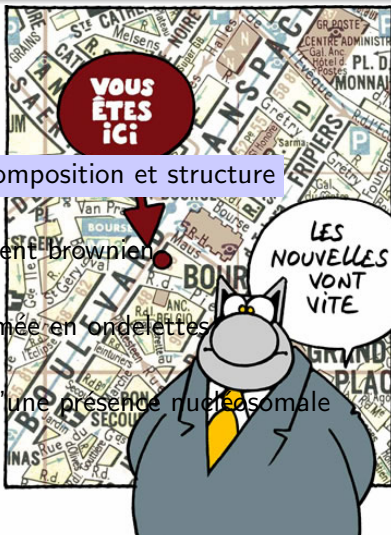


# Des maths dans l'ADN

S. Nicolay



## Nous sommes ici



- L'ADN : composition et structure
- Le mouvement brownien
- La transformée en ondelettes
- Signature d'une présence nucléaire

Pour former une macromolécule (polymère), on a besoin de briques de base (monomères)...

Pour former une macromolécule (polymère), on a besoin de briques de base (monomères)...



(4 pour l'ADN)

Pour former une macromolécule (polymère), on a besoin de briques de base (monomères)...



(4 pour l'ADN)

On assemble ces briques de base pour construire une macromolécule...

On assemble ces briques de base pour construire une macromolécule...



On assemble ces briques de base pour construire une macromolécule...



Dans la chaîne, 🤝 représente une « liaison chimique » (pour l'ADN : lien phospho-diester)



En 1953, Watson et Crick ont découvert que l'ADN possède une configuration en double hélice.

En 1953, Watson et Crick ont découvert que l'ADN possède une configuration en double hélice.

La molécule d'ADN est composée de deux chaînes (brins polynucléotidiques), allant en directions opposées.

En 1953, Watson et Crick ont découvert que l'ADN possède une configuration en double hélice.

La molécule d'ADN est composée de deux chaînes (brins polynucléotidiques), allant en directions opposées.

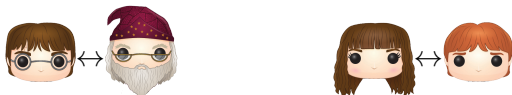
Une brique d'une chaîne ne s'apparie qu'avec une brique de l'autre chaîne spécifique (par des liaisons hydrogènes).



En 1953, Watson et Crick ont découvert que l'ADN possède une configuration en double hélice.

La molécule d'ADN est composée de deux chaînes (brins polynucléotidiques), allant en directions opposées.

Une brique d'une chaîne ne s'apparie qu'avec une brique de l'autre chaîne spécifique (par des liaisons hydrogènes).

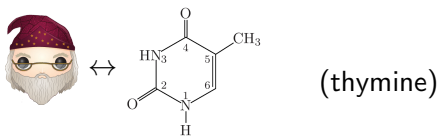
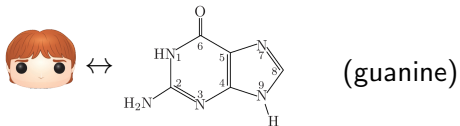
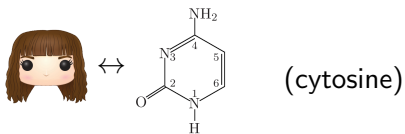
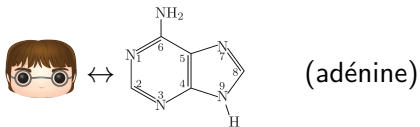


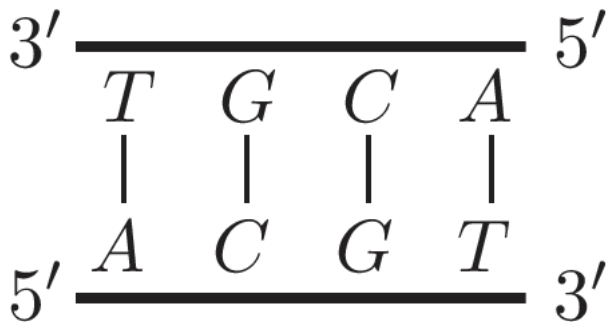
Ces deux brins sont complémentaires : la séquence d'une chaîne détermine la séquence de l'autre chaîne.



