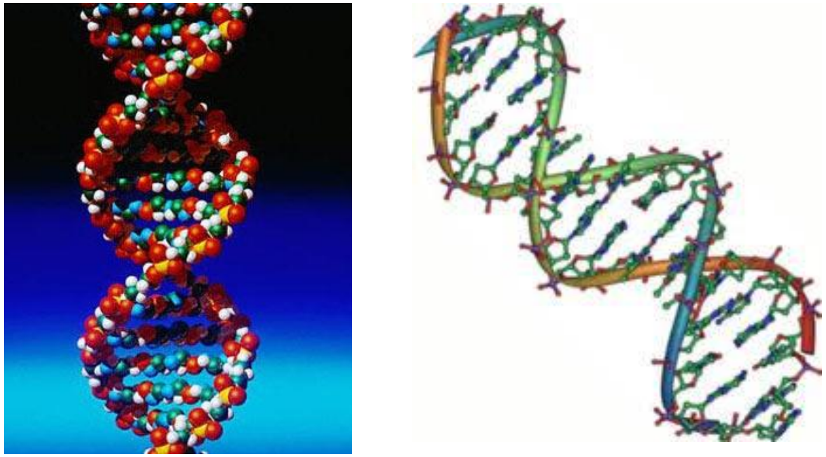
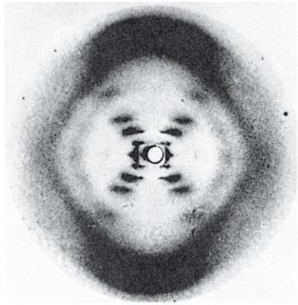
DNA replication: synthesis and repair (Ch. 4 and 14)

James Watson **Francis Crick**

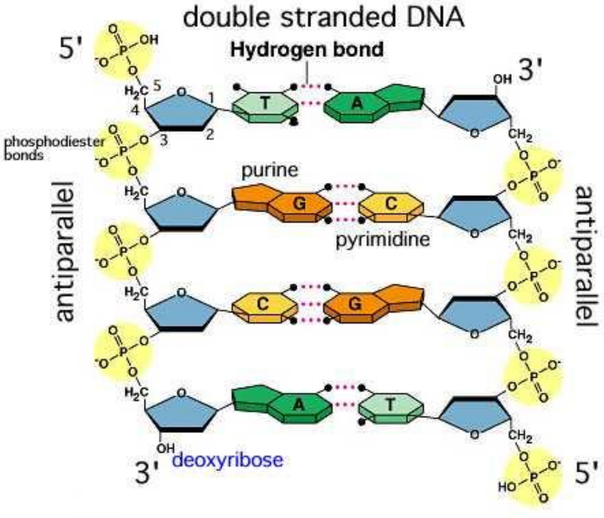


Double helix model of DNA

(a) Rosalind Franklin



(b) Franklin's X-ray diffraction photograph of DNA



Key concepts – DNA structure and function

- DNA (nucleic acid) is the genetic material (not, for example, protein)

- DNA contains the genetic “blueprint” that directs the development of all your biochemical, anatomical, physiological and behavioural traits

- You resemble your parents because of precise replication of DNA and transmission from one generation to the next

- The molecular structure of DNA (Watson & Crick’s discovery) predicts how DNA replicates

- Many proteins work together in DNA replication (e.g. primase, helicase, polymerase)

- Specific details of replication process differ between the two DNA strands that replicate (“leading” versus “”lagging” strands)

- DNA replication is *not* perfect: repair mechanisms and telomeres

DNA (Deoxyribonucleic acid) and RNA (Ribonucleic acid) are polymers of nucleotide monomers

Nuclelotides have 3 components

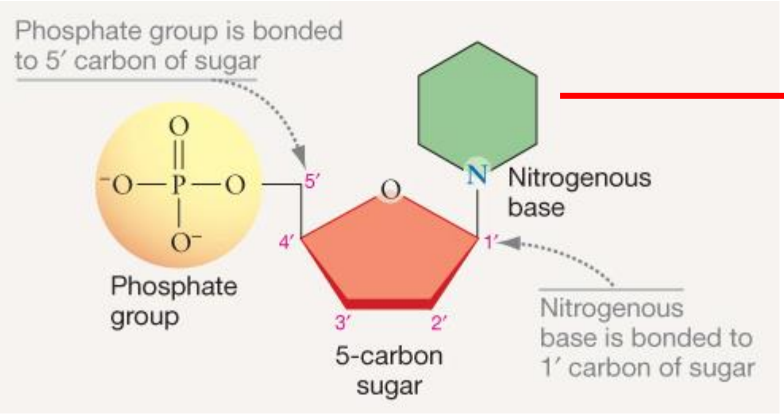
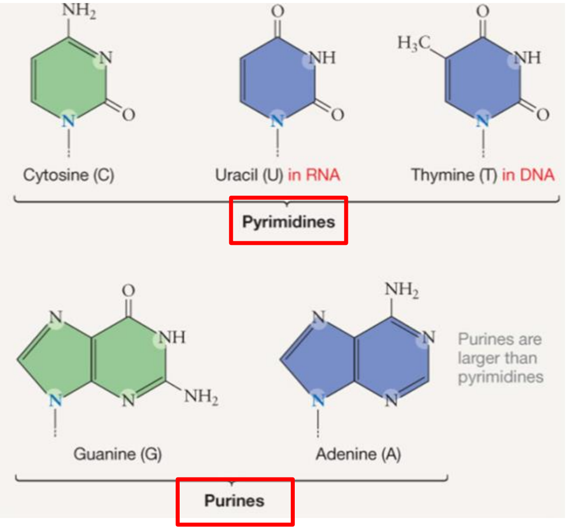


Fig. 4.1

5 different nitrogenous bases



- nitrogenous (nitrogen-containing) bases: cytosine (C), uracil (U), thymine (T), guanine (G) + adenine (A)
- DNA has deoxyribose sugar + either C, G, A or T base
- RNA has ribose sugar + C, G, A BUT U instead of T

