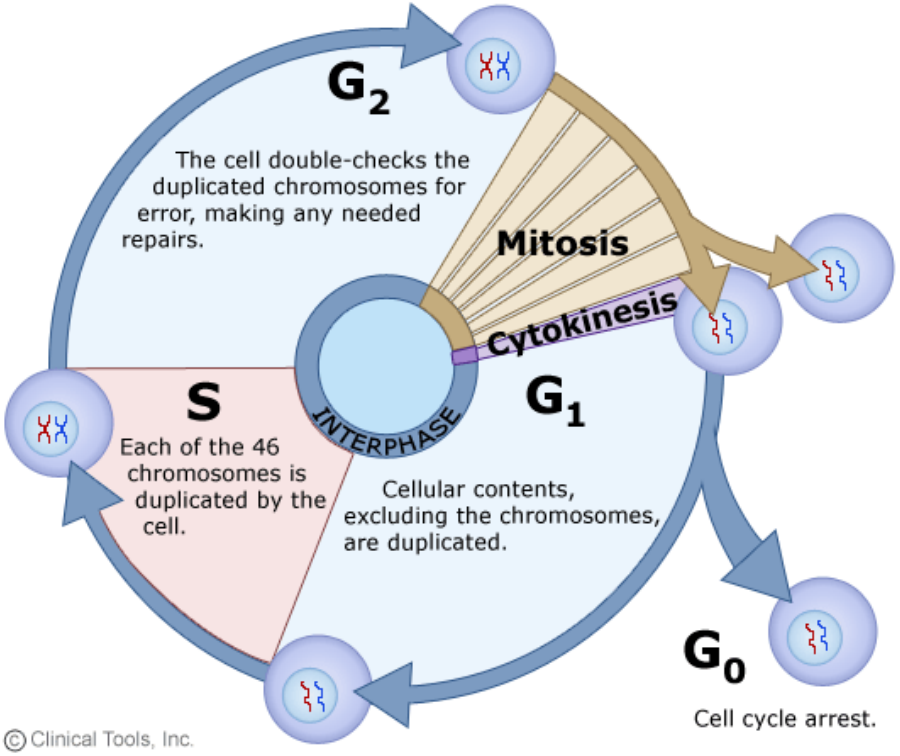
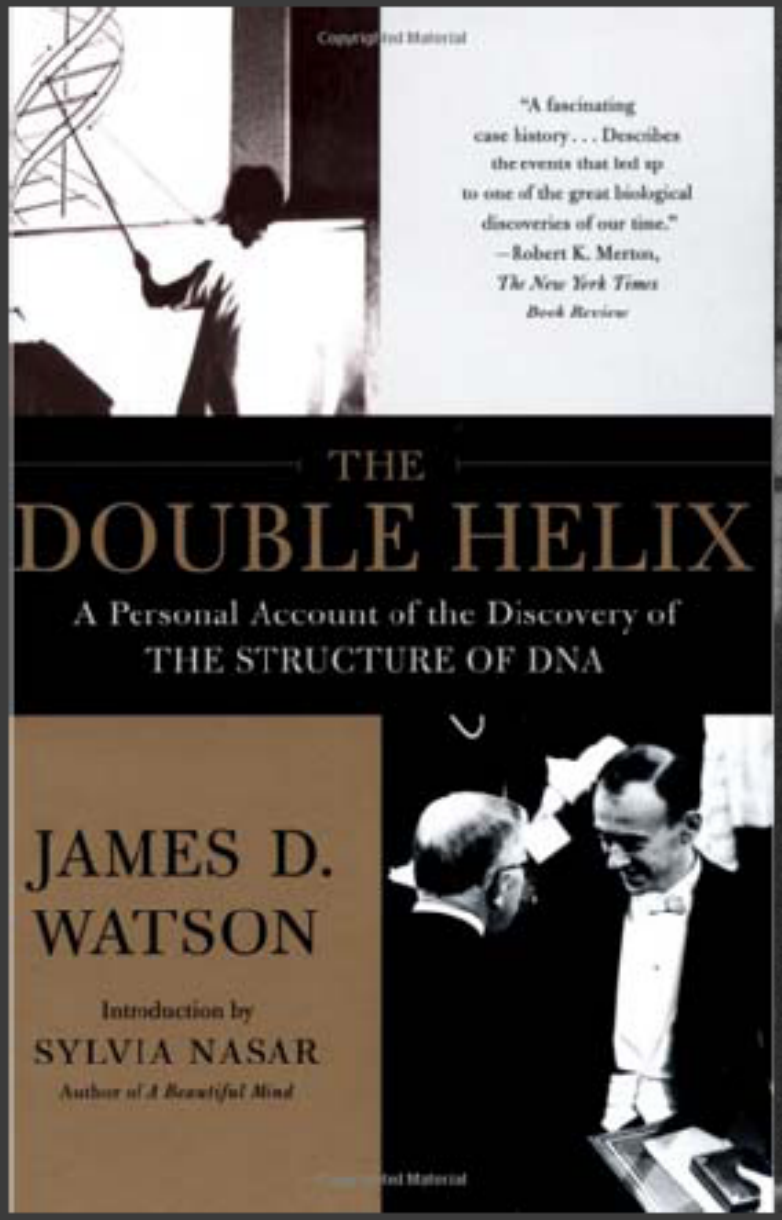
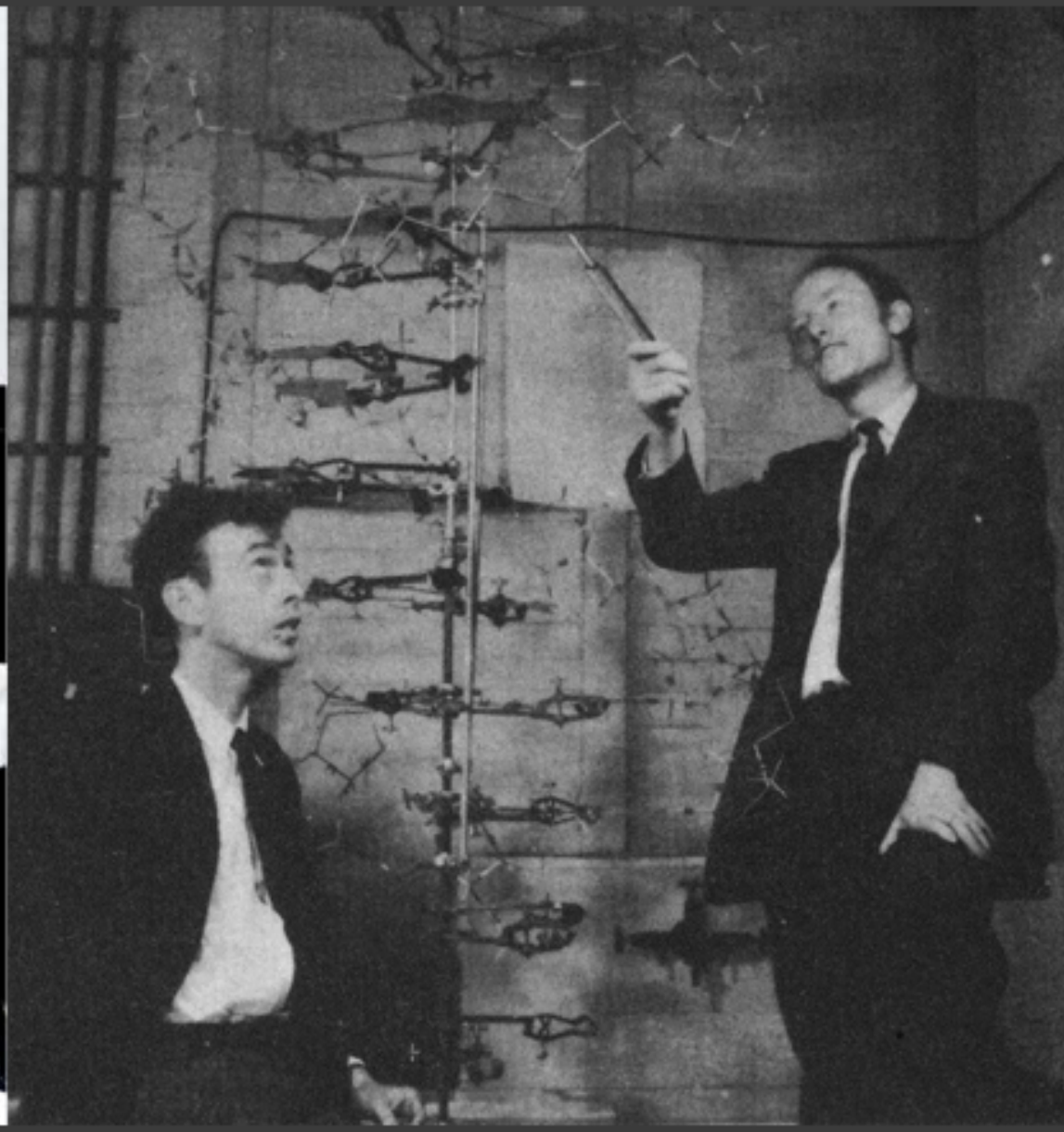
DNA REPLICATION

- DNA Replication is the process in which the DNA within a cell makes an exact copy of itself.




Copyrighted Material
"A fascinating case history . . . Describes the events that led up to one of the great biological discoveries of our time."
—Robert K. Merton, The New York Times Book Review
THE
DOUBLE HELIX
A Personal Account of the Discovery of THE STRUCTURE OF DNA
JAMES D. WATSON
Introduction by
SYLVIA NASAR
Author of A Beautiful Mind
Copyrighted Material



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.

[1] Young, F. B., Gerard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1950).

[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).

[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **14** (1) (1956).

[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (Stockholm), **2** (11) (1903).

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 Corey[1]. 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 diagrams 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.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

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-deoxyribofuranose 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 sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] 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 Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. 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 A. The distance of a phosphorus atom from the fibre axis is 10 A. 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 the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

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 King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. Watson
F. H. C. Crick

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge.

April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).

[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).

[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).

[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).

[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).

