

The Double Helix Structure of DNA

by Jean-Marc Victor
Research director at CNRS
Laboratory of Theoretical Condensed Matter Physics (CNRS UMR 7600)
Pierre-and-Marie-Curie University, Paris

If 1905 was the *annus mirabilis* of physics, with Albert Einstein's four historical papers¹, then 1953 was undoubtedly the *annus mirabilis* of biology. In the 25th April issue of the journal *Nature*, James D. Watson and Francis H.C. Crick reveal the double helix structure of DNA (Deoxyribonucleic acid). However, this "Letter", of an entire page, only contains the seeds of all molecular biology in the following fifty years, until the completion of the Human Genome Project in 2003².

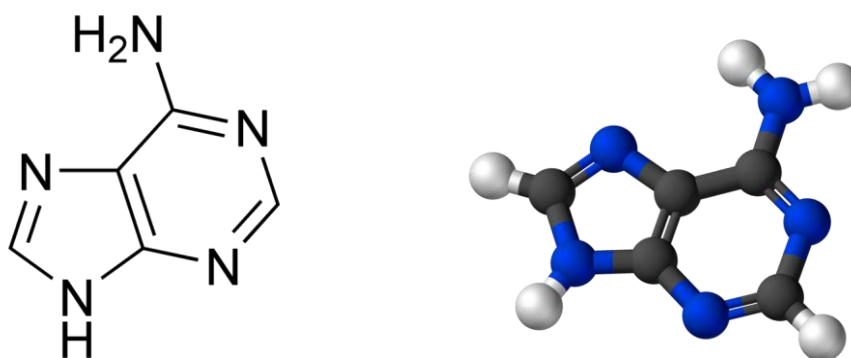


Figure 1: The different elements of a DNA molecule. In order to illustrate that chromosomes are physicochemical entities, and to better understand the following text about DNA structure, its various constituent molecular elements are demonstrated.

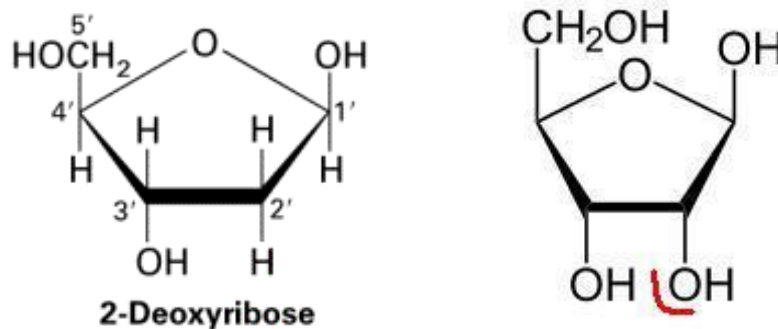
Above: Adenine, one of the four nitrogenous bases of DNA (bases A, C, G, T: adenine, cytosine, guanine, thymine). The developed formula for Adenine is $C_5H_5N_5$ (reminder of formula conventions, known as topological, like the one above on the left: we do not include C atoms – in grey on the right – and we do not include H atoms – in white on the right – when they are attached to C atoms; illustrated in blue on the right are N atoms).

1a, below (on the left): 2-deoxyribose, the DNA sugar with the developed formula $C_5H_{10}O_4$ (also not included in this formula are C atoms in position 1', 2', 3', 4'). Ribose is shown, above on the right, to better understand the name deoxyribose. The formula for Ribose is $C_5H_{10}O_5$; it belongs to the family of glucides (sugars) or carbohydrates with the formula $C_n(H_2O)_p$ (here $n = p = 5$). It is important to understand that the two formulas below are near enough the same, just by one atom: the oxygen atom below on the right

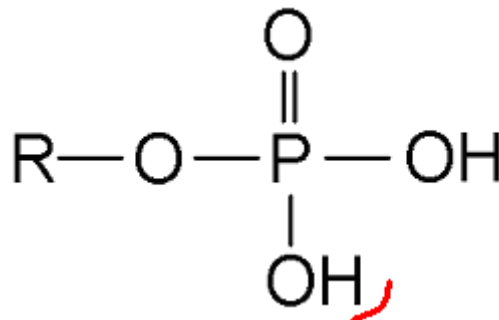
1. These four articles concern the special theory of relativity, the equivalence of mass and energy (with the famous equation $E=mc^2$), the theory of the photoelectric effect and Brownian motion.

