

# CHAPITRE 1

## Nombres

### *Des résonances numériques dans l'ADN*

#### **I. Protocoles et méthodologie de calcul des résonances**

Nous exposons ici le vocabulaire de base et les protocoles de calcul systématique des résonances du supra-code de l'ADN. Sans communiquer ici de listings de programmes, nous donnons les éléments suffisants pour que le lecteur puisse reproduire et tester lui même notre méthodologie de calcul des résonances de l'ADN.

##### ***1. Principe***

Le concept de base consiste à rechercher des structures globales entre les proportions relatives de nucléotides, qu'il s'agisse tant d'ADN que d'ARN, ou de régions codantes ou non codantes des génomes. Nous appelons cette structure de résonances le "supra-code". Rappelons qu'une résonance est une proportion relative de nucléotides selon les nombres de Fibonacci: 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610 987 1597 2584... et/ou les nombres de Lucas : 1 3 4 7 11 18 29 47 76 123 199 322 521 843 1364 2207... Par exemple, si un ensemble contigu de 89 nucléotides se partitionne en 34 bases A et 55 bases TCG (quelles que soient leurs positions relatives), nous dirons qu'il y a là une résonance A/TCG. Afin

de donner une idée du grand nombre de résonances, nous citerons le cas du génome du Sida qui, long de 9 000 bases, compte plus de 50 000 résonances significatives, plus d'un millier de résonances de longueur supérieure à 2 000 bases, et plusieurs centaines de résonances "architecturant" plus des deux tiers du génome (plus de 6 000 bases)...

Véritable "horlogerie génétique", les protocoles de calcul des résonances sont simples, déterministes et reproductibles. En voici maintenant l'exposé, austère, mais nécessaire...

## 2. Vocabulaire de base

Nous venons de proposer une définition sommaire de la résonance. D'autres concepts importants vont être : le support, le domaine, la partition, la logique de pagination et les types de résonances.

### a) Le support

Nous ne nous intéressons qu'aux seules séquences de nucléotides T C A G. On aurait pu, par exemple, étudier les séquences d'acides aminés, ce ne sera pas le cas. Les séquences de nucléotides peuvent se présenter sous de multiples formes :

- le double brin de la double hélice d'ADN;
- le simple brin des ARN messagers ou rétrovirus, ou plus simplement un seul des deux brins de la double hélice d'ADN. On sait en effet depuis Crick et Watson qu'à une base T d'un brin fait face une base A sur l'autre brin, qu'à une base C fait face une base G, qu'à une base A fait face une base T et qu'à une base G correspond une base C.

Nous appelons "support" la structure de brin(s) ainsi étudiée. Dans la présente étude, le support sera toujours un simple brin (d'ADN ou d'ARN, codant ou non codant). Nous considérerons en fin d'ouvrage des supports plus complexes (la double hélice, dans sa totalité) ou aussi plus abstraits (par exemple certaines structures particulières de cheminement topologique dans la double hélice).

### b) Le domaine

Le support sera donc pour nous un simple brin d'ADN. Sur la séquence définie sur ce simple brin, on délimite un domaine qui est une région compacte d'ADN ou d'ARN. Par exemple, un «morceau» de génome

sera un domaine (appelé «contig», pour "contigu" par les généticiens). Un gène, suite alternée d'introns et d'exons, sera aussi un domaine. Un gène de type ARN messager (suite concaténée des exons seuls), un génome entier (comme le génome du Sida) seront également des domaines. Un domaine est donc caractérisé par un début, une fin, une orientation (sens d'exploration de la séquence), et une séquence ordonnée de bases T C A G.

Un mot à propos de cette dernière notion de «sens» ou d'orientation : on sait que l'orientation a un sens (!) au niveau biologique (technologie des «antisens», etc.). Cependant, il est courant que des morceaux de séquences se recombinent en changeant de sens, c'est-à-dire en se retournant. Nous devons donc préciser que la notion de résonance est indépendante du sens d'exploration de la séquence (puisque'il s'agit de proportions relatives de nucléotides). Cependant, vis-à-vis d'une orientation donnée du brin d'ADN, la position de la résonance aura une importance (exemple des régions précurseurs aux paragraphes 3a et chapitre 3.III dont le rôle fonctionnel dépend du sens d'exploration de la séquence). Dans le même esprit, toute résonance définie sur un brin possède son alter ego ou symétrique sur l'autre brin par exemple, si 89 bases se répartissent en 34 T et 55 A C G sur le brin étudié, une résonance miroir existera sur l'autre brin : elle aura 34 A et 55 T C G. Ainsi, les tableaux de synthèse ou signatures que nous définirons au tableau de la figure 1.2 posséderont leurs symétriques sur l'autre brin. Par exemple :