Nucleotides polymerise to form nucleic acids

3’ carbon

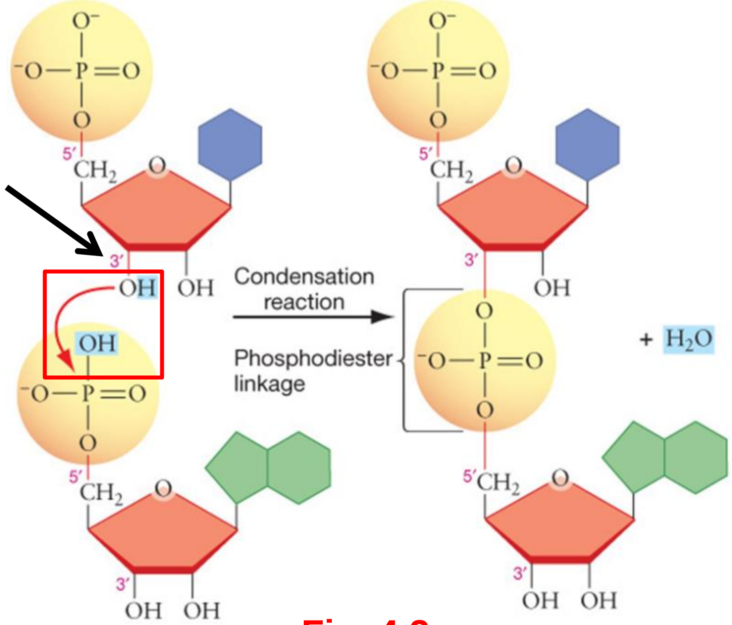
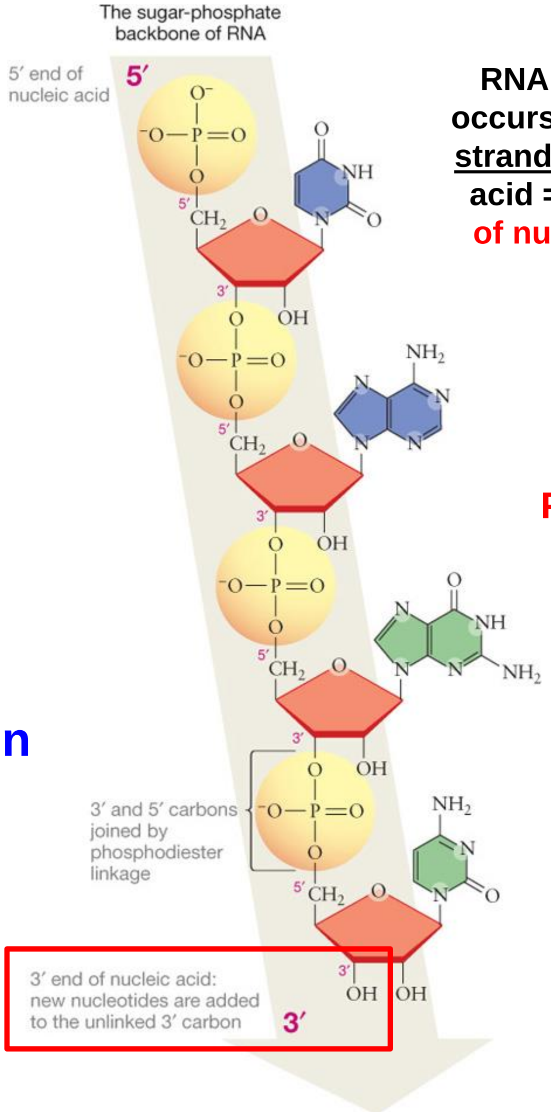


Fig. 4.2

- Nucleotides polymerize via a condensation reaction forming phosphodiester link between 3’ carbon and 5’ carbon
- Forms a sugar-phosphate “backbone” which is directional 5’ → 3’

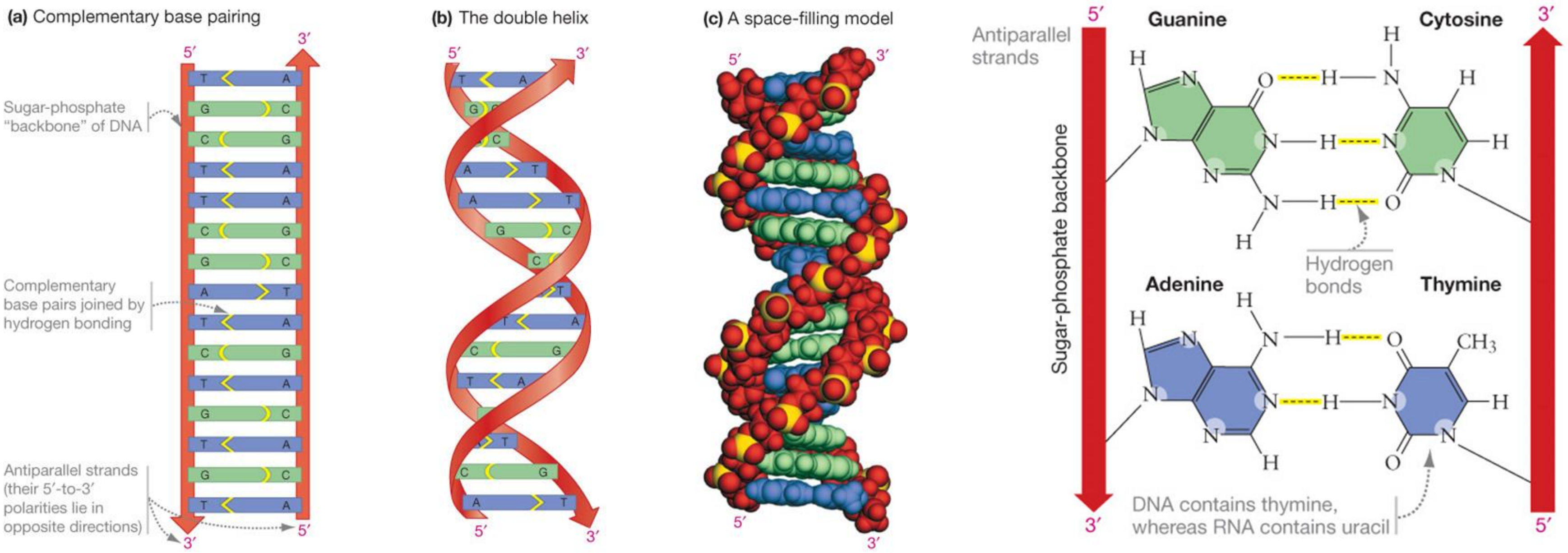


RNA typically occurs as single-stranded nucleic acid = polymer of nucleotides

Fig. 4.3

Important: new nucleotides are only added at 3’ end

DNA has a double-helix structure with two anti-parallel nucleic acid strands (Fig. 14.4 and 4.6)



- **Antiparallel strands**: one runs 5' to 3' and the other in the opposite direction 3' to 5'

- **Double helix is stabilised by hydrogen bonds between nucleotide bases**

- **Complimentary base pairing: C + G, and A + T (see Fig. 4.6a)** 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., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).

