# Bi 1: The Great Ideas of Biology
# Homework 5
# Due Date: Thursday, May 11, 2023

## 1. Capturing Light

The first task that a photosynthetic organism must undertake is to capture light so that it may be converted into a usable form of energy. In this problem, we will explore the logistics of light capture and consider how various pigments and other cellular structures aid in this process.

### The power of the sun reaching Earth

The sun emits a broad range of wavelengths at different intensities. Figure 1 shows how much sunlight of different wavelengths arrives at the earth by plotting the spectral irradiance (essentially the power per area per wavelength) for a range of electromagnetic wavelengths. Notably, the irradiance curves differ substantially depending on the part of the earth being considered; the upper atmosphere receives significantly more sunlight than the surface of the earth, and at 10 m below the surface of the ocean there is very little sunlight at all. Interestingly, all three curves have peaks in the visible range, which coincides with the relevant wavelengths for photosynthesis.

For the problems below, use street-fighting mathematics to make your estimates. Make sure to explain in simple terms how you do your estimates.

**Question 1a**

Using Figure 1, estimate the power per square meter received at the surface of the earth, which is thus available for use by terrestrial organisms. How much of this power is available in the form of visible (400-700 nm), UV ($< 400$ nm), or infrared ($> 700$ nm) light? You should perform your estimate by using simple shapes such as rectangles and triangles.

To get a better sense of scale for your answer to part **1a**, let's consider how the power of sunlight compares to the power consumption of an average American household. According to the US Energy Information Association, the average household in 2015 had an annual energy expenditure of approximately $10^4$

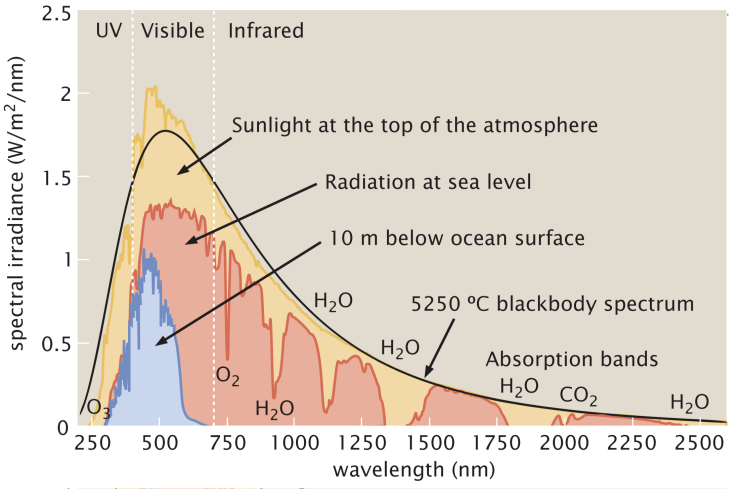


Figure 1: **Sunlight interacting with the earth at different wavelengths.** (Phillips, R., Kondev, J., Theriot, J., Garcia, G. (2012) Physical Biology of the Cell, 2nd edition. Adapted from `http://rredc.nrel.gov/solar/spectra/am1.5/` and `http://lasp.colorado.edu/bagenal/3720/CLASS4/SolarSpectrum2.jpg`)

kilowatt-hours ($\mathrm{kW} \cdot \mathrm{h}$). To convert this quantity into average power, one needs to divide it by the number of hours in a year.

**Question 1b**

Given your answer to part **1a**, how large of a surface (in $\mathrm{m}^2$) is required to collect enough solar power to power an American household? What are some reasons that an actual solar panel would need to be larger than this in order to fully power a home?

### Absorbing photons with pigments

Plants use light-absorbing pigment molecules to capture photons which will be converted into stored energy through photosynthesis. Chlorophylls (see chlorophyll *a* in Figure 2) are probably the best-known of these pigments, as they are largely responsible for the green color in leaves and other plant matter. Like other pigments, chlorophyll *a* captures energy for photosynthesis by

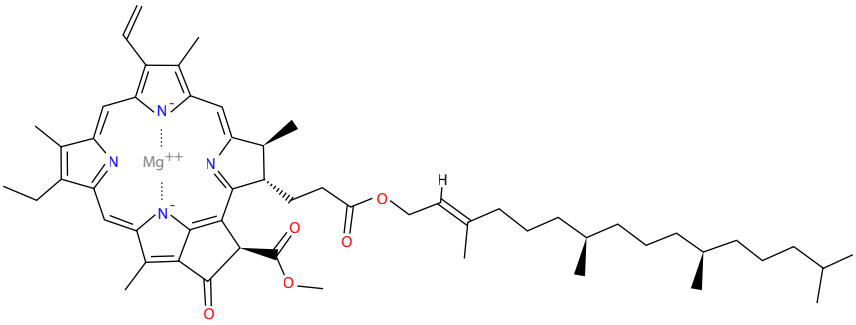


Figure 2: **Molecular structure of the light-absorbing pigment chlorophyll** *a* The light-absorbing porphyrin ring is shown on the left-hand side.

absorbing photons of favorable wavelengths. For chlorophyll $a$, light is maximally absorbed at 680 nm, but other pigments absorb photons at different wavelengths, with the result that plants can harness photons with nearly any wavelength in the visible spectrum. Figure 3 shows how the absorption spectra of different plant pigments overlap to provide full coverage of the visible spectrum.