Signature type d'un rétrovirus HIV1 (cf. 3.A, ex. HIV1-ELB)									
Brin 1 de l'ADN → Brin 2 de l'ADN									
T	C	A	G			T	C	A	G
-	-	X	-			X	-	-	-
-	X	-	-			-	-	-	X
-	-	-	X			-	X	-	-
-	-	X	-			X	-	-	-
Signature type d'un gène (cf. 2.2, ex. HUMCIA1)									
Brin 1 de l'ADN → Brin 2 de l'ADN									
T	C	A	G			T	C	A	G
-	X	-	X			-	X	-	X
X	-	X	-			X	-	X	-
-	-	-	-			-	-	-	-
-	X	-	X			-	X	-	X

Figure 1.1. Signatures types pour un rétrovirus et un gène (X symbolise un très grand nombre de résonances, x un nombre de résonance et le signe - symbolise un nombre de résonances résiduel ou nul)

Il est permis de spéculer sur ces deux séries de signatures puisque la première est asymétrique tandis que la seconde est symétrique. J'ai observé que le profil des rétrovirus est toujours asymétrique. Au contraire, le profil des gènes (ARN messenger) est tantôt symétrique tantôt asymétrique. Il est probable que cette observation gagnerait à être exploitée. On sait en effet qu'un même rétrovirus est stable dans deux milieux distincts lorsqu'il est seul (simple brin) et lorsqu'il habite en parasite l'ADN (double brin).

Dans le premier cas, sa durée de vie doit être longue (très supérieure à celle des ARN messagers dont la vie est courte... telle celle d'un messenger qui ne vivrait que le temps de transmettre le message). Il est alors possible que cette signature typique du rétrovirus renforce sa stabilité à ce stade de son évolution. Au contraire, dans le second cas, lorsqu'il habite l'ADN, il se pourrait qu'une métastructure très stable résulte de la complémentarité entre les deux profils de signatures dans les deux brins. En effet, à une structure riche en résonances A sur le premier brin fait face une structure miroir riche en résonances T (cf. figure 1.1).

### *c) La partition*

Pour un domaine donné, nous étudions systématiquement toutes les résonances tout au long du domaine pour les treize classes de résonances :

- résonances simples : T/ACG C/TAG A/TCG G/TCA
- bi-résonances : TC/AG TA/CG TG/CA
- méta-résonances : T/C T/A T/G C/A C/G A/G

Parmi ces treize classes de résonances, dix pourraient avoir une signification génétique, ainsi :

- les quatre types de résonances simples correspondent à des régions physiques topologiquement contiguës dans le génome;
- TC/AG correspond à la partition pyrimidines/purines. Il en sera de même pour T/C et A/G;
- TA/CG correspond aux affinités de paires de bases de Crick et Watson. Il en sera de même pour T/A et C/G.

Par contre, les résonances TG/CA, T/G et C/A semblent ne correspondre à rien à ce niveau physico-chimique.

Pendant longtemps, je n'ai recherché que les résonances simples. Il s'agit là de la partition la plus naturelle : proportions relatives d'une base vis-

à-vis des trois autres et vis-à-vis de l'ensemble des quatre natures de bases T C A G. Avec la notion de bi-résonance, on groupe les natures de bases 2 à 2. On analyse donc alors les résonances comme si l'ADN n'était formé que de deux natures de nucléotides. Ce type d'analyse n'est pas absurde lorsque l'on sait que, au fil de l'évolution, l'ADN semble être passé d'un ADN composé de trois types de bases à l'ADN actuel, composé de quatre types de bases. Notre analyse pourrait donc révéler d'anciennes structures entre nucléotides. Les méta-résonances constituent quant à elles une structure très abstraite puisque une telle analyse considère deux types de bases (par exemple T et C) en ignorant totalement et en faisant abstraction des deux autres types de bases (resp. A et G). Nous allons revenir plus en détail sur ces trois types d'analyses. Mais nous pouvons d'ores et déjà faire remarquer que ces trois types de résonances ne sont pas redondants. Ils sont complémentaires et traduiraient différents niveaux d'organisation de l'ADN simplement complémentaires et indépendants.

Réfléchissons un court instant sur la notion de structure ou de niveau d'organisation. Vous percevez un objet vivant : ici l'ADN, vos "lunettes" sont, par exemple ici, les séquenceurs d'ADN... Ils seraient en principe fiables... Supposons qu'ils soient effectivement fiables, même si les biologistes n'en sont pas totalement certains (cf. II). Tout naturellement, on a tendance à tenter d'expliquer du niveau où on observe. Ainsi, ici on bâtirait des explications reposant sur la séquence détaillée d'ADN. Limiter ainsi son champ de recherche scientifique constitue une erreur ! Rappelons nous que "la structure de contrôle et d'action est rarement la structure perçue et visible".

### *d) La logique de pagination*

Soit les deux suites de Fibonacci et de Lucas :

FIBONACCI : 1 1 2 3 5 8 13 21 34 55 89 144 233...

LUCAS : 1 3 4 7 11 18 29 47 76 123 199 322 521 ...

