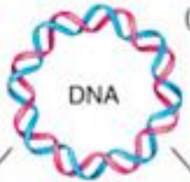
A Single-Molecule View of DNA Replication

Antoine van Oijen
Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School

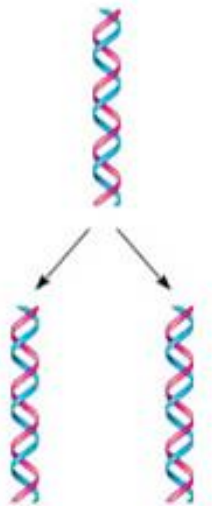
Central Dogma

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.

(a) (b)

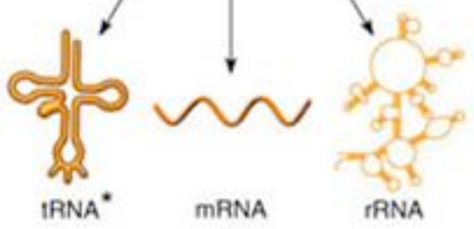


Replication of DNA

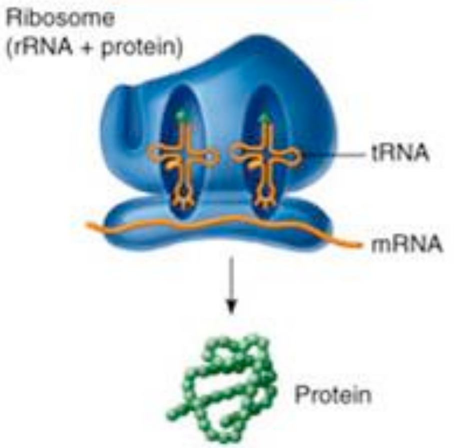


Inheritance of DNA language in daughter cells

Transcription of DNA



Translation of RNA



Expression of DNA for structure and functions of cell

*The sizes of RNA are not to scale—tRNA and mRNA are enlarged to show details.

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 [illegible] chemical arguments.

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

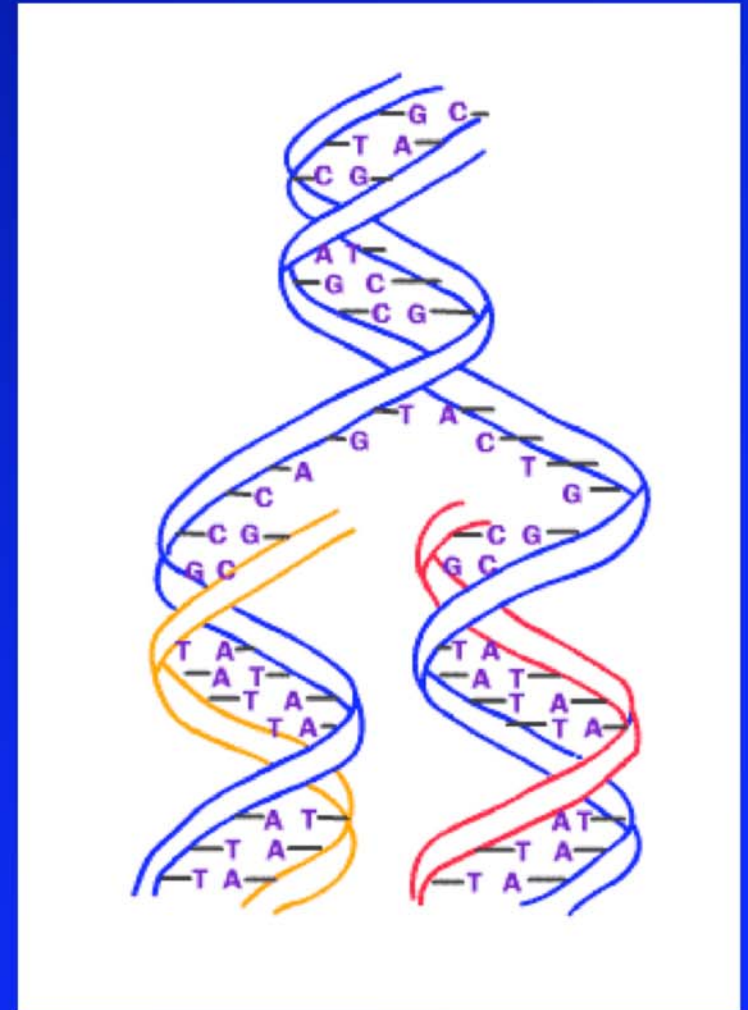
Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

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

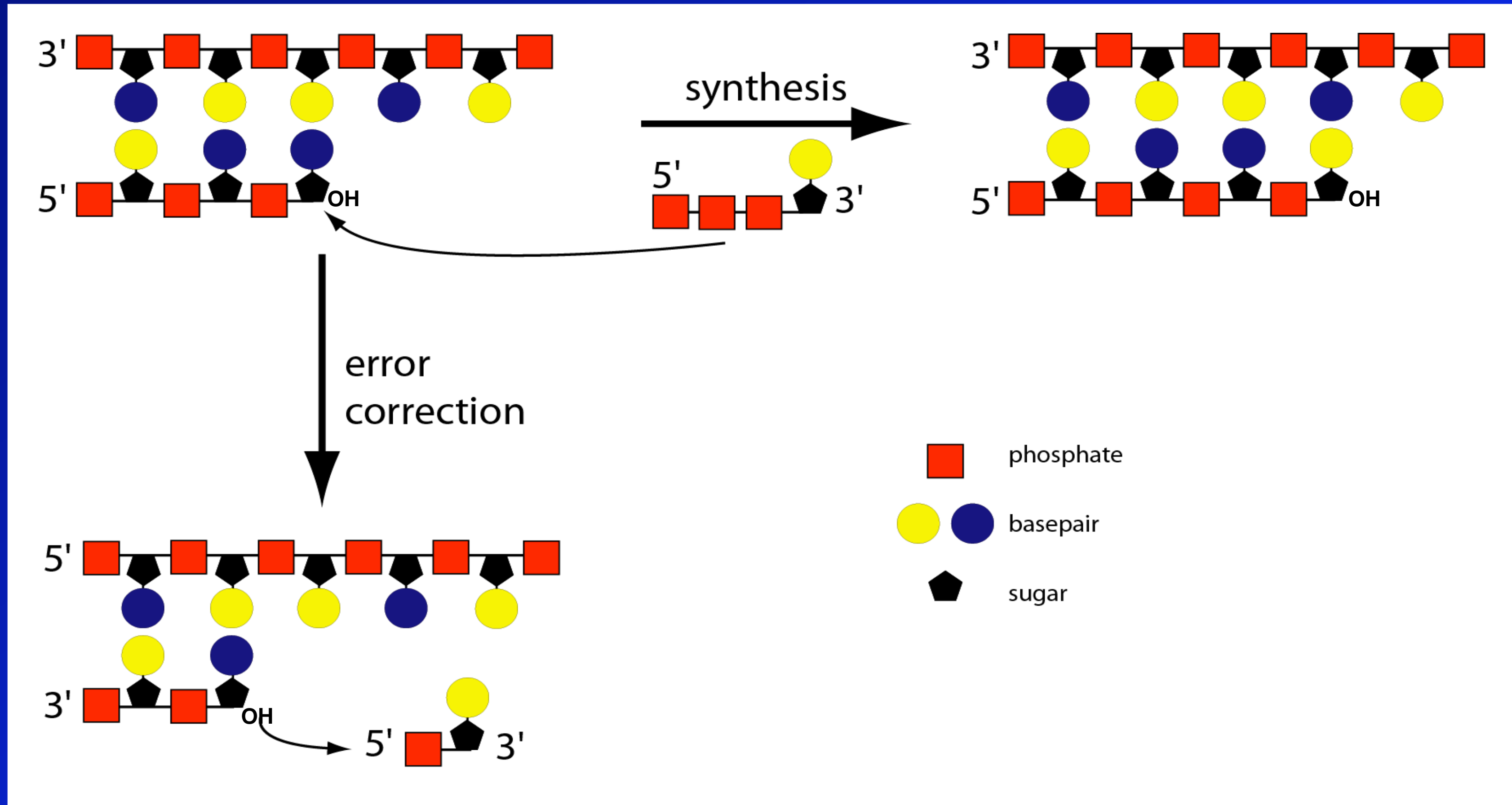
It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Semi-conservative replication

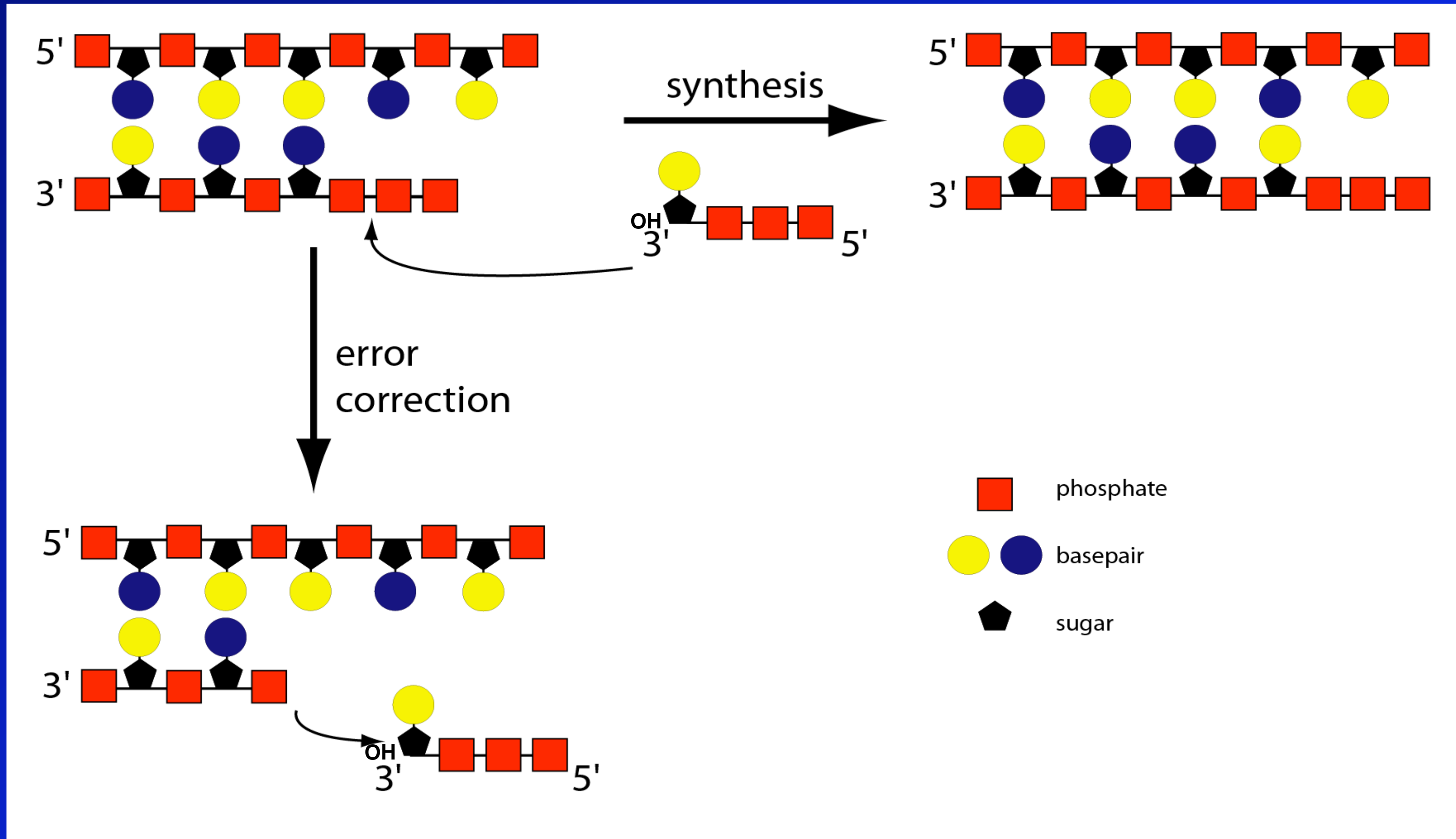
Copy each parental ssDNA strand
→ two daughter dsDNA strands



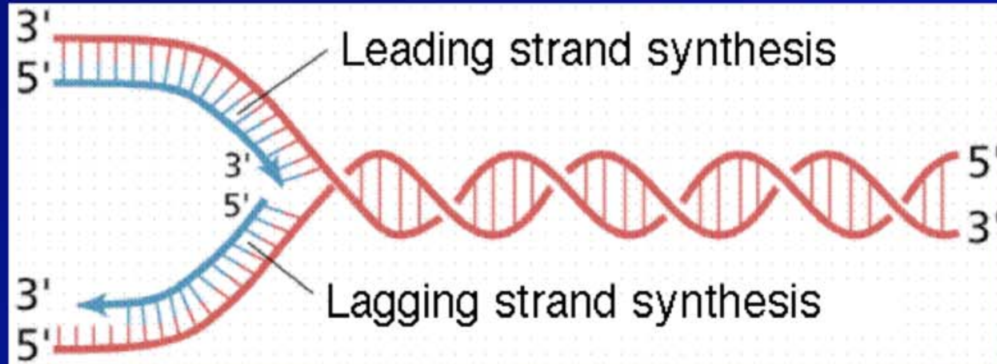
DNA is synthesized in 5' to 3' direction only

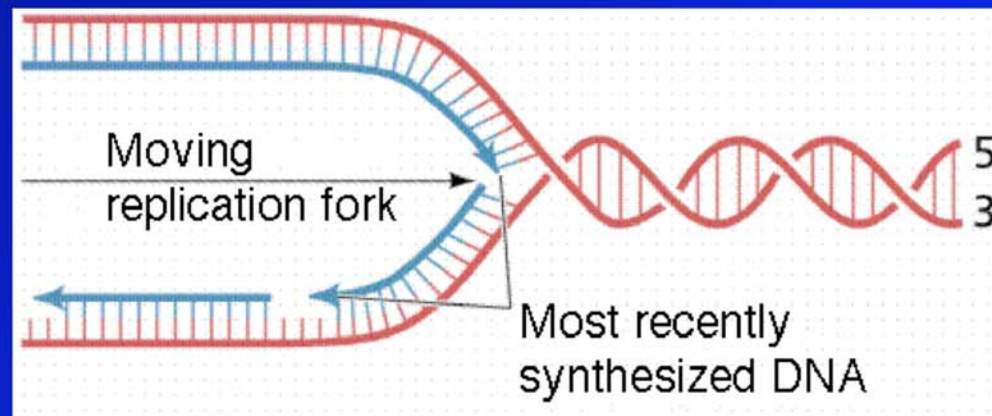


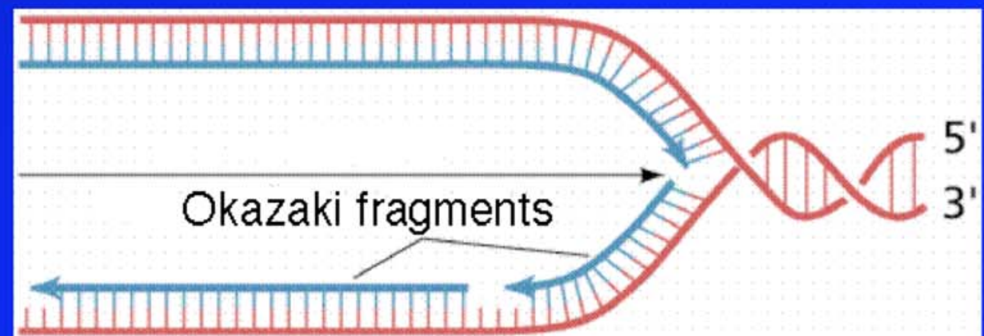
Synthesis in 3’ to 5’ direction does not allow for error correction