2. The Human Genome Project's objective was the complete sequencing of the human genome. For the record, James D. Watson's genome is one of very first individual genomes to have been completely sequenced. Today, barely ten years later, everyone can pay for the complete sequence of their genome for a few thousand dollars!

(circled in red on the right diagram in position 2'): hence the name deoxyribose (ribose, from which an oxygen atom is removed). This component, the sugar deoxyribose, therefore gives its name to DNA (Deoxyribonucleic acid).



1b, below: the third constitutive element type of DNA, the phosphate group. It is attached to sugars (deoxyribose, figure 1a on the left) at two locations: (see figure 1c, the left strand) 1°) in position 5' on the "upper" sugar, the CH₂ plays the role of the radical R below (see figure 1c); 2°) in position 3' on the "above" sugar, via the O atom (which loses its H atom during polymerisation, circled in red).

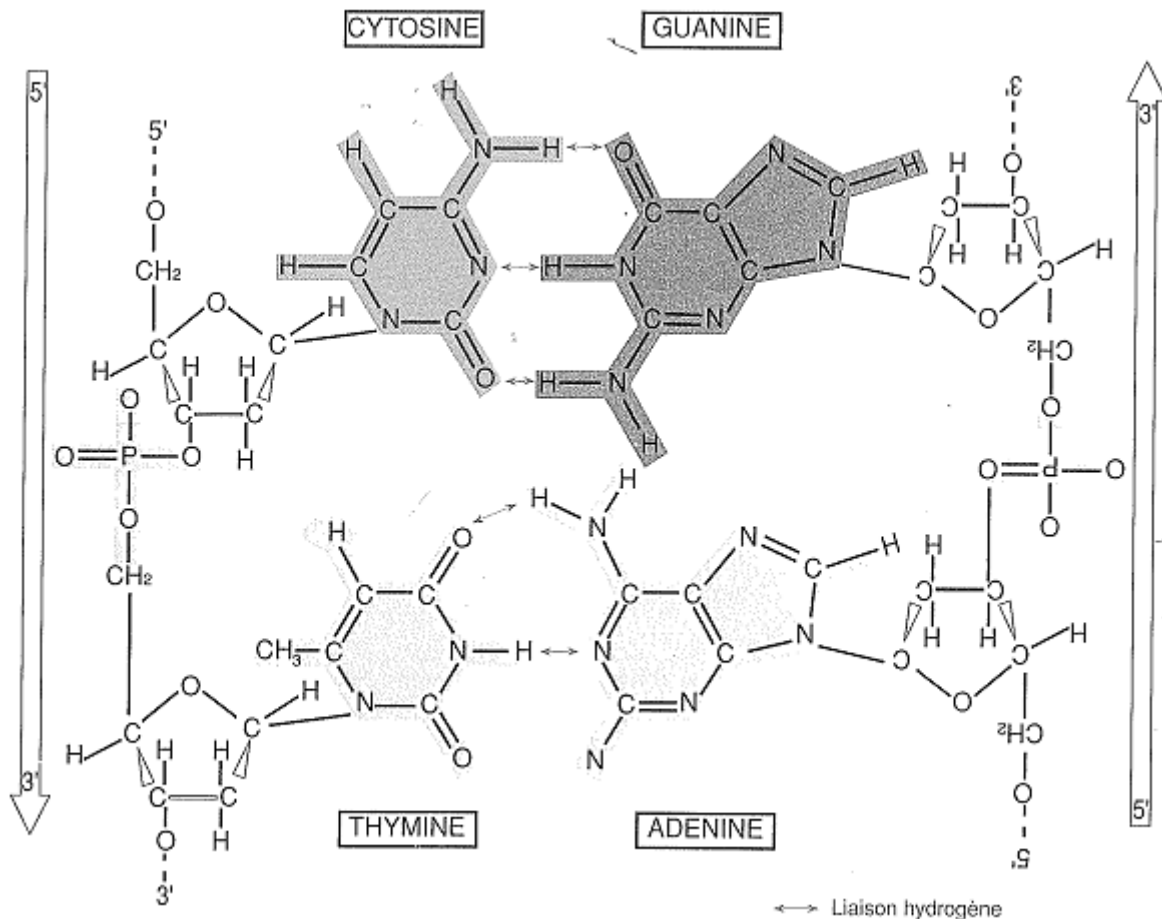


1c, below: actual complete DNA structure with base pairs (cytosine-guanine and thymine-adenine) between the two complementary strands (taken from the work *Biologie moléculaire et médecine, Jean-Claude Kaplan and Marc Delpech, Médecine-Sciences, Flammarion, second edition (1996)*).