Nous allons maintenant explorer de manière systématique le domaine de nucléotides afin d'en recenser toutes les résonances significatives. Nous nommerons ce processus d'exploration dynamique de la séquence «logique de pagination». En effet, en partant de la première base de la séquence, nous fixons une base pivot qui sera donc successivement la base 1, la base 2, la

base 3, etc. Calés sur chaque base pivot, nous allons définir des masques successifs qui seront des nombres de Fibonacci ou Lucas (cf. e). Chaque masque, par exemple 55 89 144 233... permet de sélectionner dans la séquence, à partir de la base pivot, un sous-domaine que nous nommerons "page". Ainsi, si le domaine étudié contient 250 bases, nous définirons pour les nombres de Fibonacci :

- à partir de la base pivot 1 : des pages successives de 55 (bases 1 à 55), 89 (bases 1 à 89), 144 (bases 1 à 144), 233 (bases 1 à 233);
- puis, à partir de la base pivot 2 : des pages successives de 55 (bases 2 à 56), 89 (bases 2 à 90), 144 (bases 2 à 145), 233 (bases 2 à 234)... bases...

Et ainsi de suite jusqu'à utilisation de toutes les bases pivot, soit ici, pour un domaine de 250 bases :

- à partir de la base pivot 195 : dernière page de 55 (bases 195 à 250);
- à partir de la base pivot 161 : dernière page de 89 (bases 161 à 250);
- à partir de la base pivot 106 : dernière page de 144 (bases 106 à 250);
- à partir de la base pivot 17 : dernière page de 233 (bases 17 à 250)...

Puis nous déroulerons un processus analogue pour les nombres de Lucas, soit 47 76 123 199...

**REMARQUES.-** Afin de ne retenir que les résonances significatives nous imposons un seuil à la taille de la plus petite page : ici 55 pour les pages de Fibonacci et 47 pour les pages de Lucas. Ce seuil peut varier selon les objectifs de l'analyse. De même, la plus grande longueur de page est le plus grand nombre de Fibonacci ou Lucas qui puisse entrer dans le domaine : par exemple, le domaine mesurant ici 250 bases, la borne des plus grandes pages possibles est de 233 pour Fibonacci et de 199 pour Lucas. Ainsi, lorsque dans le chapitre 3, nous étudierons la variabilité de 16 génomes du Sida, nous n'explorerons que les grandes résonances de plus de 2 000 bases et, les génomes mesurant de l'ordre de 9 000 bases, les tailles des pages successives seront donc :

- pour Fibonacci : 2584 4181 6765
- et pour Lucas : 2207 3571 5778

**e) Les types de résonances**

Nous explorons systématiquement les résonances émergeant de pages dont les longueurs sont soit des nombres de Fibonacci, soit des nombres de

Lucas. Par exemple, si une séquence de 89 bases de long se répartit en 34 bases T et 55 bases A ou C ou G, nous dirons que nous avons une résonance de type FFF (F comme Fibonacci) car aussi bien 34 que 55 ou encore 89 sont des nombres de Fibonacci.

Nous considérons donc les deux séries de Lucas et de Fibonacci :

Fibonacci : 1 12 35 8 13 21 34 55  
89 144 233 377 610 987 1597...

Lucas : 1 3 4 7 11 18 29 47 76 123  
199 322 521 843 1364 2207 3571...

Dans ces deux suites, chaque élément est la somme des deux précédents. De plus, de nombreuses relations numériques relient ces deux suites, par exemple:

$$47 = 13 + 34$$

Lucas Fibonacci

C'est une raison pour laquelle ces deux suites de Lucas et Fibonacci sont très "liées" mathématiquement. Nous développerons cet aspect dans la seconde partie de ce livre.

Nous allons maintenant explorer séquentiellement le domaine en appliquant, depuis chaque base pivot, tous les masques successifs possibles de longueurs Fibonacci et Lucas. Nous imposons un seuil à ces longueurs de masques, par exemple 47 pour Lucas et 55 pour Fibonacci. Ainsi, pour un gène long de 1 767 bases, les pages possibles seront :

Fibonacci : 55 89 144 233 377 610 987 1597

Lucas : 47 76 123 199 322 521 843 1364

A l'intérieur de ces masques successifs nous allons comptabiliser les nombres relatifs de bases T et C, A ou G, de bases C et A, G ou T, de bases A et G, T ou C, de bases G et T, C ou A. Par exemple, si, à partir de la base 18, une page de longueur 55 bases contient 21 T et 34 C, A ou G (ou 21 C et 34 A, G ou T, ou 21 A et 34 T ou C, ou 21 G et 34 T, C ou A), nous dirons que nous avons une résonance.

Nous recherchons quatre types de résonances : les résonances FFF (Fibonacci Fibonacci Fibonacci), les résonances FFL (Fibonacci Fibonacci Lucas) et les résonances LFF et LLL, pour les quatre bases. Nous pouvons alors en déduire deux matrices 4 x 4 de synthèse (Cf. figure 1.2).