[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

While the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury[1]) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the polynucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all species (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline[1-3], semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleotides along the chain, and the other by the longer spacings of the chain configuration. The sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline deoxypentose nucleic acid ('structure B' in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3·4-A. reflexion corresponded to the internucleotide repeat along the fibre axis. The ~ 34 A. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[4] (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the nth layer line being proportional to the square of J_n, the nth order Bessel function. A straight line may be drawn approximately through

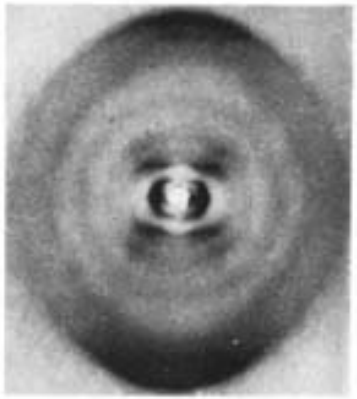


Fig. 1. Fibre diagram of deoxypentose nucleic acid from B, cell. Fibre axis vertical

the innermost maxima of each Bessel function at the origin. The angle this line makes with the equat is roughly equal to the angle between an element the helix and the helix axis. If a unit repeats n tim along the helix there will be a meridional reflexi (J_n^2) on the nth layer line. The helical configurati produces side-bands on this fundamental frequenc the effect[4] being to reproduce the intensity distributi about the origin around the new origin, on the n layer line, corresponding to C in Fig. 2.

We will now briefly analyse in physical terms som of the effects of the shape and size of the repeat un or nucleotide on the diffraction pattern. First, if t nucleotide consists of a unit having circular symmet about an axis parallel to the helix axis, the whe diffraction pattern is modified by the form factor the nucleotide. Second, if the nucleotide consists a series of points on a radius at right-angles to t helix axis, the phases of radiation scattered by t helices of different diameter passing through on point are the same. Summation of the correspondi Bessel functions gives reinforcement for the inne

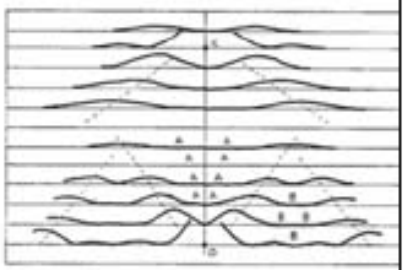


Fig. 2. Diffraction pattern of system of helices corresponding to structure of deoxypentose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third and fifth layer lines for half of the nucleotide mass at 20 A. diameter and smaller distributed along a radius, the mass at a given radius being proportional to the radius. About C on the tenth layer the similar functions are plotted for an outer diameter of 12 A.

Nature, Vol. 171, p.737, April 25, 1953

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

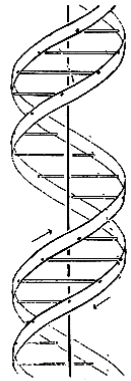
A Structure for Deoxyribose Nucleic Acid

We wish to s uggest 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 Corey (1). 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 diagrams 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 F raser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hyd rogen 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 ra dically 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 ß-D-deoxyribofuranose 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 sequences of the atoms in t he two chains run in opposite directions. Each chain loosely resembles Furberg's2 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 Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each every 3.4 A. in the z-direction. We have assumed an angle of 36° between adjacent



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. A s 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. The are joined together in pairs, 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 t o 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 th ese assumptions the other member must be thymine ; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on t he other chain is automatically determined.

It has been found experimentally (3,4) that the ratio of the amounts of adenine to thymine, and the ration of guanine to cytosine, are always bery close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact. The previously published X-ray data (5,6) on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

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.

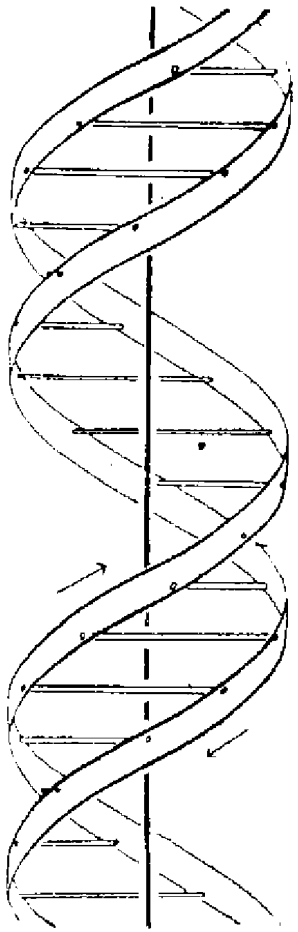
We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E . Franklin and their co-workers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK

Medical Research Council Unit for the Study of Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge.

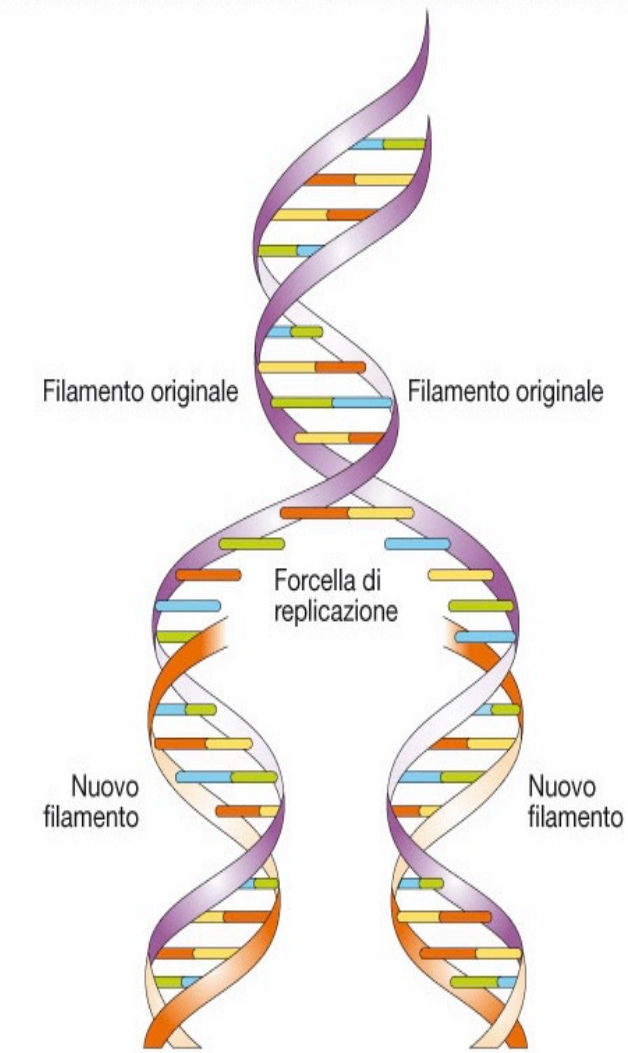
1. Pauling, L., and Corey, R. B., Nature, 171, 346 (1953); Proc. U.S. Nat. Acad. Sci., 39, 84 (1953).
2. Furberg, S., Acta Chem. Scand., 6, 634 (1952).
3. Ch argaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., Biochim. et Biophys. Acta, 9, 402 (1952).
4. Wyatt, G. R., J. Gen. Physiol., 36, 201 (1952).
5. A stbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
6. W ilkins, M. H . F., and Randall, J. T., Biochim. et Biophys. Acta, 10, 192 (1953).

Il modello a doppia elica di Crick e Watson suggerisce il meccanismo di replicazione, **semi-conservativa**



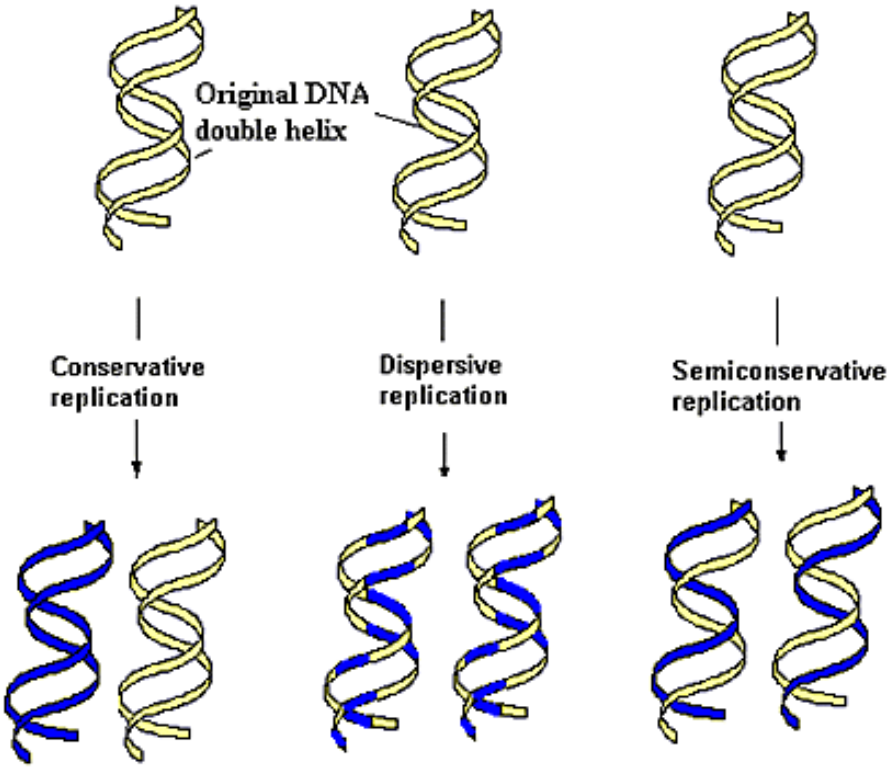
This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

I due filamenti della molecola di DNA parentale (viola) si separano e servono da stampo per la neosintesi di due filamenti complementari (arancio). La forcella di replicazione cosi' generatasi puo' propagarsi nel DNA non ancora replicato dando origine a due molecole figlie uguali alla molecola originaria e costituite da un filamento parentale e uno di neosintesi