[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., **11** (3) (1950).

[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

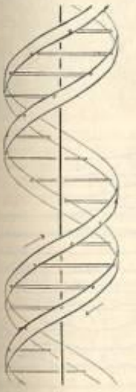
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

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 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.

Francis Crick

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[5] (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 the innermost maxima of each Bessel function and the origin. The angle this line makes with the equator is roughly equal to the angle between an element of the helix and the helix axis. If a unit repeats n times along the helix there will be a meridional reflexion (J_0^2) on the nth layer line. The helical configuration produces side-bands on this fundamental frequency, the effect[5] being to reproduce the intensity distribution about the origin around the new origin, on the nth layer line, corresponding to C in Fig. 2.

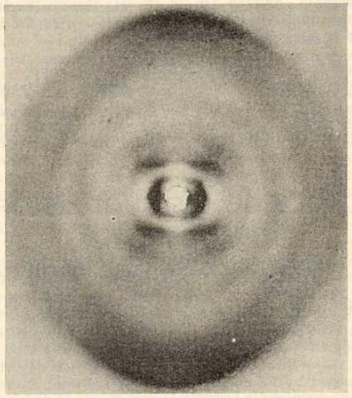


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *B. coli*. Fibre axis vertical

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

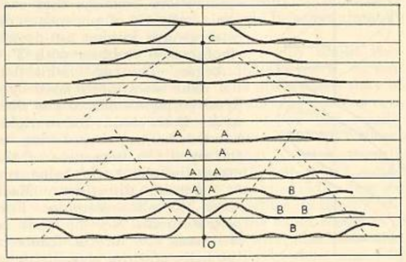
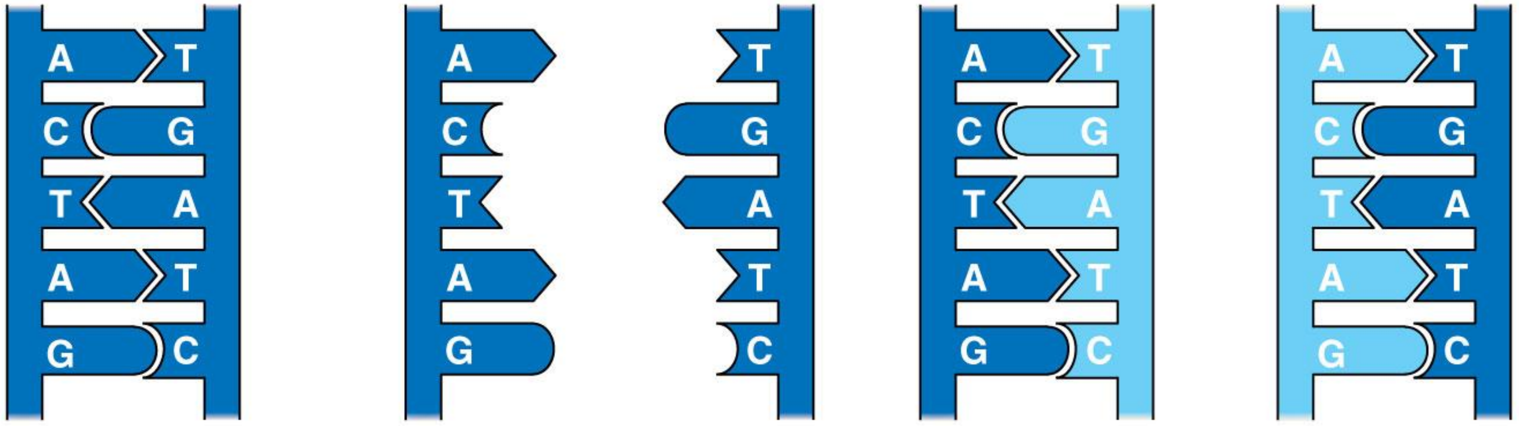


Fig. 2. Diffraction pattern of system of helices corresponding to structure of deoxypentose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third and fifth layer lines for half of the nucleotide mass at 20 A. diameter and remainder distributed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

The basic model for DNA replication (see Fig. 4.9)

Watson & Crick (1953) "*it has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism*"



(a) Parent molecule

(b) Separation of strands

(c) "Daughter" DNA molecules, each consisting of one parental strand and one new strand

Unwinding and separation ("unzipping") of complimentary DNA strands forms two template strands; then addition of appropriate, complimentary nucleotides forms (2) new, identical copies

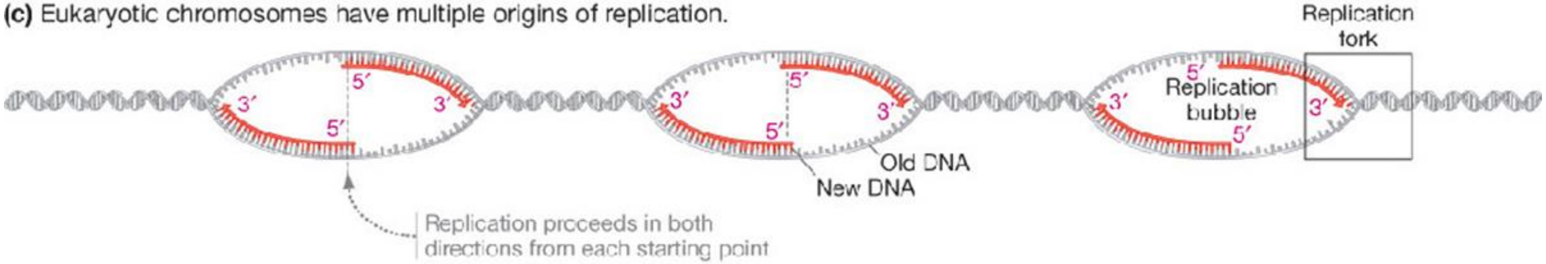
Sequence of pairs of bases is duplicated exactly

How does DNA replication get started?

(a) A chromosome being replicated

(b) Bacterial chromosomes have a single origin of replication.

(c) Eukaryotic chromosomes have multiple origins of replication.

