

La structure de l'ADN en double hélice

par Jean-Marc Victor
Directeur de recherche au CNRS
Laboratoire de Physique Théorique de la Matière Condensée (CNRS UMR 7600)
Université Pierre et Marie Curie, Paris

Si 1905 est *l'annus mirabilis* de la physique, avec les quatre papiers historiques d'Albert Einstein¹, 1953 est incontestablement celle de la biologie. Dans le numéro du 25 avril du journal *Nature*, James D. Watson et Francis H.C. Crick révèlent la structure en double hélice de l'ADN (Acide DésoxyriboNucléique). Cette « Lettre » d'une page seulement contient en germe toute la biologie moléculaire des cinquante années suivantes, jusqu'à l'accomplissement en 2003 du projet Génome Humain².

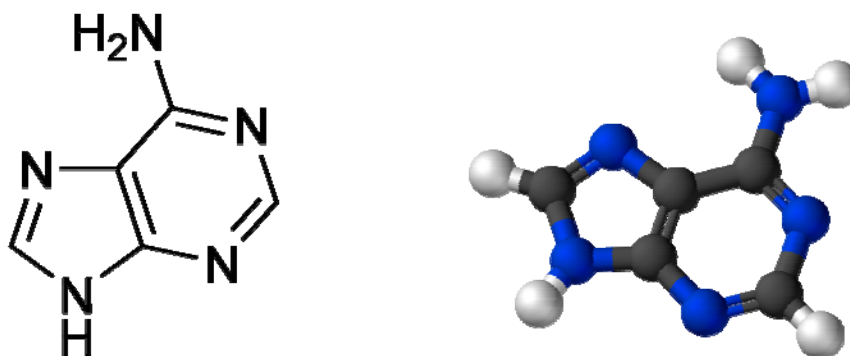


Figure 1 : Les différents éléments d'une molécule d'ADN. Pour illustrer le fait que les chromosomes sont des entités physico-chimiques, et pour mieux comprendre dans la suite du texte la structure de l'ADN, illustrons ses divers éléments moléculaires constitutifs.

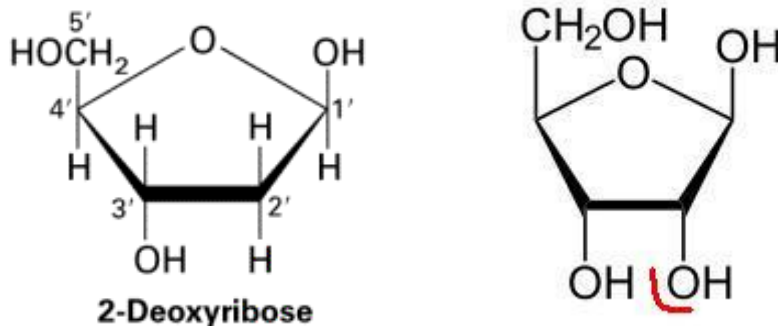
Ci-dessus : l'adénine, une des quatre bases azotées de l'ADN (les bases A, C, G, T : adénine, cytosine, guanine, thymine). L'adénine a pour formule développée $C_5H_5N_5$ (rappel des conventions pour les formules dites topologiques comme celle ci-dessus à gauche : on ne fait pas figurer les C – en gris à droite – et on ne fait pas figurer les H – en blanc à droite – quand ils sont accolés à des C ; en bleu à droite sont figurés les N).

1 bis, ci-dessous : (à gauche) le 2-désoxyribose, sucre de l'ADN. Il a pour formule développée $C_5H_{10}O_4$ (ne figurent pas sur cette formule les C en position 1', 2', 3', 4'). On a fait figurer, ci-dessous à droite, le ribose pour mieux comprendre le nom désoxyribose.

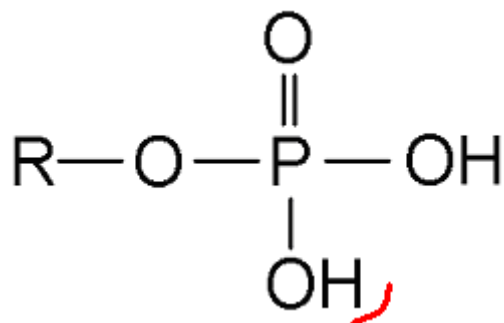
1. Ces quatre articles concernent : la théorie de la relativité restreinte, l'équivalence masse-énergie (avec la célèbre équation $E=mc^2$), la théorie de l'effet photoélectrique et celle du mouvement brownien.

2. Le projet Génome Humain avait pour objectif le séquençage complet du génome humain. Pour la petite histoire, le génome de James D. Watson est l'un des tout premiers génomes individuels à avoir été complètement séquencé. Aujourd'hui, à peine dix ans plus tard, tout individu peut se payer la séquence complète de son génome pour quelques milliers de dollars !

Le ribose a pour formule $C_5H_{10}O_5$; il appartient à la famille des glucides (sucres) ou hydrates de carbone de formule $C_n(H_2O)_p$ (ici $n = p = 5$). Il est important de comprendre que les deux formules ci-dessous sont les mêmes, à un atome près : celui d'oxygène en bas à droite (que nous avons cerclé de rouge sur la formule de droite, en position 2') : d'où le nom de désoxyribose (ribose auquel un atome d'oxygène a été ôté). Ce composant, le sucre désoxyribose, donne son nom à l'ADN acide désoxyribonucléique.