Asymmetry at the replication fork

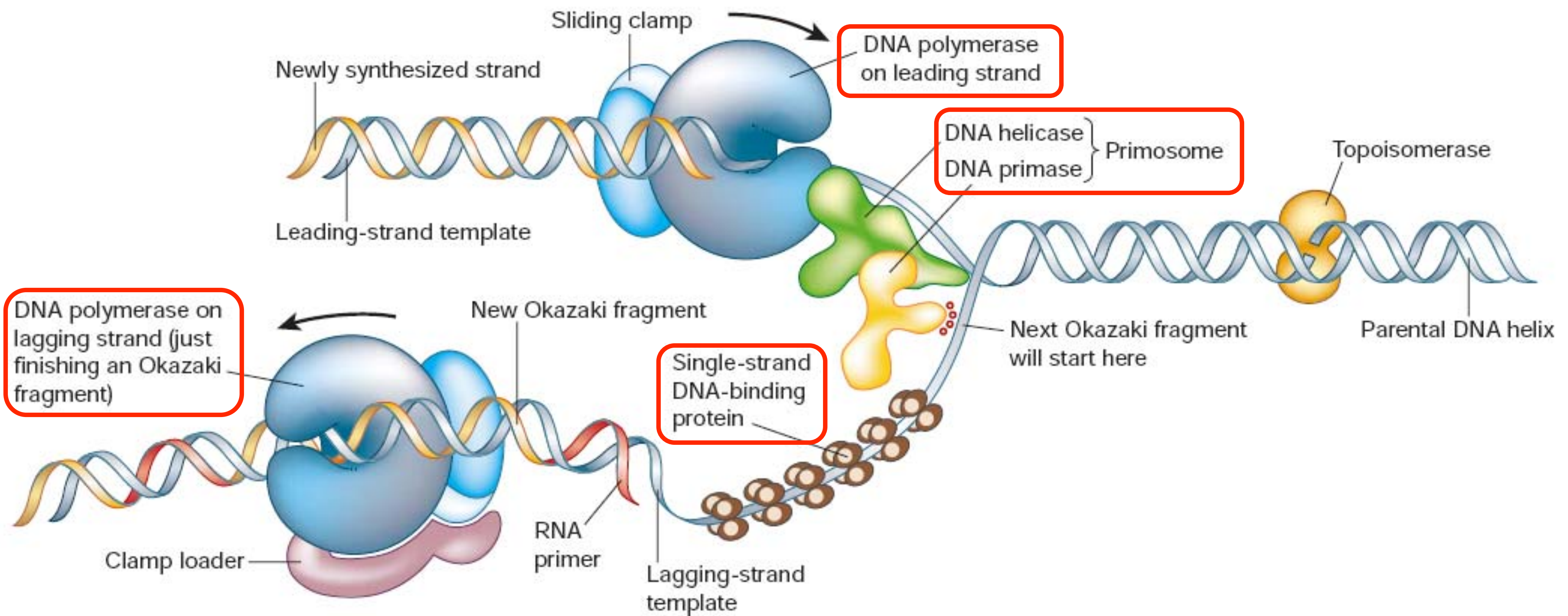




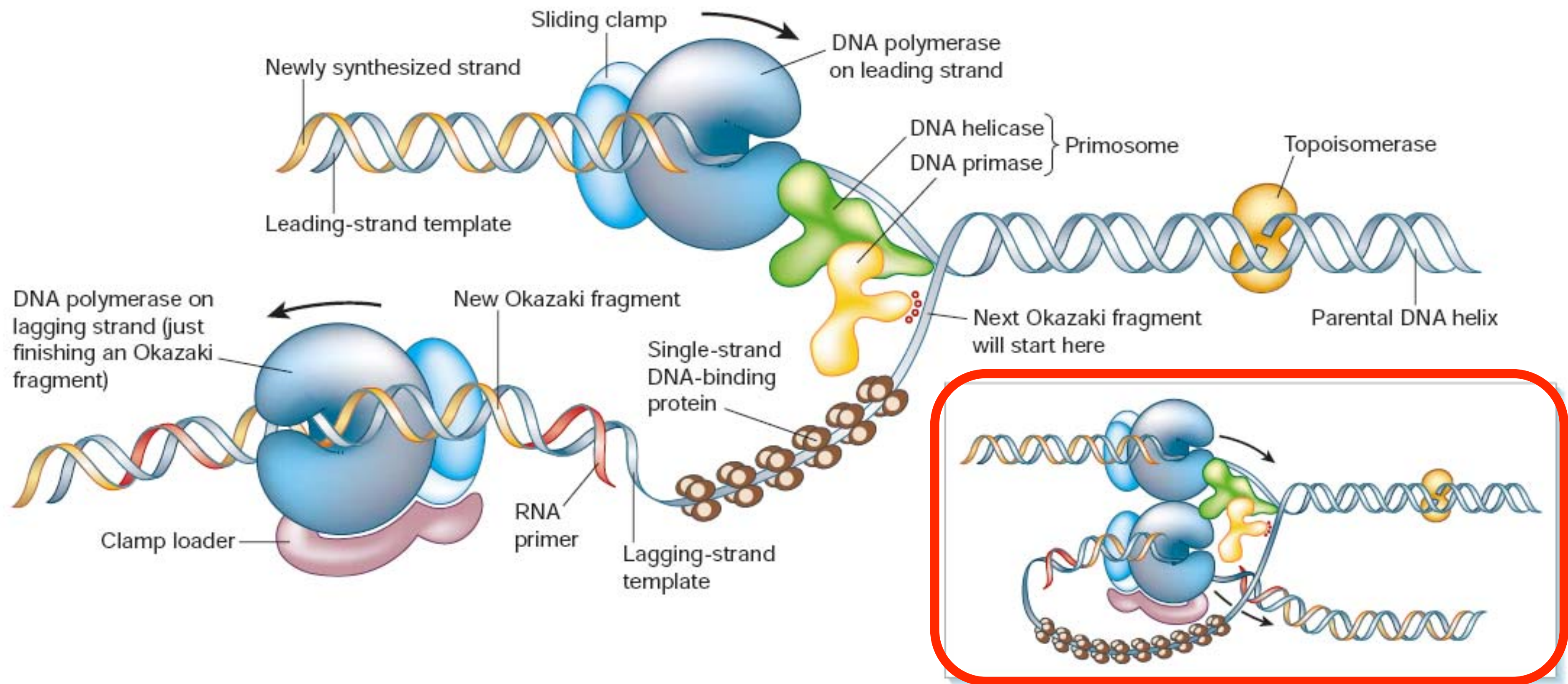


Adapted from Purvis *et al*., 'Life: the science of biology'

A generalized view of the replication fork



Formation of a replication loop



DNA Replication
(Camera: Back Left)

Duration: 0'18"
File Size: 1.2 MB
Contact: wehi-tv@wehi.edu.au

Questions

1) Is length Okazaki fragments regulated?

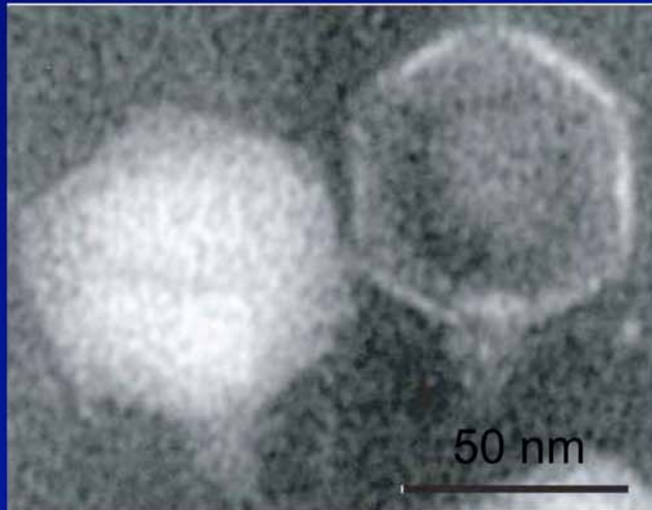
2) What triggers primer synthesis (and subsequent Okazaki fragment synthesis)?

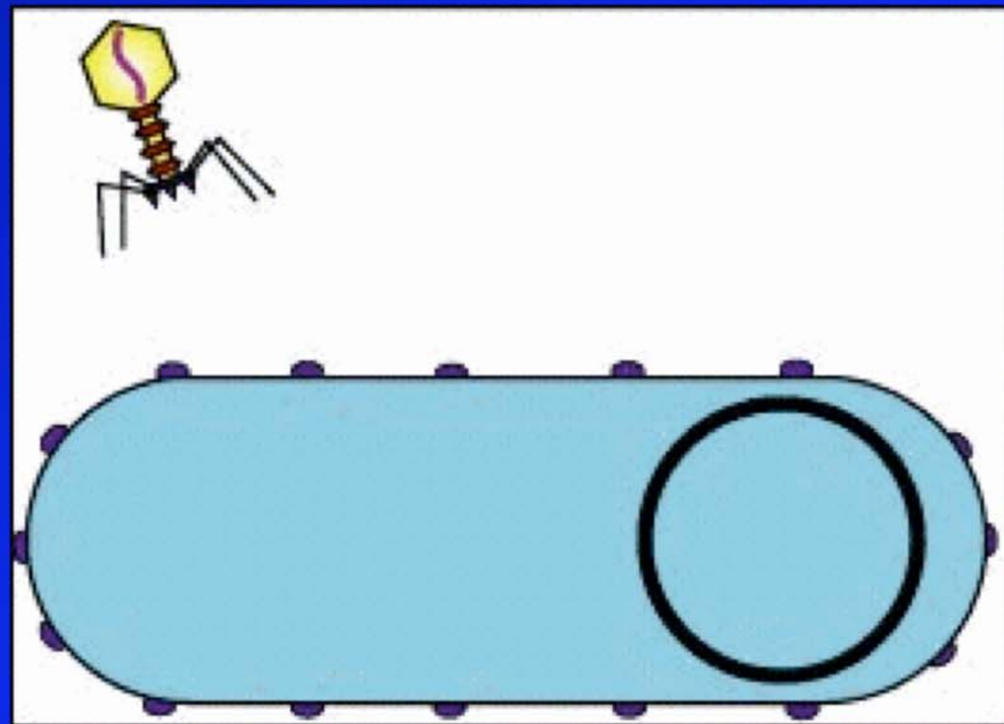
3) How is continuous leading-strand synthesis coupled with discontinuous synthesis on lagging strand?

Need to obtain timeline of (transient) events during replication cycle

Single-molecule studies

The bacteriophage T7





http://www.sp.uconn.edu/~terry/images/anim/phage.gif

The bacteriophage T7 replisome as a model system

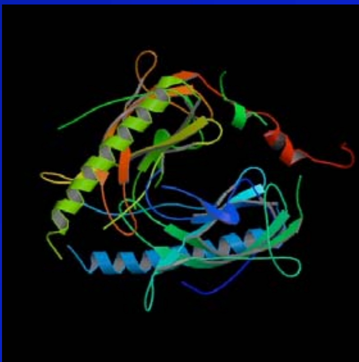
Singleton *et al.*, 2000

gp4: hexameric protein, encircling ssDNA containing both 5’-3’ helicase and primase activities

Doublie *et al.*, 1998

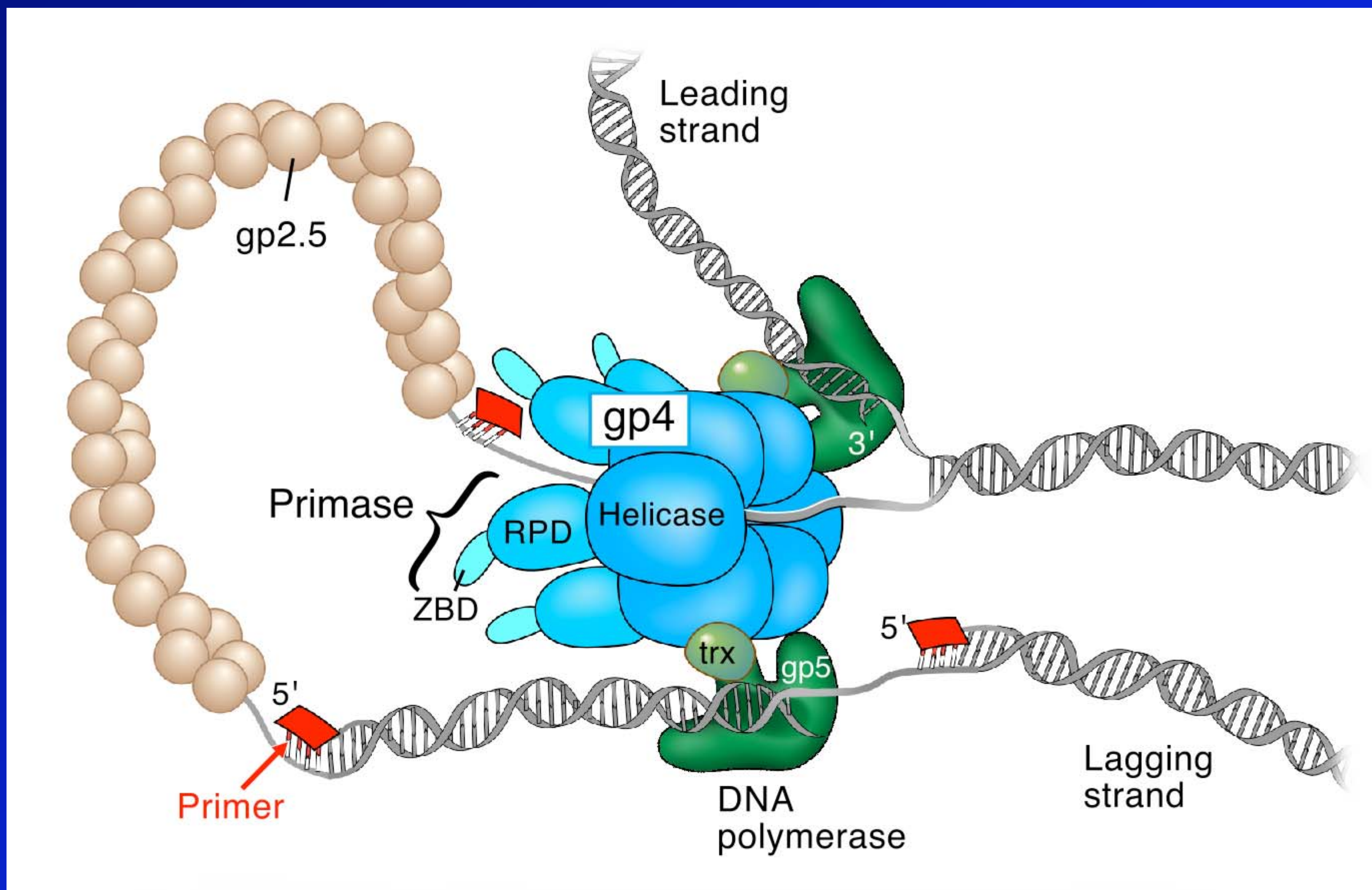
gp5: DNA polymerase, needs *E.coli* thioredoxin as cofactor to be processive