FFF (Fibonacci Fibonacci Fibonacci)								
ex.	Masque	-->	55	89	144	233	377	610...
	Bases A	-->	21	34	55	89	144	233...
	Bases T C G	-->	34	55	89	144	233	377...
FFL (Fibonacci Fibonacci Lucas)								
ex.	Masque	-->	55	89	144	233	377	610...
	Bases A	-->	8	13	21	34	55	89...
	Bases T C G	-->	47	76	123	199	322	521...
LFF (Lucas Fibonacci Fibonacci)								
ex.	Masque	-->	47	76	123	199	322	521...
	Bases A	-->	13	21	34	55	89	144...
	Bases T C G	-->	34	55	89	144	233	377...
LLL (Lucas Lucas Lucas)								
ex.	Masque	-->	47	76	123	199	322	521...
	Bases A	-->	18	29	47	76	123	199...
	Bases T C G	-->	29	47	76	123	199	322...
TYPE	NOMBRES DE RESONANCES				VOLUMES DE RESONANCES			
	T	C	A	G	T	C	A	G
FFF	3	-	-	-	199	-	-	-
FFL	-	-	-	-	-	-	-	-
LFF	-	-	-	-	-	-	-	-
LLL	-	-	-	-	-	-	-	-

Figure 1. 2. Les quatres types de résonances

Nous pouvons remarquer qu'il y a ici, pour les résonances de type FFF; et pour la base T, trois résonances :

- une résonance de 89 (34 bases T et 55 bases C, A ou G),
- deux résonances de 55 (21 bases T et 34 bases C, A ou G).

Soit un volume cumulé (somme des longueurs) qui sera égal à :  $(1 \times 89) + (2 \times 55) = 199$  pour  $1 + 2 = 3$  résonances, d'où les deux nombres 3 et 199 du tableau de synthèse (cf. figure 1.2).

On remarquera que si 199 est un nombre de Lucas, c'est également la somme de 3 nombres de Fibonacci :  $55 + 55 + 89$ , et donc aussi de 4 nombres successifs de Fibonacci :  $21 + 34 + 55 + 89$ . A nouveau nous pouvons entrevoir la richesse combinatoire des nombres de Fibonacci et de Lucas.

### 3. Des exemples typiques de représentation des résonances

Nous allons explorer ici les différentes partitions de résonances et profiterons de cette occasion pour commenter plusieurs types de présentation des résultats.

#### a) Synoptiques de synthèse des treize partitions de résonances

Une forme très synthétique est la suivante : nous allons y détailler, position de base par position de base, toutes les résonances ventilées selon les 13 partitions et en précisant, pour chaque résonance sa longueur, la position de la base, les bi-résonances (TC/AG, TA/CG, TG/CA), les résonances simples (T/CAG, C/AGT, A/GTC, G/TCA) et les méta-résonances (T/C, T/A, T/G, C/A, C/G,A/G). Ce type de présentation permet de bien suivre le flot des résonances, il y manque cependant l'information de type de résonance (FFF, FFL, LFF, LLL). Nous présentons en figure 1.4 le début d'un tel tableau pour l'oncogène NRAS, dont voici la séquence d'ADN (cf. figure 1.3), longue de 630 nucléotides. Cet oncogène est présent chez tous les humains et joue un rôle déterminant dans certains cancers.

Ici, pour un ADN de 630 bases, contenant un gène de 570 bases (car ce gène comporte de très petits introns : 60 bases de "résidu" pour 570 bases exprimées), on constate un très haut niveau de structure de supra-code. Ainsi, à partir des neuf premières bases explorées, débutent 57 résonances de plus de 47 bases.

- de la première base, partent quatre résonances de plus de 100 bases;
- des bases 1 et 2, partent sept résonances de plus de 100 bases;
- des bases 1, 2, 3 et 4 partent onze résonances de plus de 100 bases.

```

CAGGTTCTTGCTGGTGTGAAATGACTGAGTACAAACTGGTGGTGGTTGGAGCAGGTGGTG
TTGGGAAMGCGCACTGACAATCCAGCTAATCCAGAACCCTTTGTAGATGAATAIGATC
CCACCATAGAGGTGAGGCCAGTGGTAGCCGATTCCTACAGAAAACAAGTGGTTATAGA
TGGTGAACCTGTTTGTGGACATACTGGATACAGCTGGACAAGAAGAGTACAGTGCCAT
GAGAGACCAATACATGAGGACAGGCGAAGGCTTCTCTGTGTATTTGCCATCAATAATAG
CAGTCAATTTGCGGATATTAACCTCTACAGGTACTAGGAGTCATTATTTGGAGCAGATT
MGGCAGTAAMGACTCGGATGATGTACCTATGGTGTAGTGGGAACAAGTGTGATTTG
CCACAMGGACAGTTGATACAAACAGCCCAAGAACGGCCAAAGATACGGGATCCCA
TTCATTGAAACCTCAGCCAAGACCAGACAGGGTGTGAAGATGCTTTTACACACTGGTA
AGAGAAATACGCCAGTACCGAATGAAAAAAGCTCAACAGCAGTGTATGAGGACTCAGGGT
TGTATGGGATGCAATGTGTGGTGTATGTA
    
```