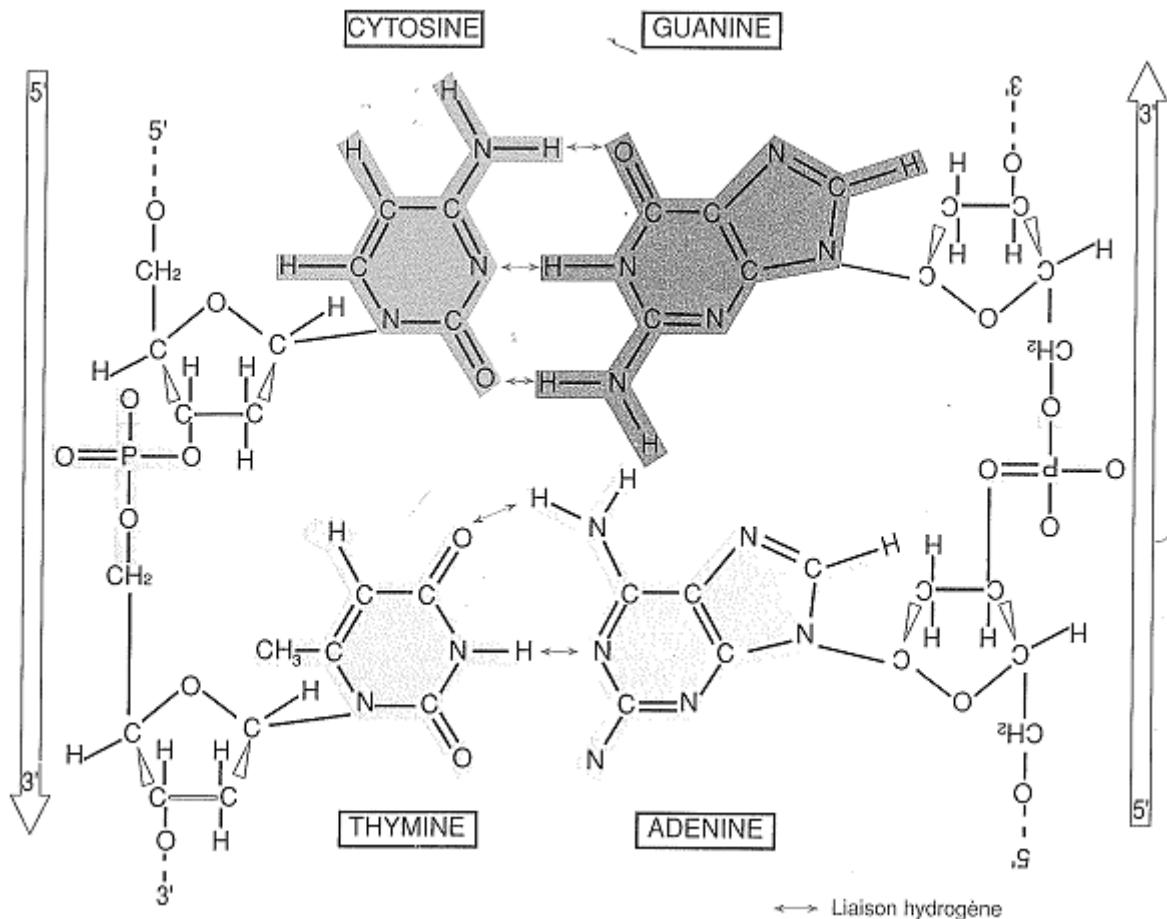
1 ter, ci-dessous : le troisième type d'élément constitutif de l'ADN, le groupe phosphate. Il se fixe aux sucres (désoxyribose, figure 1bis à gauche) à deux endroits : (voir figure 1 quater, brin de gauche) 1°) en position 5' sur le sucre « en dessous », le CH₂ joue le rôle du radical R ci-dessous (voir figure 1quater) ; 2°) en position 3' sur le sucre « au-dessus », par l'intermédiaire de l'atome O (qui a perdu son H, cerclé de rouge, lors de la polymérisation).



1 quater, ci-dessous : structure réelle complète de l'ADN avec appariement des bases (cytosine-guanine et thymine-adénine) entre les deux brins complémentaires (tiré de l'ouvrage Biologie moléculaire et médecine, Jean-Claude Kaplan et Marc Delpéch, Médecine-Sciences, Flammarion, 2ème édition (1996)).

Les bases A, T, C ou G se mettent toujours en position 1' du désoxyribose. Chaque brin est obtenu par la polymérisation alternée de sucres et de phosphates ; ce faisant il ne reste plus qu'un seul groupe OH dans chaque phosphate. Les sucres sont tous orientés de la même façon dans un brin, chaque phosphate étant ainsi entouré par un carbone 5' et un carbone 3'.

Chaque brin est orienté conventionnellement de l'extrémité 5' vers l'extrémité 3' (à gauche de haut en bas, à droite de bas en haut, les brins appariés étant ainsi antiparallèles, comme le montre la construction même de la figure). Entre A et T, C et G, sont figurées les liaisons hydrogène (voir encadré hors texte ci-après).



