

IFT 6299 AUTOMNE 2005

Informatique de l'ADN (Sujets spéciaux an bio-informatique)

Miklós Csűrös

André-Aisenstadt 3149

csuros@iro.umontreal.ca

[http ://www.iro.umontreal.ca/~csuros/IFT6299/](http://www.iro.umontreal.ca/~csuros/IFT6299/)

Plan de cours

Préalables :

(aucun formel)

- structures discrètes et algorithmes : graphes, arbres [IFT2010, IFT2121]
- probabilités : distribution, variables aléatoires, espérance, processus stochastiques [MAT1978]
- Aucune bio !
- aucun (ou très peu) de chevauchement avec IFT6291/BIN6000

Horaire :

jeudi 13 :30–15 :30	Z-350
vendredi 13 :30–15 :30	Z-350

Matériel

aucun livre, 70+ articles récents

notes de cours affichés avant ou juste après le cours

à peu près un article à lire chaque semaine

mini-questionnaires sur la lecture obligatoire : 10% de notes (pas de questions méchantes...)

projet 30%

devoirs théoriques 30%

présentation d'un sujet 30%

Sujets

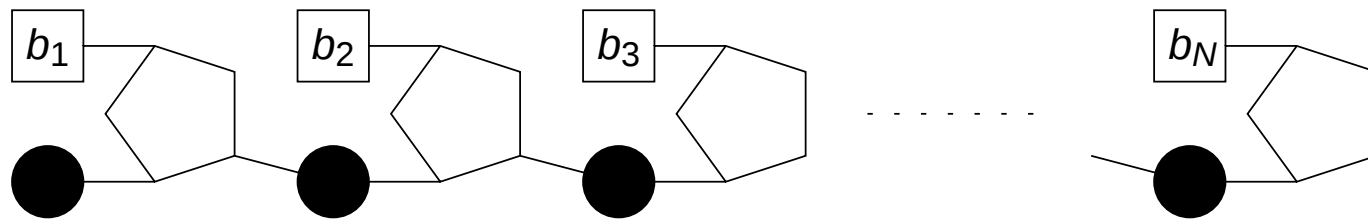
- Biotechnologie : hybridation, clonage, PCR, etc.
- Calcul moléculaire : comment placer 10^{12} ordinateurs dans une goutte d'eau ?
- Comparaison de grands ensembles de séquences : techniques de hachage, aspects pratiques.
- Alignement de grands génomes
- Génomique comparative : *footprinting*, *shadowing*, *patterns* phylétiques
- Annotation de génomes : recherche de gènes, fureteurs de génomes
- Évolution moléculaire : méthodes de vraisemblance maximale

Assemblage de séquence génomiques, cartographie physique, réseaux génétiques, ... ?

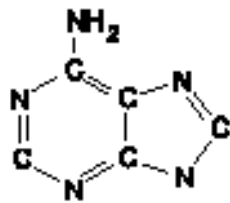
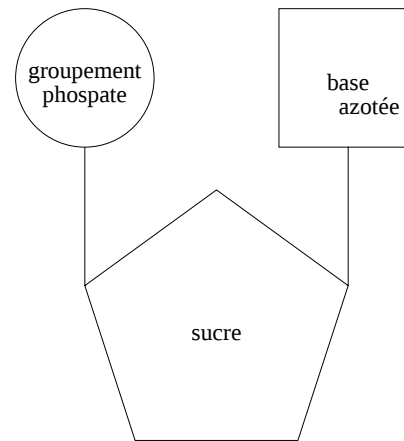
L'ADN ?

Polynucléotide :

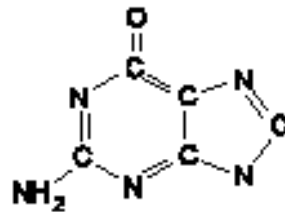
polymère (ie, chaîne) de **nucléotides**