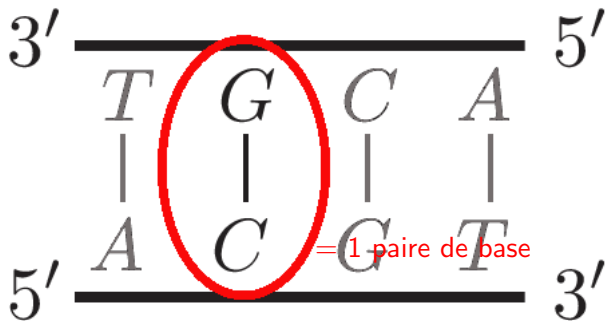


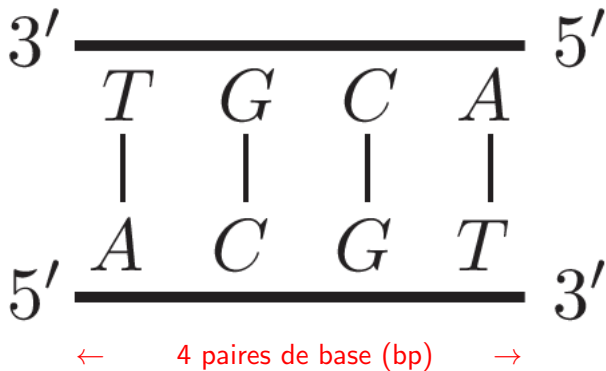
Pour l'ADN : les briques sont des nucléotides.

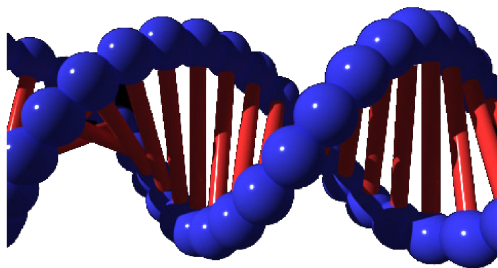












La transcription est le processus par lequel une partie de la séquence ADN est copiée pour former un brin d'ARN. Les segments d'ADN transcrits sont appelés gènes.

Ce sont les gènes qui codent pour la structure chimique des protéines, constituant fondamentaux des cellules. Plus précisément, une succession de nucléotides peut déterminer une séquence d'acides aminés formant la protéine.

La transcription est le processus par lequel une partie de la séquence ADN est copiée pour former un brin d'ARN. Les segments d'ADN transcrits sont appelés gènes.

Ce sont les gènes qui codent pour la structure chimique des protéines, constituant fondamentaux des cellules.

Plus précisément, une succession de nucléotides peut déterminer une séquence d'acides aminés formant la protéine.

Le codage de tri-nucléotides, appelés codons, en acides aminés est un principe essentiellement commun à tous les organismes, de la bactérie à l'homme.

La transcription est le processus par lequel une partie de la séquence ADN est copiée pour former un brin d'ARN.  
Les segments d'ADN transcrits sont appelés gènes.

Ce sont les gènes qui codent pour la structure chimique des protéines, constituant fondamentaux des cellules.

Plus précisément, une succession de nucléotides peut déterminer une séquence d'acides aminés formant la protéine.

La totalité du matériel génétique d'une cellule ou d'un individu est appelé génome.

Chez l'humain :  $\sim 3\,000\,000\,000$ bp de nucléotides  
pour  $\sim 30\,000$  gènes.

La transcription est le processus par lequel une partie de la séquence ADN est recopiée pour former un brin d'ARN.

Les segments d'ADN transcrits sont appelés gènes.

Même les gènes les plus longs n'utilisent qu'une faible portion de leur séquence pour coder l'information nécessaire à l'expression d'une protéine (exons).