LES OBJECTIONS A LA STRUCTURE DE PAULING & COREY

Dès la première ligne, les auteurs annoncent qu'ils souhaitent suggérer une structure pour le sel de l'acide désoxyribonucléique. Ils préviennent d'ailleurs au paragraphe suivant que la structure en triple hélice de l'acide lui-même proposée par Pauling³ et Corey n'est à leur avis pas pertinente car les diagrammes de rayons X⁴ sur lesquels sont basés toutes les structures proposées jusqu'alors ont été obtenus avec du sel d'acide et non avec l'acide lui-même :

Nous pensons que la matière qui conduit à ces diagrammes de rayons X est le sel, et non l'acide libre.

L'ADN se distingue de son sel comme suit : dans l'ADN sous forme acide, les phosphates (figure 1ter ci-dessus) possèdent tous un groupe OH et sont *neutres*

3. Linus Pauling (1901-1994) est un des rares savants à avoir reçu deux prix Nobel : celui de chimie en 1954, un an tout juste après la découverte de la double hélice, et celui de la paix en 1962 pour sa campagne contre les essais atomiques.

4. La diffraction des rayons X par un cristal ou par un pseudo-cristal (poudre de microcristaux ou encore fibres constituées de macromolécules identiques, parallèles entre elles mais de positions désordonnées) produit ce qu'on appelle un diagramme de rayons X. Dans le cas de l'ADN, ce sont des diagrammes de fibres, obtenus par Rosalind Franklin d'une part et Maurice Wilkins d'autre part, qui ont permis à Watson et Crick de découvrir la structure en double hélice.

(notons bien que les phosphates, une fois polymérisés dans la chaîne sucro-phosphate, ne possèdent plus qu'un seul groupe OH) ; dans un sel d'ADN, au contraire, ces OH ont perdu leur hydrogène (hydrogène dit acide⁵) qui est parti en solution saline sous forme H⁺, laissant sur le groupe phosphate une charge négative (sur le O qui a perdu son H et qui est devenu O⁻).

Conséquence pour la structure de Pauling et Corey : les groupes *phosphate* de l'ADN – possédant chacun une charge électrique négative dans le sel d'acide – se trouvent placés, dans la structure de Pauling et Corey, à l'intérieur, donc à proximité de l'axe de la triple hélice. Or ces charges électriques se repoussent fortement et on ne voit pas quelles forces pourraient les tenir en place dans cette structure en triple hélice :

Sans les atomes d'hydrogène acides, on ne voit pas bien quelles sont les forces qui tiennent la structure, surtout compte tenu du fait que les phosphates chargés, proches de l'axe, vont se repousser mutuellement

La structure erronée imaginée par Pauling et Corey ne pourrait donc exister qu'avec de l'ADN sous forme acide. Or ce n'est pas le cas dans les échantillons observés avec les rayons X.

On trouve d'ailleurs souvent mise en débat la question du caractère acide de l'ADN. Si les mots ont un sens, une solution d'acide désoxyribonucléique devrait en effet être acide ! Or ce n'est pas l'ADN qui est en solution mais le sel de l'ADN, c'est-à-dire que les hydrogènes associés aux groupes phosphates dans la forme acide sont remplacés dans le sel par des ions sodium. C'est donc la forme basique (avec les phosphates chargés négativement) qui est présente en solution dans les conditions usuelles d'extraction (c'est-à-dire à pH physiologique, proche de 7).

LA STRUCTURE EN DOUBLE HELICE

Revenons à la première phrase :

We wish to suggest a structure for the salt of DNA.

Elle sonne rétrospectivement comme la première phrase de la Genèse ; elle est en quelque sorte le « Bereshit » de la biologie moléculaire⁶. La formulation

5. Définition d'un hydrogène acide: dans une molécule possédant « une fonction acide », l'atome d'hydrogène dit *acide* est l'atome d'hydrogène qui peut être libéré dans l'eau sous forme d'ion H⁺. C'est donc le H d'une molécule notée AH, dont la réaction avec l'eau est schématisée par l'équation bilan : AH + H₂O = A⁻ + H₃O⁺.

6. « Bereshit », en hébreu « au commencement », est le premier mot de la Genèse.

quelques années plus tard du « dogme central de la biologie moléculaire⁷ » par Francis Crick s'inscrit d'ailleurs dans cette métaphore biblique.

Au quatrième paragraphe, les auteurs reprennent la phrase liminaire sous une forme plus agressive :

We wish to put forward a radically different structure for the salt of DNA.

Puis ils révèlent ce qui va immédiatement devenir l'icône de toute la biologie : la figure qui accompagne la publication, annoncée comme « purement schématique » est en fait extraordinairement proche de la structure atomique réelle qui ne sera obtenue que bien plus tard. Ce quatrième paragraphe est entièrement consacré à la description précise des caractéristiques géométriques de la structure. Les auteurs commencent par décrire les deux chaînes sucre-phosphate, enroulées de manière hélicoïdale autour d'un même axe et formant donc une double hélice.