Figure 1. 3. La séquence d'ADN de 630 bases de l'oncogène NRAS

BASE:	TCAG	TACG	TGCA	T/..	C/..	A/..	G/..	T/C	T/A	T/G	C/A	C/G	A/G
1	0	0	0	0	0	199	76	0	0	0	144	123	0
1	0	0	0	0	0	123	55	0	0	0	89	89	0
2	76	0	0	0	0	89	199	55	0	0	144	0	0
2	0	0	0	0	0	0	123	0	0	0	89	0	0
2	0	0	0	0	0	0	0	0	0	0	55	0	0
2	0	0	0	0	0	0	0	0	0	0	47	0	0
3	76	0	0	0	0	89	199	47	0	0	0	89	0
3	0	0	0	0	0	0	0	0	0	0	0	76	0
4	0	0	0	0	0	0	199	76	0	0	0	123	47
4	0	0	0	0	0	0	123	0	0	0	0	89	0
5	0	0	0	0	0	0	123	123	47	0	0	0	0
5	0	0	0	0	0	0	0	55	0	0	0	0	0
6	76	0	76	0	0	0	123	47	0	0	0	0	0
7	76	0	76	47	0	0	0	47	0	0	144	123	0
7	0	0	0	0	0	0	0	0	0	0	89	89	0
7	0	0	0	0	0	0	0	0	0	0	47	55	0
8	76	0	76	47	0	0	47	47	0	0	0	0	0
8	47	0	0	0	0	0	0	0	0	0	0	0	0
9	89	0	0	0	0	0	0	0	0	47	0	0	0
9	76	0	0	0	0	0	0	0	0	0	0	0	0
9	55	0	0	0	0	0	0	0	0	0	0	0	0
10	89	0	0	0	0	0	0	0	0	47	0	55	0

Figure 1. 4. Synoptique des treize partitions de résonances de NRAS

N'oublions pas que cette séquence ne comporte que 630 bases et que ce tronçon des toutes premières bases (des bases 1 à 15) constitue le précurseur du gène. C'est donc une région non traduite en acides aminés, donc en principe peu significative puisque non codante. On observe pourtant un très haut niveau de structure qui relie cette région au gène proprement dit. Nous reviendrons sur cette apparente contradiction en 3.III à propos de l'interleukine, gène essentiel dans le traitement thérapeutique des cancers

**b. Modes de représentation typiques des résonances simples**

Nous proposons trois formes plus ou moins synthétiques de représentation des résonances simples. D'autres formes pourraient d'ailleurs être imaginées comme les "flots" de résonances entrants et sortants pour un sous tronçon d'ADN donné. Les trois modes de représentation les plus usuels seront :

- la liste détaillée des résonances;
- la synthèse par nombres et volumes de résonances;
- et les représentations mixtes "tableaux et courbes".

• *Les listes détaillées des résonances*

La forme la plus détaillée de présentation des résonances (limitée ici aux résonances simples) est constituée d'une liste séquentielle à raison d'une

ligne par résonance. Cette liste est triée selon l'ordre croissant des adresses de départ des résonances. Chaque ligne comporte:

- l'adresse de la base de début de la résonance,
- son type (FFF, FFL, LFF, LLL),
- sa longueur en nombre de bases,
- le détail de la base en résonance et de ses trois bases complémentaires.

BASE:	TCAG	TACG	TGCA	T/..	C/..	A/..	G/..	T/C	T/A	T/G	C/A	C/G	A/G
1	0	0	0	0	0	199	76	0	0	0	144	123	0
1	0	0	0	0	0	123	55	0	0	0	89	89	0
2	76	0	0	0	0	89	199	55	0	0	144	0	0
2	0	0	0	0	0	0	123	0	0	0	89	0	0
2	0	0	0	0	0	0	0	0	0	0	55	0	0
2	0	0	0	0	0	0	0	0	0	0	47	0	0
3	76	0	0	0	0	89	199	47	0	0	0	89	0
3	0	0	0	0	0	0	0	0	0	0	0	76	0
4	0	0	0	0	0	0	199	76	0	0	0	123	47
4	0	0	0	0	0	0	123	0	0	0	0	89	0
5	0	0	0	0	0	0	123	123	47	0	0	0	0
5	0	0	0	0	0	0	0	55	0	0	0	0	0
6	76	0	76	0	0	0	123	47	0	0	0	0	0
7	76	0	76	47	0	0	0	47	0	0	144	123	0
7	0	0	0	0	0	0	0	0	0	0	89	89	0
7	0	0	0	0	0	0	0	0	0	0	47	55	0
8	76	0	76	47	0	0	47	47	0	0	0	0	0
8	47	0	0	0	0	0	0	0	0	0	0	0	0
9	89	0	0	0	0	0	0	0	0	47	0	0	0
9	76	0	0	0	0	0	0	0	0	0	0	0	0
9	55	0	0	0	0	0	0	0	0	0	0	0	0
10	89	0	0	0	0	0	0	0	0	47	0	55	0

Figure 1. 5. Un exemple de résonance par blocs dans HIV1-LAV1A

La figure 1.5. représente ces résonances pour les virus HIV1-LAV1A du Sida. Je n'y ai représenté que les résonances longues ou moyennes, en l'occurrence les résonances de plus de 610 bases.