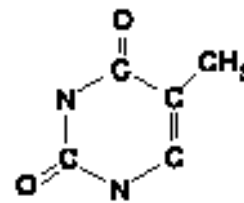
Nucléotides



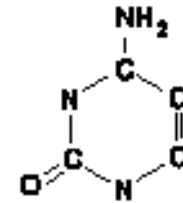
Adénine



Guanine



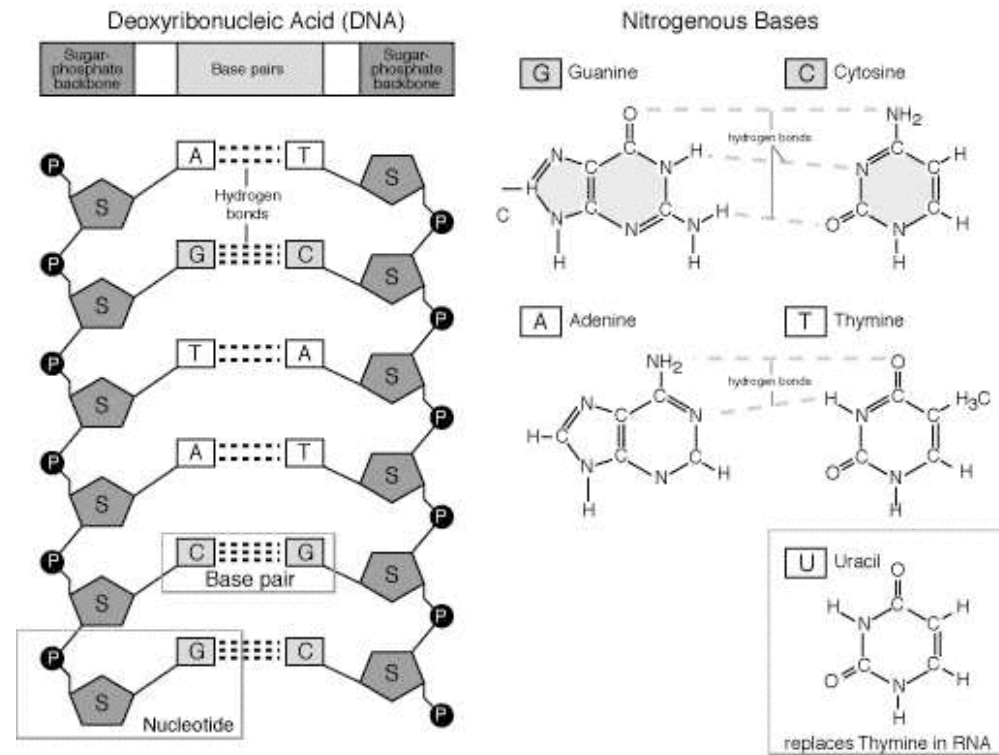
Thymine



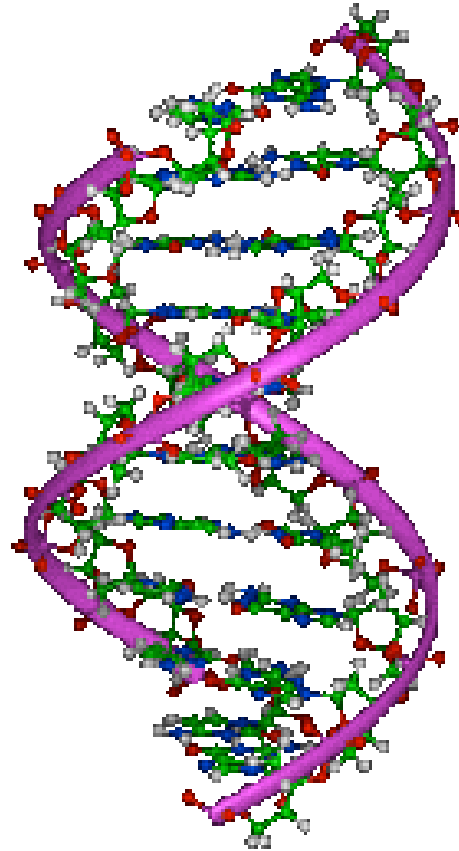
Cytosine

4 bases :

ADN : bases complémentaires



Watson et Crick : double hélice



1 pas de l'hélice : 10 paires de base, 3.4 nm (forme usuelle)

L'article

no. 4206 April 25, 1953

NATURE

737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

¹Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **46**, 149 (1900).

²Laguarda-Hugues, M. S., *Mon. Not. Roy. Astr. Soc., Geophys. Supp.*, **8**, 285 (1905).

³Von Arz, W. S., *Wegede Hole Papers in Phys. Oceanogr. Meteor.*, **11** (5) (1905).

⁴Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (Stockholm), **8**(11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey¹. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagram is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β -D-deoxy-ribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequence of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furbert's model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furbert's 'standard configuration', the sugar being roughly perpendicular to the attached base. There



This figure is purely diagrammatic. The two ribbons represent the two deoxyribose-phosphate chains, and the rungs connect the pairs of bases holding the chains together. The vertical line marks the fibre axis.

is a residue on each chain every 3.4 Å. in the z -direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 Å. The distance of a phosphorus atom from the fibre axis is 10 Å. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z -co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair on either chain, then on these assumptions

