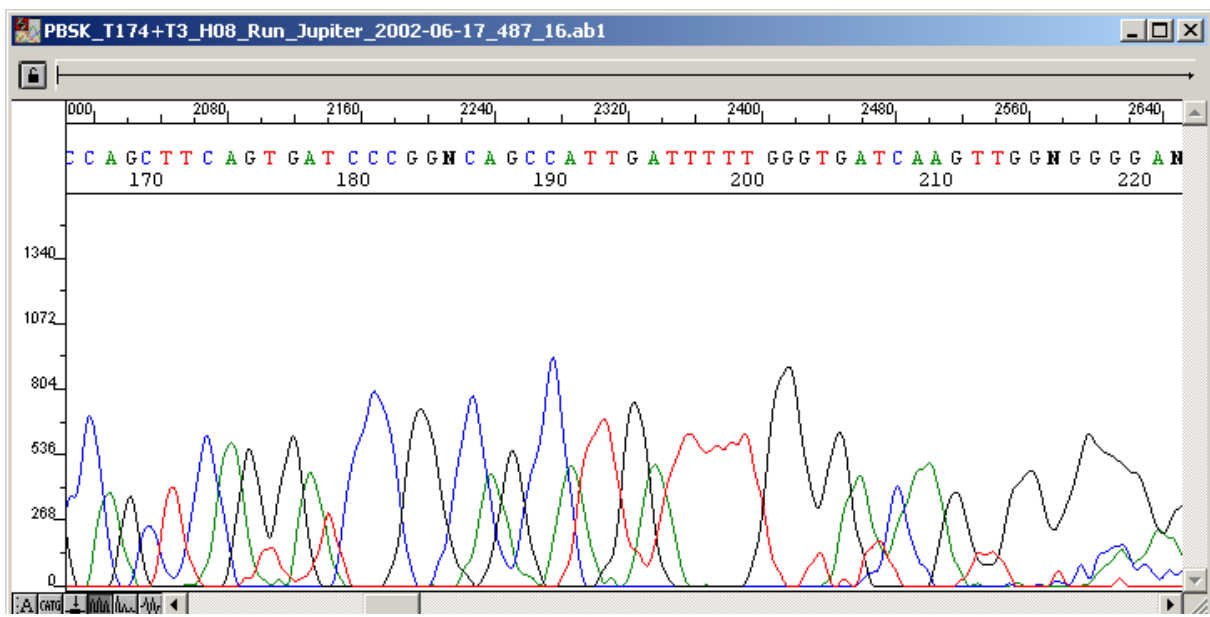
Données analysées de séquences effectuées à partir du kit DYEnamic ET Terminator Cycle sequencing

c. Au niveau technique (machine)

a.1 *Problèmes d'électro-injection*

Ces problèmes d'électroinjection se produisent de manière aléatoire a certains endroits de la plaque de séquence.

Sur le chromatogramme on observe de larges pics mal résolus du même type que les séquences où il y a des problèmes de résolution. Cet étalement du signal ne vient pas d'une diminution de la résolution mais d'un défaut dans l'électroinjection.



Fenêtre de présentation des données analysées d'une séquence présentant des problèmes d'électroinjection

Les causes de ce défaut d'électroinjection sont souvent reliés à un manque de volume au moment de l'électroinjection (doit être de 20µl), à la présence d'une bulle dans le puit ou à la présence de contaminants au sein des produits de la réaction de séquence.

On arrive le plus souvent à y remédier en injectant à nouveau les produits de la réaction de séquence. Ensuite on effectue une nouvelle migration électrophorétique des puits concernés.

Si la séquence n'a toujours pas un aspect normal après cette seconde migration, il est nécessaire de re-préparer la matrice et refaire la réaction de séquence.