On remarquera, avec l'imbrication par blocs homogènes des libellés FFF, LLL, etc., que ces résonances s'auto-organisent par paquets de résonances de même type. Nous reviendrons sur ce point en 4a. On remarquera aussi que, dès le début du génome, démarrent de très longues résonances recouvrant la moitié du génome (4 181 bases pour un génome de 9 000 bases). Enfin certaines résonances se présentent par "binômes" ou par séquences de longueurs croissantes, par exemple :

- de la base 152, partent simultanément 2 résonances de 3 571 et 4 181 ;
- même phénomène à partir des bases 284 puis 285 avec des résonances de 843 et 987.

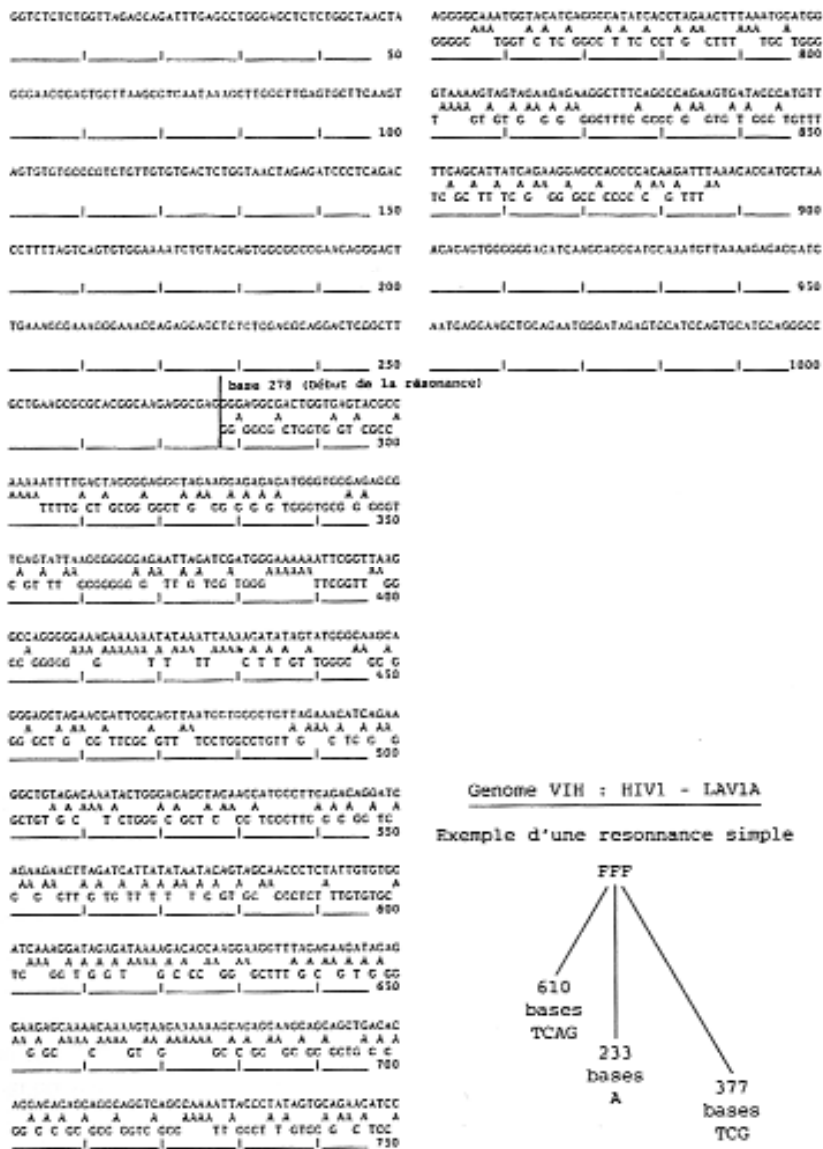


Figure 1. 6. Le détail d'une résonance FFF 610/223/377

Détaillons maintenant la structure et la composition d'une de ces résonances simples, nucléotide par nucléotide.

On a représenté en figure 1.6 la résonance de type FFF de longueur 610 qui part de la base 278. Cette résonance longue de 610 bases contient 233 A et 377 T C G.

• Les synthèses par nombres et volumes de résonances

Il s'agit d'un type de synthèse souvent utilisé, qui ne tient pas compte des positions des résonances mais simplement de leurs nombres, du cumul de leurs longueurs (volumes) et éventuellement de leurs longueurs moyennes. Ces données sont ventilées dans des matrices de 4 x 4 cases par type de résonances (FFF, FFL, LFF, LLL) et par bases (T C A G).