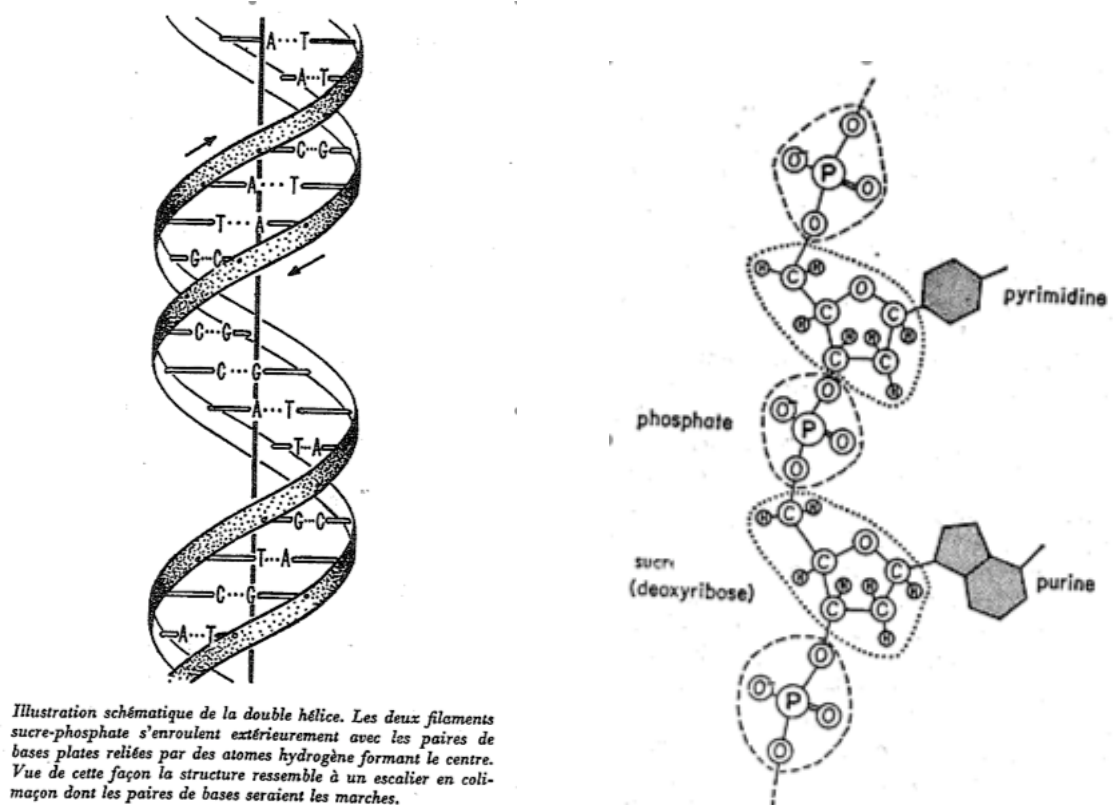


Figure 2 : (à gauche) illustration de la publication de Watson et Crick. La légende est celle de la publication traduite en français. La légende du texte anglais est tout aussi épurée, un peu différente : « La figure est purement schématique. Les deux rubans symbolisent les chaînes sucre-phosphate, et les barres horizontales sont les paires de bases unissant [litt : "faisant tenir ensemble"] les chaînes.

7. Le « dogme central de la biologie moléculaire » énoncé par Francis Crick en 1958 stipule que l'ADN est le support de l'information génétique et que cette information est transférée à la cellule après *transcription* de l'ADN en ARN messager puis *traduction* de l'ARN messager en protéine. Ce transfert est irréversible, l'information ne pouvant jamais être transférée des protéines vers l'ADN. Les progrès récents de l'épigénétique ont relativisé le dogme en montrant que des modifications biochimiques (dites épigénétiques) acquises par certaines protéines (en particulier les histones) peuvent se transmettre aux générations suivantes.

(à droite) détail d'une des chaînes sucre-phosphate (branche de l'hélice) (illustration tirée du récit La double hélice de James Watson). On voit l'enchaînement des sucres et des phosphates. Les molécules figurées en gris sur la droite de cette figure sont les bases azotées, soit pyrimidines (bases T ou C, cf. figure 1), soit purines (bases A ou G). Une pyrimidine est toujours liée à une purine, suivant deux seules configurations possibles : A-T et C-G. Rappelons que la répartition de ces paires est totalement aléatoire dans l'hélice : plusieurs A-T ou C-G de suite ou non – il n'y a pas d'ordre (cf. figure 2 à gauche)

Puis ils annoncent, au vu de la symétrie des clichés de cristallographie, que la structure possède un axe d'ordre 2, qui n'est pas l'axe commun aux deux hélices mais qui lui est perpendiculaire⁸. Ils en déduisent que l'ordre des atomes est inverse d'une chaîne à l'autre (flèches en sens inverse sur les deux hélices de la Figure 2 ; voir aussi la figure 1 quater montrant les orientations inverses 5' -->3' sur un brin et 3'-->5' sur le complémentaire).

Le cinquième paragraphe aborde le contenu en eau de la structure. Les auteurs prédisent que la structure partiellement déshydratée devrait être plus compacte grâce à l'inclinaison des bases par rapport à l'axe de la double hélice. Cette prédiction sera confirmée trois mois plus tard par Rosalind Franklin dans le numéro du 25 juillet de *Nature*. On distingue aujourd'hui pour l'ADN les formes B (celle de la publication de Watson et Crick), A (celle qui est déshydratée), et même Z (avec une double hélice enroulée à gauche) dont les illustrations comparées sont données sur la figure ci-dessous.