Bases A, T, C or G are always put in position 1' of deoxyribose.

Each strand is obtained from the alternating polymerisation of sugars and phosphates; in doing so, there only remains one OH group in each phosphate. The sugars are all positioned in the same way in a strand, each phosphate thus being surrounded by a carbon 5' and carbon 3'.

Each strand is positioned conventionally from end point 5' towards end point 3' (top to bottom on the left, bottom to top on the right, the paired strands are therefore antiparallel, as shown by the figure's same construction). The hydrogen bonds are shown between A and T, and C and G (see upcoming boxed text).



PAULING & COREY'S OBJECTIONS TO THE STRUCTURE

Right from the first line, the authors announce their wish to suggest a structure for the *salt* of deoxyribonucleic acid. They also warn, in the following paragraph, that the triple helix structure of the acid itself, proposed by Pauling³ and Corey, is in their opinion irrelevant because the X-ray diagrams⁴, on which all previously proposed structures are based, were obtained with acid salt and not with the acid itself:

We believe that the material which gives the X-ray diagrams is the salt, not the free acid.

DNA differs from its salt as follows: in DNA, in acid form, the phosphates (figure 1b above) all have an OH group and are *neutral* (note that the phosphates, once polymerised in the sugar-phosphate chain, no longer have *only*

3. Linus Pauling (1901-1994) is one of the few scholars to have received two Nobel prizes: the Chemistry Prize in 1954, just one year after the discovery of the double helix, and the Peace Prize in 1962 for his campaign against atomic testing.

4. The diffraction of the X-rays by a crystal or a pseudo-crystal (powder of microcrystals or fibres consisting of identical macromolecules, mutually parallel but disorderly positioned) produces what is called an X-ray diagram. In the case of DNA, these are diagrams of fibres, firstly obtained by Rosalind Franklin and secondly by Maurice Wilkins, which allowed Watson & Crick to discover the double helix structure.

one OH group); on the contrary, in a DNA salt, these OH groups have lost their hydrogen (hydrogen known as acidic⁵) which, when left in H⁺ saline solution, leaves a negative charge in the phosphate group (on the O atom which loses its H atom and becomes O⁻).

The consequence for Pauling and Corey's structure is that DNA *phosphate* groups – each having a negative charge in the acid salt – are found placed *inside* Pauling and Corey's structure and therefore near the triple helix axis. Yet, these electric charges strongly repel each other; as a result, it is unclear which forces in this triple helix structure could hold them in place:

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

Consequently, the incorrect structure imagined by Pauling and Corey could only exist with DNA in acid form. Yet, this is not the case in the X-ray samples.

Additionally, the acidic nature of DNA is often under debate. If words have any meaning, a deoxyribonucleic acid solution should indeed be acidic! Yet, it is not DNA that is in the solution but DNA salt, i.e., hydrogen atoms associated with phosphate groups in acid form are replaced in the salt by sodium ions. So, it is the base form (with the negatively charged phosphates) that is present in the solution in usual conditions of extraction (i.e., at physiological PH, close to 7).

THE DOUBLE HELIX STRUCTURE

Back to the first sentence:

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

It sounds retrospectively like the first sentence of Genesis; it is somehow the "Bereshit" of molecular biology⁶. Years later, Francis Crick's wording of the "central dogma of molecular biology"⁷ also fits into this biblical metaphor.

In the fourth paragraph, the authors resume the opening phrase but in a more aggressive form:

5. Definition of a hydrogen acid: in a molecule having "an acidic function", the hydrogen atom is known as *acidic* when released into water as an H⁺ ion. Therefore, it is the H of an AH molecule whose reaction with water is illustrated by the chemical equation: $AH + H_2O = A^- + H_3O^+$.

6. "Bereshit", in Hebrew "in the beginning", is the first word of Genesis.

7. The "central dogma of molecular biology" stated by Francis Crick in 1958 stipulates that DNA is the carrier of genetic information and that this information is transferred to the cell after *transcription* of DNA into ARN messenger, then *translation* of ARN messenger into protein. This transfer is irreversible. The information can never be transferred from proteins into DNA. Recent advances in epigenetics have relativized the dogma by showing that biochemical modifications (known as epigenetic) acquired by certain proteins (particularly histones) can be transmitted to the following generations.