Environ 80% des gènes sont non-codants (introns).

Plus l'organisme est complexe, plus la quantité et la taille de ces régions sont importantes.

La transcription est le processus par lequel une partie de la séquence ADN est copiée pour former un brin d'ARN.  
Les segments d'ADN transcrits sont appelés gènes.

Dans l'ADN, il y a beaucoup de nucléotides qui ne servent à rien !  
= poubelle ?



La transcription est le processus par lequel une partie de la séquence ADN est recopiée pour former un brin d'ARN.  
Les segments d'ADN transcrits sont appelés gènes.

Dans l'ADN, il y a beaucoup de nucléotides qui ne servent à rien !  
= poubelle ?

Notre but : étudier ces régions.

## Nous sommes ici

- L'ADN : composition et structure
- Le mouvement brownien
- La transformée en ondelettes
- Signature d'une présence nucléaire



La mouvement brownien : le point de vue d'un marcheur ivre

La mouvement brownien : le point de vue d'un marcheur ivre

Supposons qu'un promeneur ivre marche dans une rue bordée de lampadaires (à intervalles réguliers).

La mouvement brownien : le point de vue d'un marcheur ivre

Supposons qu'un promeneur ivre marche dans une rue bordée de lampadaires (à intervalles réguliers).

À chaque fois que le marcheur arrive au niveau d'un lampadaire, il s'agrippe à celui-ci.

La mouvement brownien : le point de vue d'un marcheur ivre

Supposons qu'un promeneur ivre marche dans une rue bordée de lampadaires (à intervalles réguliers).

À chaque fois que le marcheur arrive au niveau d'un lampadaire, il s'agrippe à celui-ci.

Toutefois, comme il est ivre, au moment de repartir, il ne se souvient plus de l'endroit d'où il vient.

La mouvement brownien : le point de vue d'un marcheur ivre

Supposons qu'un promeneur ivre marche dans une rue bordée de lampadaires (à intervalles réguliers).

À chaque fois que le marcheur arrive au niveau d'un lampadaire, il s'agrippe à celui-ci.

Toutefois, comme il est ivre, au moment de repartir, il ne se souvient plus de l'endroit d'où il vient.

Il repart donc dans un sens ou dans l'autre de manière aléatoire ;

La mouvement brownien : le point de vue d'un marcheur ivre

Supposons qu'un promeneur ivre marche dans une rue bordée de lampadaires (à intervalles réguliers).

À chaque fois que le marcheur arrive au niveau d'un lampadaire, il s'agrippe à celui-ci.

Toutefois, comme il est ivre, au moment de repartir, il ne se souvient plus de l'endroit d'où il vient.

Il repart donc dans un sens ou dans l'autre de manière aléatoire ; il a une chance sur deux de revenir sur ses pas et une sur deux de continuer dans la même direction.



Une illustration...

Video

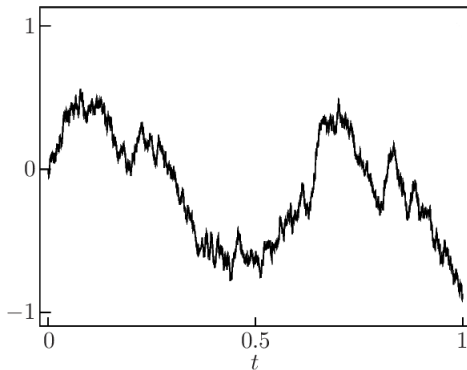
# Le mouvement brownien : le point de vue du mathématicien