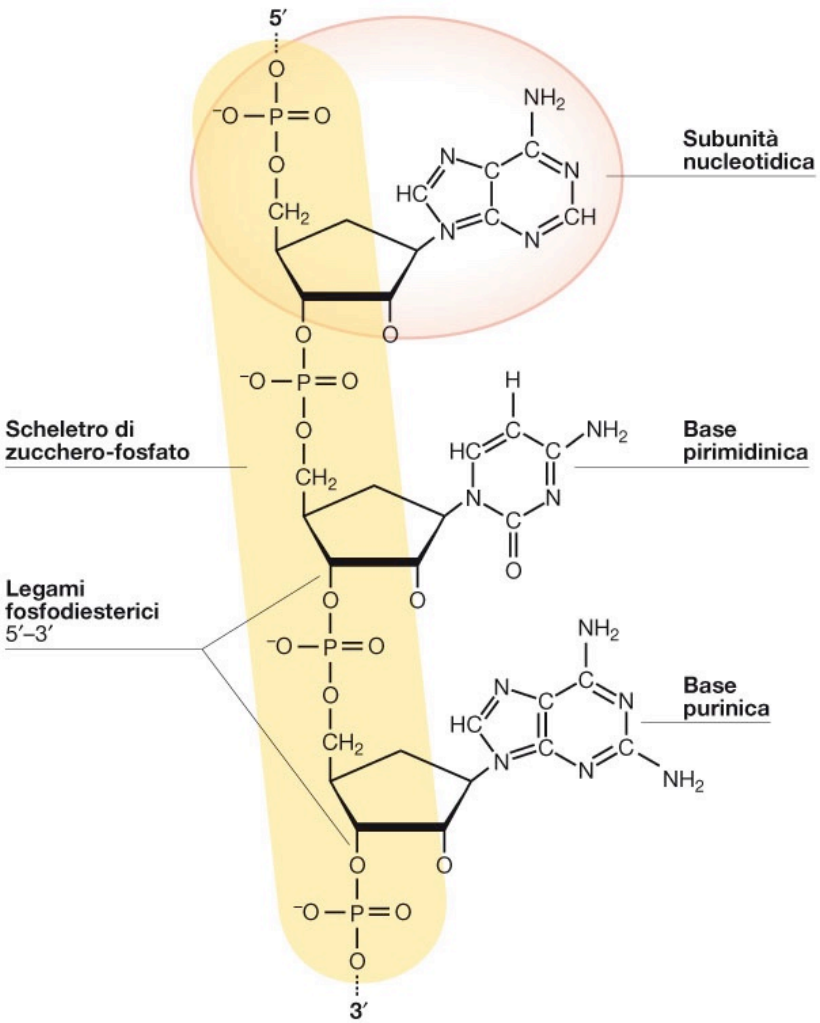
The Three Possible DNA Replication Models

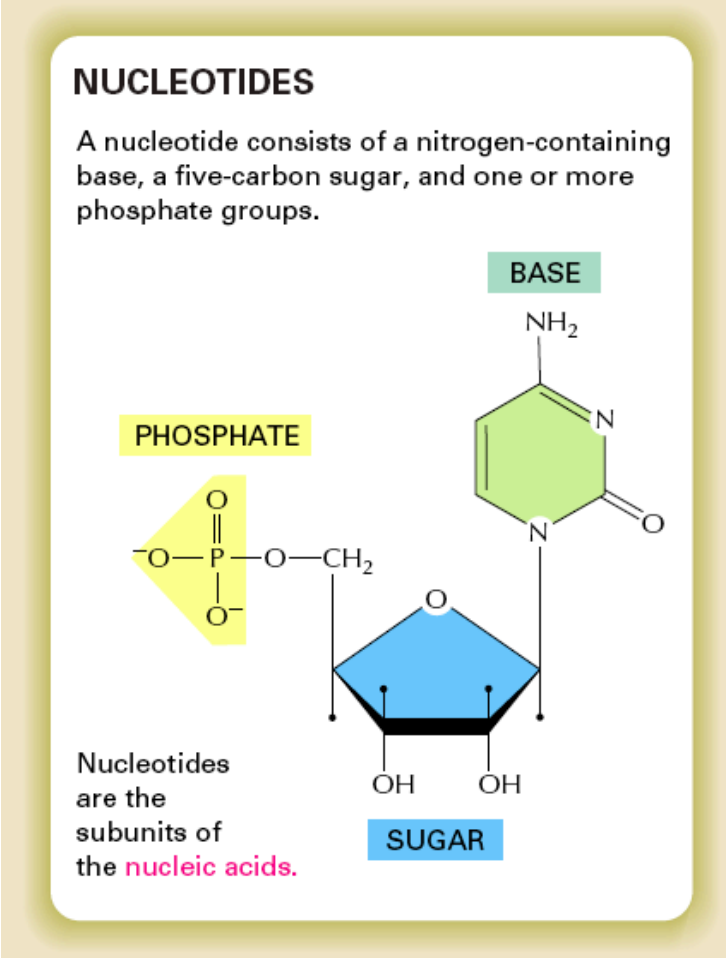
- Conservative- would leave the original strand intact and copy it.
- Dispersive-would produce two DNA molecules with sections of both old and new along each strand.
- Semiconservative – would produce DNA molecules with both one old strand and one new strand.



Possible Models of DNA Replication

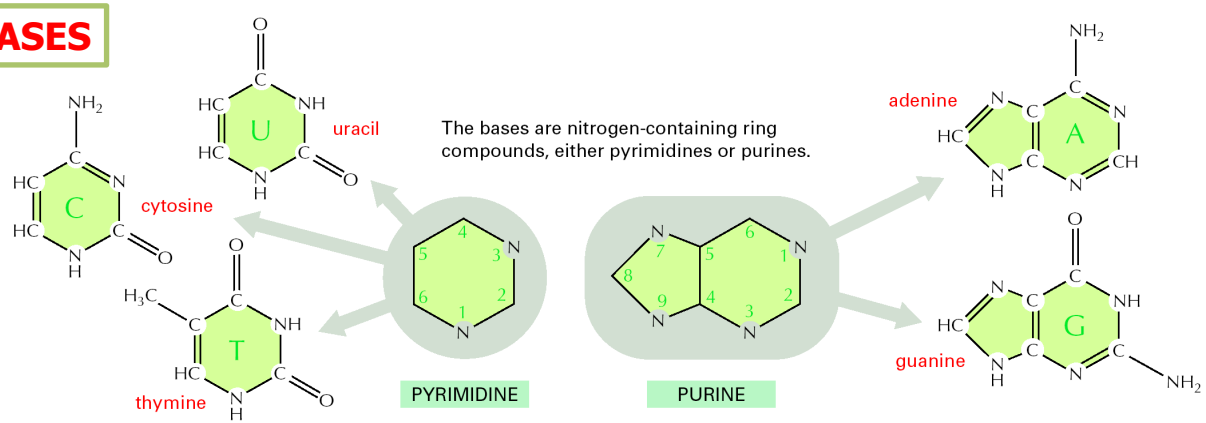
The DNA building blocks are the NUCLEOTIDES





Nucleotides structure

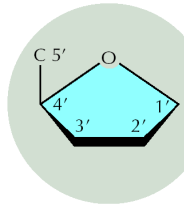
BASES



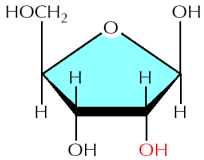
The bases are nitrogen-containing ring compounds, either pyrimidines or purines.

SUGARS

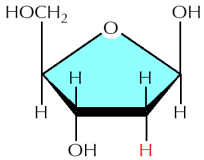
PENTOSE
a five-carbon sugar



two kinds are used



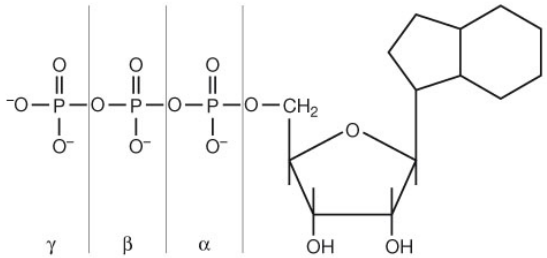
β-D-ribose
used in ribonucleic acid

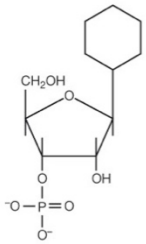


β-D-2-deoxyribose
used in deoxyribonucleic acid

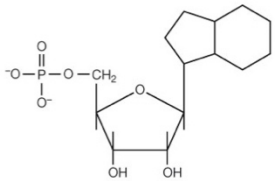
Each numbered carbon on the sugar of a nucleotide is followed by a prime mark; therefore, one speaks of the "5-prime carbon," etc.

PHOSPHATES





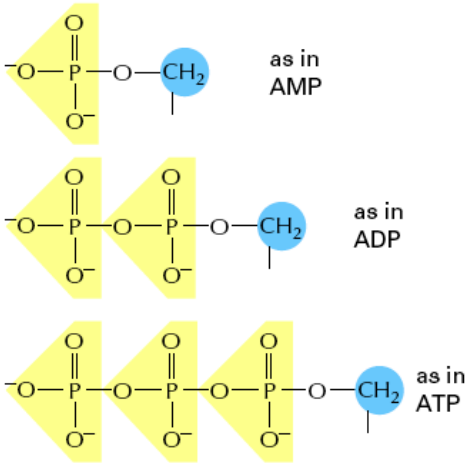
3′-monofosfato
(Pirimidina)



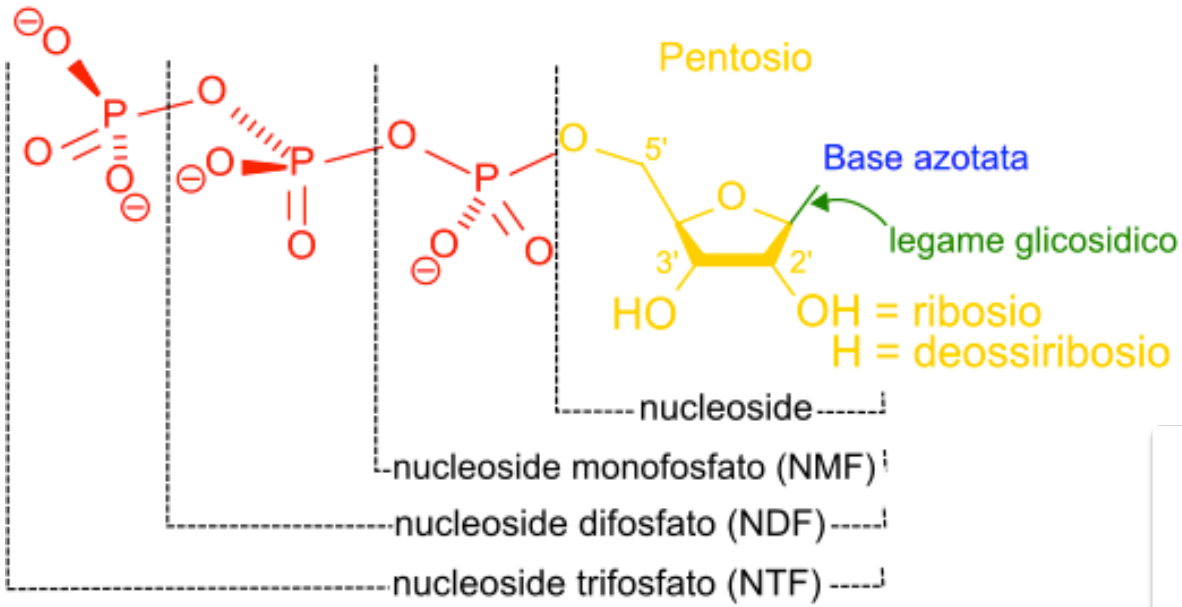
5′-monofosfato
(Purina)