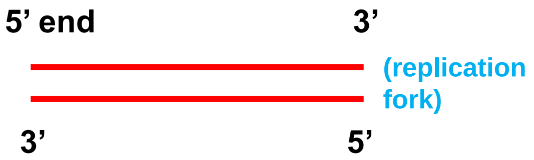


- DNA strands start to separate at replication “bubbles” which form at the origin of replication = stretches of DNA with specific nucleotide sequences

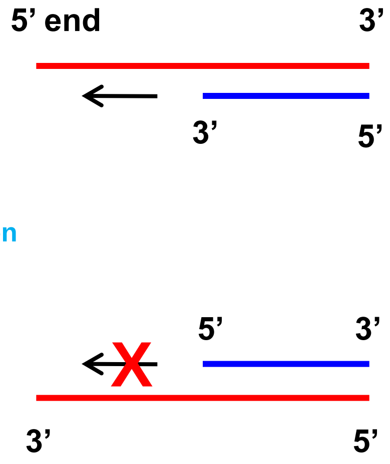
- replication bubbles grow as replication proceeds because synthesis of new complimentary DNA strands is bidirectional

- once a replication bubble opens a suite of enzymes initiates replication; this starts at the origin and moves towards replication fork

Why does DNA replication seem so complicated? The problem of the leading versus lagging template strand

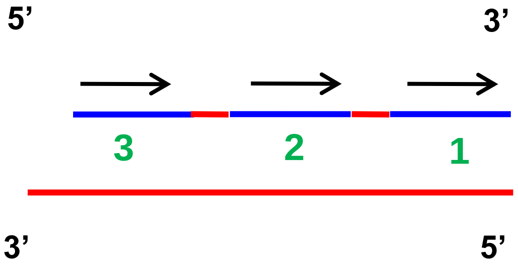


Antiparallel DNA



Leading strand – nucleotides added to 3’ end of complimentary strand = continuous

Red = parental (template) strand
Blue = new, complimentary strand



Lagging strand

Lagging strand – nucleotides added to 3’ end of complimentary strand in short segments *away* from replication fork = Okazaki fragments

- this way nucleotides can be added at 3’ end and the short segments are then joined together to complete lagging strand = discontinuous

How is the double helix opened and stabilised?

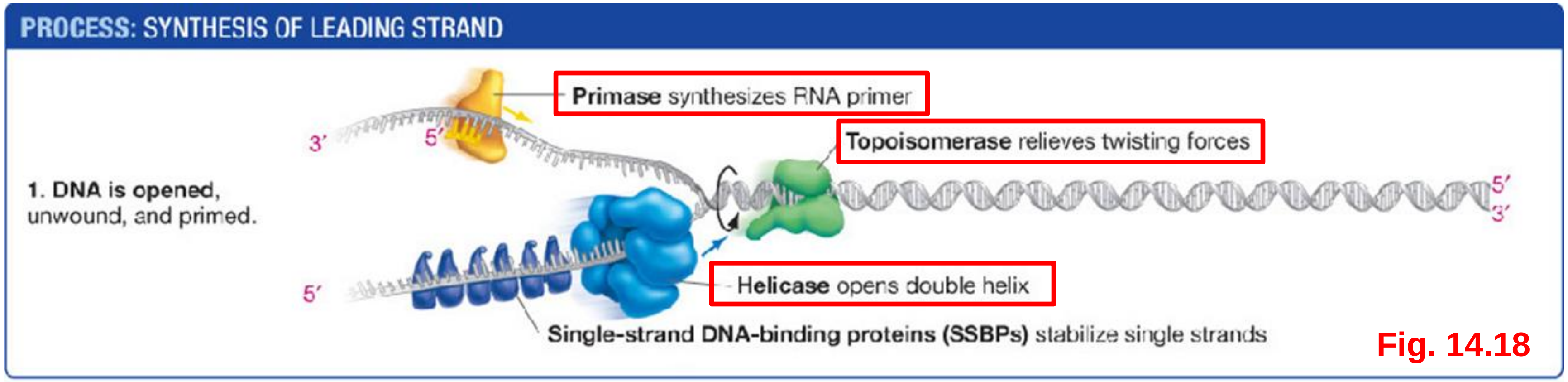
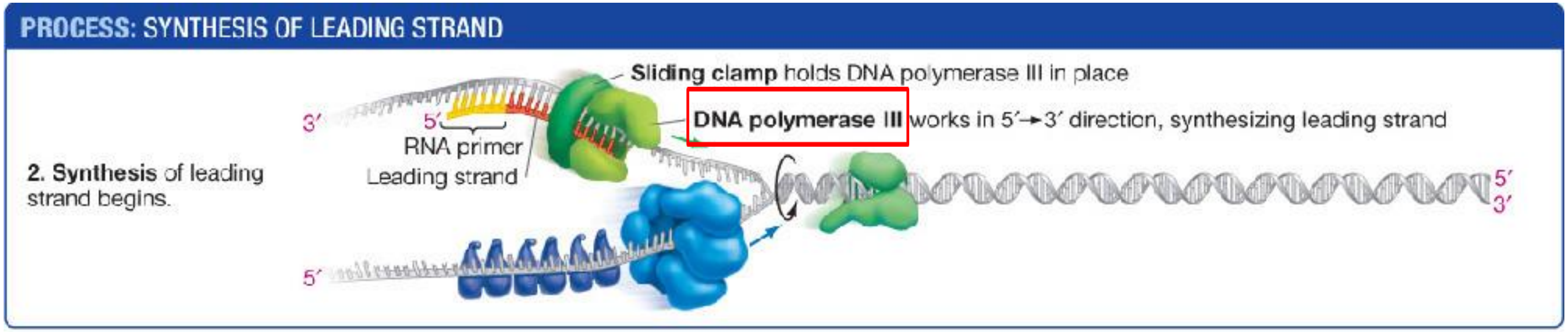