## Le mouvement brownien : le point de vue du mathématicien

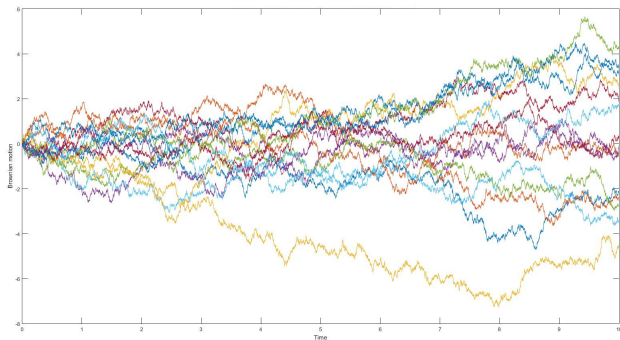
Un mouvement brownien est un processus stochastique (i.e. une « fonction à valeurs aléatoires »)  $B$  tel que

- $B(0) = 0$ ,
- $B$  est (presque sûrement) continu,
- $B(t+h) - B(t)$  est « indépendant de tout ce qui s'est passé avant  $t$  » ( $h > 0$ ),
- $B(t+h) - B(t)$  suit une loi normale de moyenne nulle et de variance  $h$  ( $h > 0$ ).

Une représentation d'une réalisation d'un mouvement brownien



## Une représentation d'une réalisation d'un mouvement brownien



Pour faire le malin...

## Théorème

Soit le cylindre

$$C_{B_n}(x_1, \dots, x_n) = \{f \in C^0([0, +\infty[) : (f(x_1), \dots, f(x_n)) \in B_n\},$$

pour tout ensemble borélien  $B_n$  à  $n$  dimensions, tous les réels  $0 \leq x_1 < x_2 < \dots < x_n$  et notons  $\mathcal{B}$  la plus petite sigma-algèbre contenant les sous-ensembles

$$\{C_{B_n}(x_1, \dots, x_n) : 0 \leq x_1 < x_2 < \dots < x_n\}$$

de  $C^0([0, +\infty[)$ . Il existe une mesure de probabilité unique  $\mathcal{W}$  sur  $(C^0([0, +\infty[), \mathcal{B})$  satisfaisant

$$P((B(t_1), \dots, B(t_n)) \in B_n) = \mathcal{W}(C_{B_n}(t_1, \dots, t_n))$$

pour tout borélien  $B_n$  et tous réels  $0 \leq t_1 < t_2 < \dots < t_n$ .

# Le mouvement brownien fractionnaire

## Le mouvement brownien fractionnaire

Rappel : pour un mouvement brownien  $B$  :  $B(t+h) - B(t)$  suit une loi normale de moyenne nulle et de variance  $h$  ( $h > 0$ ).

Si on regarde les incréments  $B(t+h) - B(t)$  d'un mouvement brownien, ces valeurs ne sont pas corrélées.



## Le mouvement brownien fractionnaire

Rappel : pour un mouvement brownien  $B$  :  $B(t+h) - B(t)$  suit une loi normale de moyenne nulle et de variance  $h$  ( $h > 0$ ).

Si on regarde les incréments  $B(t+h) - B(t)$  d'un mouvement brownien, ces valeurs ne sont pas corrélées.

Pour un mouvement brownien fractionnaire d'indice  $H$  ( $0 < H < 1$ ),  $B_H$ , on demande que  $B_H(t+h) - B_H(t)$  suive une loi normale de moyenne nulle et de variance proportionnelle à  $h^{2H}$ .

## Le mouvement brownien fractionnaire

Rappel : pour un mouvement brownien  $B$  :  $B(t+h) - B(t)$  suit une loi normale de moyenne nulle et de variance  $h$  ( $h > 0$ ).

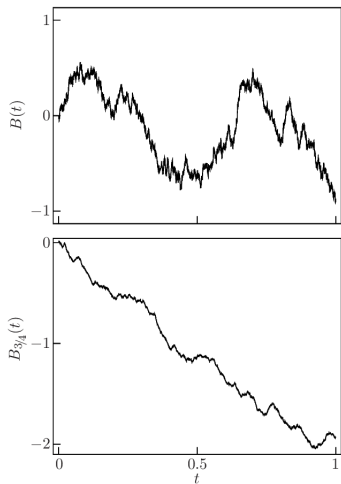
Si on regarde les incréments  $B(t+h) - B(t)$  d'un mouvement brownien, ces valeurs ne sont pas corrélées.