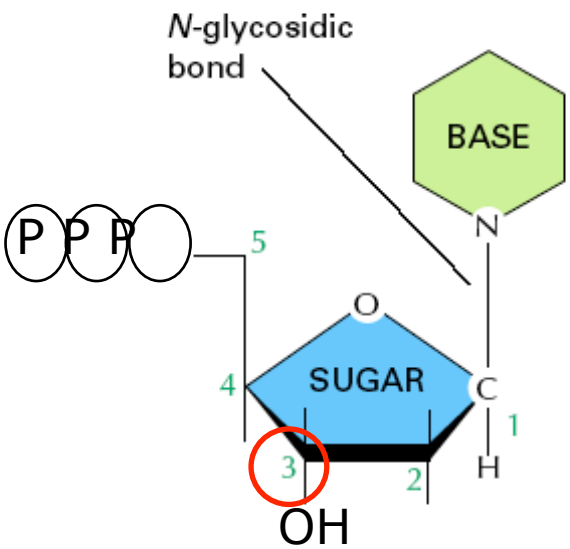
PHOSPHATES

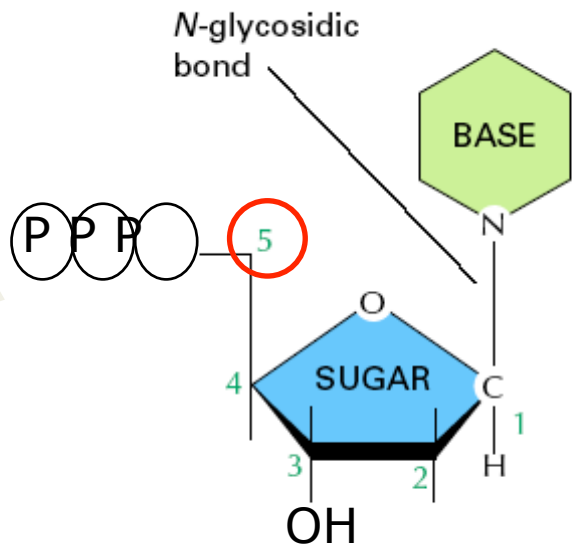
The phosphates are normally joined to the C5 hydroxyl of the ribose or deoxyribose sugar (designated 5'). Mono-, di-, and triphosphates are common.

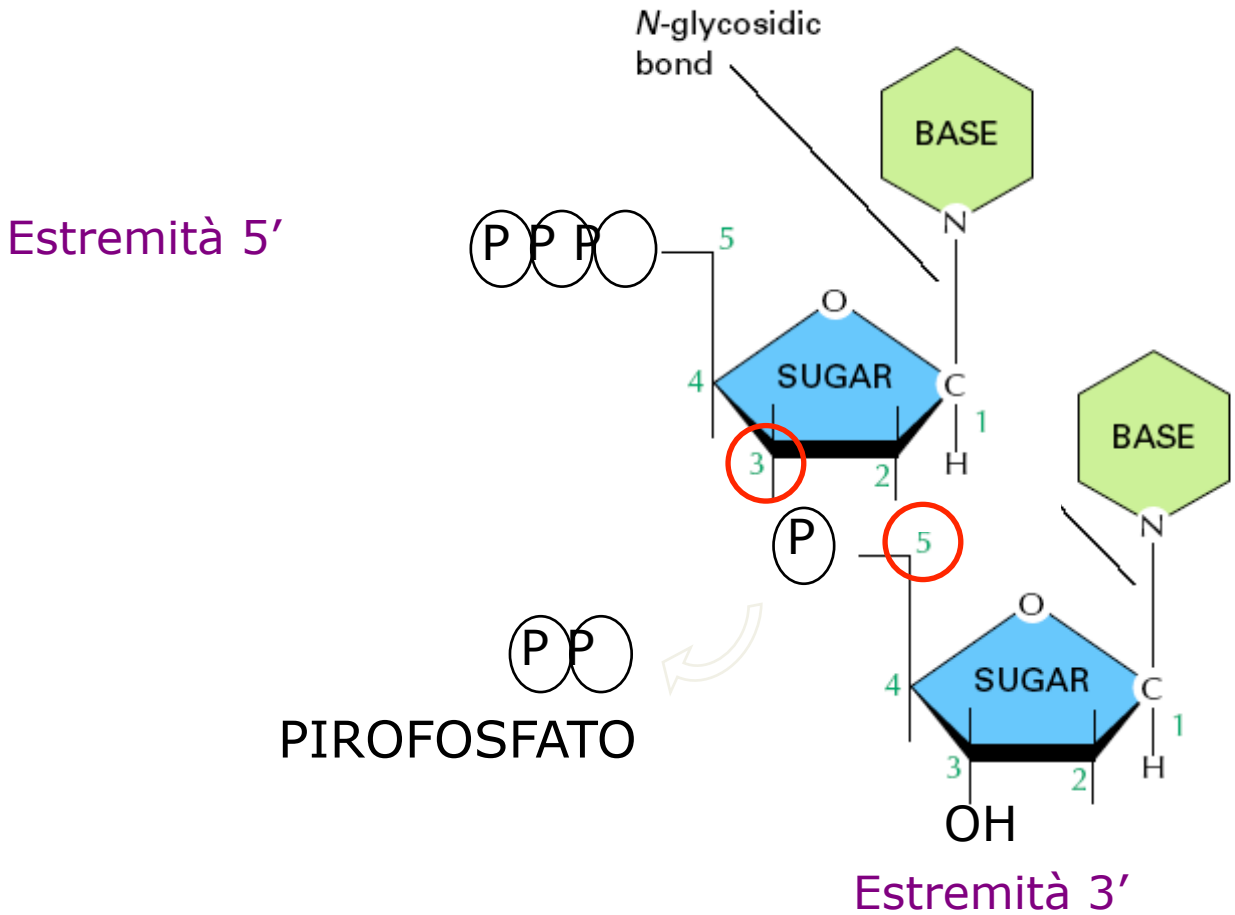


The phosphate makes a nucleotide negatively charged.