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

They then reveal what immediately became the icon of all biology: the figure accompanying the publication, announced as “purely schematic”, is in fact exceptionally close to the actual atomic structure that was not achieved until much later on. This fourth paragraph is entirely devoted to the precise description of the structure’s geometric characteristics. The authors begin by describing the two sugar-phosphate chains, helically wound around an identical axis and therefore forming a double helix.

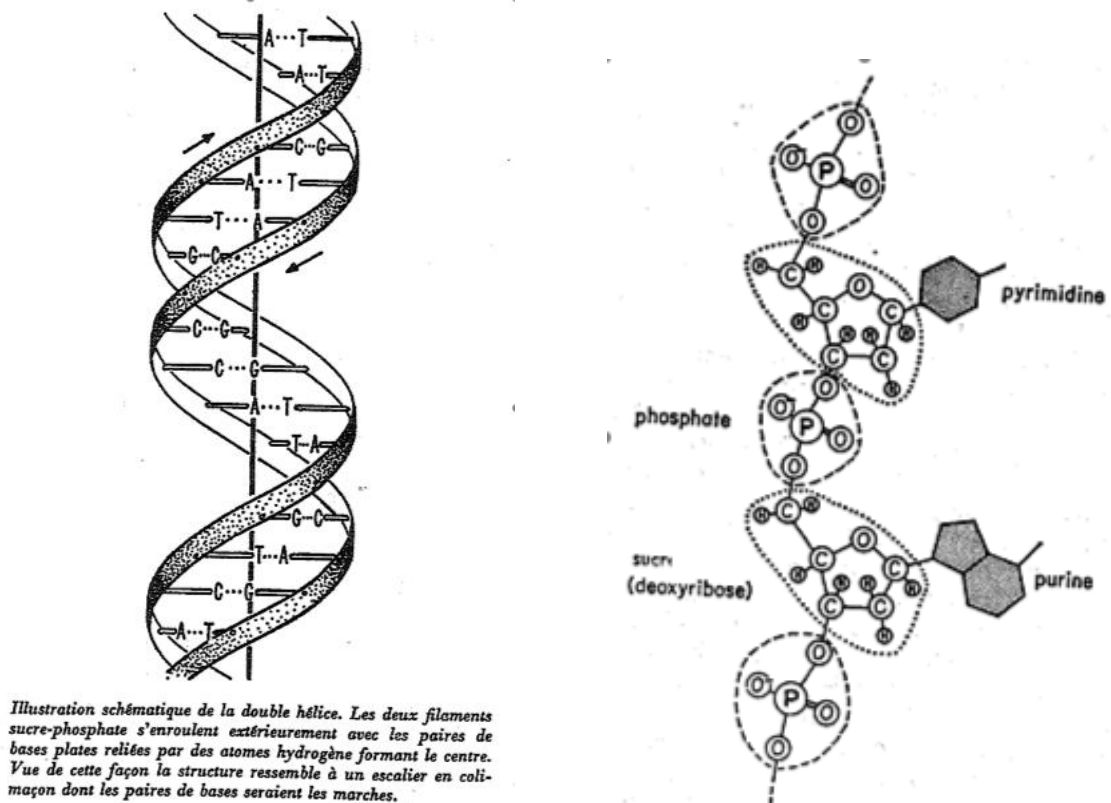


Figure 2: (on the left) illustration from Watson & Crick's publication. The caption is from the publication, translated into French. The caption from the English text is equally refined, but slightly different: "The two ribbons represent the phosphate-sugar chains and the horizontal rods represent the bonding between the pair of bases."

(on the right) detail of one of the sugar-phosphate chains (the helix branch) (illustration taken from James Watson's account The Double Helix). We see the bonds of sugars and phosphates. The molecules shown in grey on the right of this figure are nitrogenous bases, either pyrimidines (bases T or C, cf. figure 1), or purines (bases A or G). A pyrimidine is always linked to a purine, following only two possible configurations: A-T and C-G. Remember: the distribution of these pairs in the helix is completely random: several A-T or C-G, either straight away or not- there is no order (cf. figure 2 on the left)

Then they announce, given the symmetry of the crystallography images, that the structure has an axis of order 2, which is not the common axis of the

two helices but perpendicular to it⁸. From this, they deduce that the order of atoms is the opposite from one chain to another (arrows in opposite directions on the two helices of the figure 2; see also figure 1c showing the opposite directions 5' -->3' on one strand and 3'-->5' on the other).