Hollis *et al.*, 2001

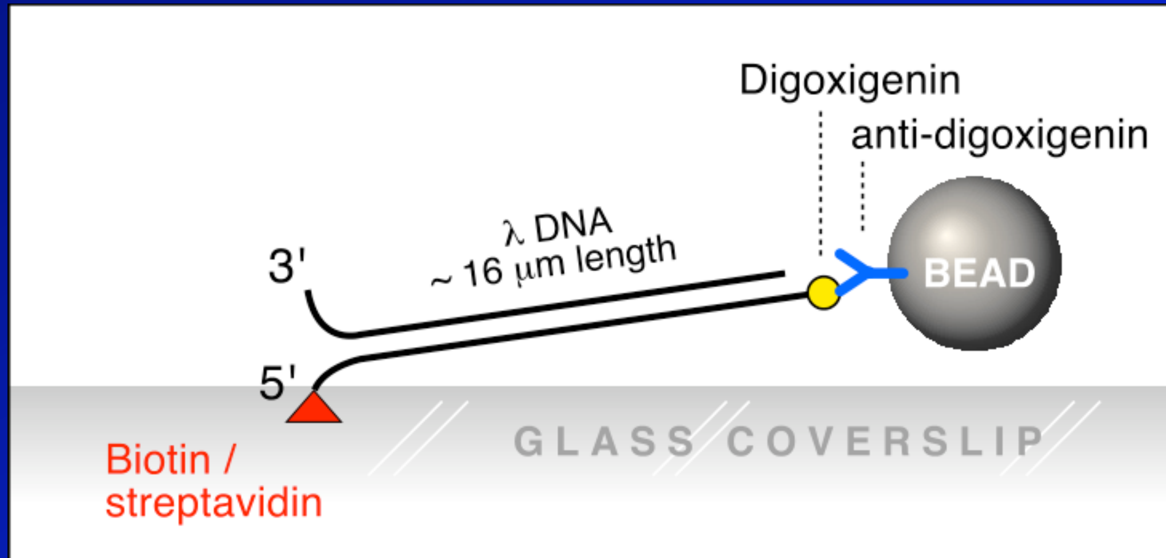


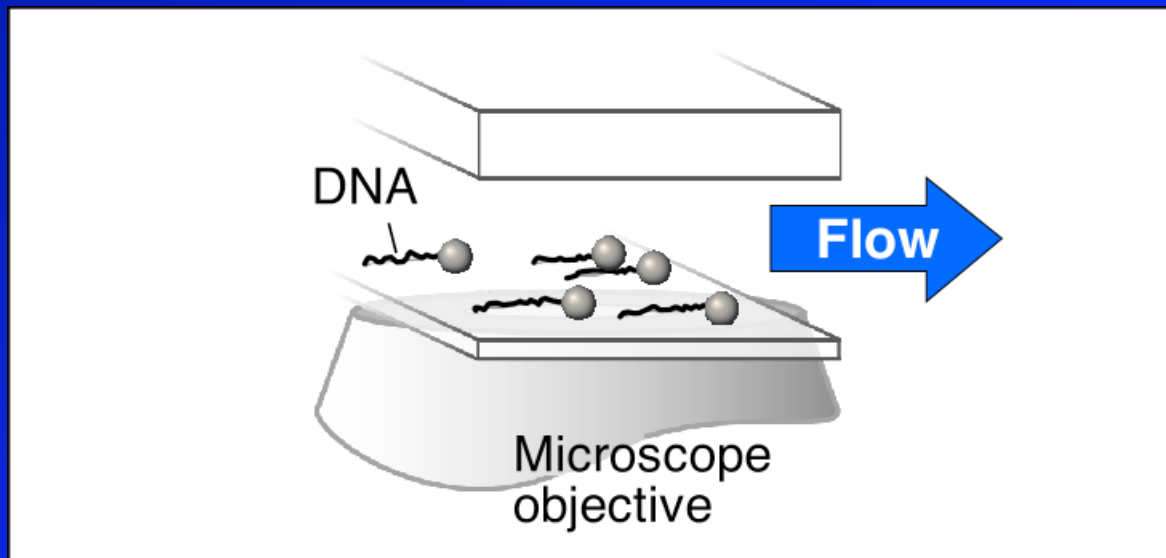
gp2.5: ssDNA-binding protein

The bacteriophage T7 replisome

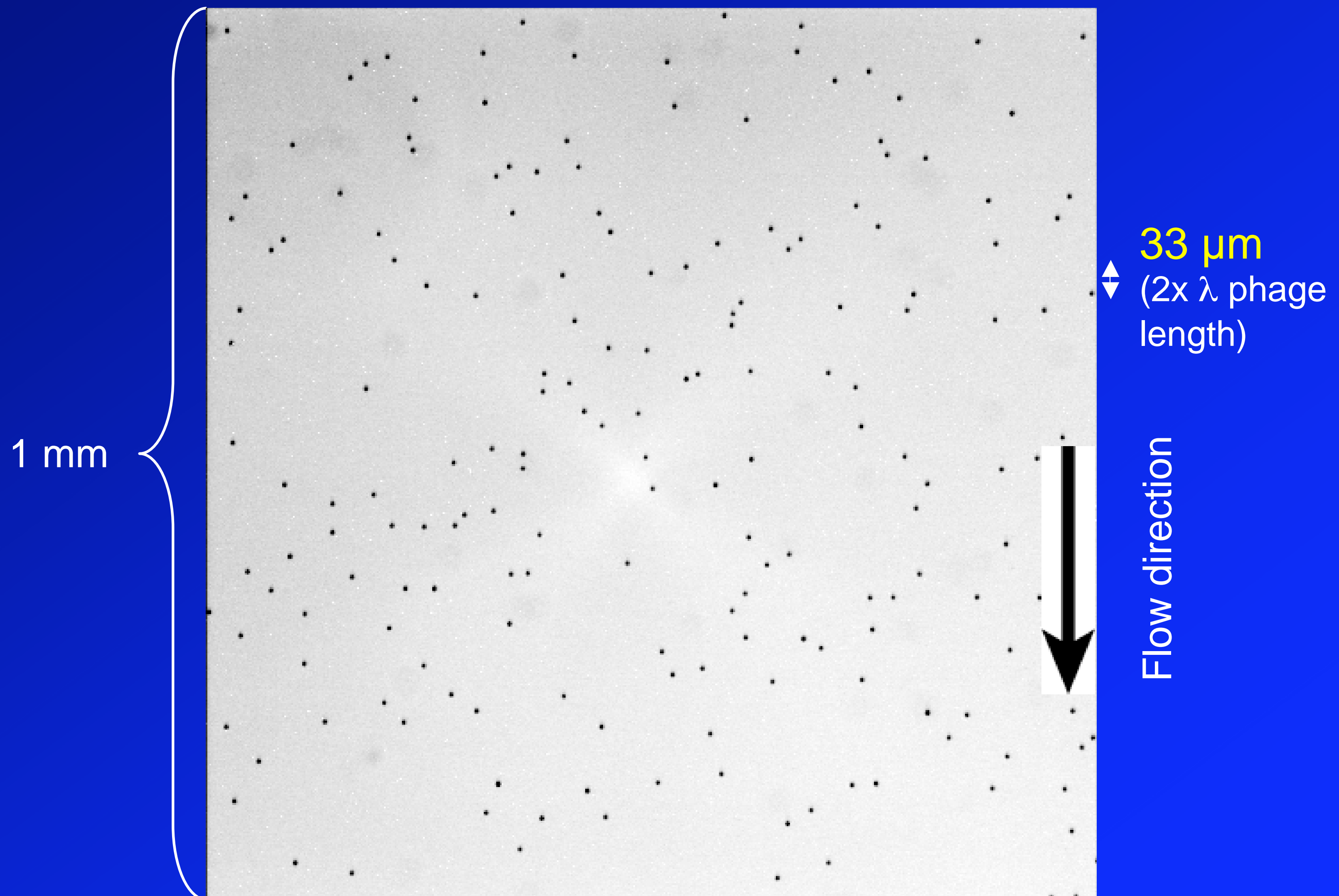


Flow stretching DNA molecules



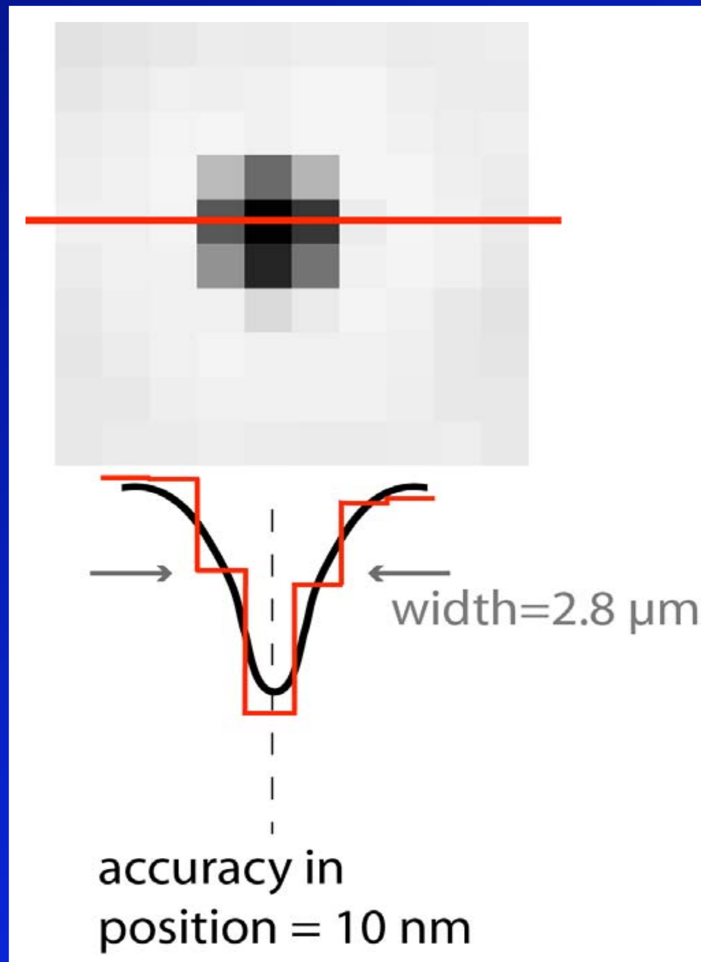


Flow stretching DNA molecules



Particle tracking

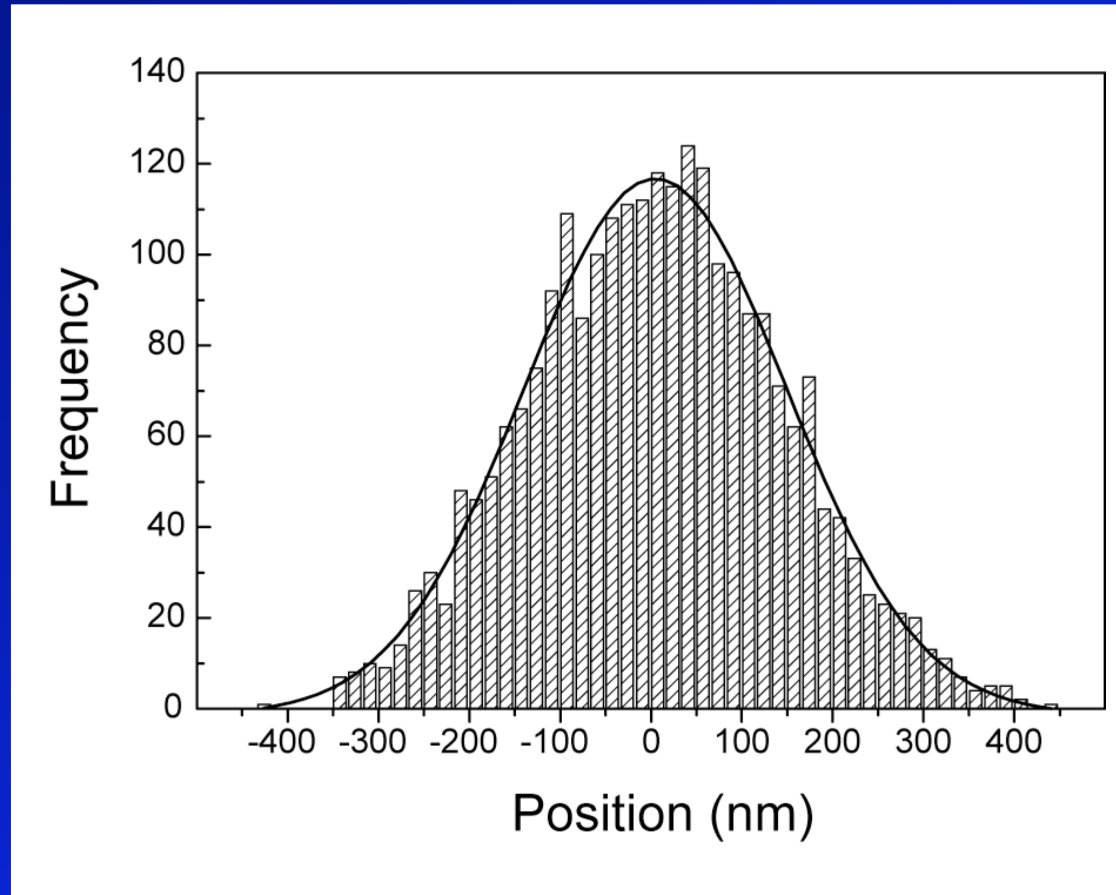
Bead positions can be found with high accuracy by fitting individual images with Gaussians



→ gives DNA length with 10 nm precision

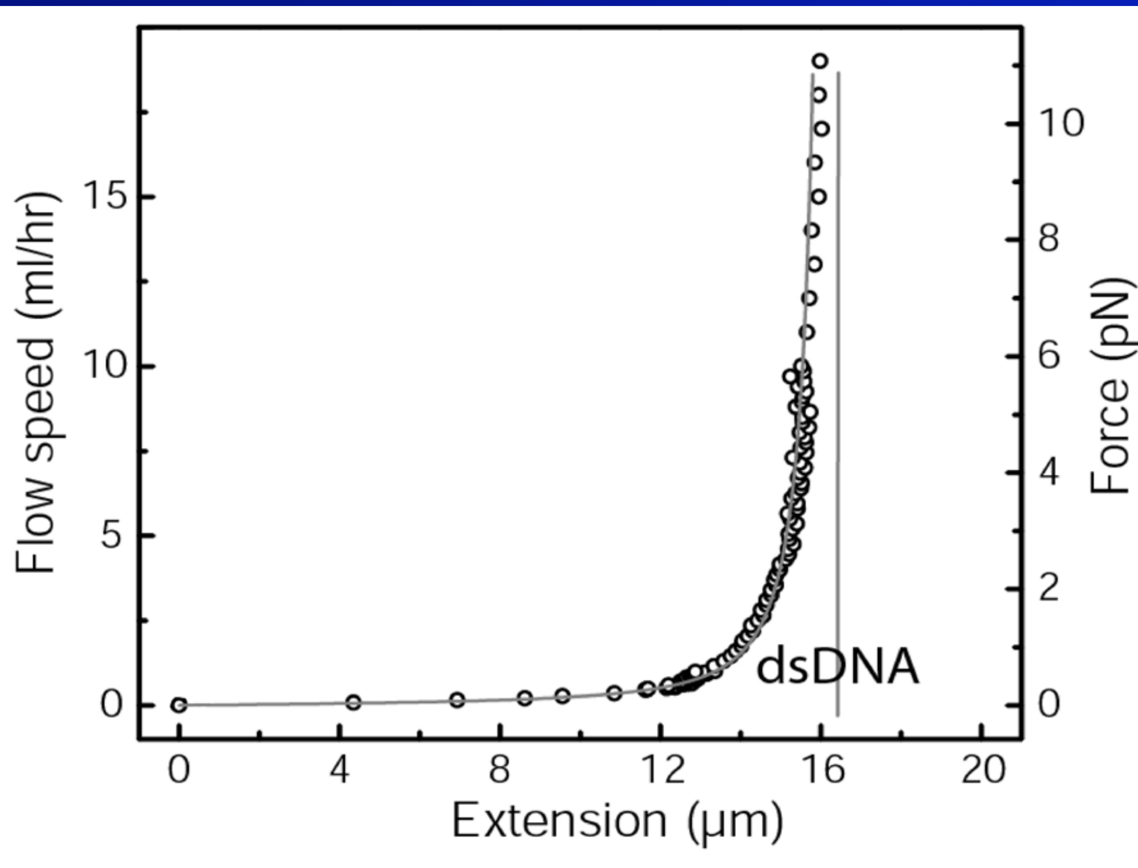
Force calibration

Displacement of bead perpendicular to stretching force



Obtain force through equipartition theorem: $F = k_B T\, L/\langle\delta x^2\rangle$

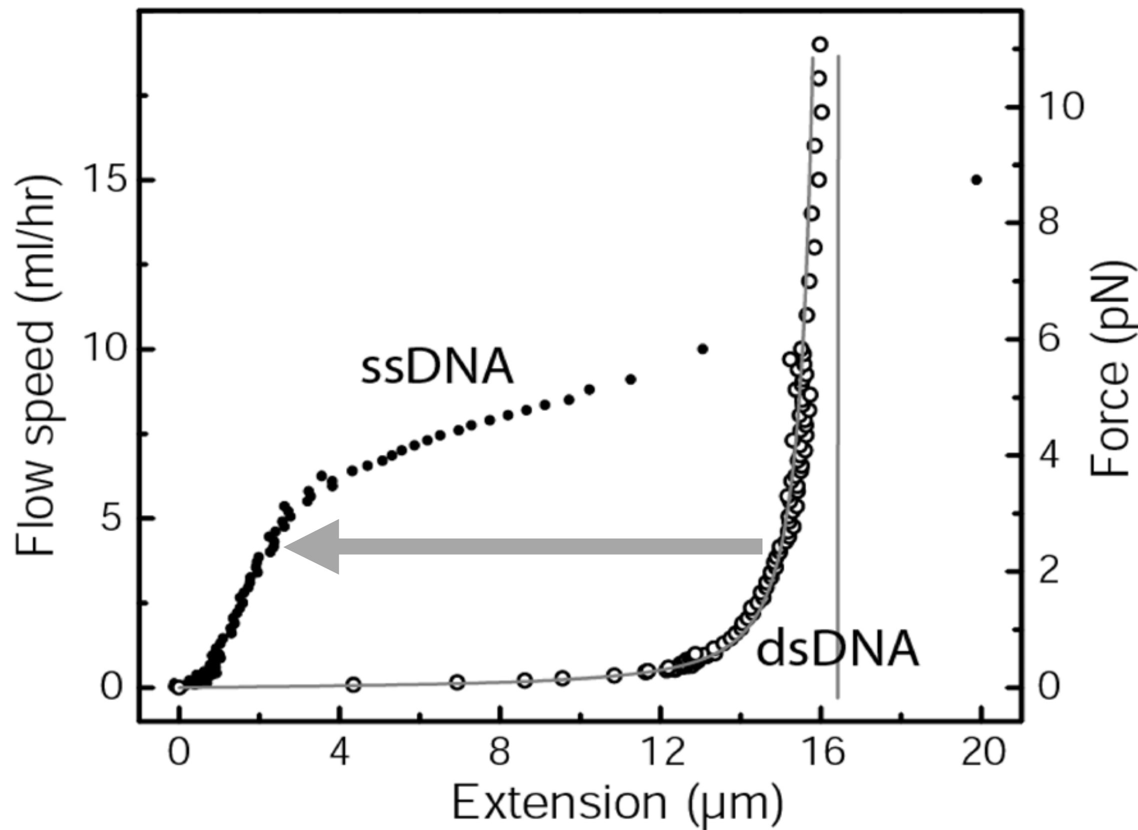
Drag force vs. extension of individual DNAs



dsDNA extension described by Worm-Like Chain model:

$$F = \frac{k_B T}{4\xi}\left\{\left(1-\frac{L}{L_0}\right)^{-2}+\frac{4L}{L_0}-1\right\}$$