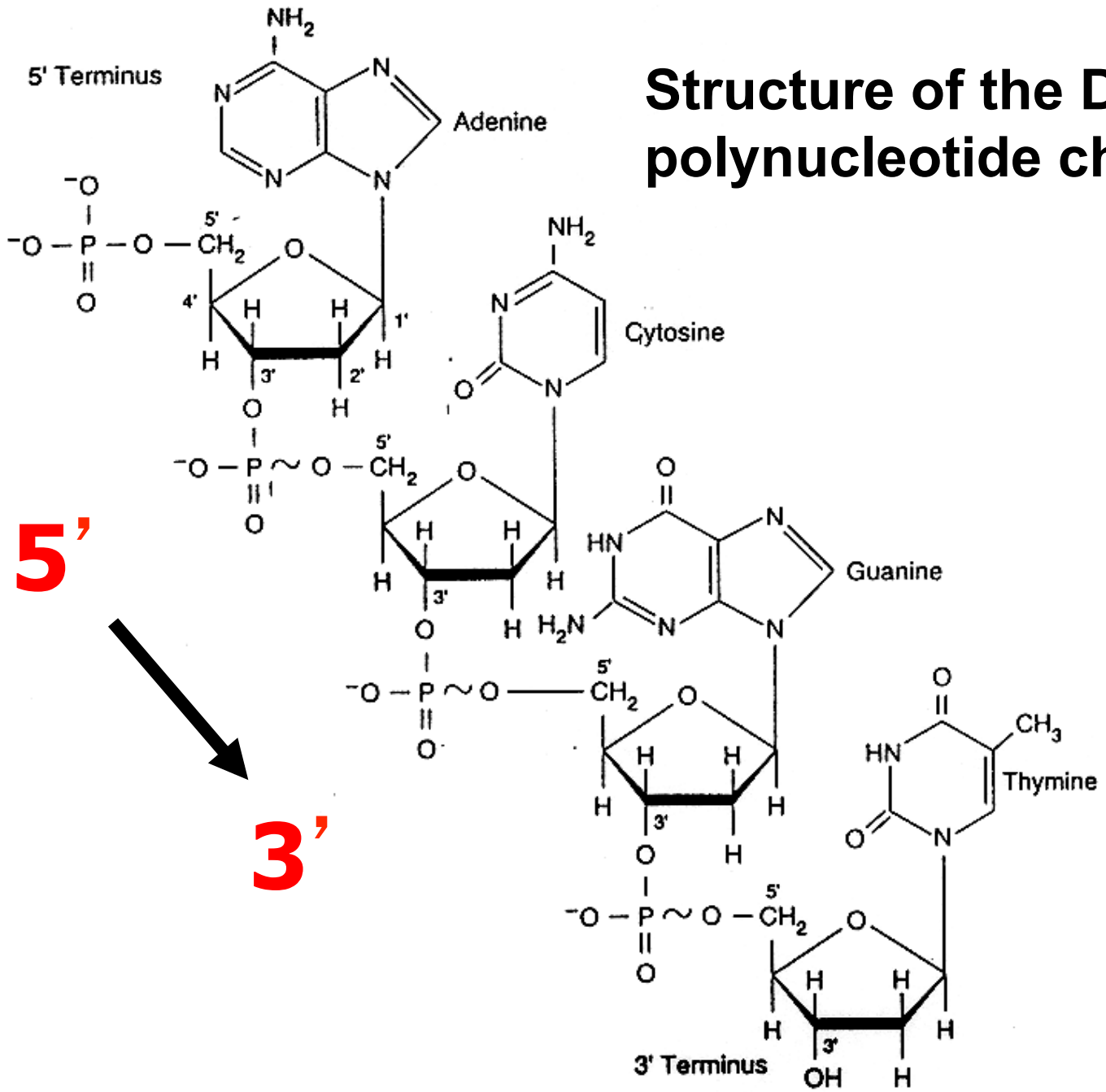

N-glycosidic
bond
BASE
P P P
5
N
O
SUGAR
4
C
1
3
2
H
OH


N-glycosidic
bond
BASE
P P P
5
N
O
SUGAR
4
C
1
3
2
H
OH

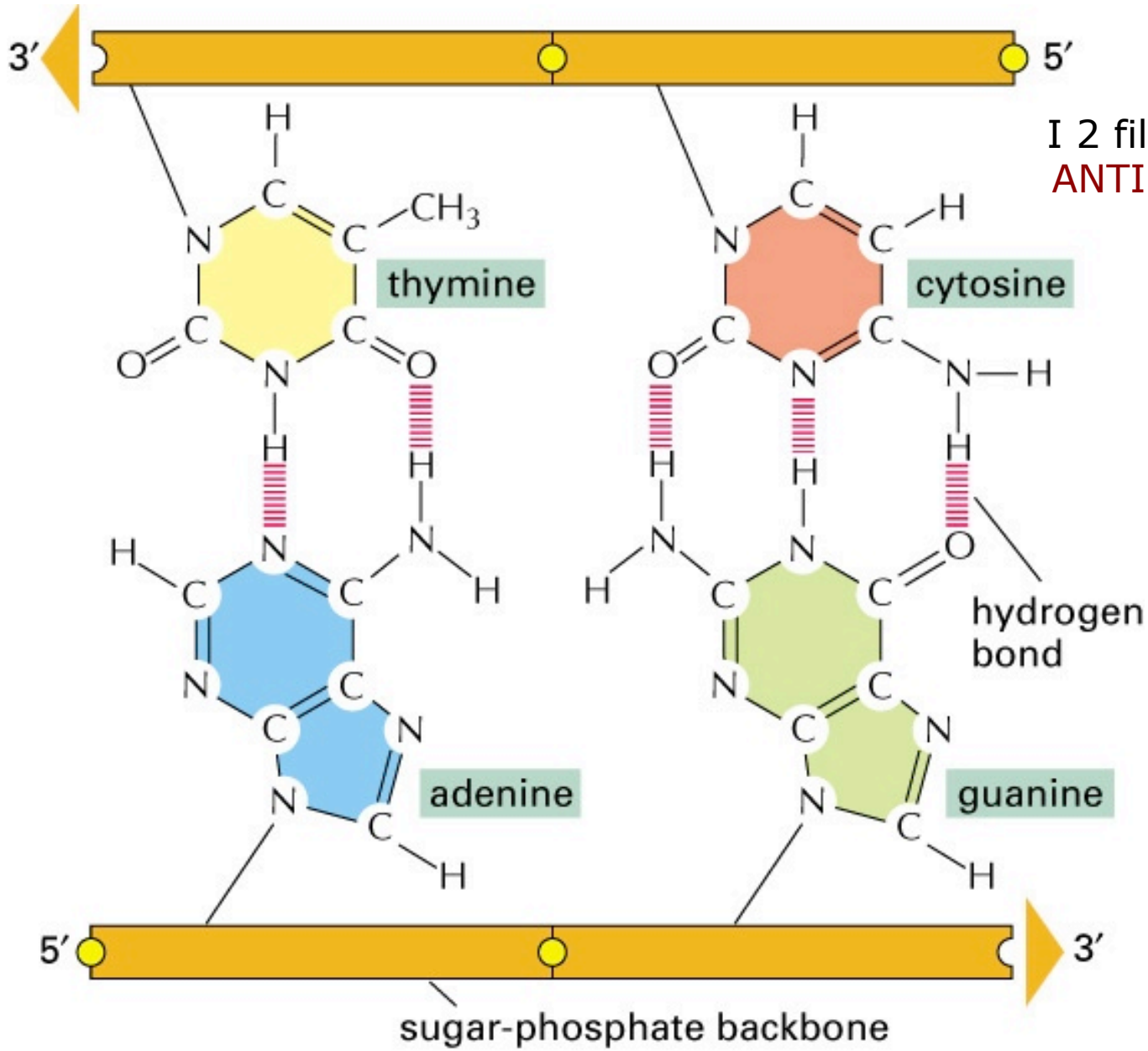


I filamenti di DNA ed RNA hanno una polarità

Structure of the DNA polynucleotide chain



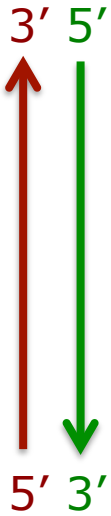
Hydrogen bonding of the bases

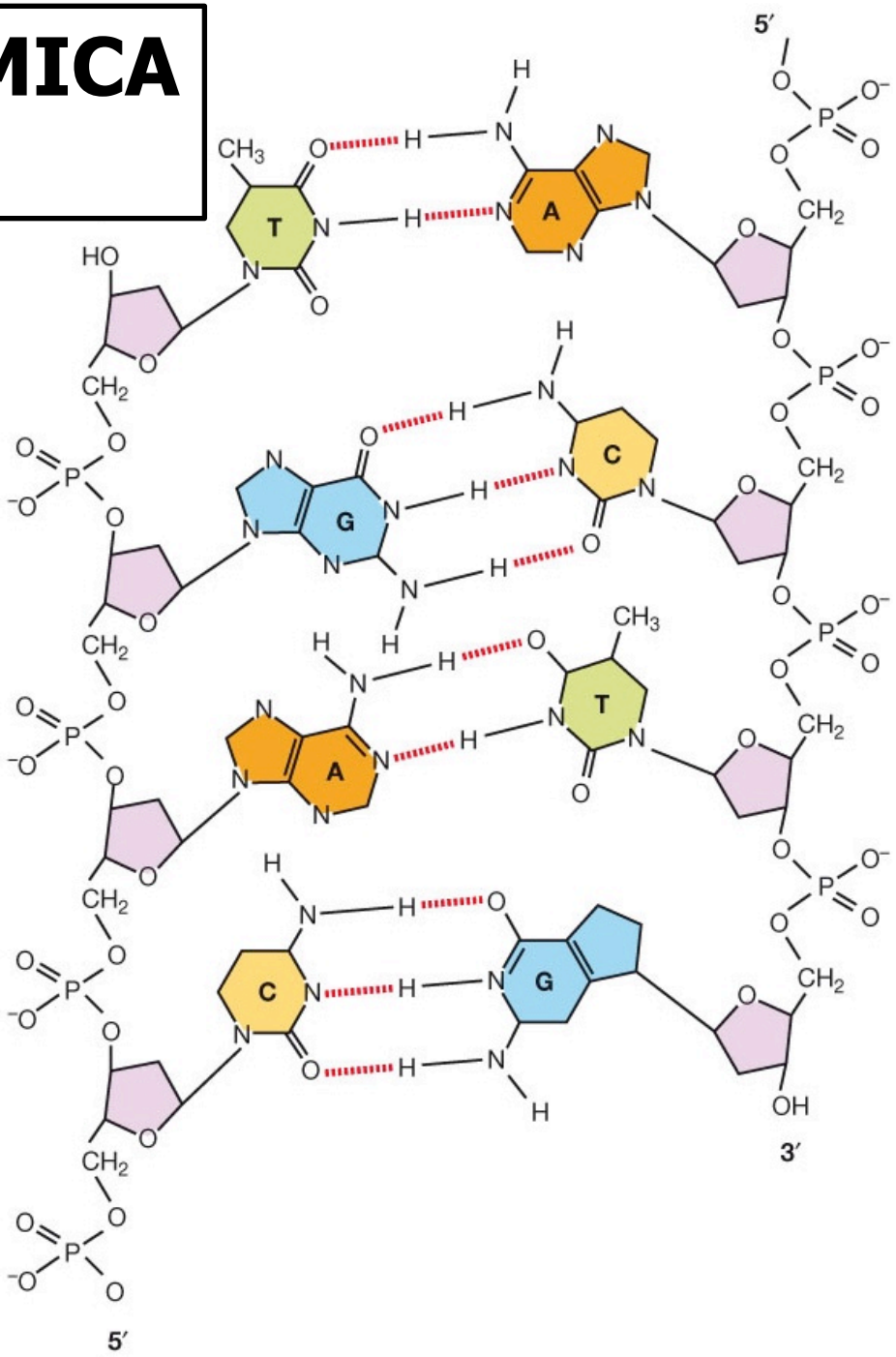


I 2 filamenti sono ANTIPARALLELI!