Pour un mouvement brownien fractionnaire d'indice  $H$  ( $0 < H < 1$ ),  $B_H$ , on demande que  $B_H(t+h) - B_H(t)$  suive une loi normale de moyenne nulle et de variance proportionnelle à  $h^{2H}$ .

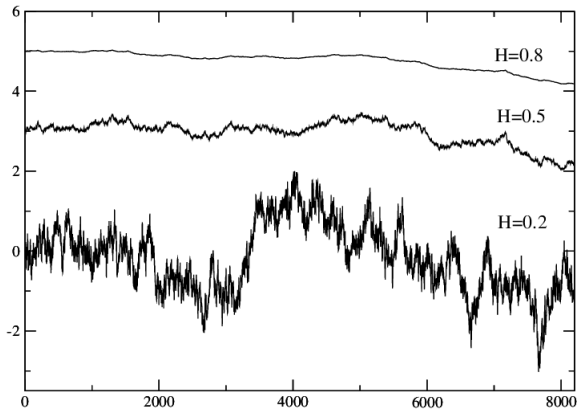
Si  $H > 1/2$ , les valeurs  $B_H(t+h) - B_H(t)$  les valeurs sont légèrement corrélées : la corrélations tend vers 0 à l'infini, mais très doucement.

On parle de corrélations à longue portée.

## Représentation d'une réalisation...



## Représentation d'une réalisation...



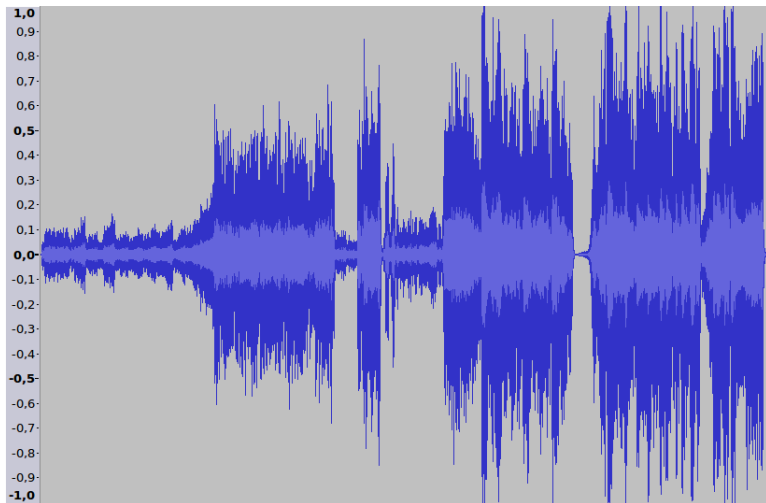
## Nous sommes ici

- L'ADN : composition et structure
- Le mouvement brownien
- La transformée en ondelettes
- Signature d'une présence nucléaire





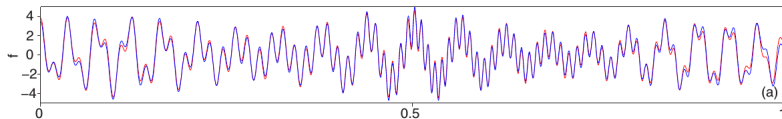
La représentation de la musique :



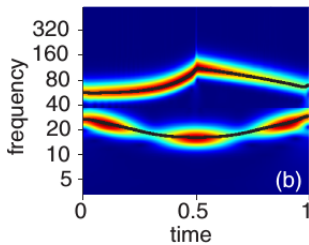
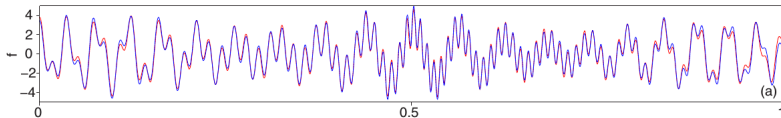