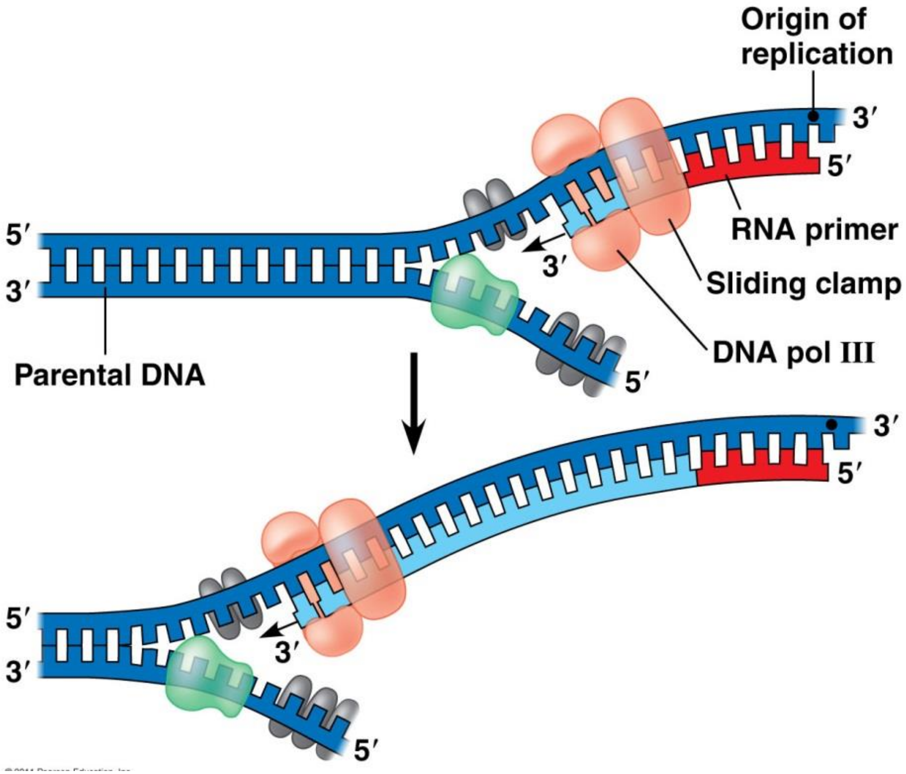
Fig. 14.18

- A battery of enzymes and other proteins converge on the point where the double helix opens and stabilize the separating single DNA strands
- Now the single template strands are ready to be copied



DNA polymerase III is the enzyme that adds nucleotides at the 3’ end to the template strand

Synthesis of the leading strand is straightforward after an RNA primer is in place – a summary (see Fig. 14.8)

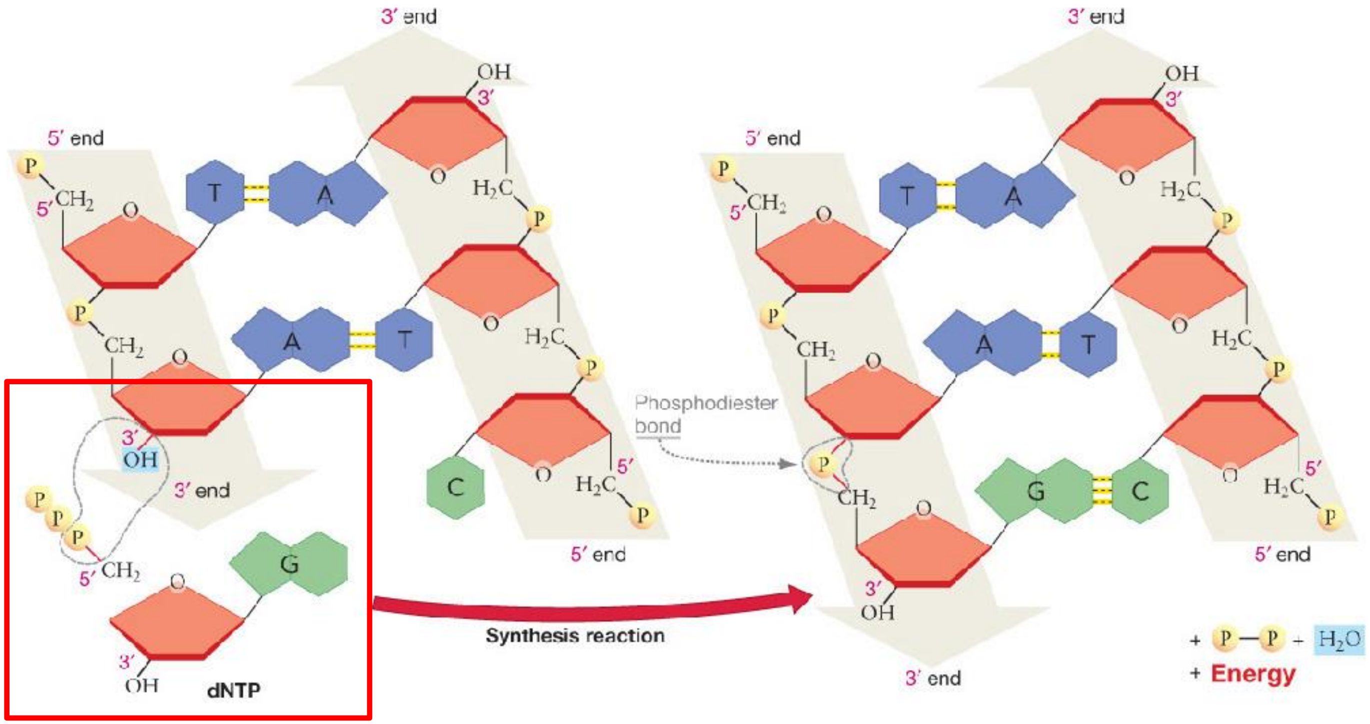


- **Primase** synthesises the initial nucleotide chain = an **RNA primer**
- [nucleotides can only be added to an existing nucleotide chain]
- Primer is 5-10 RNA nucleotides long, base-paired to template strand
- New **DNA strand** starts at 3’ end of primer

DNA polymerase III is moved along the DNA template strand by the associated **“sliding clamp” protein**

- catalyses synthesis of DNA by adding nucleotides to pre-existing chain at 3’ end (in a 5’ to 3’ direction) at a rate of 50 per second

The DNA synthesis reaction: incorporating a complimentary nucleotide into a growing DNA strand