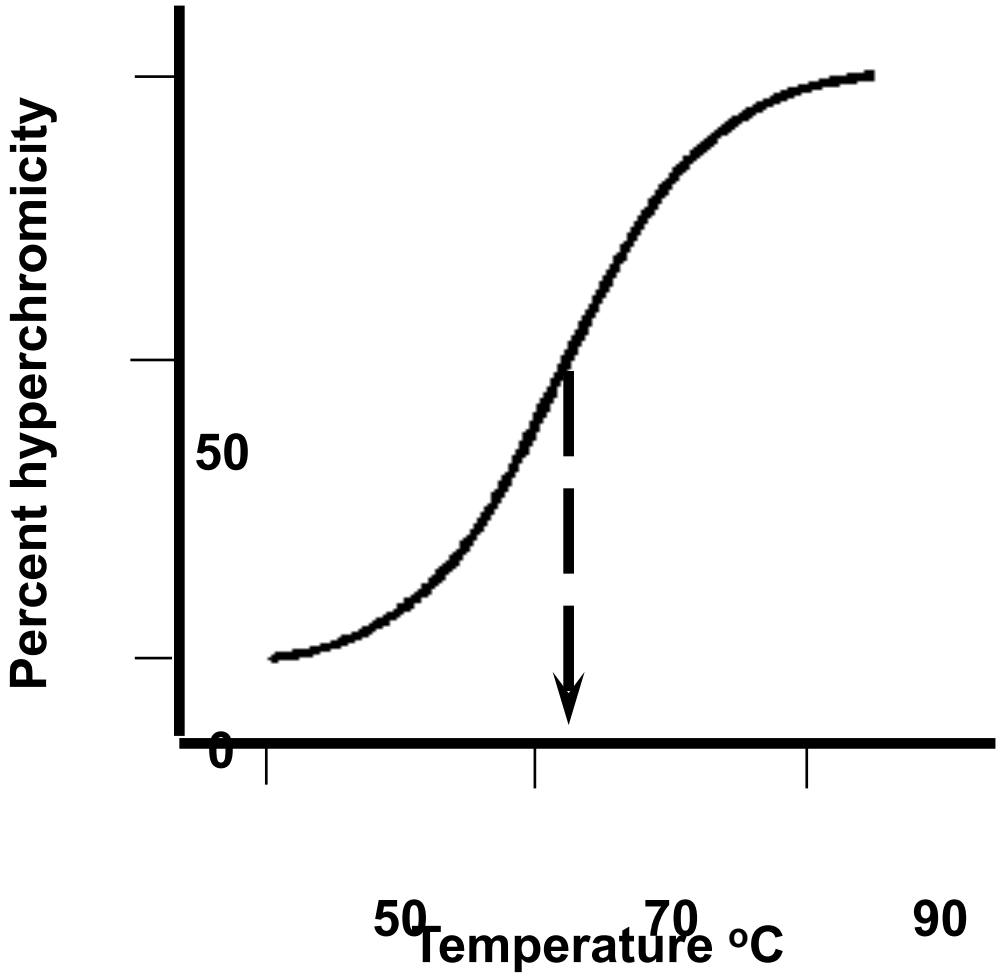
LA STRUTTURA CHIMICA DEL DNA

I 2 filamenti sono
ANTIPARALLELI



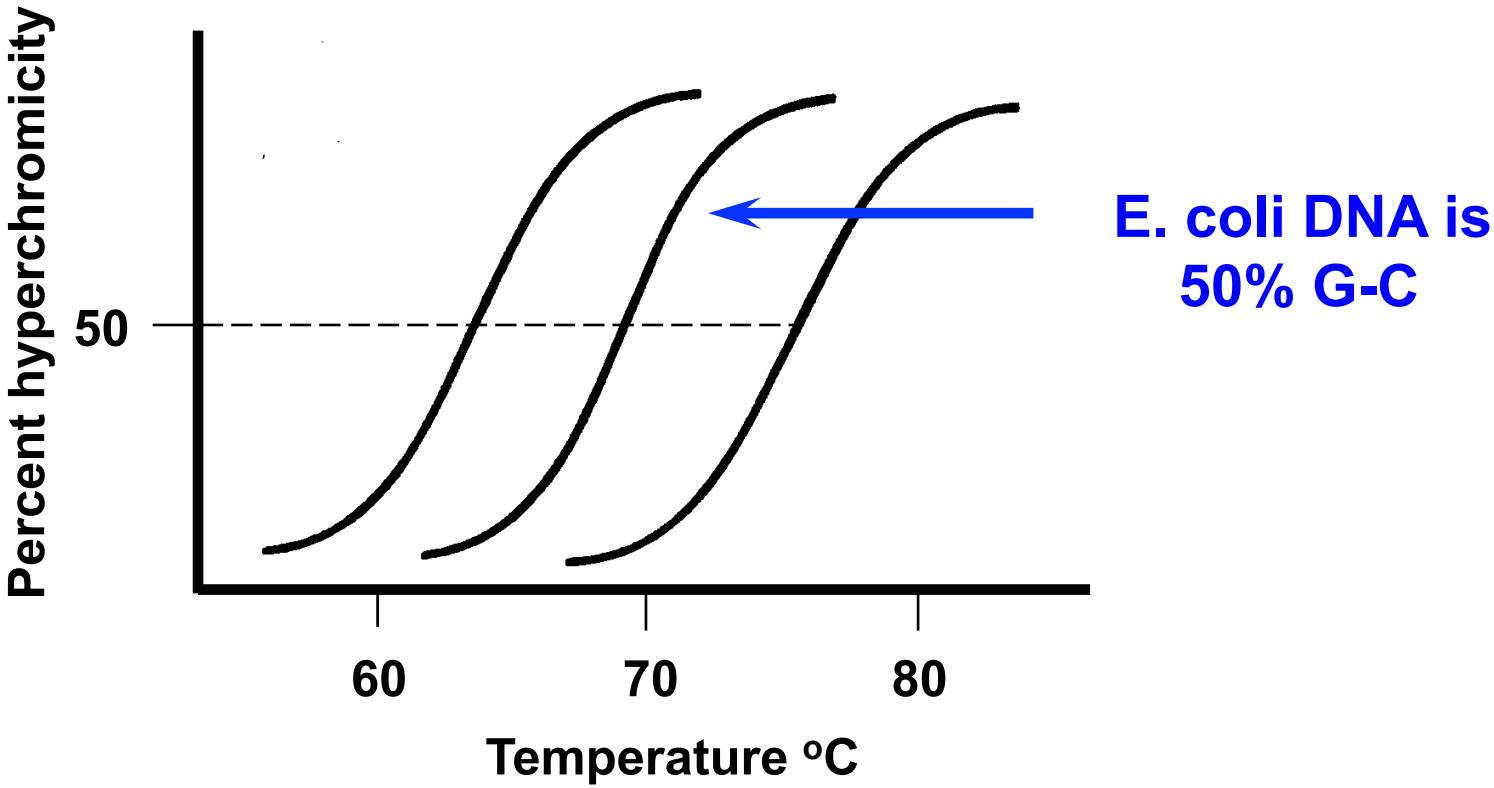


DNA melting curve



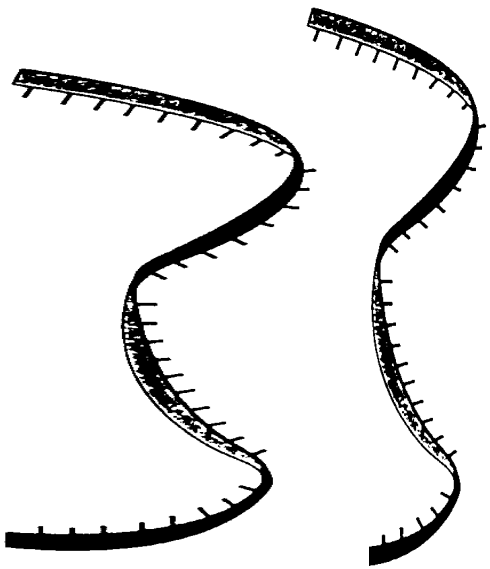
- T_m is the temperature at the midpoint of the transition

T_m is dependent on the G-C content of the DNA



Average base composition (G-C content) can be determined from the melting temperature of DNA

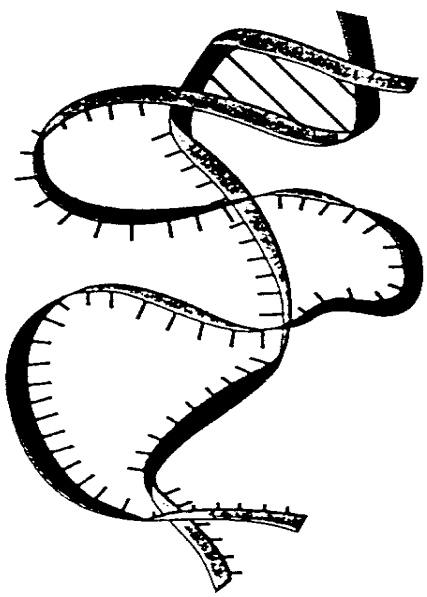
DNA reassociation (renaturation)



Denatured, single-stranded DNA

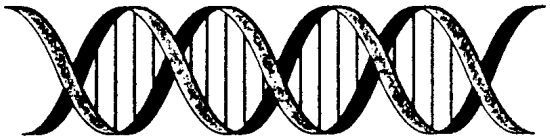
k_2

Slower, rate-limiting, second-order process of finding complementary sequences to nucleate base-pairing