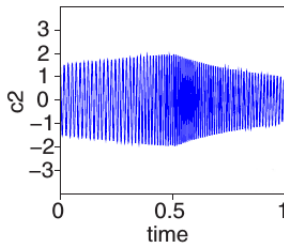
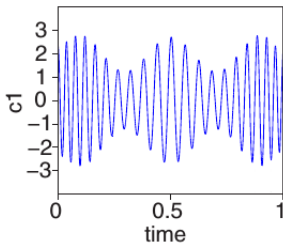
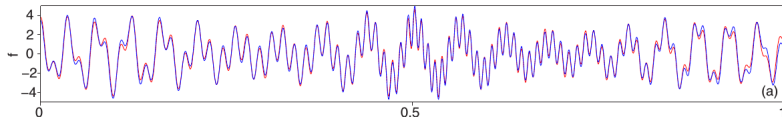
On cherche souvent à associer plusieurs représentations à une fonction...



On cherche souvent à associer plusieurs représentations à une fonction...



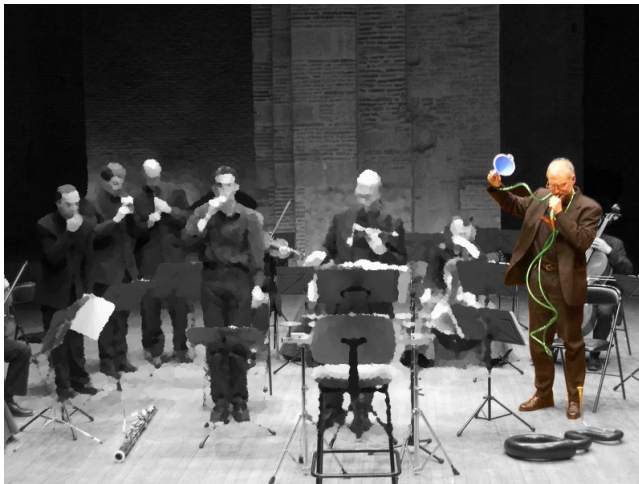
On cherche souvent à associer plusieurs représentations à une fonction...



La transformée en ondelettes permet de se concentrer sur certaines particularités de la fonction



La transformée en ondelettes permet de se concentrer sur certaines particularités de la fonction



L'idée est de décomposer une fonction en atomes élémentaires (les notes de musique) appelés ondelettes.

L'idée est de décomposer une fonction en atomes élémentaires (les notes de musique) appelés ondelettes.

Grâce aux ondelettes, on peut estimer l'indice  $H$  d'un mouvement brownien fractionnaire  $B_H$ .

## Nous sommes ici

- L'ADN : composition et structure
- Le mouvement brownien
- La transformée en ondelettes
- Signature d'une présence nucléosomale





En étude du signal, on considère un brin d'ADN comme un mot formé à partir de 4 lettres (A, C, G et T).

ACGAACGTTACGGAATGCCAGGGTACCC...

En étude du signal, on considère un brin d'ADN comme un mot formé à partir de 4 lettres (A, C, G et T).

ACGAACGTTACGGAATGCCAGGGTACCC...

Cela représente encore trop d'information :

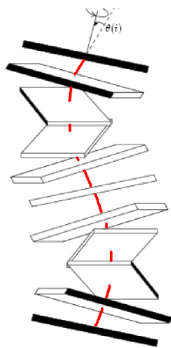
On va coder ce mot fabriqué avec 4 lettres pour en faire une fonction.

Le codage PNuc est obtenu par l'observation du mode de positionnement préférentiel des nucléosomes sur des séquences ADN.

Il tente de rendre compte de la courbure spontanée de la double hélice (dans un complexe nucléosomal).

Le codage PNuc est obtenu par l'observation du mode de positionnement préférentiel des nucléosomes sur des séquences ADN.

Il tente de rendre compte de la courbure spontanée de la double hélice (dans un complexe nucléosomal).