The fifth paragraph addresses the structure's water content. The authors predict that the partially dehydrated structure should be more compact as a result of the inclination of the bases in relation to the double helix. Rosalind Franklin confirmed this prediction three months later in the 25th July issue of *Nature*. We distinguish today for DNA the forms B (from Watson & Crick's publication), A (the one which is dehydrated), and even Z (with a double helix wound *to the left*) whose compared illustrations are shown in the figure below.

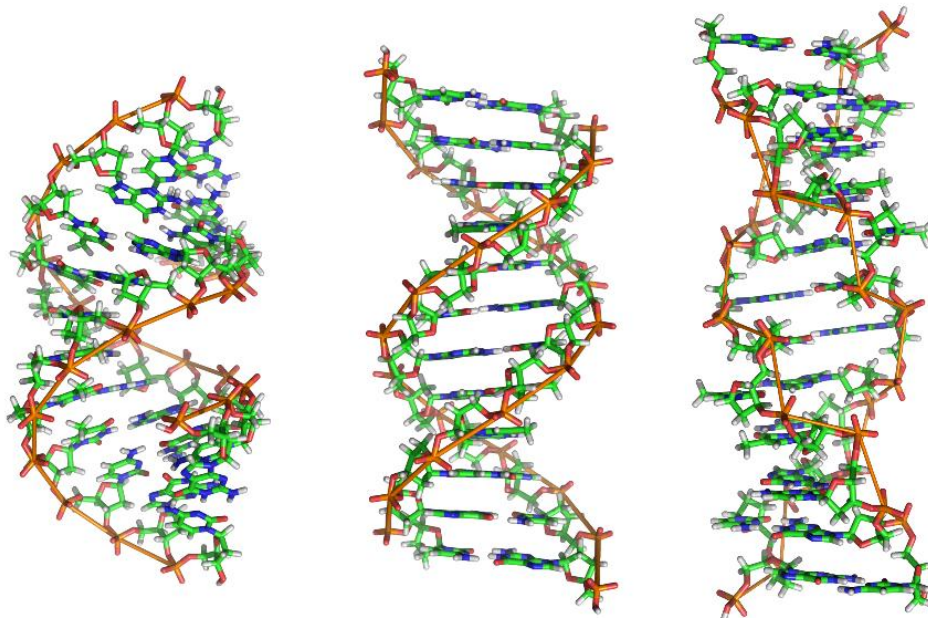


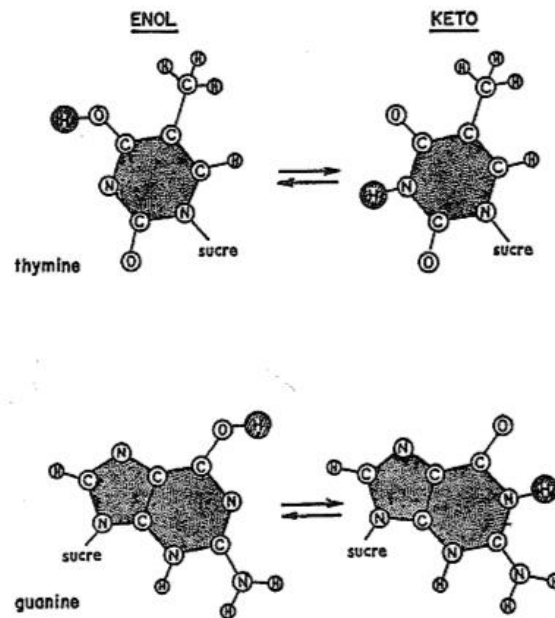
Figure 3: from left to right, the 3 structures of DNA in form A, B and Z. Watson & Crick's structure is the central structure (B).

THE BONDS BETWEEN BASES: WATSON & CRICK'S RADICAL NOVELTY

The authors then address what constitutes the radical novelty of their structure, which is in how two chains are held together by bases. In fact, the key to the structure came to Watson one morning while he was playing with cut-out cardboard shapes, representing the 4 bases: adenine (A), guanine (G), cytosine (C), thymine (T).

8. A molecule has an axis of order 2 if a rotation of 180° around this axis leaves the molecule unchanged. The axis of order 2 of DNA passes between the two helices and cuts their common axis perpendicularly.

Having learnt from a crystallographer colleague (Jerry Donohue) that thymine and guanine bases should adopt the tautomeric form "keto" rather than "enol", favoured at the time by organic chemists (figure 4), Watson immediately constructed the corresponding forms and tried to assemble them to form base pairs held together by hydrogen bonds.



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: From the account The Double Helix.