Observing ssDNA ⟷ dsDNA conversions

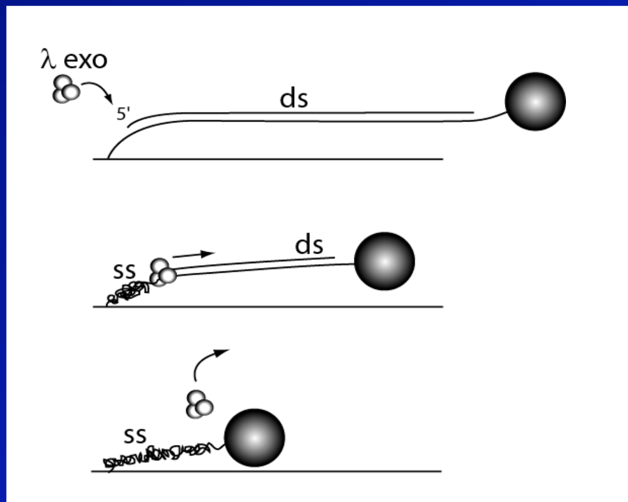


Elastic properties ss, ds DNA different

For $F_{stretch}$ < 6 pN:

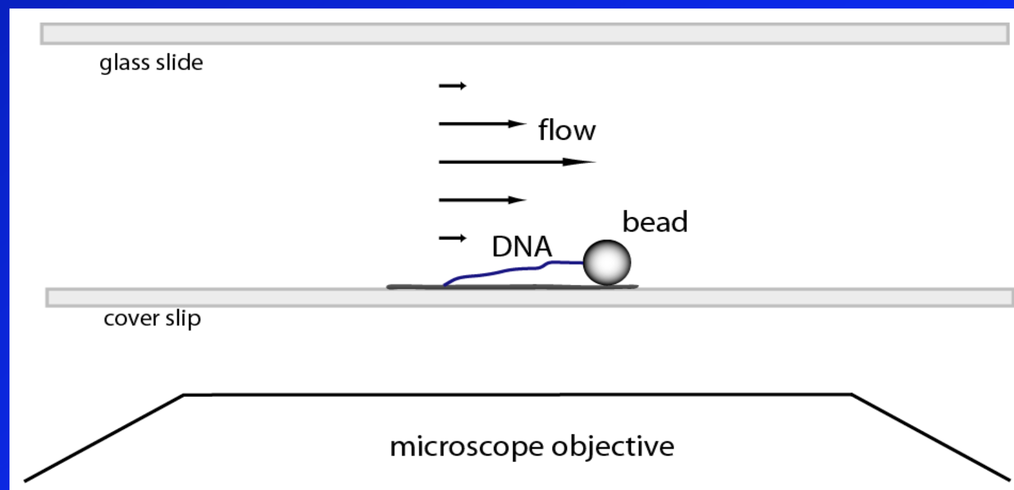
$l_{ss} < l_{ds}$

(see also Wuite *et al.*, 2000, and Maier *et al.*, 2000)

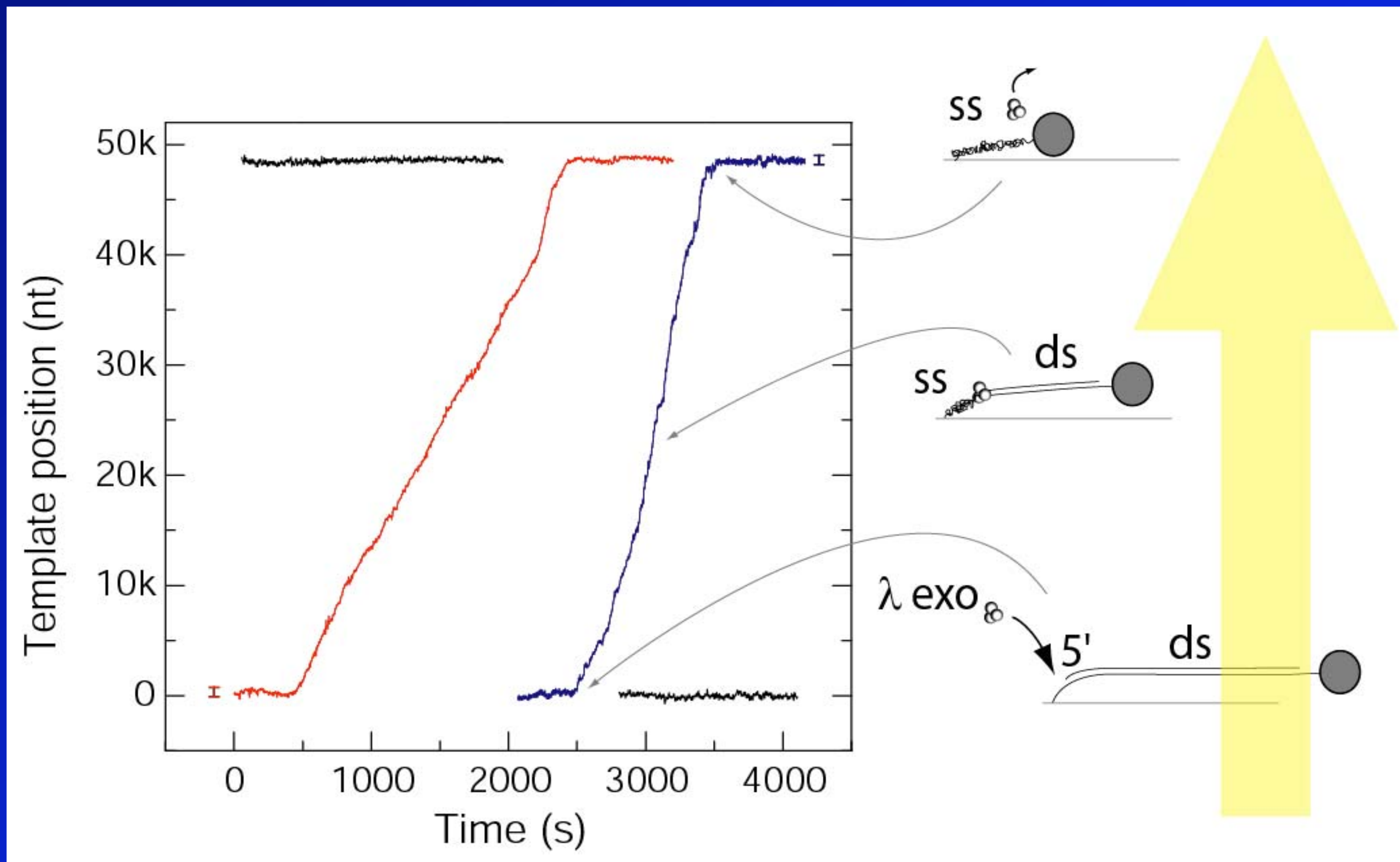


flow

Full-length dsDNA

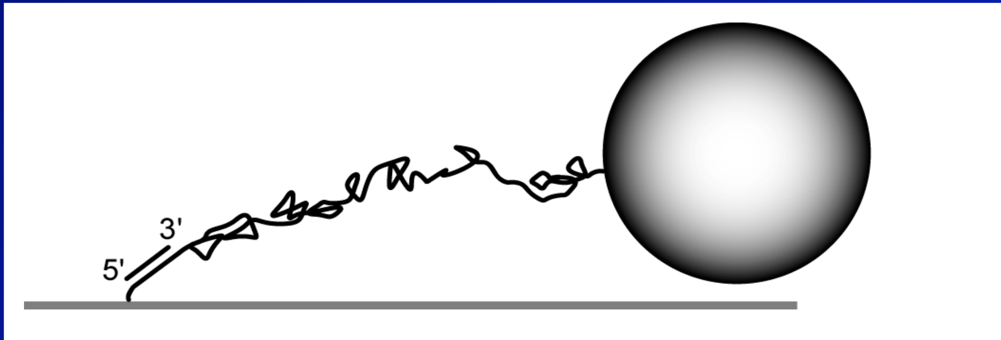


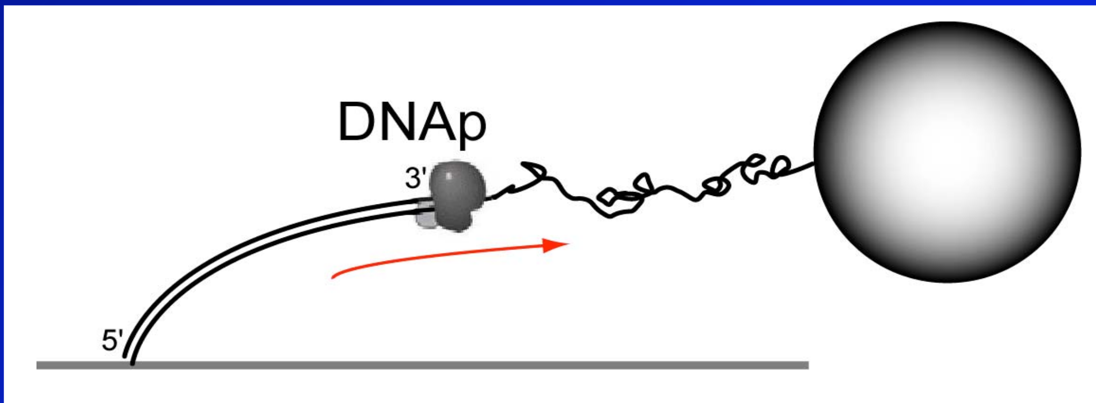
Single λ exonuclease activity



van Oijen *et al.*, Science 301 (2003) pp. 1235-8

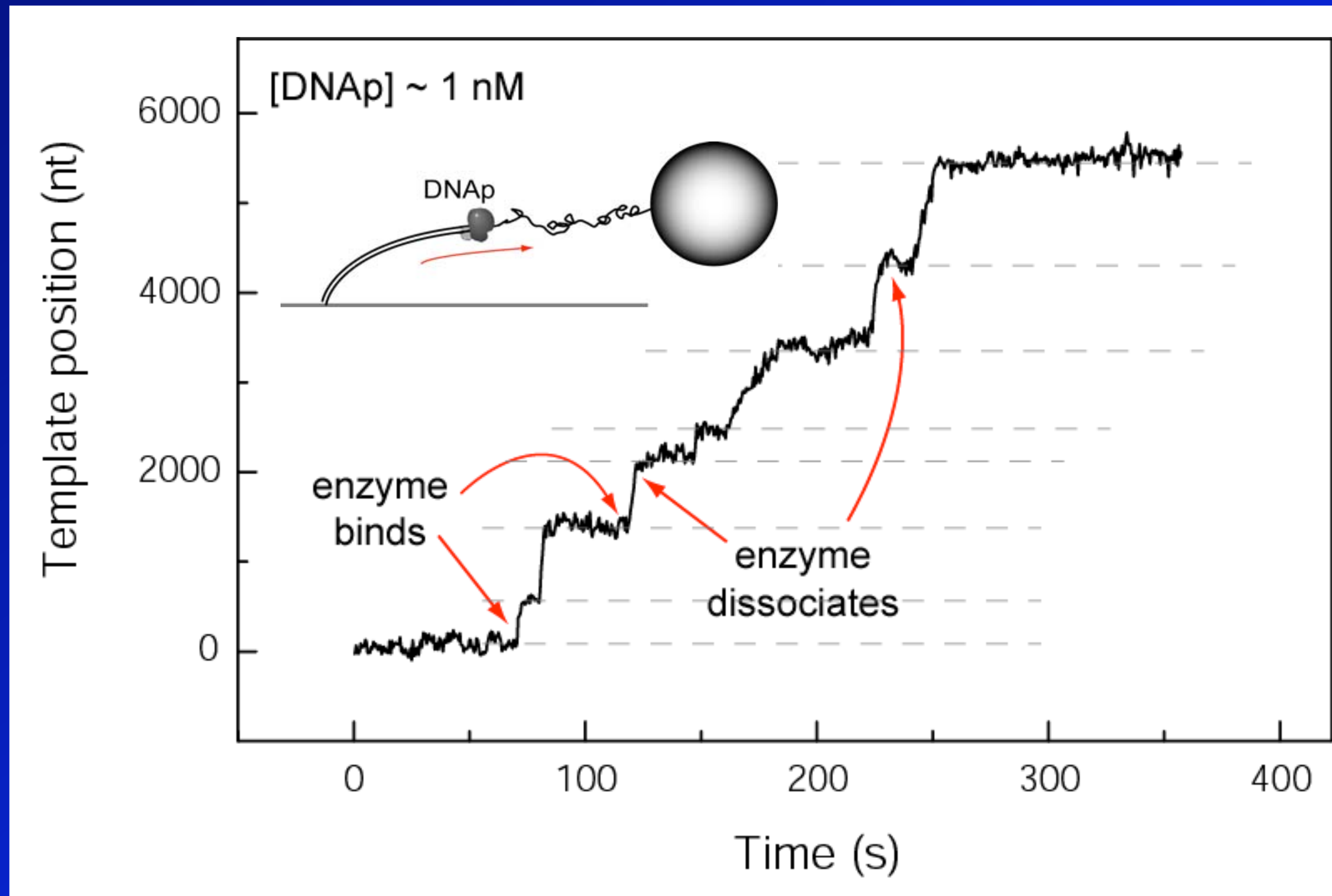
T7 DNA polymerase





ssDNA → dsDNA:
DNA polymerase
activity observable
as lengthening of DNA

T7 DNA polymerase



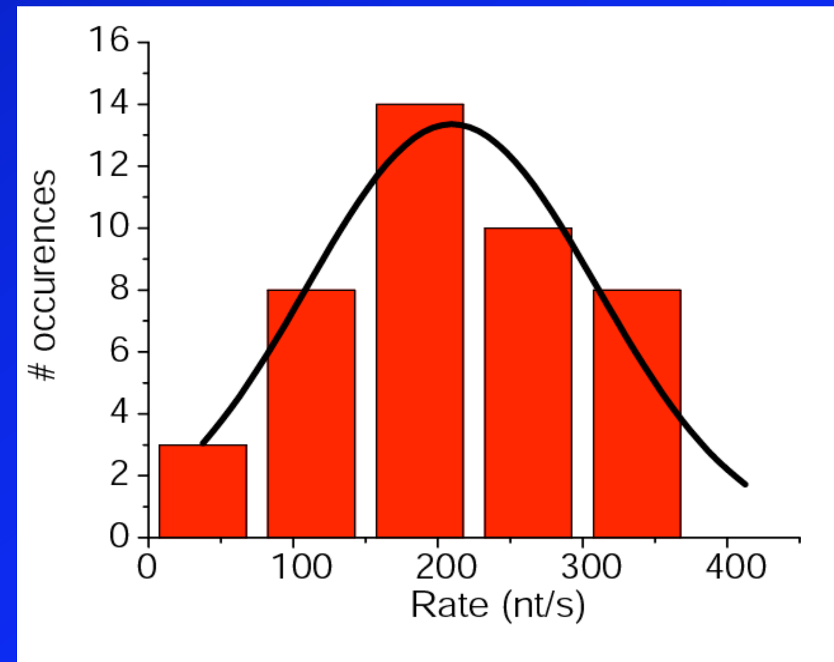
Enzymatic rate

How fast? → Rate distribution

Every datapoint is average rate of single enzyme:

Gaussian distribution (central limit theorem)