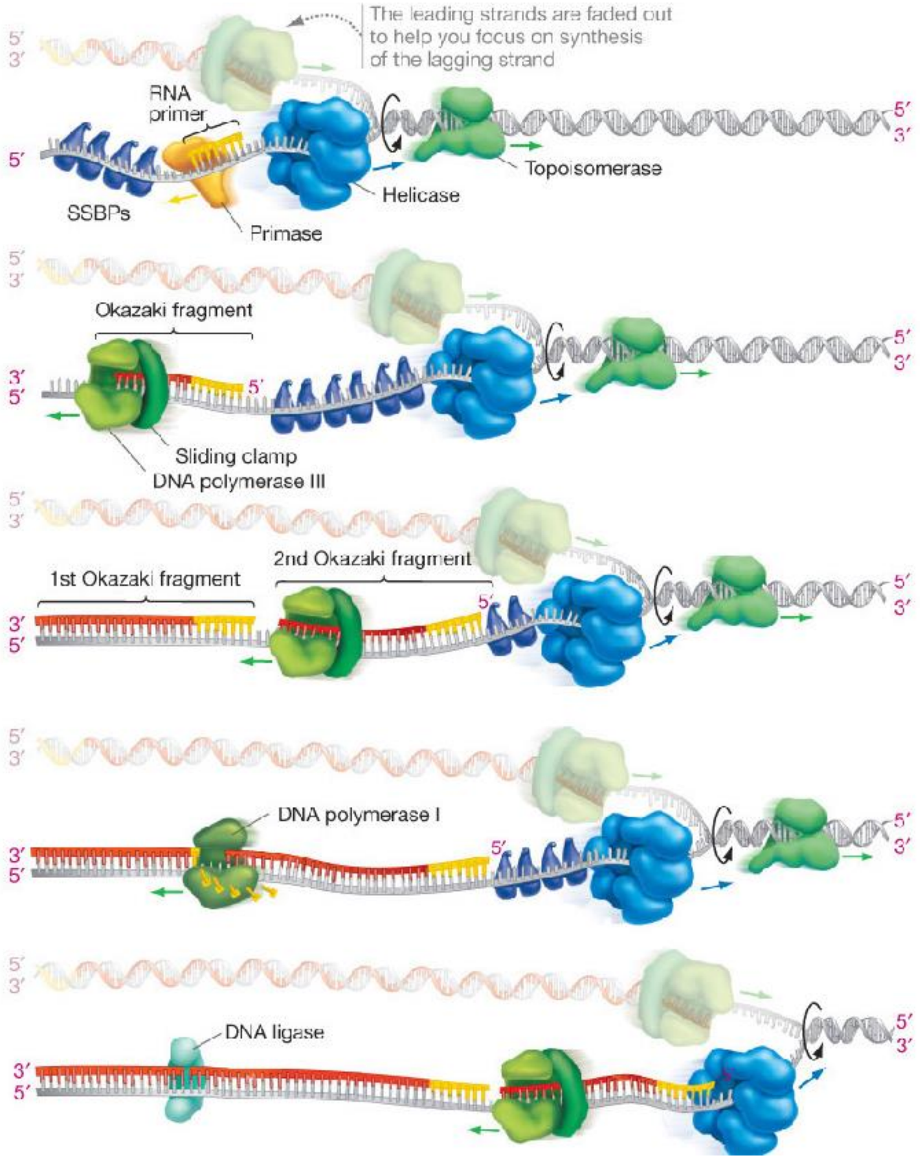


DNA polymerase catalyzes synthesis of new DNA strand by adding nucleotides to a pre-existing chain – why only at the 3’ end?

- **Nucleotides come from nucleoside triphosphates = base + sugar + 3 P**
- **Triphosphates (like ATP) are chemically reactive**
- **Loss of 2 P is energetically favourable (exergonic) and drives polymerisation**

Synthesis of the lagging strand during DNA replication

Fig. 14.10 (see Table 14.1 for summary of proteins required for DNA synthesis)



The leading strands are faded out to help you focus on synthesis of the lagging strand

1. **Primase** forms RNA primer

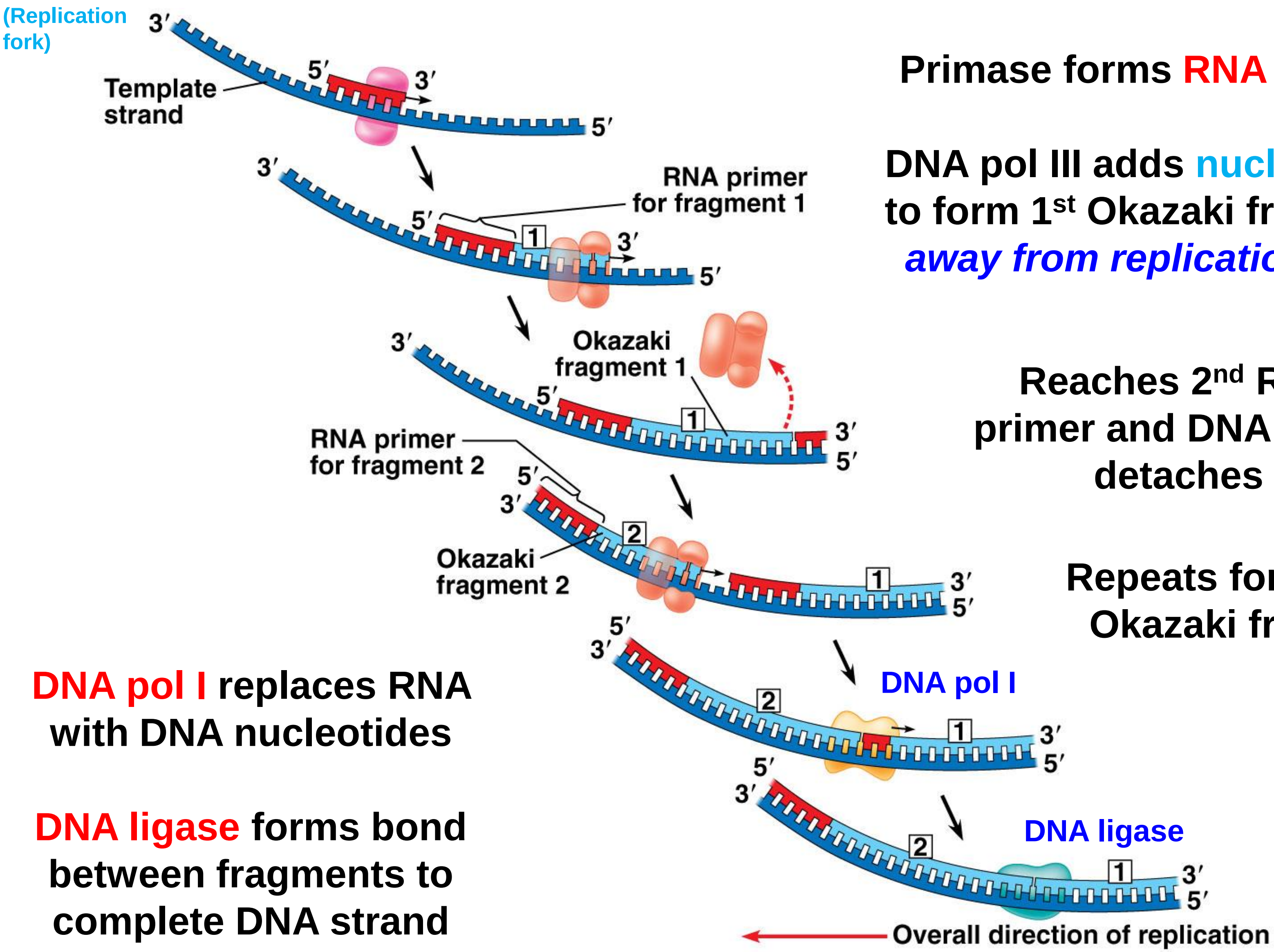
2. **DNA pol III** synthesises 1st Okazaki fragment in 5’ → 3’ durection (away from replication fork)