The Hydrogen Bond

It is a fundamental element of chemistry which occurs here in DNA structure. A hydrogen bond is around twenty times weaker than a classical covalent bond - but is ten times stronger than a traditional van der Waals force.

This connection is made between two atoms. One - known as *donor* - has a hydrogen atom to which it bonds covalently; the other - known as *recipient* - interacts with this hydrogen in a configuration where the three atoms (in order: donor, hydrogen and recipient) are practically aligned. This directionality brings closer the hydrogen bond to a covalent bond; yet, the interaction is essentially electrostatic in nature between the hydrogen (positive) and the receiver atom (negative).

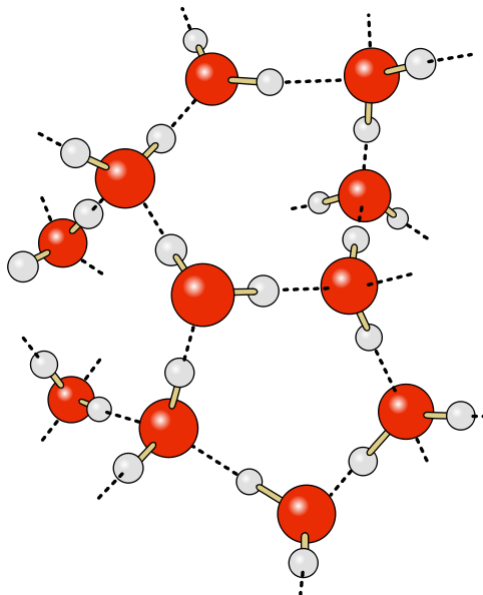


Figure 5: Hydrogen bonds between water molecules (shown in dotted lines). In fact, a covalent OH bond is polarised: the two electrons shared by O (in red) and H (in white) are between O and H, which moves the centre of gravity of negative charges, making the bond end (hydrogen side) positively charged. Conversely, the oxygen atom's area is negatively charged instead, and will thus electrically attract the H of another molecule. This type of bond may exist between the H of a covalent bond (with C, N or O, for example) and a strongly electronegative atom implicated for its part in another covalent bond (for example, C, N, O or F, on the right on the second row of the periodic table, which are among the most electronegative elements). The hydrogen bond may occur between different molecules. In the case of water, the hydrogen bond (in this case between identical molecules) explains many physical properties of water (boiling point, liquid nature, etc.)

There are two types of hydrogen bonds that provide DNA nitrogen base pairs A-T or G-C (the dotted line is the hydrogen bond, and the colon is a free electric duplet) (see these developed bonds in figure 7 below):

- Either N-H..... :O
- Or N-H.....:N

@@@@@@



Figure 6: Rosalind Franklin (1920-1958), British biophysicist.

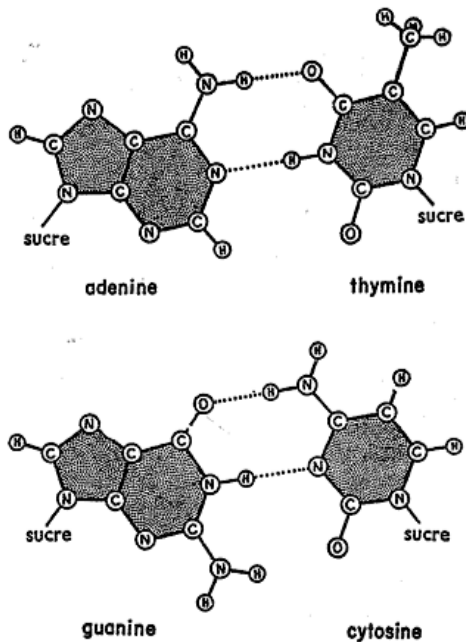
Watson recounts this crucial episode in his autobiographical account *The Double Helix*:

When I got to our still empty office the following morning, I quickly cleared away the papers from my desk top so that I would have a large, flat surface on which to form pairs of bases held together by hydrogen bonds. (...) Suddenly I became aware that an adenine-thymine pair held together by two hydrogen bonds was identical in shape to a guanine-cytosine pair held together by at least two hydrogen bonds. All the hydrogen bonds seemed to form naturally; no fudging was required to make the two types of base pairs identical in shape.