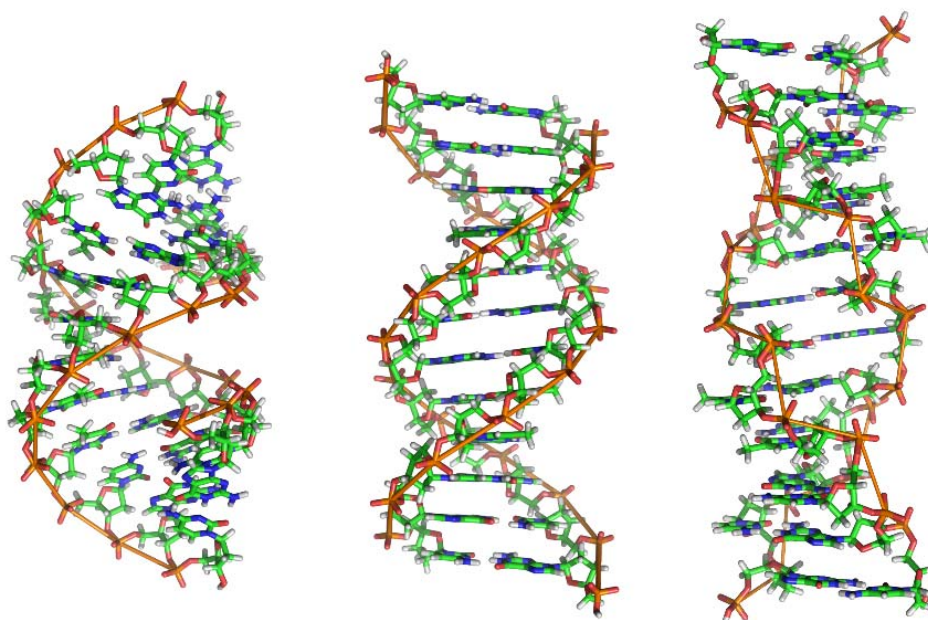


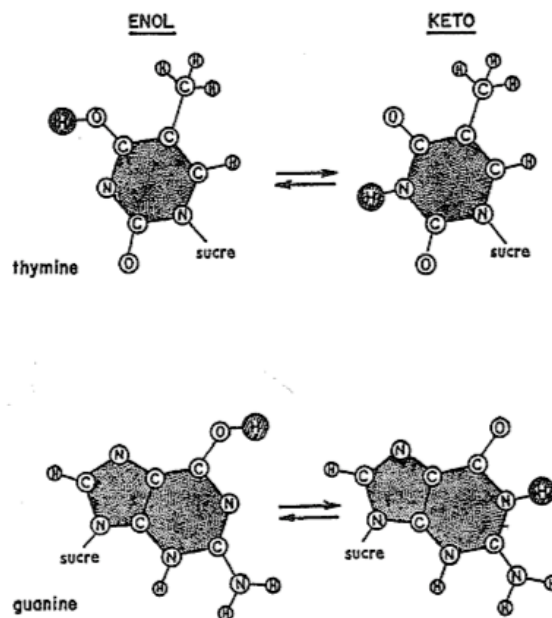
Figure 3 : de gauche à droite, les 3 structures de l'ADN en forme A, B et Z. La structure de Watson et Crick est la structure centrale (B).

8. Une molécule possède un axe d'ordre 2 si une rotation de 180° autour de cet axe laisse la molécule inchangée. L'axe d'ordre 2 de l'ADN passe entre les deux hélices et coupe leur axe commun perpendiculairement.

LES LIENS ENTRE LES BASES : UNE NOUVEAUTE RADICALE DE CRICK ET WATSON

Puis les auteurs abordent ce qui constitue la nouveauté radicale de leur structure. Celle-ci tient dans la façon dont les deux chaînes sont tenues ensemble par les bases. La clé de la structure est en effet venue un matin à Watson alors qu'il jouait avec des formes découpées dans du carton, représentant les 4 bases : adénine (A), guanine (G), cytosine (C), thymine (T).

Ayant appris d'un collègue cristallographe (Jerry Donohue) que les bases thymine et guanine devraient adopter la forme tautomérique « keto » plutôt que la forme « enol » favorisée à l'époque par les chimistes organiciens (figure 4), Watson avait immédiatement construit les formes correspondantes et essayé de les assembler pour former des paires de bases tenues ensemble par des liaisons hydrogènes.



Formes tautomériques de la guanine et de la thymine telles qu'elles peuvent se présenter dans l'ADN. Les atomes d'hydrogène qui peuvent subir des changements de position (une transformation réversible) sont hachurés.

Figure 4 : La légende (traduite en français) est tirée du récit *La double hélice*.

La liaison hydrogène

C'est un élément fondamental de la chimie qui intervient ici dans la structure de l'ADN. Une liaison hydrogène est environ vingt fois plus faible qu'une liaison covalente classique – mais est dix fois plus forte qu'une force de van der Waals traditionnelle.