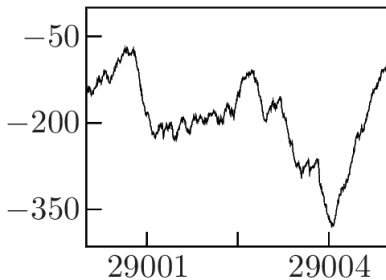


Le codage PNuc est obtenu par l'observation du mode de positionnement préférentiel des nucléosomes sur des séquences ADN.

Il tente de rendre compte de la courbure spontanée de la double hélice (dans un complexe nucléosomal).

Tri-base	PNuc	Tri-base	PNuc	Tri-base	PNuc	Tri-base	PNuc
<i>AAA</i>	0	<i>CAA</i>	3.3	<i>GAA</i>	3	<i>TAA</i>	2
<i>AAC</i>	3.7	<i>CAC</i>	6.5	<i>GAC</i>	5.4	<i>TAC</i>	3.7
<i>AAG</i>	5.2	<i>CAG</i>	4.2	<i>GAG</i>	5.4	<i>TAG</i>	2.2
<i>AAT</i>	0.7	<i>CAT</i>	6.7	<i>GAT</i>	5.3	<i>TAT</i>	2.8
<i>ACA</i>	5.2	<i>CCA</i>	5.4	<i>GCA</i>	6	<i>TCA</i>	5.4
<i>ACC</i>	5.4	<i>CCC</i>	6	<i>GCC</i>	10	<i>TCC</i>	3.8
<i>ACG</i>	5.4	<i>CCG</i>	4.7	<i>GCG</i>	7.5	<i>TCG</i>	8.3
<i>ACT</i>	5.8	<i>CCT</i>	5.4	<i>GCT</i>	7.5	<i>TCT</i>	3.3
<i>AGA</i>	3.3	<i>CGA</i>	8.3	<i>GGA</i>	3.8	<i>TGA</i>	5.4
<i>AGC</i>	7.5	<i>CGC</i>	7.5	<i>GGC</i>	10	<i>TGC</i>	6
<i>AGG</i>	5.4	<i>CGG</i>	4.7	<i>GGG</i>	6	<i>TGG</i>	5.4
<i>AGT</i>	5.8	<i>CGT</i>	5.4	<i>GGT</i>	5.4	<i>TGT</i>	5.2
<i>ATA</i>	2.8	<i>CTA</i>	2.2	<i>GTA</i>	3.7	<i>TTA</i>	2
<i>ATC</i>	5.3	<i>CTC</i>	5.4	<i>GTC</i>	5.4	<i>TTC</i>	3
<i>ATG</i>	6.7	<i>CTG</i>	4.2	<i>GTG</i>	6.5	<i>TTG</i>	3.3
<i>ATT</i>	0.7	<i>CTT</i>	5.2	<i>GTT</i>	3.7	<i>TTT</i>	0

Signal PNuc d'une partie du chromosome 21 (en milliers de pb) :



Grâce à la transformée en ondelettes, on peut montrer que l'exposant  $H$  mesuré sur ces marches dépend du type d'organisme considéré.

Grâce à la transformée en ondelettes, on peut montrer que l'exposant  $H$  mesuré sur ces marches dépend du type d'organisme considéré.

Les organismes eucaryotes sont associés à une valeur  $H$  voisine de 0.6.



Grâce à la transformée en ondelettes, on peut montrer que l'exposant  $H$  mesuré sur ces marches dépend du type d'organisme considéré.

Les organismes eucaryotes sont associés à une valeur  $H$  voisine de 0.6.

Suivant le modèle brownien, ces organismes présentent des corrélations à longue portée.

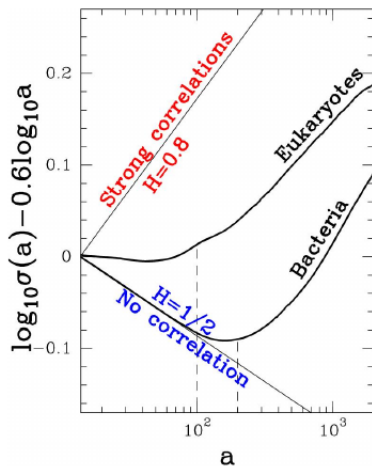
Grâce à la transformée en ondelettes, on peut montrer que l'exposant  $H$  mesuré sur ces marches dépend du type d'organisme considéré.