Faster, zippering reaction to form long molecules of double-stranded DNA

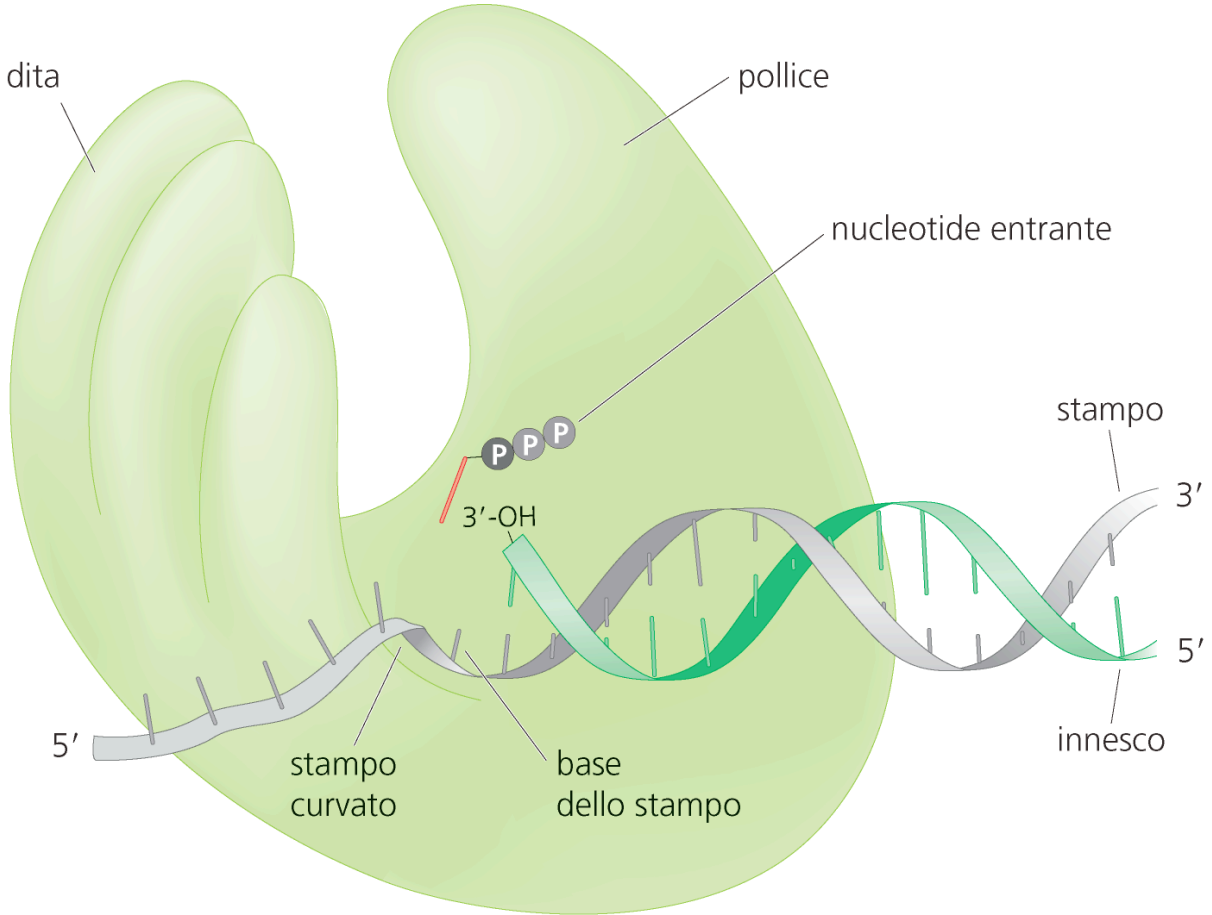
Double-stranded DNA



Le DNA Polimerasi

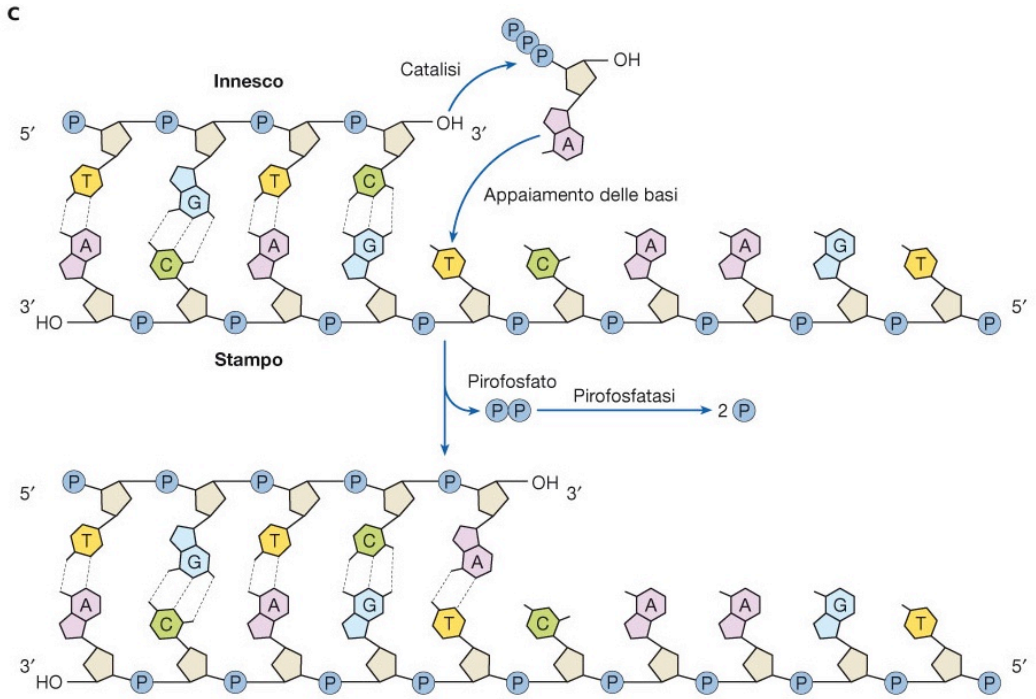
La DNA polimerasi e' un complesso multimerico formato da diverse unita' enzimatiche con diverse attivita' catalitiche.

Tutte le DNA polimerasi condividono la stessa anatomia molecolare, paragonabile a quella di una mano destra. Il complesso "innesco-stampo" attraversa il "palmo", dove viene controllata l'accuratezza dell'appaiamento tra le basi, in una scanalatura creata dalle "dita" e dal "pollice".

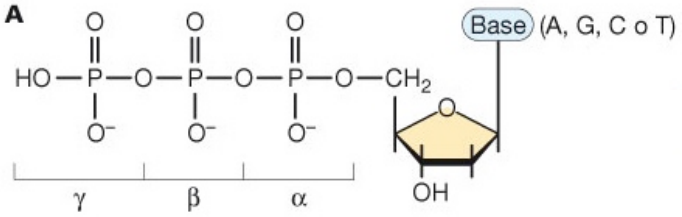


L'azione della DNA Polimerasi

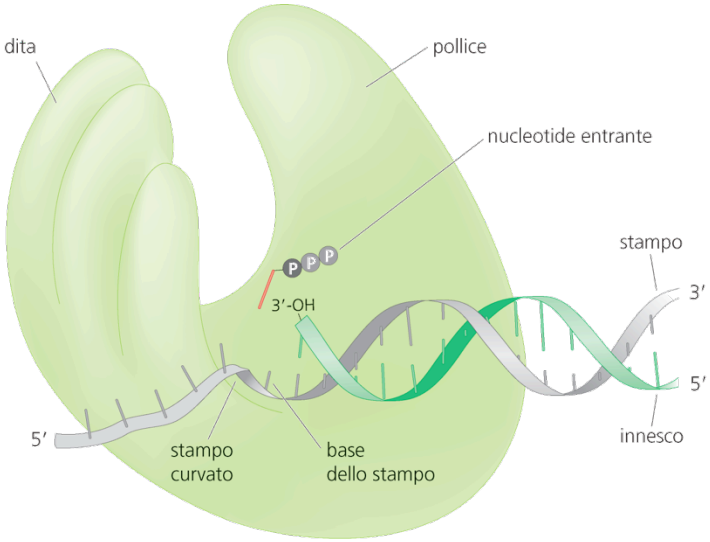
La DNA polimerasi catalizza l'aggiunta sequenziale di deossiribonucleotidi all'estremita' 3'-OH di una catena polinucleotidica accoppiata ad un filamento stampo, formando un legame fosfodiestereo con il gruppo fosfato al 5' del nucleotide aggiuntivo.



I nucleotidi entrano come nucleosidi trifosfati, apportando l'energia necessaria per la polimerizzazione, fornita dall'idrolisi di un legame fosfoanidride, con liberazione di pirofosfato. Il pirofosfato viene poi idrolizzato ulteriormente a fosfato inorganico, rendendo la reazione del tutto irreversibile.



Le DNA Polimerasi



Le diverse isoforme di DNA polimerasi condividono le seguenti caratteristiche:

1. Le DNA polimerasi non sono in grado di rompere i legami idrogeno per separare i due filamenti di una doppia elica di DNA.

2. Tutte le DNA polimerasi necessitano di uno stampo da copiare, fornito dai filamenti di un'elica preesistente.

3. Tutte le DNA polimerasi sono in grado di allungare un filamento di DNA o RNA che funge da innesco, ma non possono iniziare la sintesi di una catena ex novo.

4. I due filamenti di una elica di DNA sono antiparalleli (5'— 3' e 3' — 5') e tutte le DNA polimerasi catalizzano solo l'aggiunta di un nucleotide all'estremità 3' di una catena nascente.

5. In questo modo le catene possono crescere solo in direzione 5' — 3'.

6. Tutte le DNA polimerasi utilizzano come substrato solo i quattro nucleotidi trifosfato.

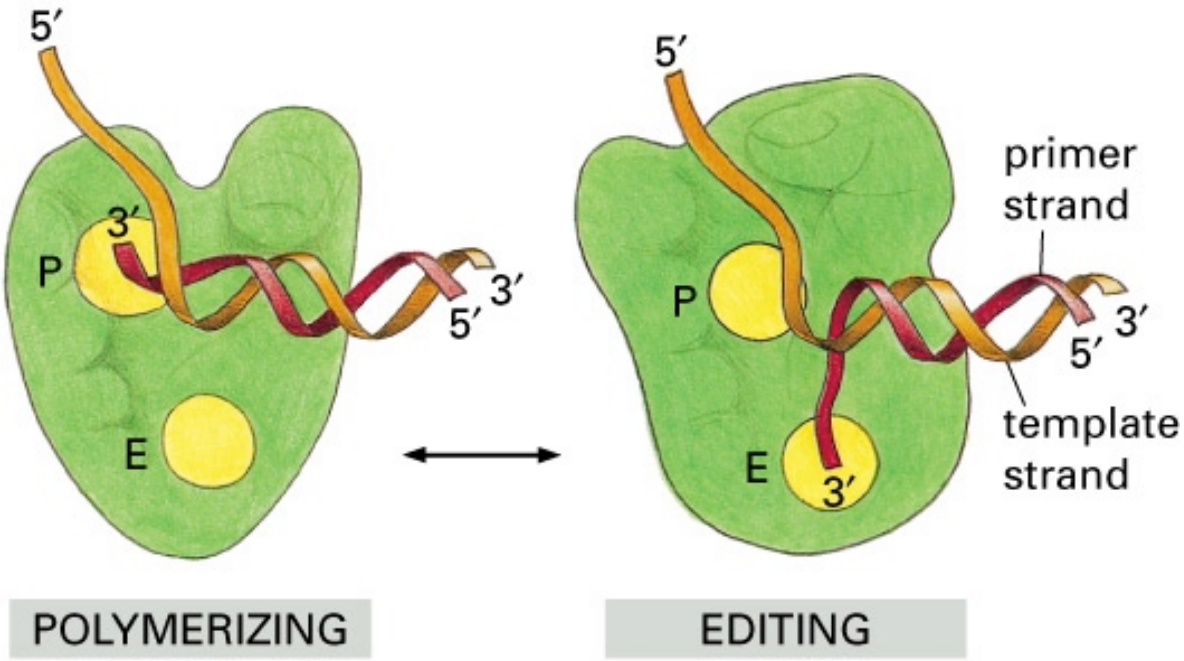
E se vengono commessi degli errori di incorporazione di nucleotidi?

Vari meccanismi di '**proofreading**' intervengono a vari livelli determinando l'elevata fedeltà di copiatura del DNA (1 errore ogni 10^9 nucleotidi copiati).

1. Dopo l'appaiamento del nuovo nt allo stampo, ma prima del legame covalente alla catena crescente
2. Dopo il legame covalente del nuovo nt alla catena crescente
3. Sul DNA già replicato

La DNA polimerasi si autocorregge (proofreading activity)

Prima di aggiungere un nuovo nucleotide, la DNA polimerasi controlla se quello precedente e' appaiato correttamente, altrimenti lo rimuove.



Dopo che un nucleotide e' entrato nel sito catalitico dell'enzima, ma **prima** che si crei il legame fosfodiesterico, la DNA Polimerasi va incontro ad un cambiamento di conformazione. Qualora il nucleotide non si appai in maniera corretta al filamento stampo, un cambio conformazionale ne determina il distacco dall'enzima.

E se vengono commessi degli errori di incorporazione di nucleotidi?

Vari meccanismi di '**proofreading**' intervengono a vari livelli determinando l'elevata fedeltà di copiatura del DNA (1 errore ogni 10^9 nucleotidi copiati).