Cette liaison se fait entre deux atomes. L'un – appelé *donneur* – possède un atome d'hydrogène qui lui est lié de façon covalente ; l'autre – appelé *receveur* – interagit avec cet hydrogène dans une configuration où les trois atomes (dans l'ordre : donneur, hydrogène et receveur) sont pratiquement alignés. Ce caractère directionnel rapproche la liaison hydrogène d'une liaison covalente ; pourtant l'interaction est de nature essentiellement électrostatique entre l'hydrogène (chargé positivement) et l'atome receveur (chargé négativement).

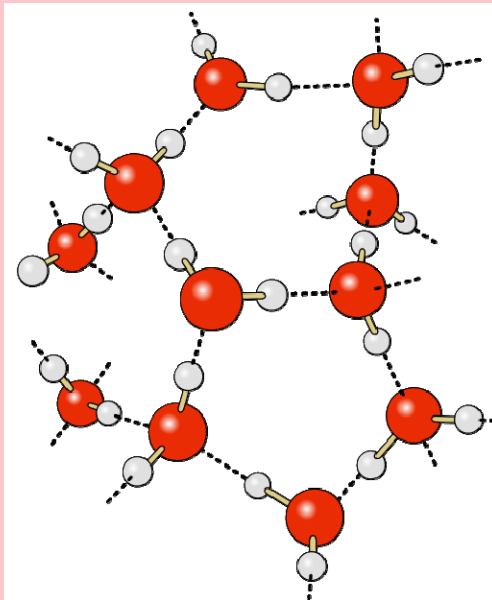


Figure 5 : Liaisons hydrogène entre molécules d'eau (notées en pointillés). En effet, une liaison covalente O — H est polarisée : les deux électrons mis en commun par O (en rouge) et H (en blanc) se situent entre O et H, ce qui déplace le centre de gravité des charges négatives, et rend l'extrémité de liaison (côté hydrogène) chargée positivement. Inversement, la région de l'atome d'oxygène est plutôt chargée négativement, et va donc attirer électriquement le H d'une autre molécule. Ce type de liaison peut exister entre le H d'une liaison covalente (avec C, N ou O par exemple) et un atome fortement électronégatif impliqué de son côté dans une autre liaison covalente (par exemple C, N, O ou F, à droite sur la seconde ligne du tableau périodique, qui figurent parmi les éléments les plus électronégatifs). La liaison hydrogène peut se faire entre molécules différentes. Dans le cas de l'eau, la liaison hydrogène (entre molécules identiques dans ce cas) explique de nombreuses propriétés physiques de l'eau (point d'ébullition, caractère liquide, etc.)

Les liaisons hydrogène qui assurent l'appariement des bases azotées A—T ou G—C de l'ADN sont deux types (le pointillé figure la liaison hydrogène, et les deux points un duplet électronique libre) (voir ces liaisons développées en figure 7 ci-dessous) :

- soit N—H..... :O
- soit N—H.....:N

@@@@@@



Figure 6 : Rosalind Franklin (1920-1958), biophysicienne britannique.

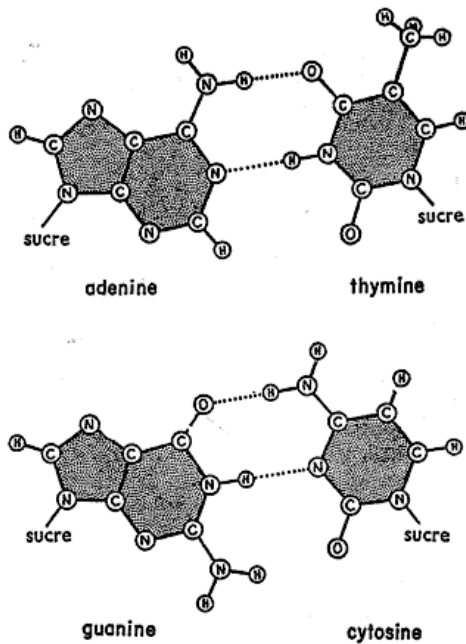
Watson raconte cet épisode crucial dans son récit autobiographique *La double hélice* :

Le matin suivant, quand je pénétrai dans notre bureau encore vide, j'écartai rapidement les papiers qui encombraient ma table pour avoir une grande surface plate sur laquelle combiner des paires de bases reliées par des ponts hydrogène. (...) Soudain, je compris qu'une paire adénine-thymine, maintenue par deux liaisons hydrogène, avait la même forme qu'une paire guanine-cytosine maintenue par deux liaisons hydrogène au moins. Tous les ponts hydrogène semblaient se former naturellement, nul besoin de tricher pour donner une forme identique à deux types de paires de bases.