La figure 1.7 propose une telle synthèse pour le gène HUMC1A1. Deux tableaux résument les résonances de son ADN long de plus de 6 000 bases (introns plus exons), et celles du gène proprement dit : 1 767 bases d'ARN réparties sur 25 exons.

En comparant ces deux séries de tableaux, on notera une distribution types/bases très hétérogène pour l'ADN, tandis qu'elle est très homogène, compacte et typée pour l'ARN : dans ce dernier cas, les profils ou signatures tels que définis précédemment, ont beaucoup de relief (valeurs "0" sur les tableaux).

SYNTHESE COMPAREE DES RESONANCES DU GENE EUCARYOTE HUMC1A1...

...POUR SON ADN...

\*\*\*\*\*

SYNTHESE DE TOUTES LES RESONANCES HUMC1A1 ADN...

NOMBRES...				VOLUMES...				MOYENNES DES RESONANCES...			
T	C	A	G	T	C	A	G	T	C	A	G
257	959	57	605	23175	114402	3645	84254	91	120	64	123
955	290	2333	658	92590	18916	421988	68401	97	66	181	104
1706	1755	898	2244	380809	519622	82029	835249	224	297	92	373
301	1067	64	806	27850	109540	3501	99557	93	103	55	124

ET POUR SON ARN...

\*\*\*\*\*

SYNTHESE DE TOUTES LES RESONANCES HUMC1A1 ARN...

NOMBRES...				VOLUMES...				MOYENNES DES RESONANCES...			
T	C	A	G	T	C	A	G	T	C	A	G
0	271	0	412	0	29053	0	37535	0	100	0	92
271	7	679	0	30711	385	148577	0	114	55	219	0
126	372	90	345	6618	33622	5863	31514	53	91	66	92
0	323	4	454	0	32680	188	40782	0	102	47	90

Figure 1. 7. Résonances simples pour l'ADN et l'ARN du gène HUMC1A1

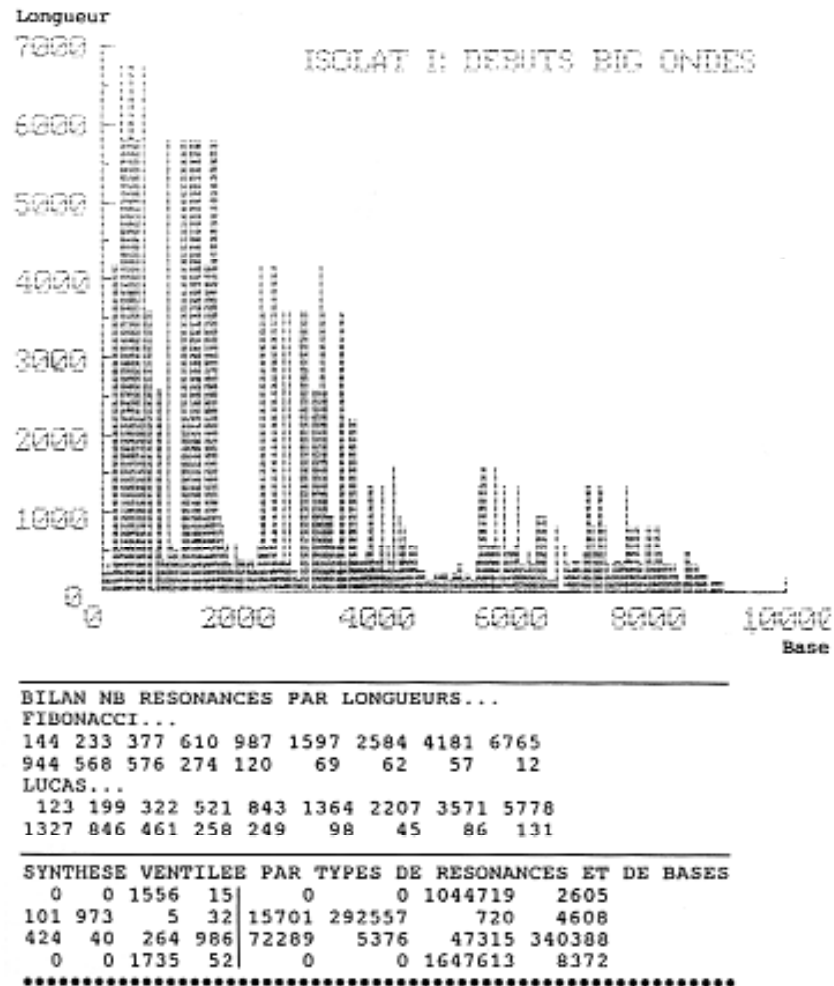


Figure 1. 8. Courbes de résonances d'un génome chimère d'HIV1

Dans la partie droite des tableaux, nous représentons les longueurs moyennes des résonances (note 1). Ces valeurs sont difficilement comparables entre les deux tableaux puisque les longueurs des séquences sont fort différentes entre l'ADN (plus de 6 000 bases) et l'ARN (1 767 bases). Cependant, sans considérer ces valeurs, une analyse de contraste ou de différenciation ("rugosité" en jargon fractal) met en évidence un plus fort relief pour l'ARN que pour l'ADN.