3. **Primase and DNA pol III** synthesise 2nd Okazaki fragment (closer to replication fork)

4. **DNA pol I** removes RNA primer and replaces with DNA nucleotides

5. **DNA ligase** closes gaps in sugar-phosphate backbone

Synthesis of the lagging strand during DNA replication



Primase forms RNA primer

DNA pol III adds nucleotides to form 1st Okazaki fragment *away from replication fork*

Reaches 2nd RNA primer and DNA pol III detaches

Repeats forming 2nd Okazaki fragment

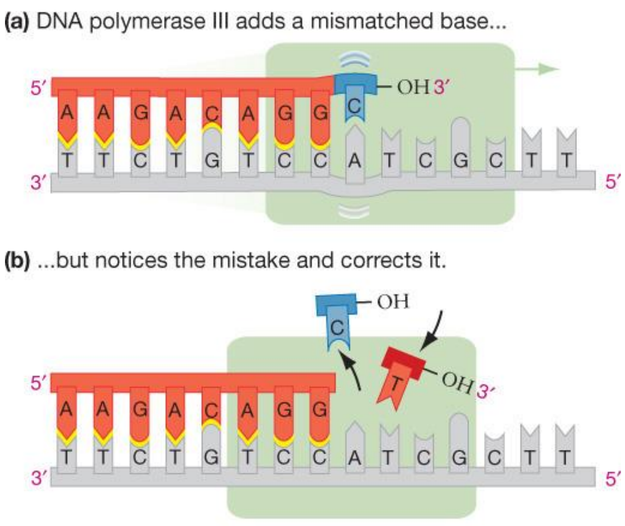
DNA pol I replaces RNA with DNA nucleotides

DNA ligase forms bond between fragments to complete DNA strand

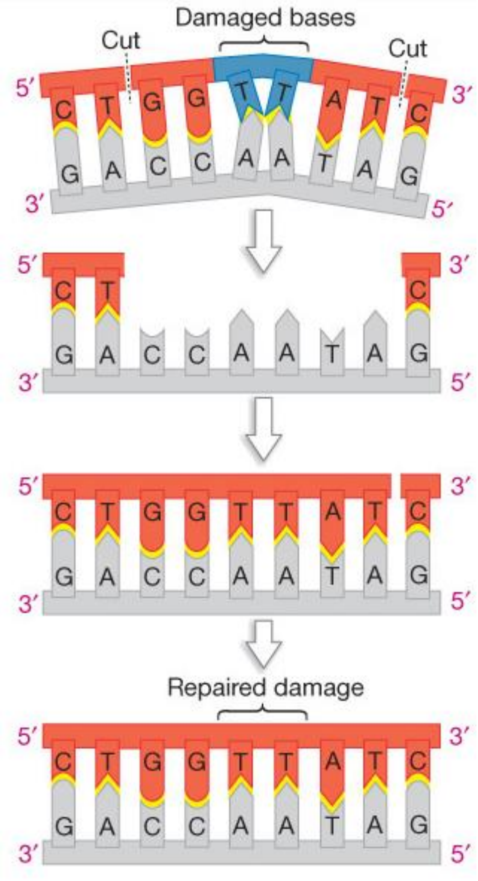
Repairing mistakes and damage in DNA synthesis

(Fig. 14.14 and 14.16)

Initial pairing errors during DNA replication occur 1:100,000, but permanent errors occur 1:10 billion = mutations



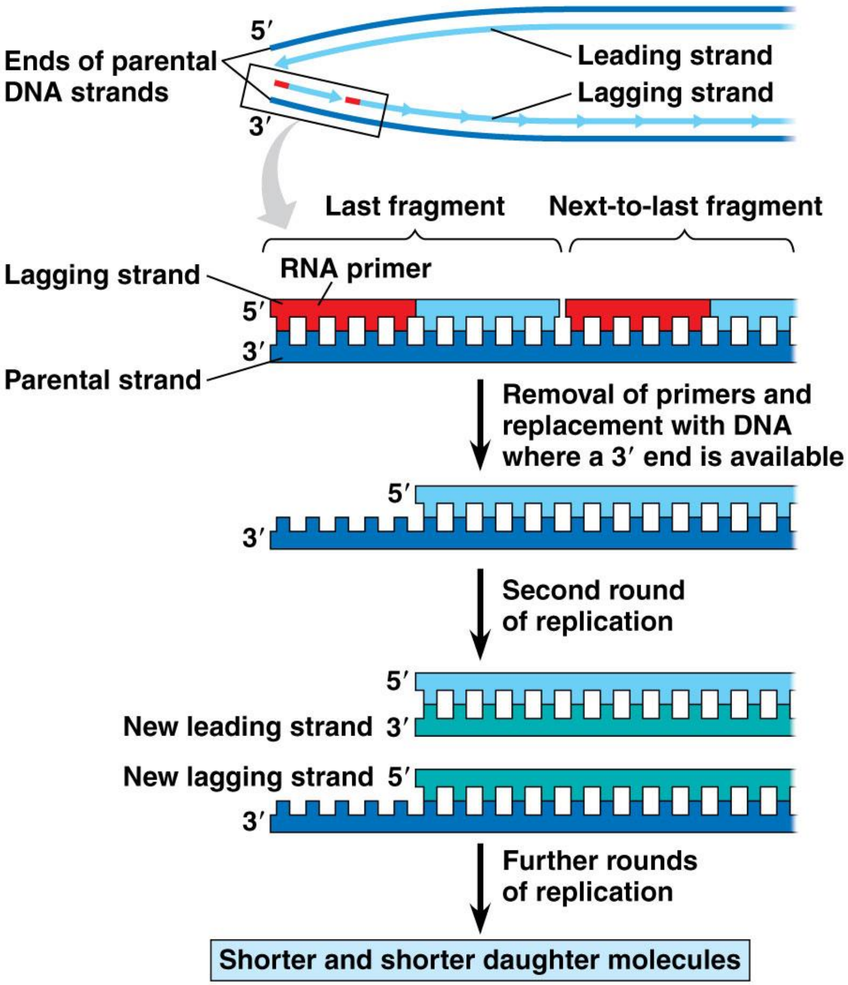
DNA polymerase can proofread – correcting mistakes as it goes