The specific base pairs – A with T and G with C (figure 7) – immediately explain the mysterious stoichiometric correspondence, firstly between A and T and secondly between G and C, known then as “Chargaff law”, named after the biochemist who had discovered it (the proportion of A and T, or C and G, is always the same⁹):

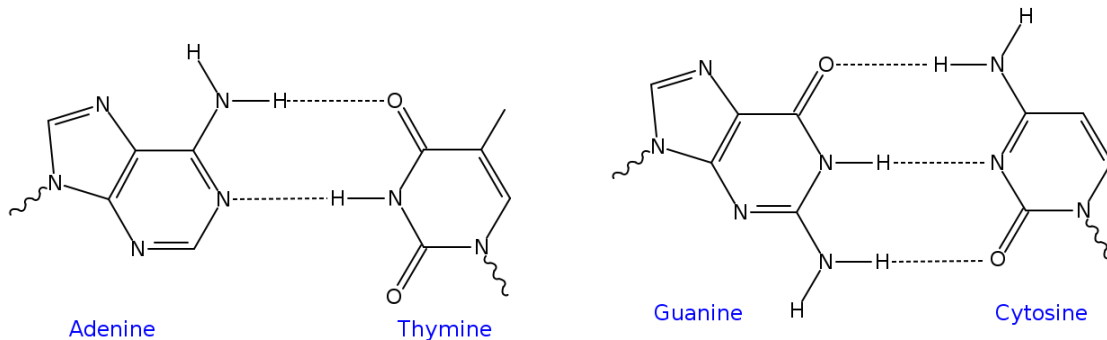
[...] if an adenine forms one member of a pair, on either chain, then on these assumptions the other one must be thymine; similarly for guanine and cytosine [...] It has been found experimentally that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very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

9. There are as many of A as T (and of C as G) in a DNA chain. On the other hand, the link, firstly between A-T bonds and secondly C-G bonds, varies between 25% and 75% according to the types.



Les paires de base adenine-thymine et guanine-cytosine utilisées pour la construction de la double hélice (les ponts hydrogène sont en pointillé). 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: The caption is taken from the account *The Double Helix*. **7a, below, modern diagrams.** The hydrogen bonds are also shown by dotted lines. The A-T and C-G bonds between nitrogenous bases thus constitute the "steps" of "the spiral staircase" that is the DNA double helix. Note that a bond between two purines (A and G) would be too long, while between two pyrimidines (C and T) would be too short: A-T and C-G bonds are similar in length. Nowadays, we consider that there are three hydrogen bonds between C and G (Watson & Crick only proposed two of them).



The authors mention that the proposed structure likely cannot be obtained with ribose in the place of deoxyribose. In fact, RNA double helices will be looked at but in a form very like form A of DNA, i.e., the dehydrated form, observed by Rosalind Franklin and published in the 25th July 1953 issue of *Nature*.

From a functional point of view, the specific base pairs allow a cell to encode in its genome any sequence of bases on a strand; without any structural constraint, the sequence on the other strand then adapts perfectly to the first.

@@@@@@

The experimental confirmation of the double helix structure is provided by two other published articles in the same 25th April 1953 issue of *Nature*, placed immediately after Watson & Crick's article. The crystallography images, firstly obtained by Maurice H.F. Wilkins (1916-2004, Nobel Prize in 1962 with Watson & Crick) and collaborators, and secondly by Rosalind E. Franklin (1920-1958) and R.G. Gosling (born in 1926) – of a very high quality – permit the identification of certain characteristic properties of the DNA molecule structure:

- The piling up of the bases like in a pile of coins ("pile of pennies" according to Wilkins).
- The pace of the helix: 3.4 nm
- The position of the bases inside the double helix

But the limited resolution of these images does not permit the recreation of the all-atom structure. Watson & Crick's unpublished exploit – and to this day still unpublished – has led to the true structure of the DNA double helix via theoretical arguments¹⁰. Essentially, there are two types:

- Geometric, with the help of models¹¹;
- Physical: by placing the charged groups on the periphery of the double helix and not in the centre as Linus Pauling had proposed in his triple helix model – despite being the greatest chemist of that time; additionally, by fully exploiting the properties of symmetry: therefore Francis Crick immediately understands, by seeing the base pairs discovered by Watson, that the two helices of DNA must be placed in opposite directions to each other (figure 1).

Regarding the in vivo pertinence of the double helix structure proposed by Watson & Crick, it is given by Maurice Wilkins in the article following Watson & Crick's, in which he gives images which are similar but obtained with different organisms. Maurice Wilkins also pays homage to Watson & Crick for having noted the first biological importance – the very meaning - of the double helix structure. At the end of their article, Watson & Crick indeed discuss it:

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism¹² for the genetic material.