Average rate = 220 ± 80 nt/s



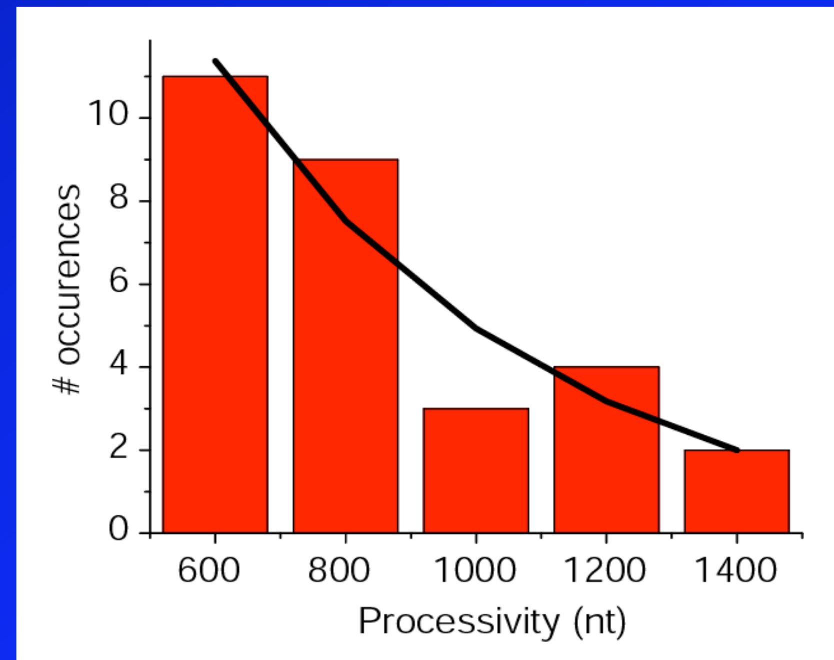
Enzymatic processivity

How long? → Processivity distribution

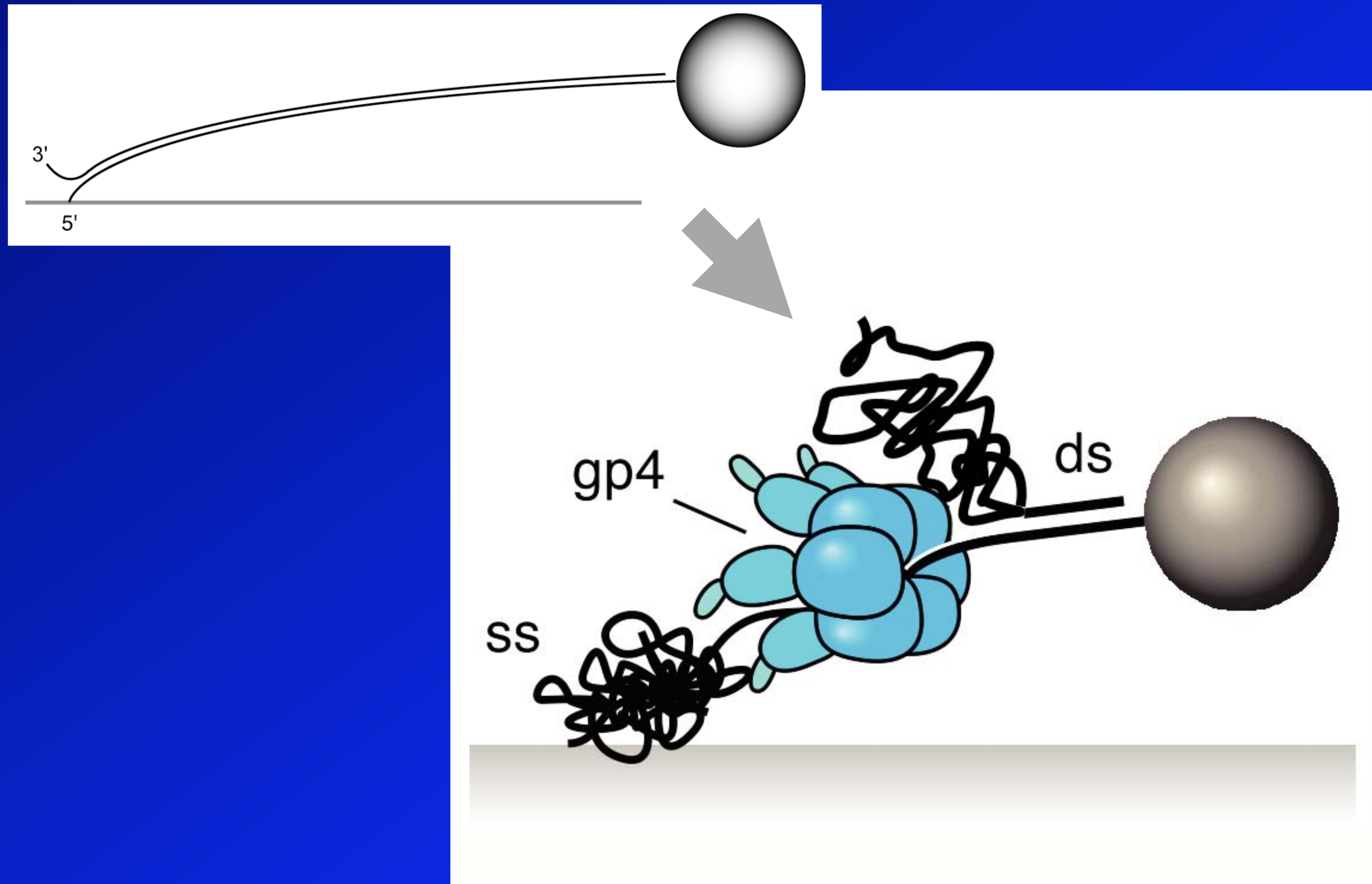
Every datapoint represents single dissociation rate single enzyme:

Exponential distribution

Average processivity = 700 ± 300 nt

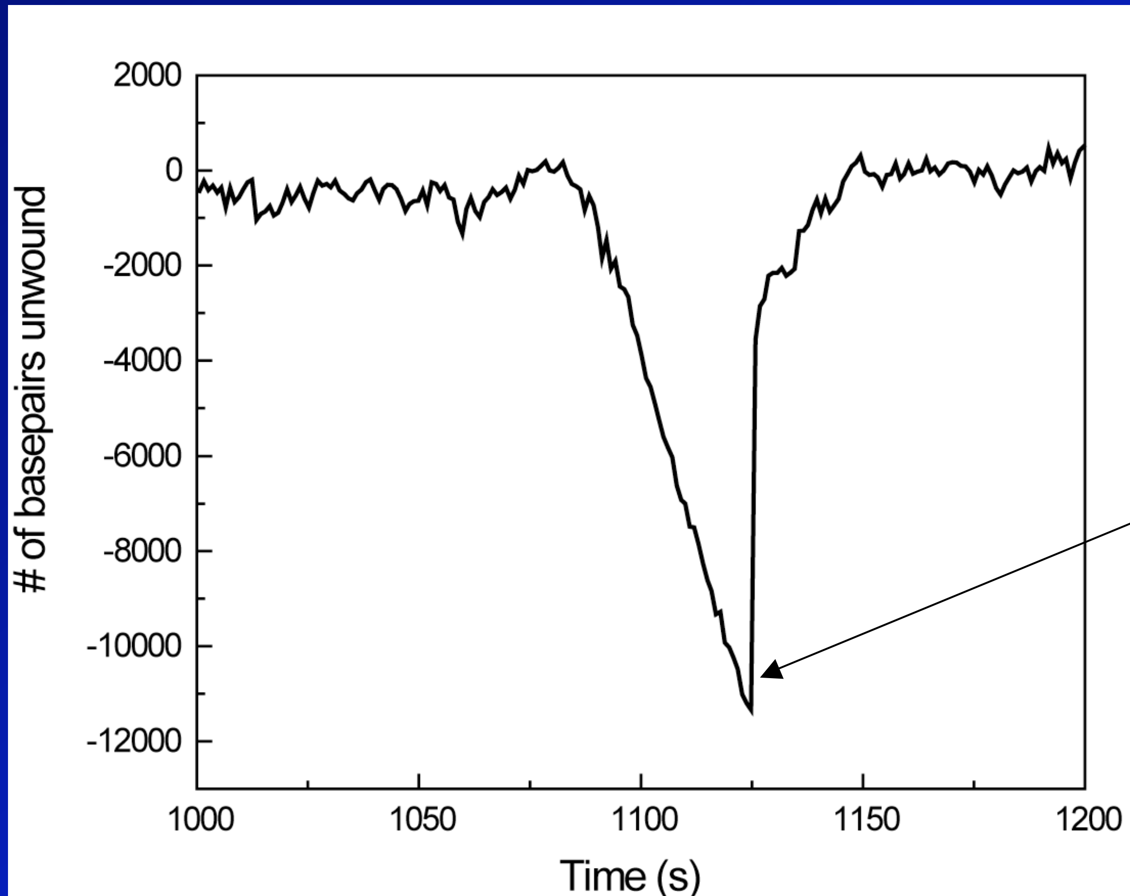


T7 helicase



dsDNA → ssDNA:
unwinding by helicase observable as shortening of DNA

T7 helicase



Two ssDNA strands re-anneal into dsDNA

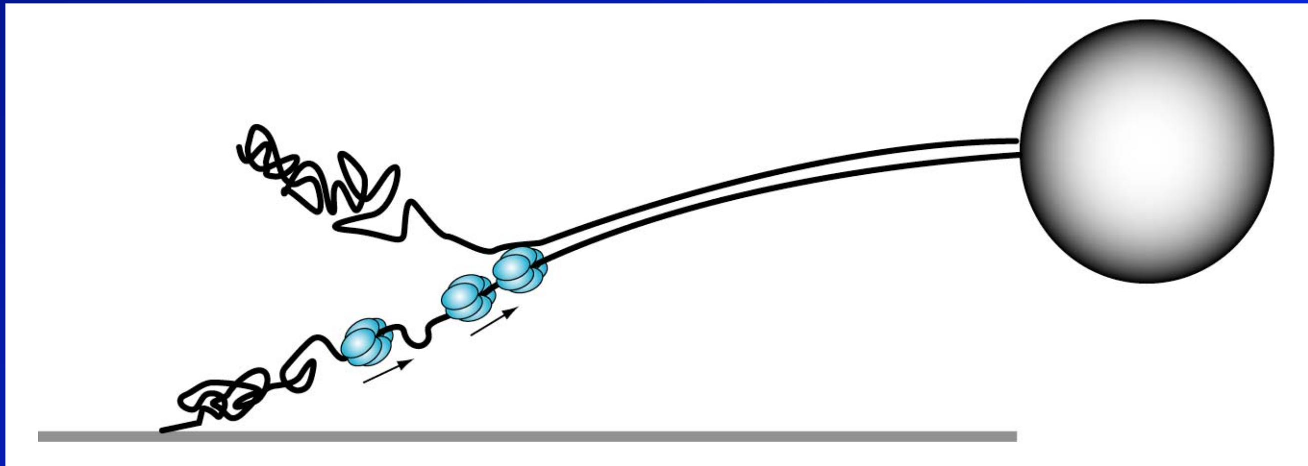
Average rate:
171 ± 11 nt/s

Average processivity:
5300 ± 600 nt

Rate much higher than measured in bulk…
(~ 70 bp/s; Patel *et al.*)

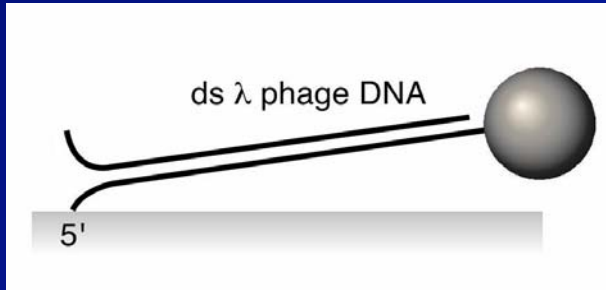
T7 helicase

Multiple enzymes in solution, multiple enzymes at fork?



How do we measure a true single-helicase event?

Single T7 helicases



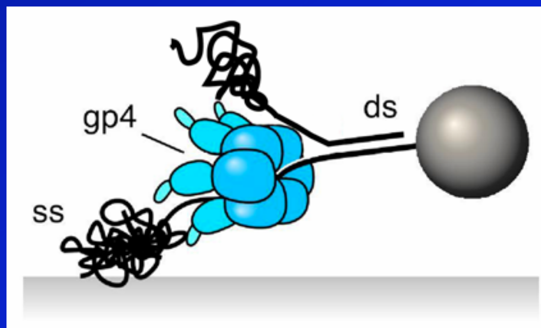
gp4, dTTP

Preassemble gp4 with dTTP

dTTP

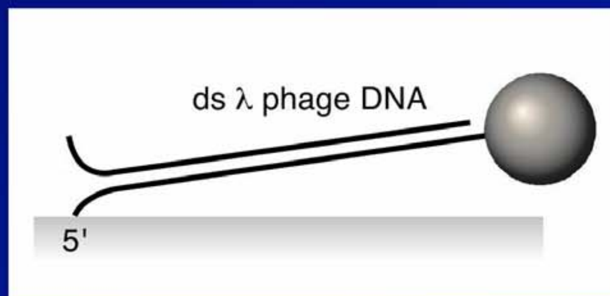
Wash out free enzymes

Mg^{2+}, dTTP



Pre-assemble gp4 hexamers in absence of Mg^{2+}, wash out all free enzymes and initiate unwinding reaction with Mg^{2+}.