Les organismes eucaryotes sont associés à une valeur  $H$  voisine de 0.6.

Suivant le modèle brownien, ces organismes présentent des corrélations à longue portée.

Pour les bactéries, les valeurs de  $H$  ne sont pas significativement différentes de  $1/2$ .

Dans ce cas, on ne peut donc pas mettre en évidence de corrélation.



Dans les mesures, il y avait quelques bactéries présentant un  $H$  différent de  $1/2$ .

Dans les mesures, il y avait quelques bactéries présentant un  $H$  différent de  $1/2$ .

Une étude plus poussée, révéla que ce sont des génomes archæbactériens.

Dans les mesures, il y avait quelques bactéries présentant un  $H$  différent de  $1/2$ .

Une étude plus poussée, révéla que ce sont des génomes archæbactériens.

Comme les bactéries, ces organismes sont des organismes procaryotes,

Dans les mesures, il y avait quelques bactéries présentant un  $H$  différent de  $1/2$ .

Une étude plus poussée, révéla que ce sont des génomes archæbactériens.

Comme les bactéries, ces organismes sont des organismes procaryotes, mais ces archæbactéries présentent un mécanisme d'empaquetage de leur ADN semblable à celui des organismes eucaryotes : ils font intervenir des protéines semblables aux histones.

Dans les mesures, il y avait quelques bactéries présentant un  $H$  différent de  $1/2$ .

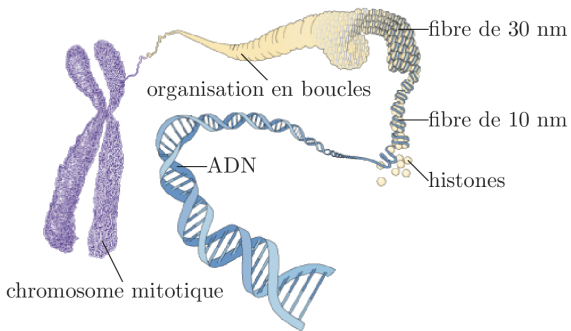
Une étude plus poussée, révéla que ce sont des génomes archæbactériens.

Comme les bactéries, ces organismes sont des organismes procaryotes, mais ces archæbactéries présentent un mécanisme d'empaquetage de leur ADN semblable à celui des organismes eucaryotes : ils font intervenir des protéines semblables aux histones.

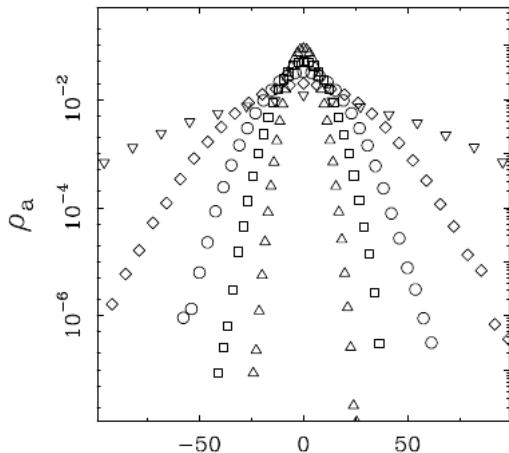
L'observation d'une valeur de  $H$  supérieure à  $1/2$  est donc la signature de la présence de nucléosomes !



## Les diverses étapes de l'empaquetage de l'ADN chez les eucaryotes :



La méthode pour trouver  $H$  : trouver le poids à placer sur les fréquences pour obtenir une invariance des distributions.



La méthode pour trouver  $H$  : trouver le poids à placer sur les fréquences pour obtenir une invariance des distributions.