10. To induce the all-atom structure of a molecule or complex from the poorly resolved X-ray images (when the molecule crystals, or of the molecular complex, are unavailable) or electron microscopy is an on-going problem. The most emblematic example is chromatin fibre's still unresolved structure: if this fibre's structural polymorphism is nowadays well established, the different structures are yet to be determined and many competing models continue to exist.

11. The method consisting of building models was very popular at the time. It had already allowed Linus Pauling to discover the fundamental structural units of the protein structure (alpha helix and beta sheet). Today, this method is largely replaced by computer graphics tools to visualise molecule structures in 3 dimensions.

Figure 8 perfectly illustrates this fundamental functional property.

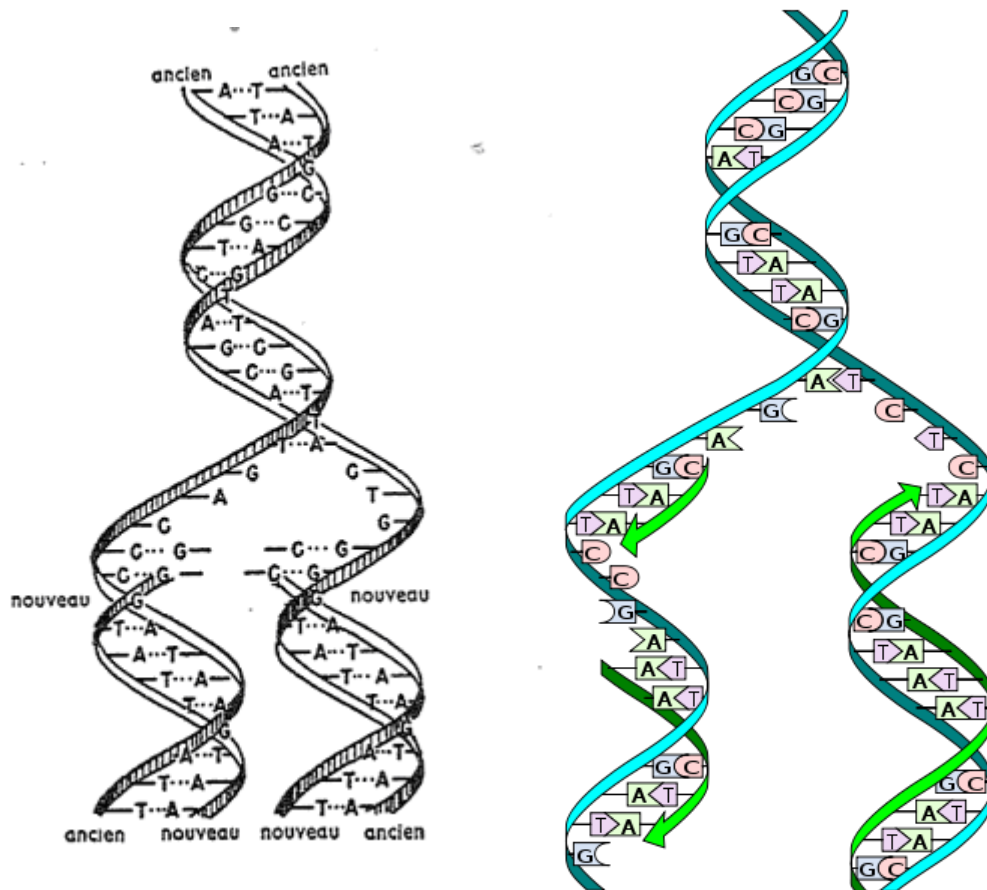


Figure 8: DNA replication (on the left, diagram taken from the account *The Double Helix*; on the right, modern image WikiCommons, author Madprime). The DNA molecule is opened like a zip fastener by breaking the hydrogen bonds between paired bases A-T and C-G. The bases of each parent strand (blue) are then paired with available bases dispersed in the nucleus. The parent strand acts as a model for the synthesis of a complementary strand. Each new double strand has a sequence A, C, T, G (remember: this sequence in a given DNA is random, see NBdP 9), identical to a double-stranded parental sequence: and so the genetic material is reproduced with accuracy.



(February 2012)

(translated in English by John Moran, published July 2013)

12. *Replication* (copying) and not *transcription* as we find it in the translation of Watson & Crick's first article in the scientific dossier accompanying James Watson's account "The Double Helix", in the 1984 edition published by Hachette.