Single T7 helicases



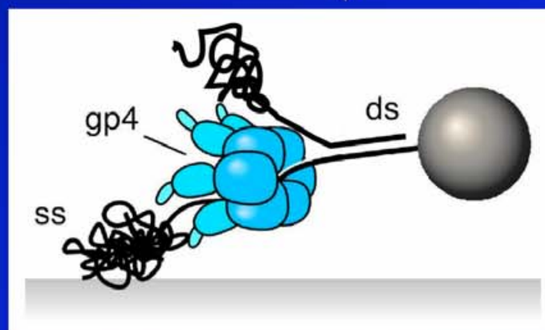
gp4, dTTP ↓

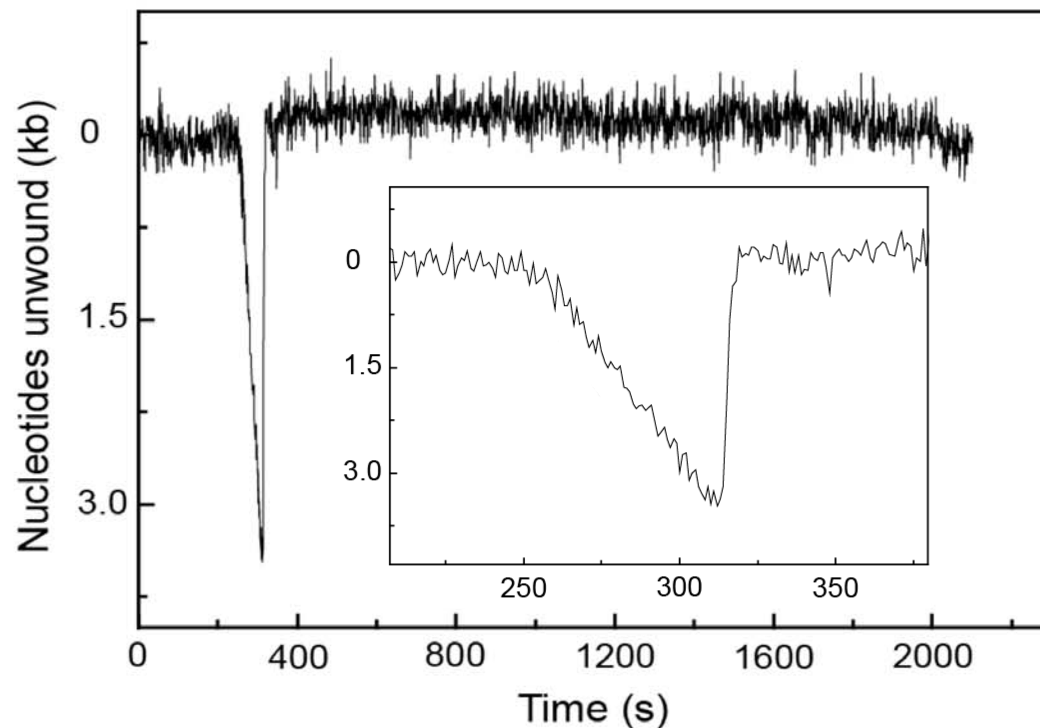
Preassemble gp4 with dTTP

dTTP ↓

Wash out free enzymes

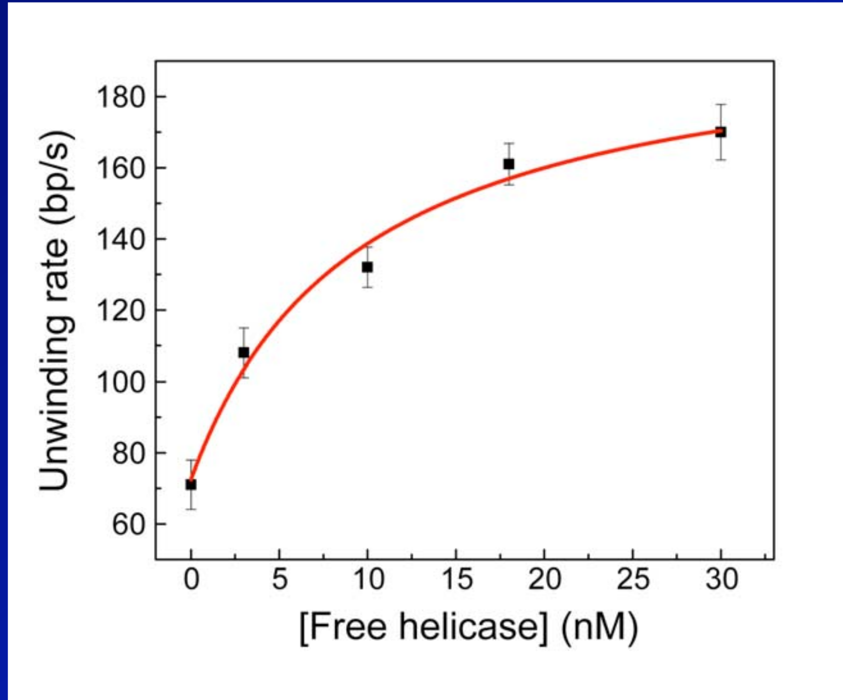
Mg^{2+}, dTTP ↓



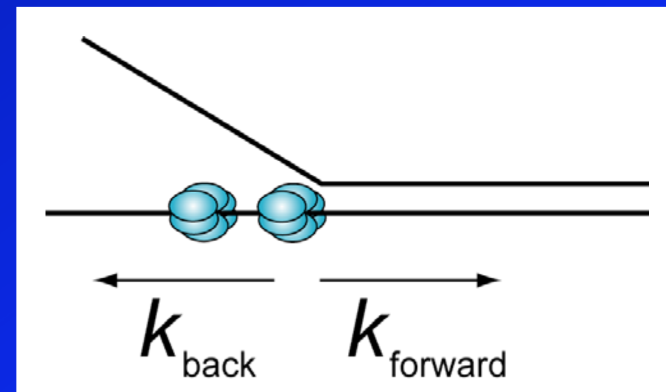


Average unwinding rate : 72 ± 11 bp/s

T7 helicase



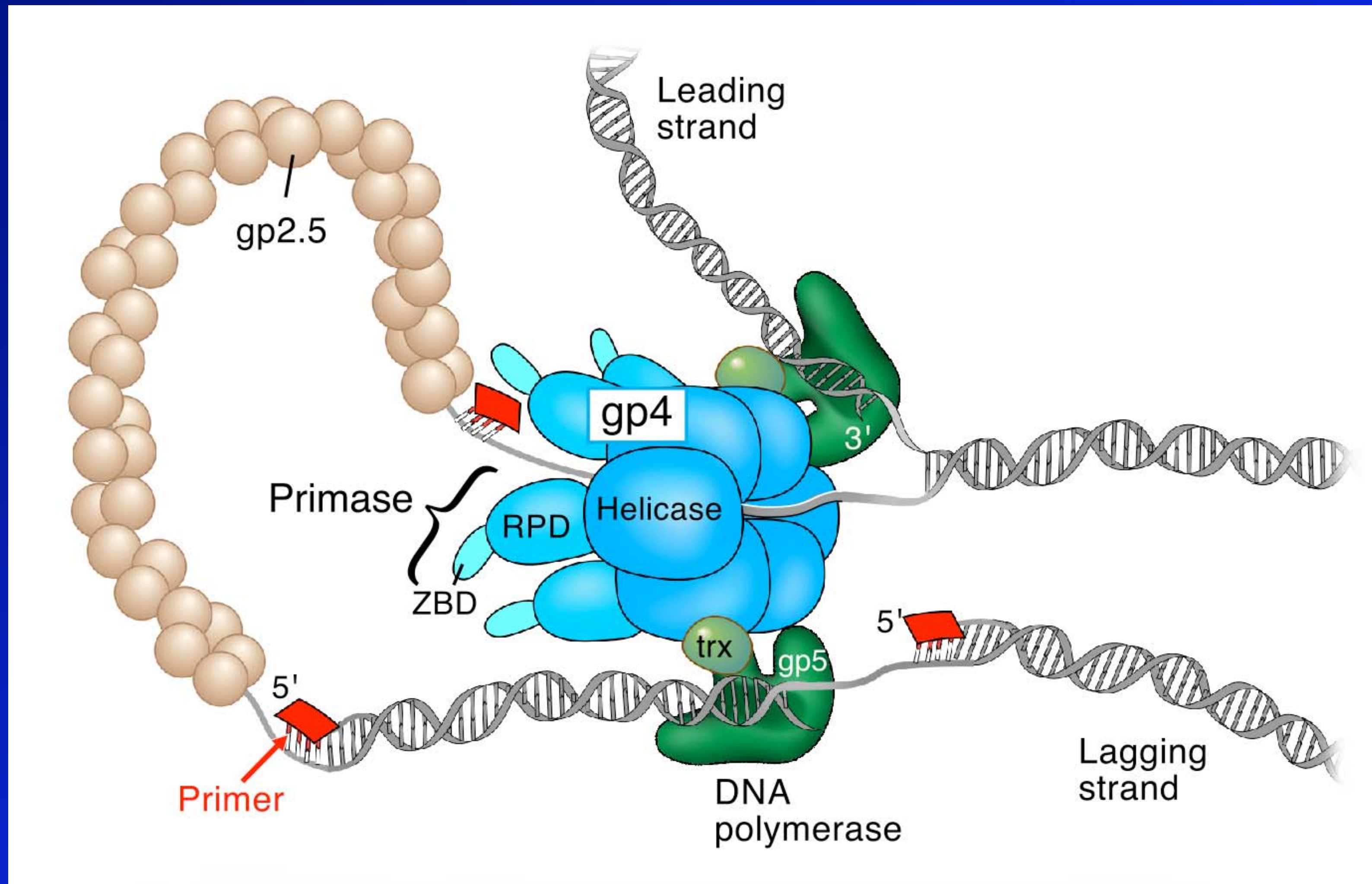
Unwinding rate of individual helicases depends on free enzyme concentration!



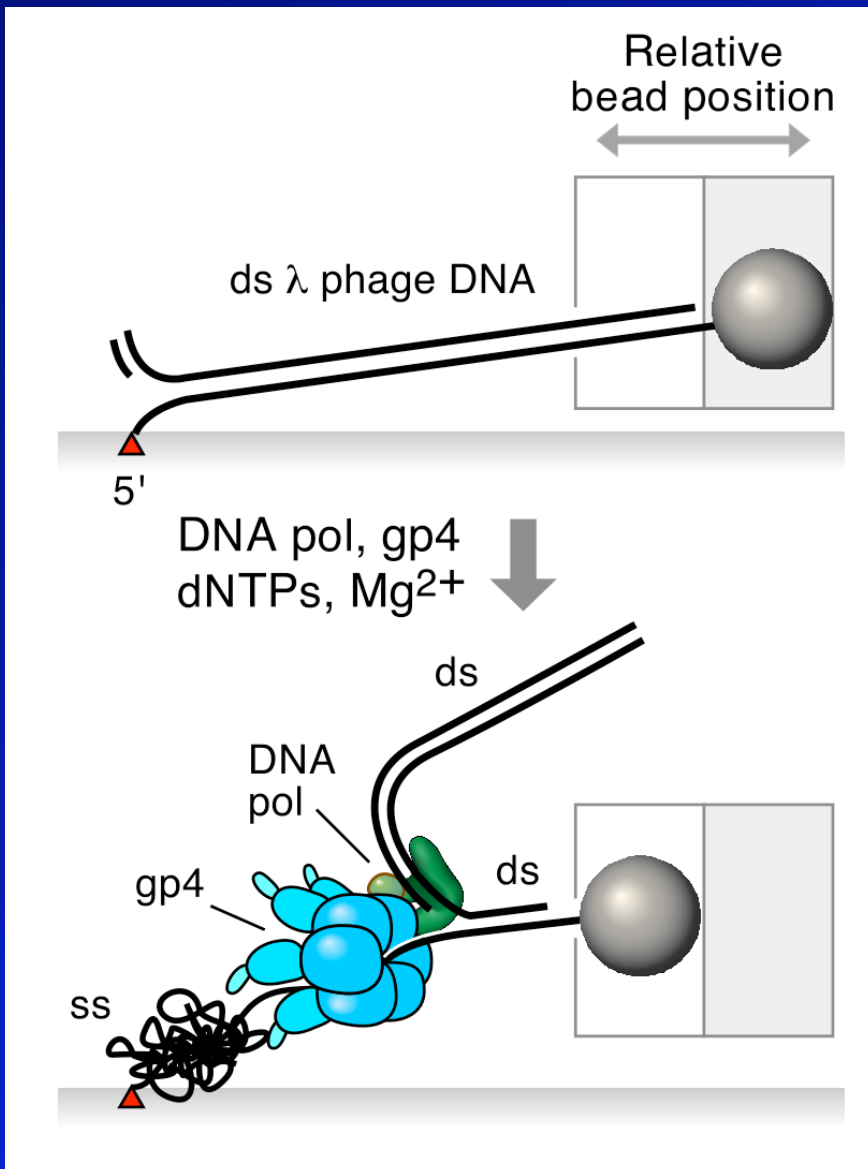
Presence of other helicases behind unwinding helicases reduces k_{back} (Brownian Ratchet)

→ Speeds up unwinding ($=k_{forward}-k_{back}$)

Towards a complete replisome: leading-strand synthesis



DNA polymerase + helicase



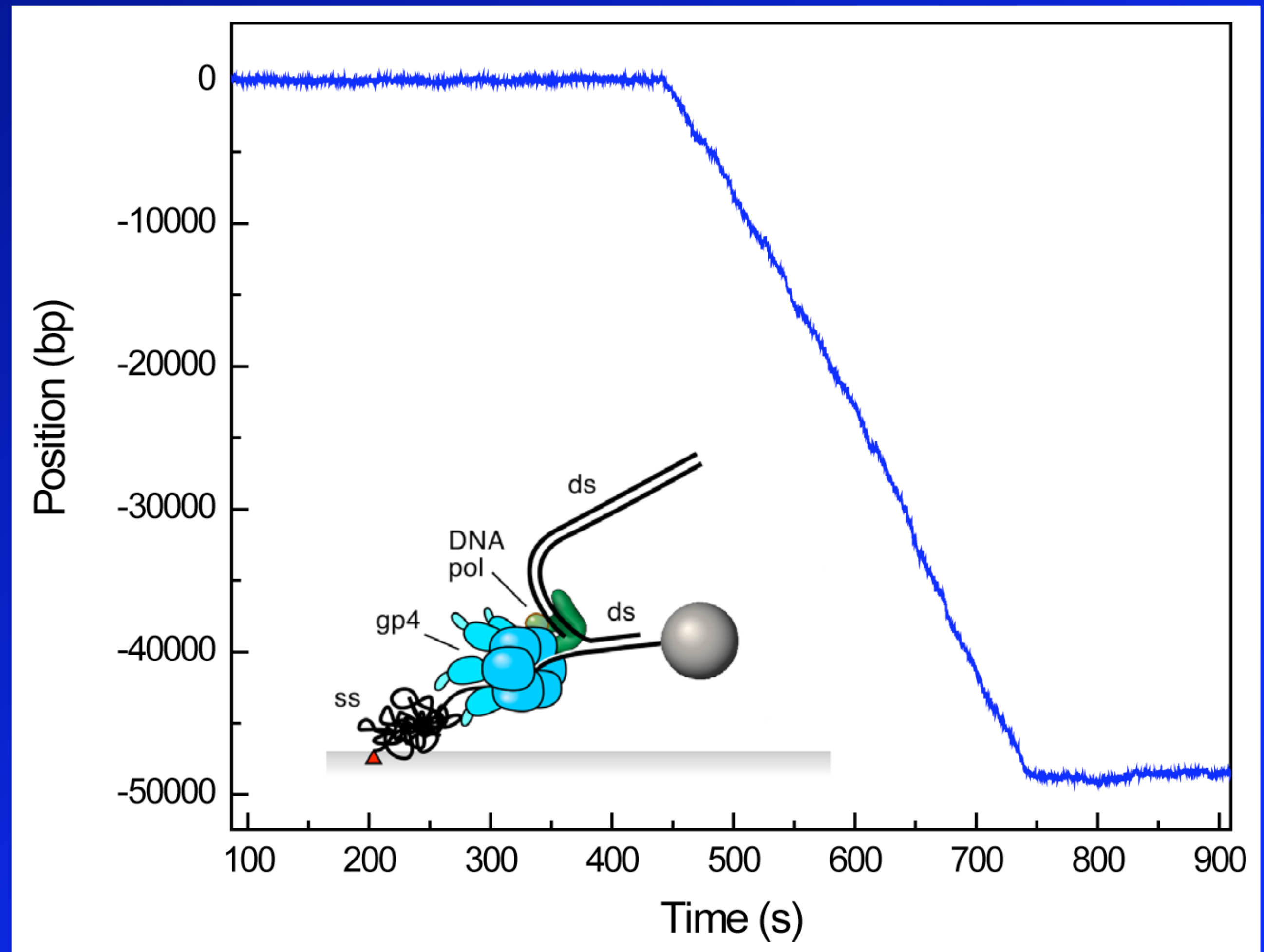
DNA polymerase + helicase only replicate ‘leading’ strand;
Lagging strand will go from dsDNA → ssDNA:

observable as shortening of DNA

DNA polymerase + helicase: T7 leading-strand synthesis

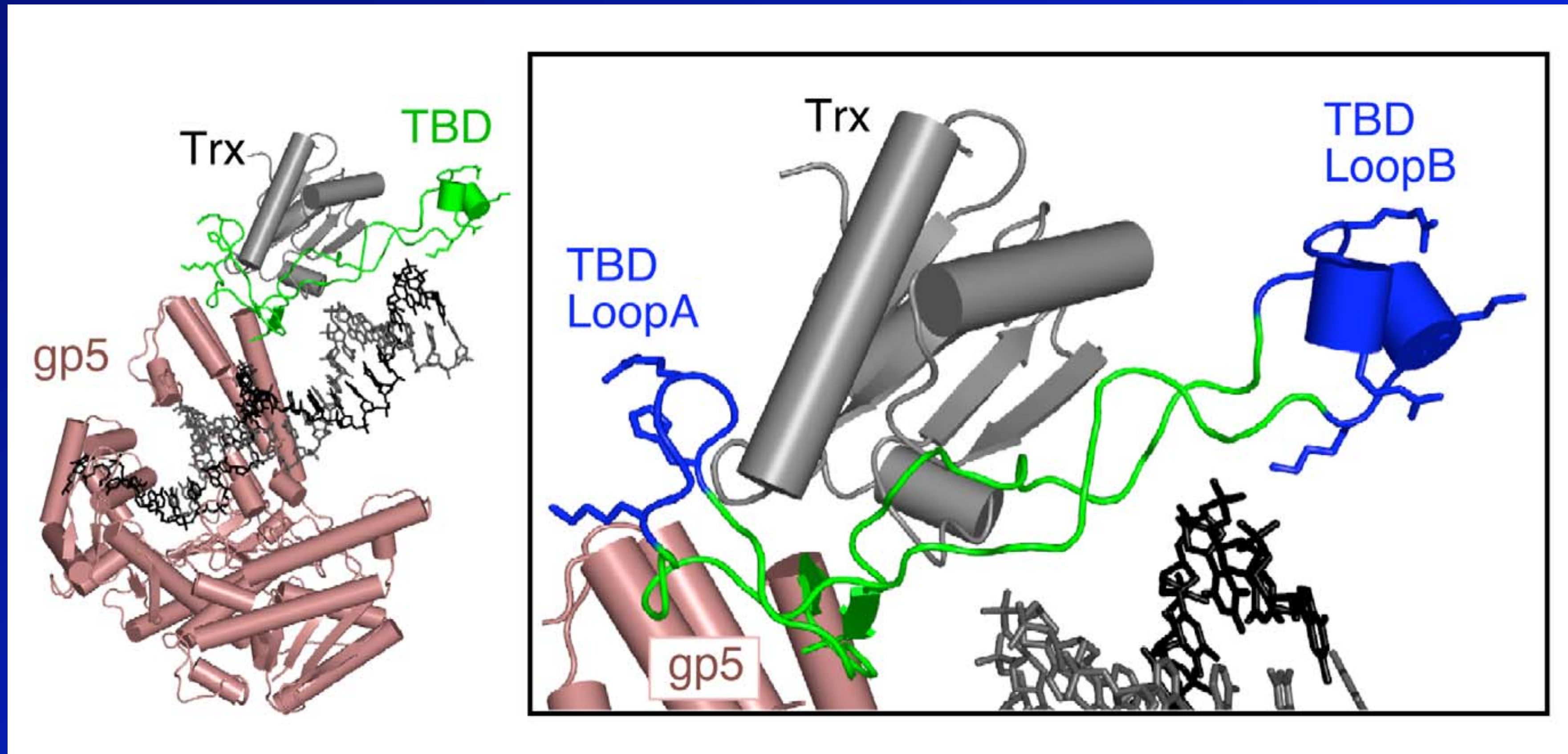
Average rate:
162 ± 36 nt/s

Average processivity:
16,000 ± 4,000 nt



→ Dramatic increase in processivity!