L'appariement spécifique des bases – A avec T et G avec C (Figure 7) – explique immédiatement la mystérieuse correspondance stœchiométrique entre A et T d'une part et G et C de l'autre, connue jusqu'alors sous le nom de « loi de Chargaff », du nom du biochimiste qui l'avait découverte (la proportion de A et T, ou de C et G, est toujours la même⁹) :

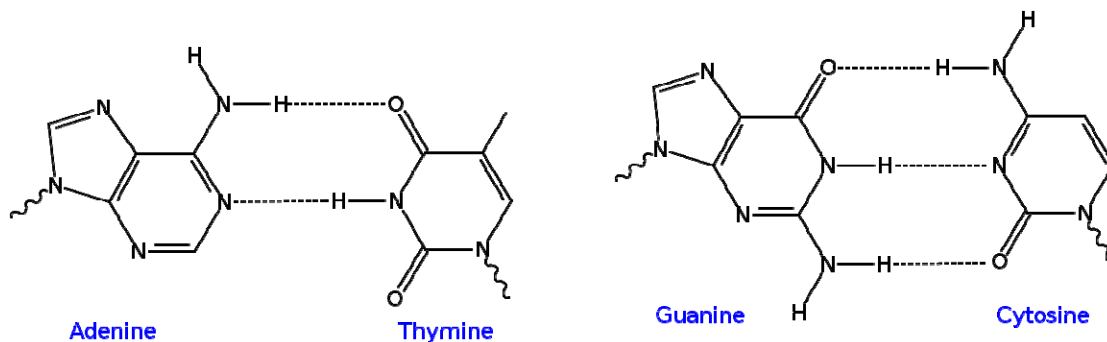
[...] if an adenine forms one member of a pair, on either chain, then on these assumptions the other one must be thymine; similarly for guanine and cytosine [...] It has been found experimentally that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribonucleic acid [texte BibNum]

9. Il y a autant de A que de T (et de C que de G) dans une chaîne ADN. En revanche, le rapport entre liaisons A-T d'une part et C-G d'autre part varie entre 25 et 75% selon les espèces.



Les paires de base adenine-thymine et guanine-cytosine utilisées pour la construction de la double hélice (les ponts hydrogène sont en pointillés). La formation d'un 3^e pont hydrogène fut envisagée puis rejetée, une étude cristallographique de la guanine ayant montré qu'il serait très faible. On sait maintenant que cette vue est fautive. Trois liaisons hydrogène peuvent être établies entre la guanine et la cytosine.

Figure 7 : La légende (traduite en français) est tirée du récit La double hélice. 7bis, ci-dessous, schémas modernes. Les liaisons hydrogène sont là aussi figurées en pointillés. Les liaisons A – T et C – G entre bases azotées constituent les « marches d'escalier » de « l'escalier en colimaçon » que constitue la double hélice d'ADN. On remarquera qu'une liaison entre deux purines (A et G) serait trop longue, et entre deux pyrimidines (C et T) trop courte : les liaisons A-T et C-G sont, elles, de longueur comparable. On considère de nos jours qu'il y a trois liaisons hydrogène entre C et G (Watson & Crick n'en proposaient que deux).



Les auteurs mentionnent que la structure proposée ne peut probablement pas être obtenue avec des riboses en lieu et place des désoxyriboses. En fait, on observera plus tard des double hélices d'ARN mais sous une forme proche de la forme A de l'ADN, c'est-à-dire la forme déshydratée, observée par Rosalind Franklin et publiée dans le numéro du 25 juillet 1953 de *Nature*.

D'un point de vue fonctionnel, l'appariement spécifique des bases permet à une cellule de coder dans son génome n'importe quelle séquence de bases sur un

brin, sans aucune contrainte structurale, la séquence sur l'autre brin s'adaptant alors parfaitement à la première.

@@@@@@

La confirmation expérimentale de la structure en double hélice est apportée par deux autres articles publiés dans le même numéro du 25 avril 1953 de *Nature* et placés immédiatement à la suite de l'article de Watson et Crick. Les clichés de cristallographie obtenus d'une part par Maurice H.F. Wilkins (1916-2004, prix Nobel en 1962 avec Watson et Crick) et collaborateurs, et d'autre part par Rosalind E. Franklin (1920-1958) et R.G. Gosling (né en 1926) – d'une qualité très supérieure – permettent de cerner certaines propriétés caractéristiques de la structure de la molécule d'ADN :

- l'empilement des bases comme dans une pile de pièces de monnaie (« pile of pennies » selon Wilkins).
- le pas de l'hélice : 3,4 nm
- la position des bases à l'intérieur de la double hélice

Mais la résolution limitée de ces clichés ne permet pas de remonter à la structure atomique. L'exploit inédit – et à ce jour non réédité – de Watson et Crick est d'avoir induit la vraie structure de la double hélice d'ADN par des arguments théoriques¹⁰. Ceux-ci sont essentiellement de deux types :

- géométriques, en s'aidant de maquettes¹¹ ;
- physiques : en plaçant les groupes chargés à la périphérie de la double hélice et non au centre comme l'avait proposé dans son modèle en triple hélice Linus Pauling – pourtant le plus grand chimiste de l'époque ; en exploitant aussi au maximum les propriétés de symétrie : c'est ainsi que Francis Crick comprend immédiatement, en voyant l'appariement des bases découvert par Watson, que les deux hélices d'ADN doivent être disposées en sens inverse l'une de l'autre (Figure 1).