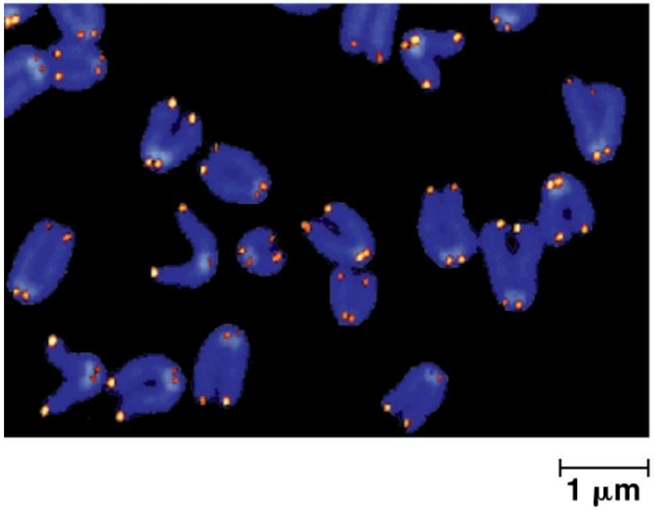


- **Enzymes detect and cut DNA removing damaged section (nuclease enzymes)**
- **DNA polymerase fills in the gap by adding in missing, complimentary nucleotides**
- **DNA ligase seals the free ends of the new DNA to the old DNA**

DNA replication is ***not*** perfect: telomeres (see Fig. 14.12)

Normal replication machinery provides no way to complete the 5’ end of lagging strands: genes near the end of DNA could be eroded away!



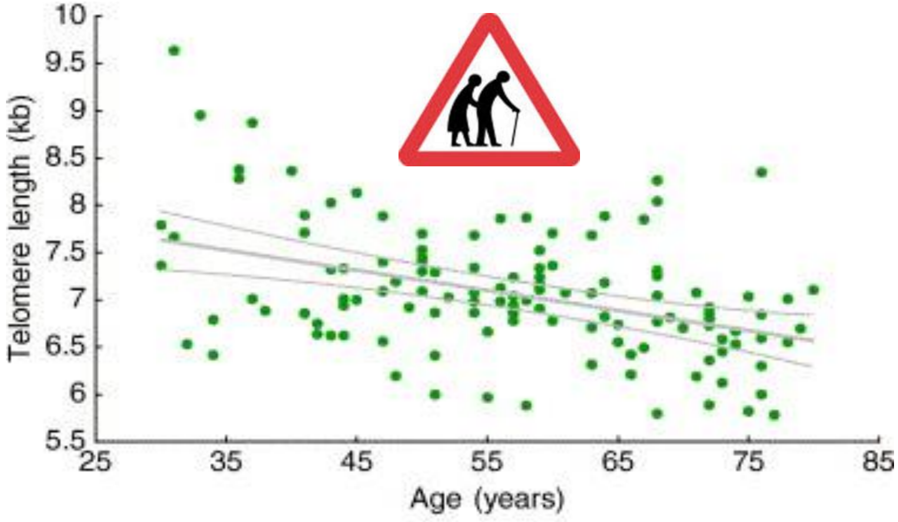


Telomeres = repetitive, non-coding nucleotide sequences at the ends of DNA which protect genes (“protective caps”)

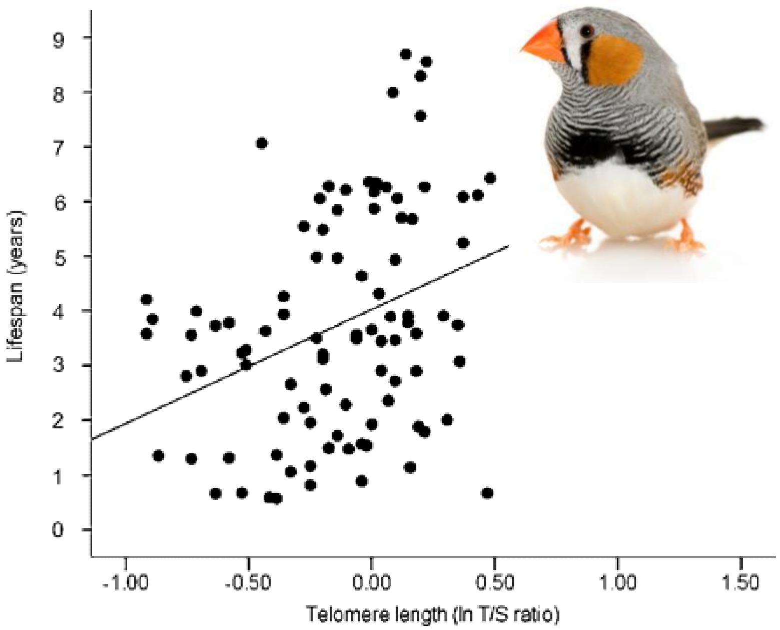
e.g. in human DNA the six-nucleotide sequence TTAGGG is repeated 100-1000 times

Telomeres: ageing, lifespan, disease and cancer

Humans >60 years old, with relatively short telomeres have higher mortality from heart and infectious disease



TRENDS in Ecology & Evolution



Telomeres predict lifespan

Telomere loss might protect organisms from excessive proliferation (cancer) by limiting number of cell divisions that cells can undergo

Cancer cells express telomerase an enzyme which rebuilds telomeres; does this allow cancer cells to persist and divide repeatedly? Box 14.1