1. Dopo l'appaiamento del nuovo nt allo stampo, ma prima del legame covalente alla catena crescente
2. Dopo il legame covalente del nuovo nt alla catena crescente
3. Sul DNA già replicato

L'attivita' *proof-reading*

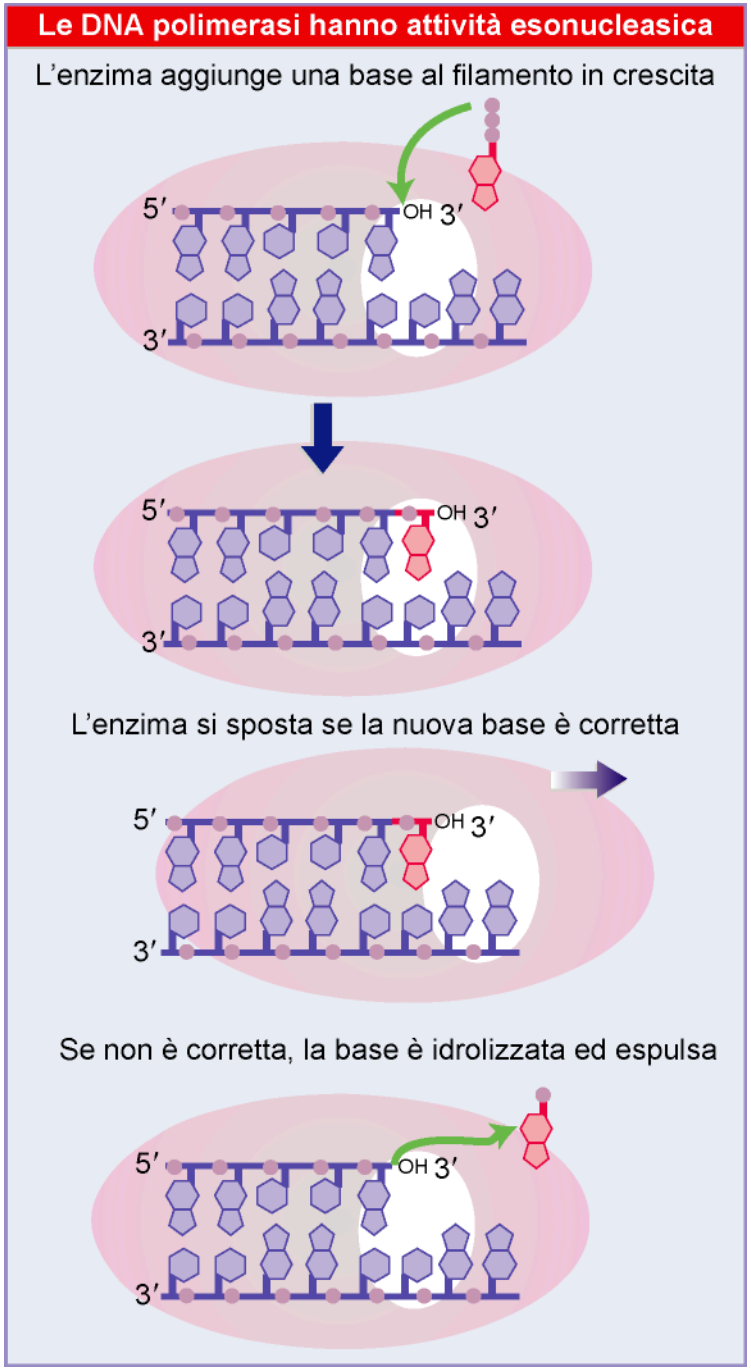
La DNA polimerasi possiede due attivita' enzimatiche distinte, espletate da due domini diversi:

- attivita' polimerasica in direzione 5' -3'

- attivita' esonucleasica in direzione 3' -5'

Quando la polimerasi rileva un errore di appaiamento tra le basi, dopo che il legame PDE e' stato creato, il complesso stampo-innesco si avvicina al dominio con attivita' esonucleasica, dove viene eliminato il nucleotide errato, permettendo alla polimerasi di riprendere velocemente la sintesi, senza provocare la dissociazione dell'intero complesso.

Questa attivita' *proof reading* riduce la frequenza di errore a uno ogni 10^7 nucleotidi



E se vengono commessi degli errori di incorporazione di nucleotidi?

Vari meccanismi di '**proofreading**' intervengono a vari livelli determinando l'elevata fedeltà di copiatura del DNA (1 errore ogni 10^9 nucleotidi copiati).

1. Dopo l'appaiamento del nuovo nt allo stampo, ma prima del legame covalente alla catena crescente
2. Dopo il legame covalente del nuovo nt alla catena crescente
3. Sul DNA già replicato

SISTEMA DI RIPARAZIONE DEGLI APPAIAMENTI SBAGLIATI
(Strand-directed mismatch repair system)

Sistema di enzimi che:

1. Riconosce la distorsione dell'elica dovuta al mismatch (distinguendo il filamento nuovo dal vecchio);

2. Taglia il segmento di DNA contenente il mismatch;

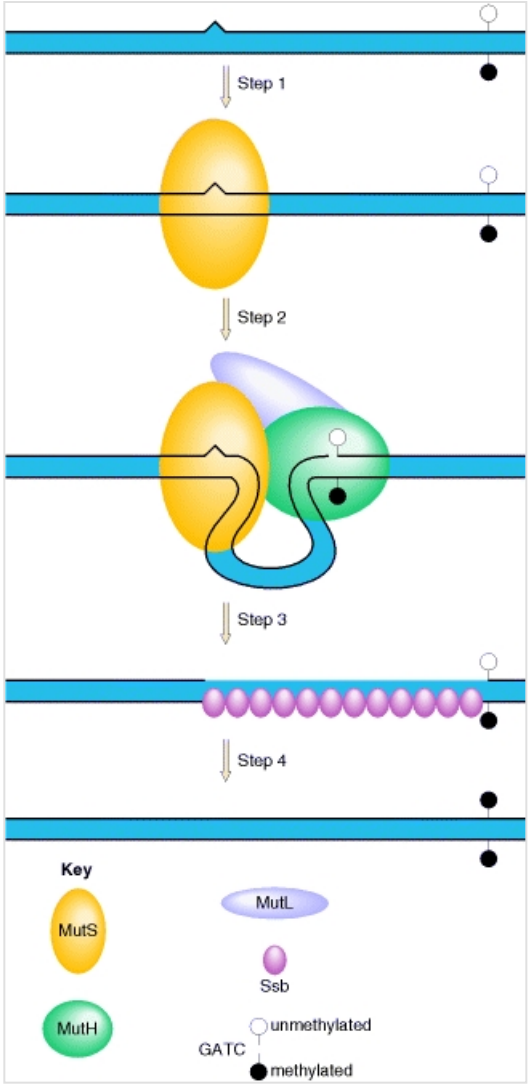
3. Risintetizza il segmento tagliato usando il filamento vecchio come stampo.

SISTEMA DI RIPARAZIONE DEGLI APPAIAMENTI SBAGLIATI (Strand-directed mismatch repair system)

Nei batteri **MutS** scorre lungo il DNA e riconosce il mismatch

MutH e **MutL** vengono reclutate

MutH taglia un segmento di filamento neosintetizzato, il gap viene riempito dalla DNA polimerasi e le estremità saldate dalla DNA ligasi.

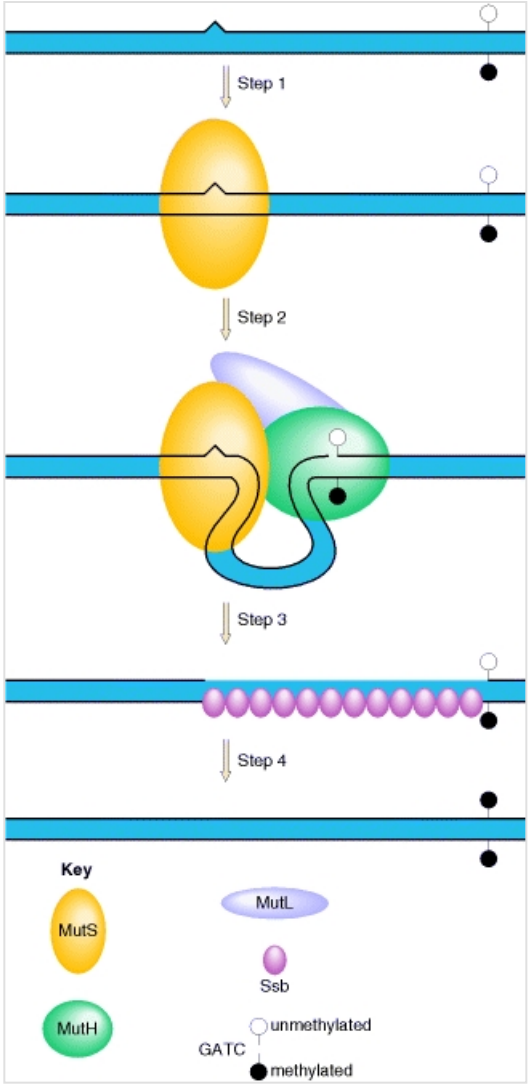


SISTEMA DI RIPARAZIONE DEGLI APPAIAMENTI SBAGLIATI (Strand-directed mismatch repair system)

Nei batteri **MutS** scorre lungo il DNA e riconosce il mismatch

MutH e **MutL** vengono reclutate

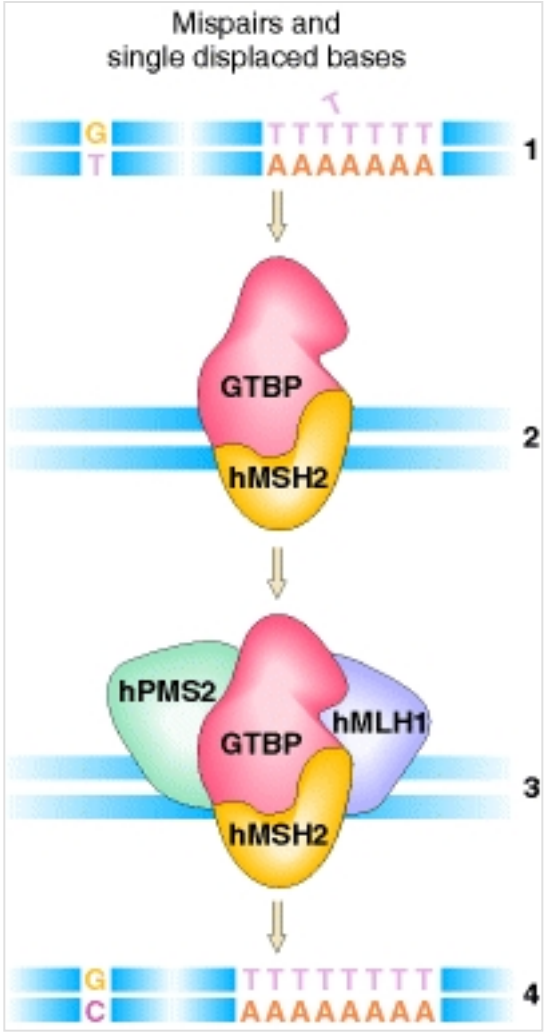
MutH taglia un segmento di filamento neosintetizzato, il gap viene riempito dalla DNA polimerasi e le estremità saldate dalla DNA ligasi.



In bacteria the strand distinction mechanism depends on the selected **methylation of A residues.**
Methyl groups are added to all A in the sequence **<u>GATC</u>** but not until some time after A incorporation - the recognition of unmethylated GATC allows the new DNA to be transiently distinguished.

SISTEMA DI RIPARAZIONE DEGLI APPAIAMENTI SBAGLIATI (Strand-directed mismatch repair system)

Nelle cellule umane hMSH2 e hMLH1, sono simili a MutS e MutL rispettivamente, ma intervengono tante altre proteine



Mutazioni nei geni del mismatch repair system causano la predisposizione a certi tipi di cancro (es. HNPCC, carcinoma colonrettale ereditario non poliposico, noto anche come sindrome di Lynch; test predittivo: mutazioni nei geni MSH2 e MLH1).