Quant à la pertinence in vivo de la structure en double hélice proposée par Watson et Crick, elle est fournie par Maurice Wilkins dans l'article suivant celui de

10. Induire la structure atomique (« all-atom structure » en anglais) d'une molécule ou d'un complexe à partir de clichés de rayons X mal résolus (quand les cristaux de la molécule, ou du complexe moléculaire, ne sont pas disponibles) ou de microscopie électronique est un problème toujours d'actualité. L'exemple le plus emblématique est la structure toujours non résolue de la fibre de chromatine : si le polymorphisme structural de cette fibre est aujourd'hui bien établi, les différentes structures ne sont pas encore déterminées et de nombreux modèles concurrents continuent d'exister.

11. La méthode consistant à construire des maquettes était très en vogue à l'époque. Elle avait déjà permis à Linus Pauling de découvrir les motifs structuraux fondamentaux de la structure des protéines (l'hélice alpha et le feuillet bêta). Cette méthode est aujourd'hui largement remplacée par des outils d'infographie permettant de visualiser la structure des molécules en 3 dimensions.

Watson et Crick, dans lequel il mentionne des clichés semblables obtenus avec différents organismes. Maurice Wilkins rend d'ailleurs hommage à Watson et Crick pour avoir noté les premiers l'importance – le sens – biologique de la structure en double hélice. A la fin de leur article, Watson et Crick ont en effet écrit :

Il n'a pas échappé à notre attention que l'appariement spécifique des bases que nous avons proposé suggère immédiatement un mécanisme possible de réplication¹² pour le matériel génétique.

La Figure 8 illustre parfaitement cette propriété fonctionnelle fondamentale

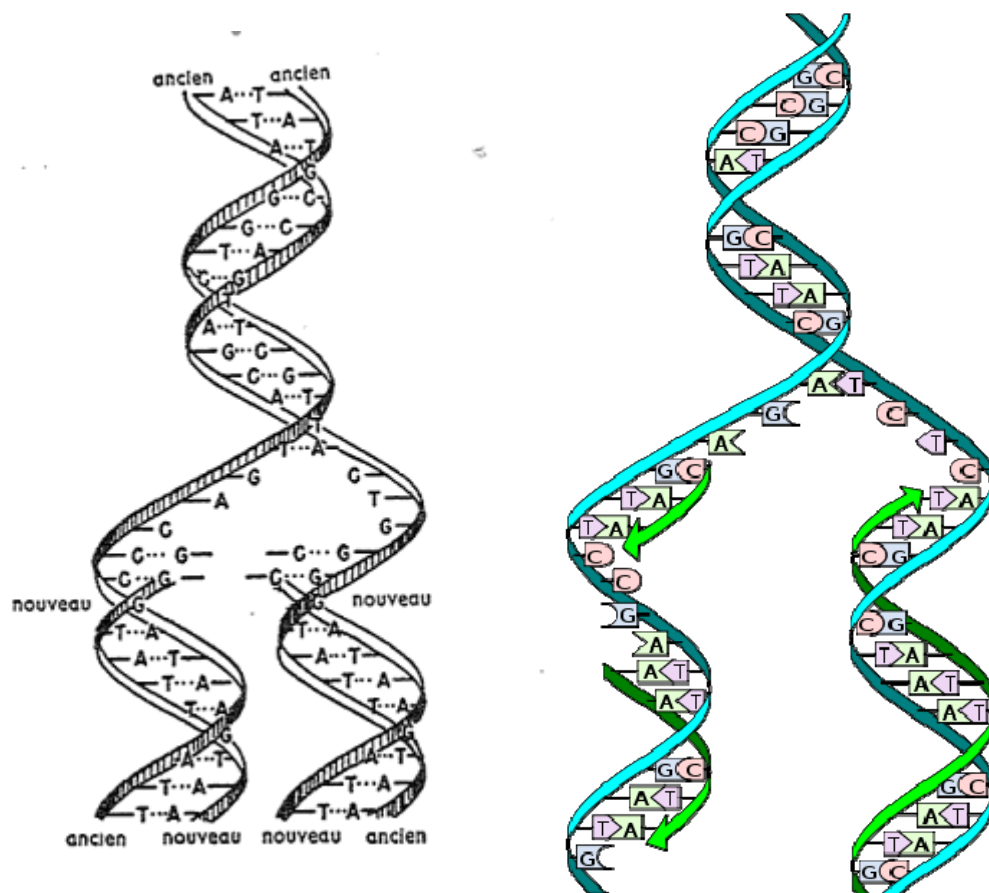


Figure 8 : Réplication de l'ADN (à g., schéma tiré du récit La double hélice ; à dr. image moderne WikiCommons, auteur Madprime). La molécule d'ADN s'ouvre comme une fermeture éclair par rupture des liaisons hydrogènes entre bases appariées A-T et C-G. Les bases de chaque brin parent (bleu) s'apparient alors avec des bases libres dispersées dans le noyau. Le brin parent se comporte comme un modèle pour la synthèse d'un brin complémentaire, dit brin fils. Chacun des double-brins fils a une séquence A, C, T, G (rappel : cette séquence est aléatoire dans un ADN donné, voir NBdP 9) identique à celle du double brin parental : ainsi le matériel génétique est reproduit avec exactitude.

(février 2012)

12. *Réplication* (copying) et non pas *transcription* comme on le trouve dans la traduction de l'article princeps de Watson et Crick dans le dossier scientifique accompagnant le récit de James Watson « La double hélice », dans l'édition de 1984 publiée chez Hachette.