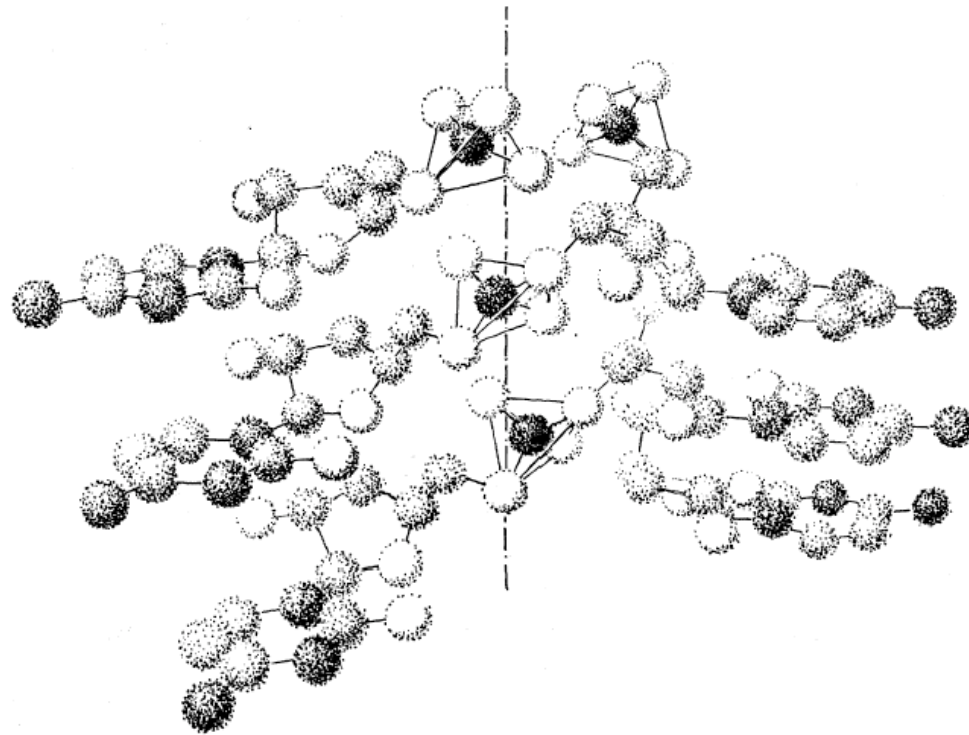
It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

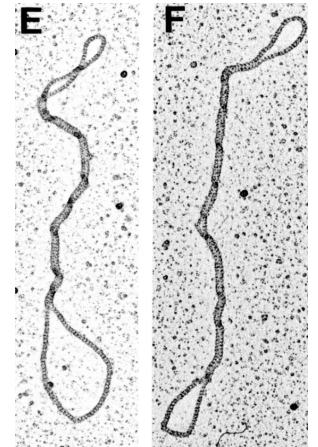
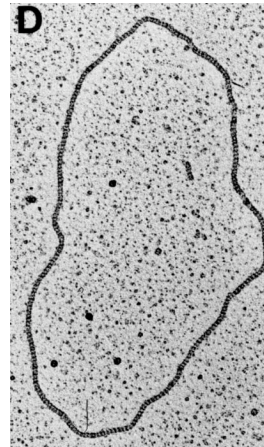
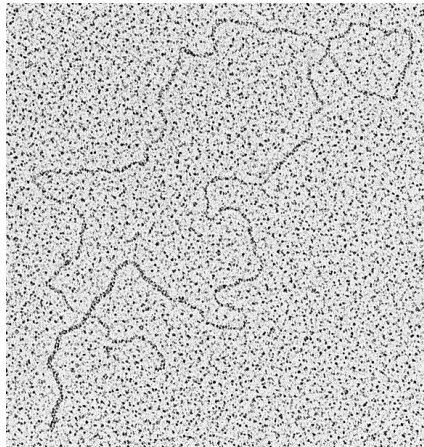
Un autre modèle — le triple hélice



Pauling & Corey, *PNAS* 39 : 84 (1953)

ADN - structure

la molécule d'ADN peut être
linéaire (nos chromosomes), ou
circulaire (bactéries, organelles)



Ross Inman

ADN

stockage de l'ADN : dans le noyau de la cellule, organisé en chromosomes (en eukaryotes)

1. duplication de l'information en ADN : hérédité
2. duplication de l'information en ARN (souvent traduit en protéines)

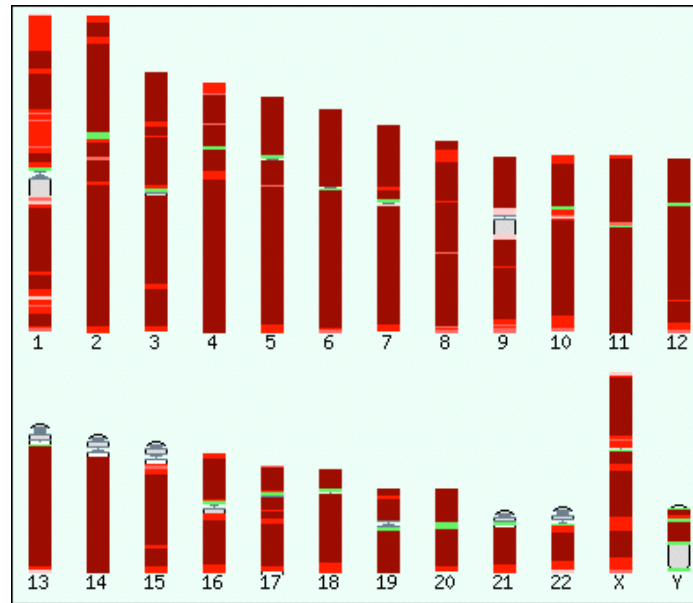
Replication

[animation]

[animation]

Lodish et al., *Molecular Biology of the Cell*, 2002

Chromosomes



22 chromosomes diploïdes et 2 chromosomes sexuels

tailles : $250 \cdot 10^6$ – $23 \cdot 10^6$ paires de base

taille totale du génome humaine : $3 \cdot 10^9$ pb

Structure de chromosomes et replication

[animation]

Dolan DNA Learning Center