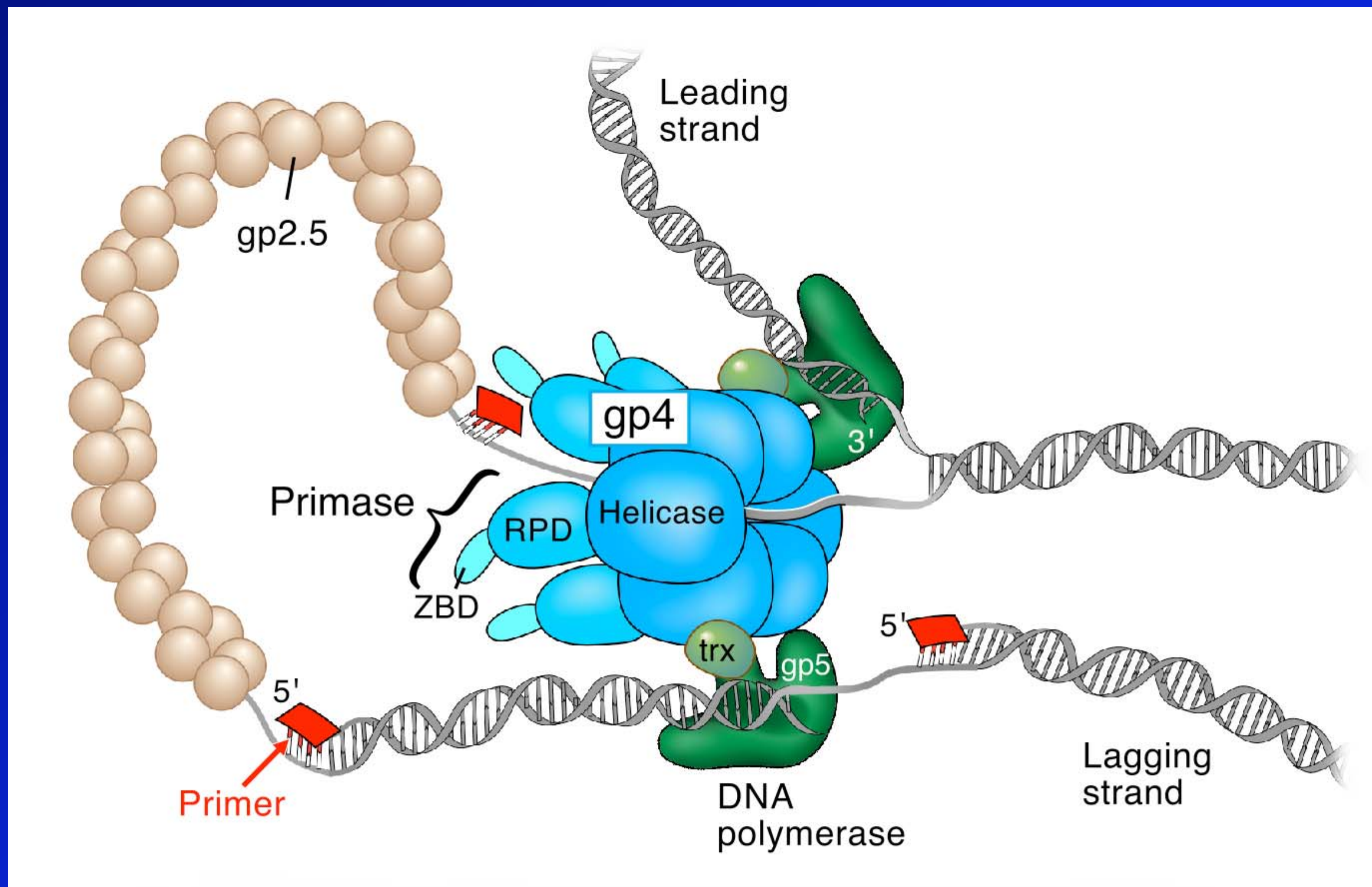
DNA polymerase structure



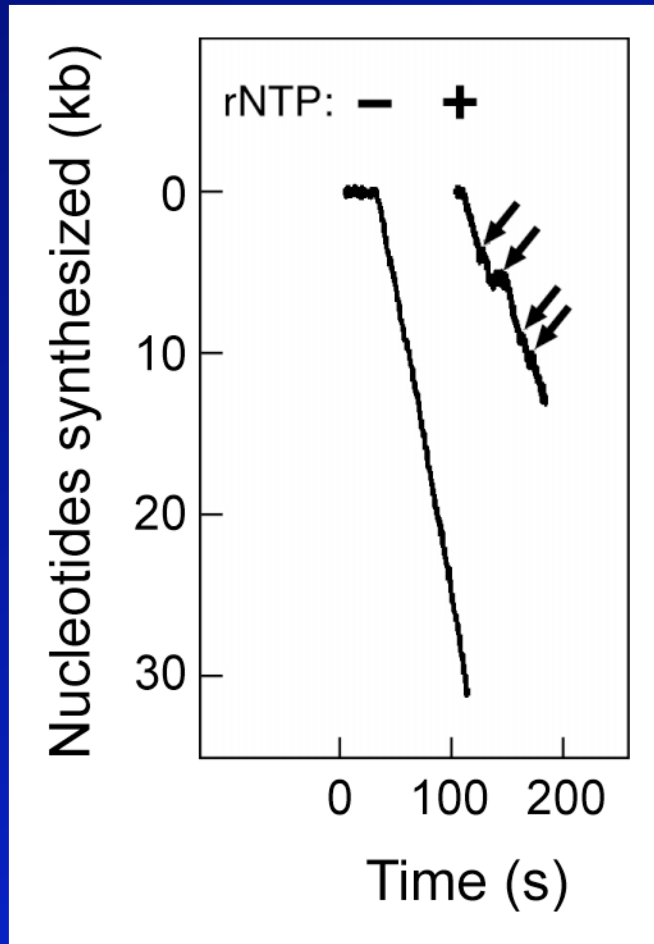
Doublie *et al.*, 1998

Basic residues in the thioredoxin binding domain (TBD) of gp5 bind to acidic C-terminal tail of gp4 helicase

Towards a complete replisome: Primase



Primase activity causes pausing?

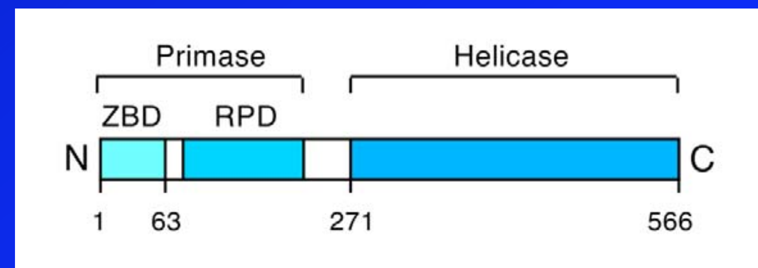


Lee *et al*., Nature 439 (2006) pp. 621-4

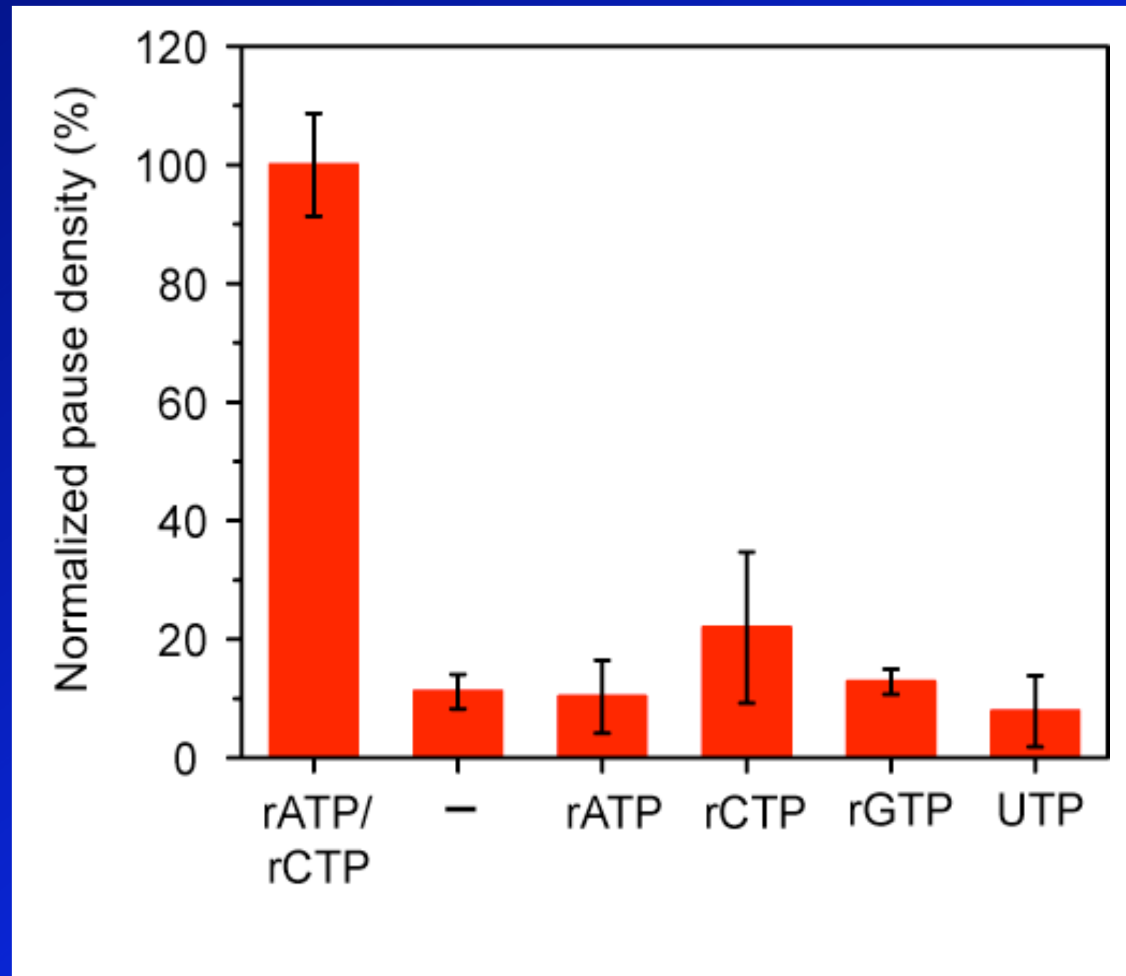
Controls:

Ribonucleotides rATP, rCTP are necessary for primase activity

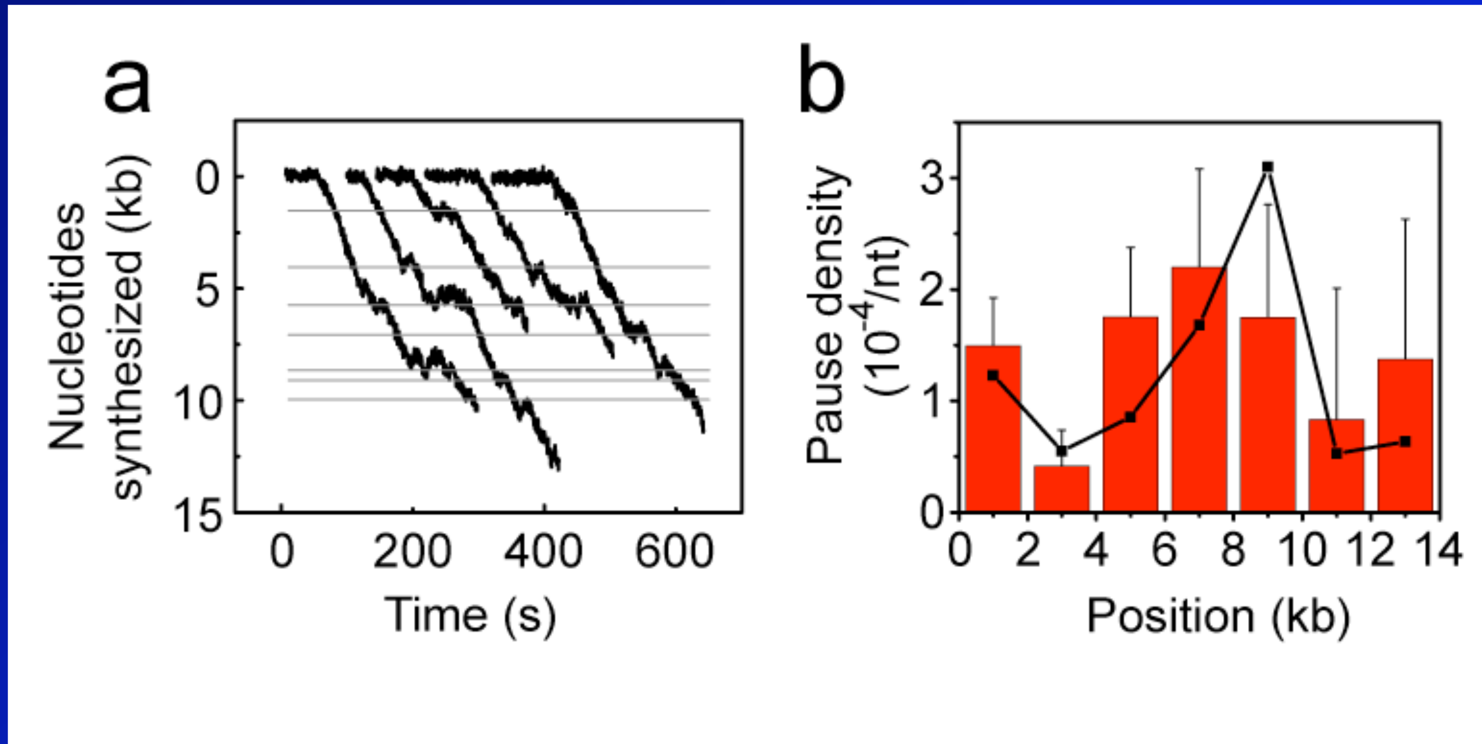
gp4's zinc-binding domain (ZBD) is necessary for primase activity



Both rATP and rCTP are required

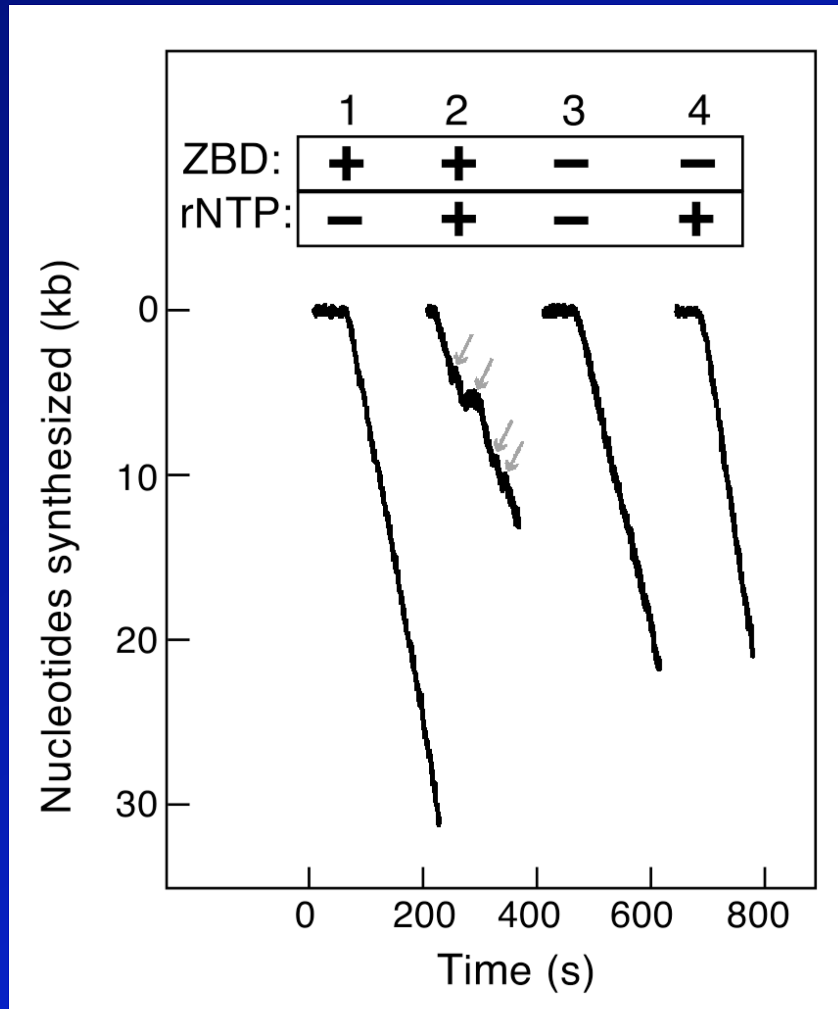


Pause sites correlate with primase recognition sites



Pause sites observed in experiment correlate well with known primase recognition sites on lambda phage genome

Zinc-binding domain is essential for pausing



gp4's zinc-binding domain (ZBD) is necessary for primase activity

Lee *et al.*, Nature 439 (2006) pp. 621-4

What's going on?