• *Les représentations mixtes "tableaux et courbes"*

Une autre forme de présentation, relativement synthétique, est illustrée par les figures 1.8 et 1.9. Il s'agit ici de génomes entiers de virus chimères du Sida que j'avais étudiés avec le Pr. Jean-Claude Chermann. Le virus représenté ici est très fortement pathogène in vitro. Nous l'avons baptisé "I" pour des raisons de confidentialité. La figure 1.8 représente la totalité des résonances simples de plus de 123 bases de long. La figure 1.9 est un zoom sur un gène très important du virus (cette importance est corrélée entre les mesures biologiques et les analyses par le supra-code). Ce gène, situé vers le début du génome, est le gène GAG.

Nous avons donc représenté une nouvelle forme de ventilation des résonances : la ventilation des nombres de résonances par longueurs et par grands types (Fibonacci ou Lucas). Elle permet d'identifier immédiatement les très longues résonances ainsi que leurs proportions relatives. Ainsi, la figure 1.8 nous apprend qu'il y a plus de résonances de 5 778 bases (131) qu'il n'y a de résonances de 987 bases (120).

Au niveau des représentations graphiques, on voit ici deux variantes de représentations. La figure 1.8 trace verticalement les longueurs de résonances et horizontalement leurs positions. Cela permet de visualiser les flots et concentrations dynamiques de résonances. La figure 1.9 est sensiblement différente : la courbe du bas donne les positions de début des résonances, tandis que la courbe du haut, c'est-à-dire le nuage de points, donne les positions de fin des résonances. L'écart entre les deux courbes est donc l'amplitude des résonances. Ainsi, le léger liseré blanc qui sépare les deux courbes représente les petites résonances (celles de 123 ou 144 bases).



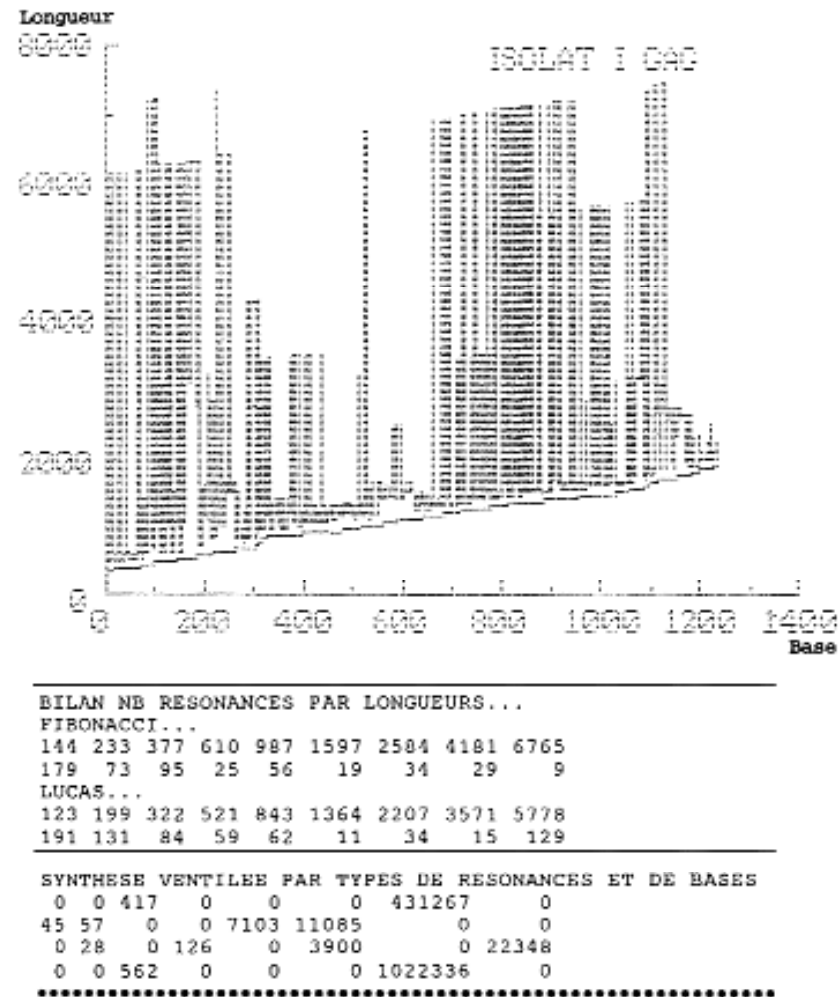


Figure 1. 9. Résonances simples issues du gène GAG du virus HIV1

Enfin, en comparant, pour les deux figures 1.8 et 1.9, les nombres relatifs de très longues résonances (5 778 et 6 765 bases), on s'apercevra que la majeure partie des très longues résonances (couvrant les deux tiers du génome) partent presque toutes du gène GAG : 9 résonances de 6 765 bases pour un total de 12, et 129 résonances de 5 778 bases pour un total de 131.