- Leading-strand synthesis continuous
- Lagging-strand synthesis discontinuous
- Primase activity is very slow

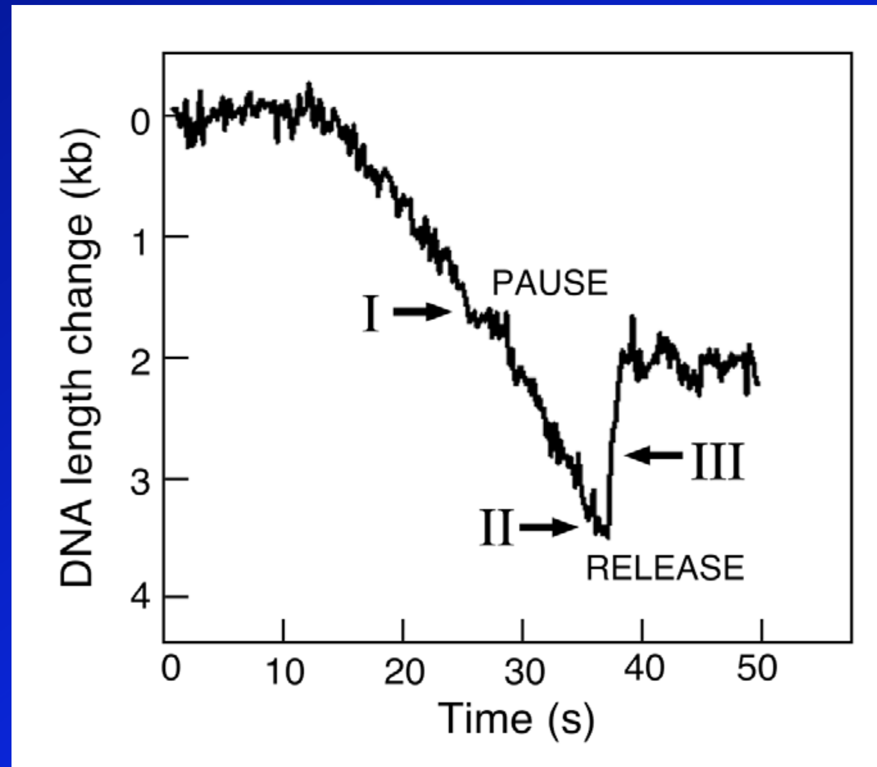
How does lagging-strand synthesis keep up with leading strand???

Hypothesis:

Primase activity on the lagging strand transiently stalls leading-strand synthesis

Leading- and lagging-strand synthesis

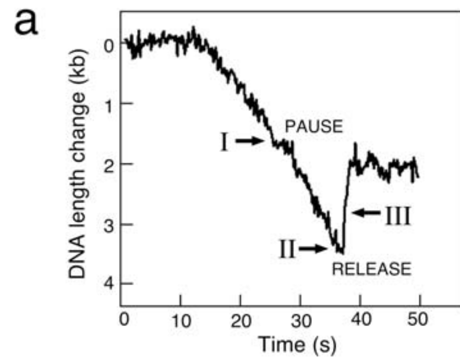
Excess DNA polymerase during replication reaction
→ lagging-strand synthesis

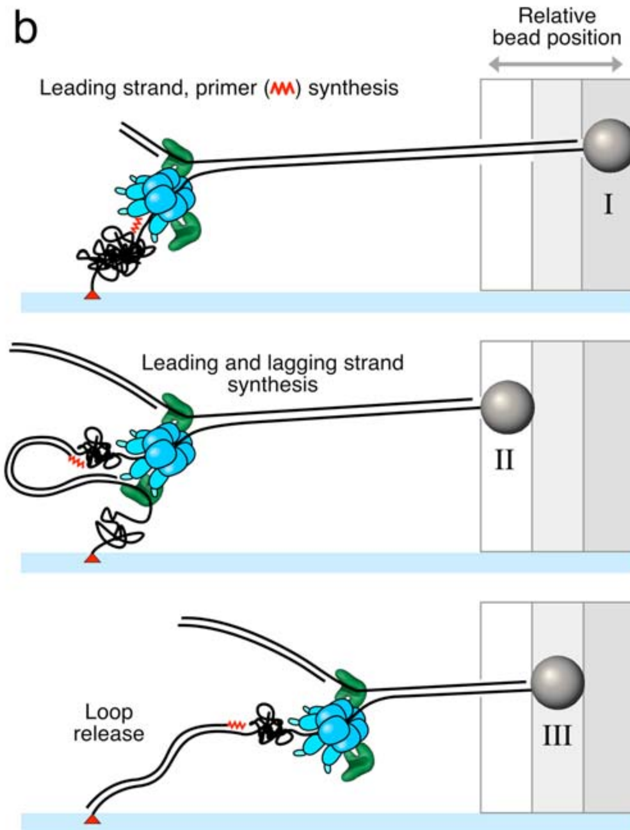


Lee *et al.*, Nature 439 (2006) pp. 621-4

Observation: sudden lengthening of DNA

Leading- and lagging-strand synthesis





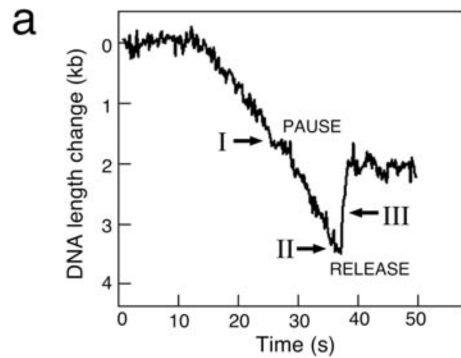
Lee *et al.*, Nature 439 (2006) pp. 621-4

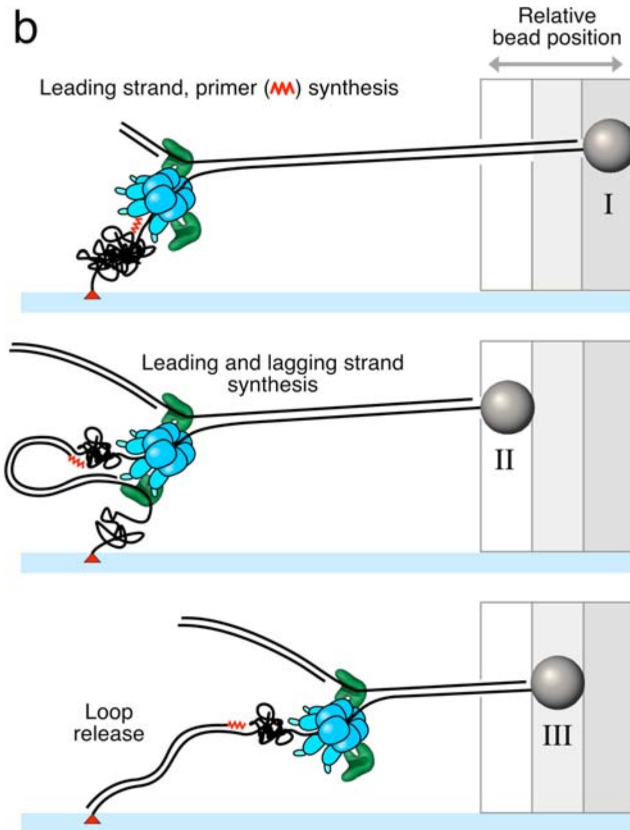
→ Replication loop formation!

Also, pausing is still there:

Primase acts as a molecular brake in replication

Leading- and lagging-strand synthesis





Lee *et al.*, Nature 439 (2006) pp. 621-4

Leading-strand synthesis:

$$\delta l = k_{pol} l_{ss} - k_{pol} l_{ds}$$

Leading- and lagging-strand synthesis (no loop formation):

$$\delta l = (k_{pol} l_{ss} - k_{pol} l_{ds}) + (-k_{pol} l_{ss} + k_{pol} l_{ds}) = 0$$

Leading- and lagging-strand synthesis (with loop formation):

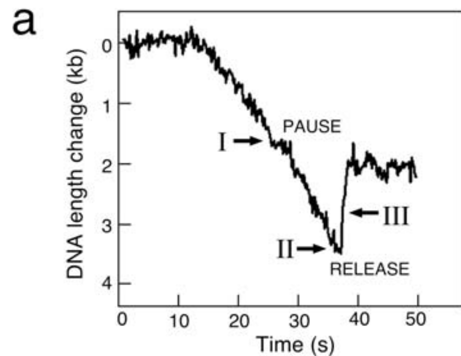
$$\delta l = (\cancel{k_{pol} l_{ss}} - k_{pol} l_{ds}) + (-k_{pol} l_{ss} + \cancel{k_{pol} l_{ds}})$$

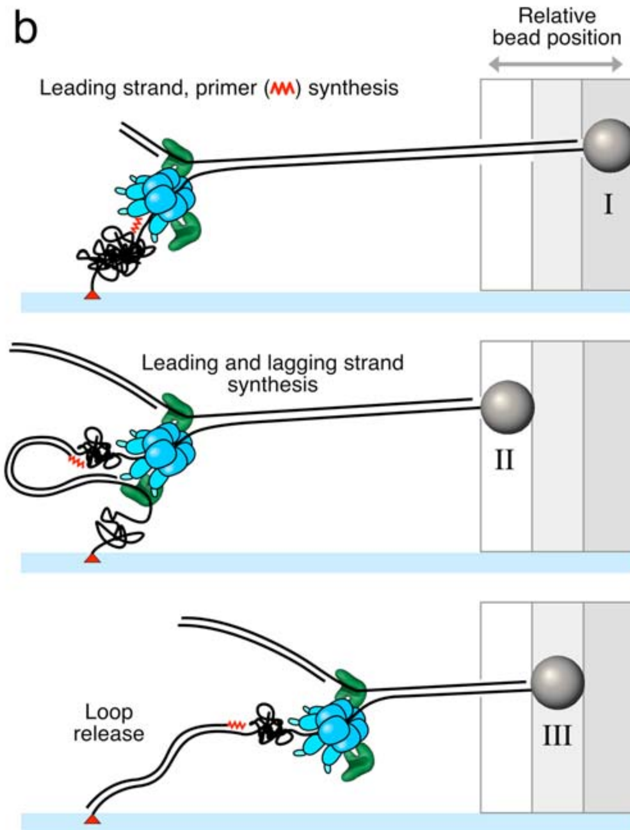
ssDNA helicase product — dsDNA lag. DNApol product

Into loop

$$\longrightarrow \quad \delta l = -(k_{pol} l_{ds} + k_{pol} l_{ss})$$

Leading- and lagging-strand synthesis





Lee *et al.*, Nature 439 (2006) pp. 621-4

→ Replication loop formation!

Leading-strand synthesis:

$\delta l = k_{pol} l_{ss} - k_{pol} l_{ds}$

Leading- and lagging-strand synthesis (with loop formation):

$\delta l = -(k_{pol} l_{ds} + k_{pol} l_{ss})$

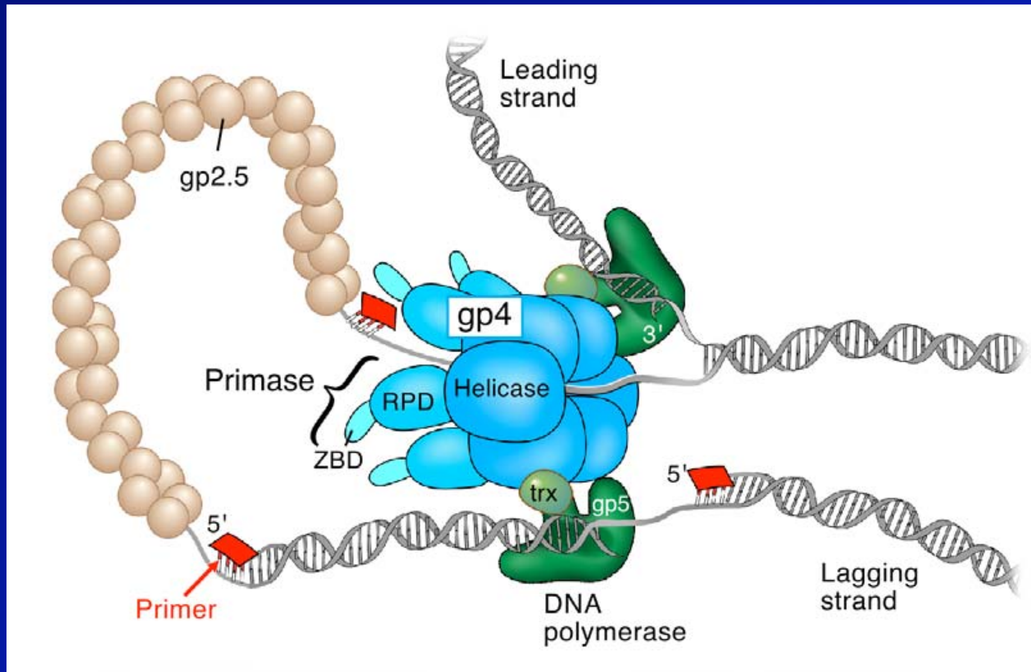
$$\frac{\text{slope loop formation}}{\text{slope leading-strand synthesis}} = \frac{l_{ds} + l_{ss}}{l_{ds} - l_{ss}}$$

Predicted:
1.2 (l_{ds}=0.32 nm/nt, l_{ss}=0.03 nm/nt @ 3 pN)

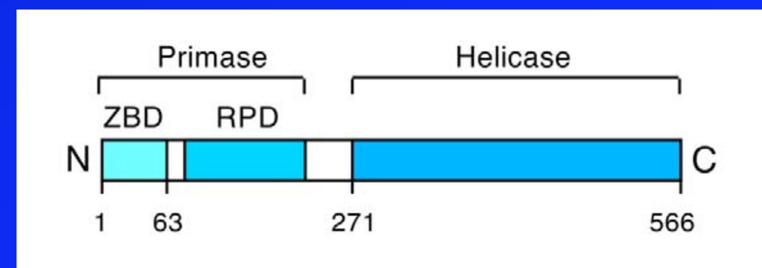
Measured:
1.3 ± 0.2

→ Both DNA pols equally fast

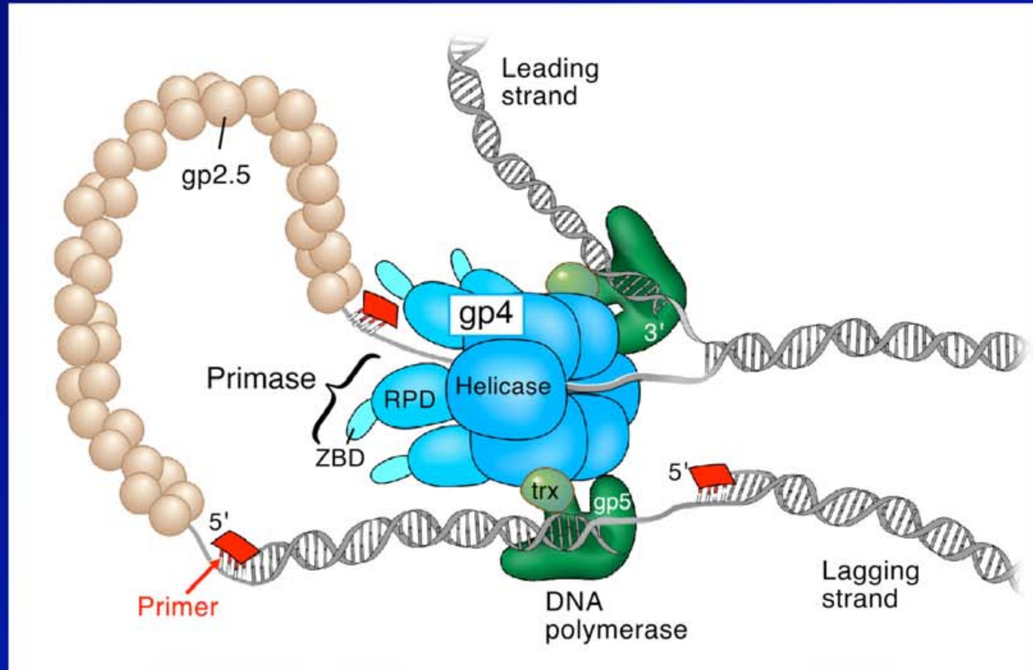
A molecular model for primase-induced braking

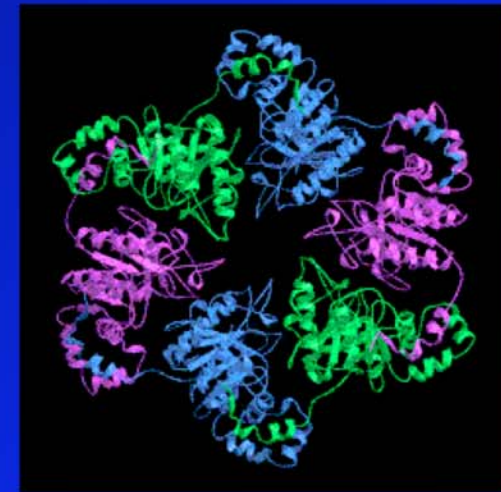


1) During primer synthesis, zinc-binding domain (ZBD) interacts with RNA-pol domain (RPD) of neighboring subunit in gp4 hexamer (*i.e., in trans* instead of *in cis*)
(Lee, Richardson, 2002)



A molecular model for primase-induced braking





gp4: Singleton *et al*., 2000

2) Helicase activity is facilitated by large conformational motions in hexamer
(Singleton *et al*.)

Summary

- Single-molecule enzymology on multi-enzyme complexes
- Helicase rate modulated by presence of second helicase
- Helicase-DNApol interaction as ball-bearing
- Primase activity on lagging strand stalls leading strand synthesis

Van Oijen Group

Candice Etson
Biophysics graduate student

Dan Floyd
BBS graduate student

Sam Hamdan
Postdoc

Satoshi Habuchi
Postdoc

J.B. Lee
Postdoc

Nathan Tanner
BBS graduate student

Collaboration with Charles Richardson, Dept. of BCMP