As with many pigments, chlorophyll $a$ absorbs light via a porphyrin ring, which is the structure shown on the left-hand side of the molecule in Figure 2. For the purpose of this problem we can simply think of the porphyrin ring as a light-absorbing disk, though the quantum mechanics of such light absorption is a fascinating topic in its own right. During this part of the problem we will extend our exploration of solar power by considering the capability of a porphyrin ring to capture sunlight.

**Question 1c**

A porphyrin ring, as shown in Figure 2, acts as a molecular light-absorbing disk that is about 1 nm in diameter. Given your answer to part **1a**, how many photons can be expected to impinge upon the porphyrin ring every second? To obtain your answer you will need to use the equation for the energy of a photon, $E = \frac{hc}{\lambda}$, where $h$ is Planck's constant, $c$ is the speed of light, and $\lambda$ is the wavelength of the photon.

We note that the porphyrin ring is not a perfectly efficient light absorber. Because it preferentially absorbs light at specific wavelengths, only about 1 in 100 of the photons that collide with the molecule will actually be absorbed.

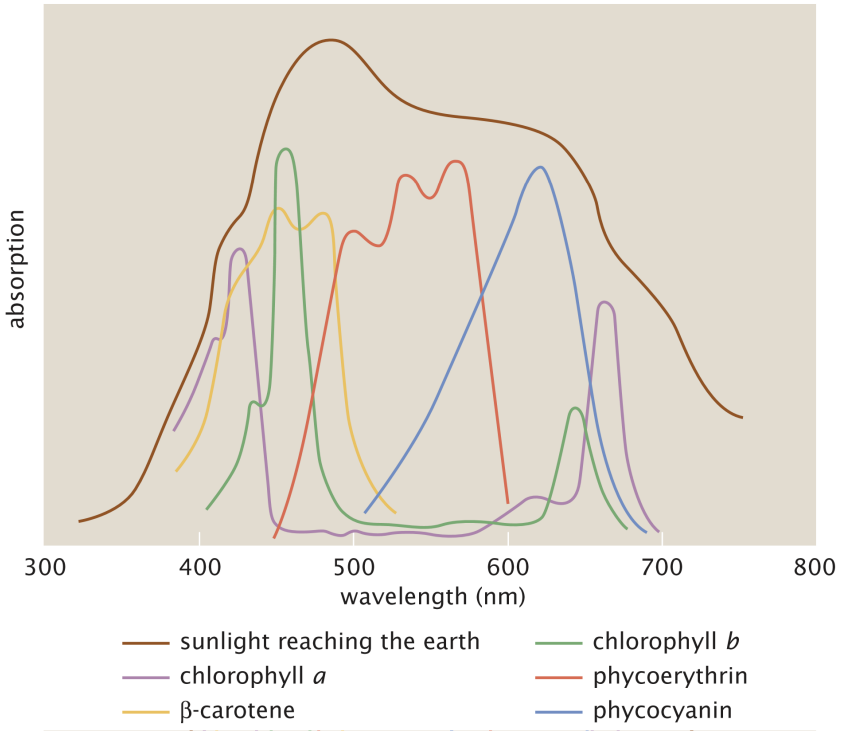


Figure 3: **Absorption spectra for a number of different photosynthetic pigments.** This plot shows absorption spectra for the portion of the electromagnetic spectrum that falls in the visible range of approximately 300-700 nm. Note that the y-axis is shown in arbitrary units, which accounts for the fact that the "sunlight reaching the earth" curve appears distorted relative to the analogous curve in Figure 1. (Phillips, R., Kondev, J., Theriot, J., Garcia, G. (2012) Physical Biology of the Cell, 2nd edition)

**Question 1d**

Given the inefficiency of absorption discussed above, how many photons per second are actually *absorbed* by the porphyrin?

As mentioned before, chlorophyll $a$ preferentially absorbs photons with a wavelength 680 nm, as can be seen in Figure 3. We will now explore the energy associated with each of these photons. The energy of a photon is given by the equation $E_p = \frac{hc}{\lambda}$, where $h$ is Planck's constant, $c$ is the speed of light, and $\lambda$ is the wavelength of the photon.

**Question 1e**

What is the energy of the average photon absorbed by chlorophyll $a$? How much energy on average will the porphyrin ring of chlorophyll $a$ absorb every second? Write your answer in $k_BT$ units using the conversion factor 1 $k_BT = 4.11 \times 10^{-21}$ J.

It can be difficult to say how much energy is "a lot" in a biological context, so it would be helpful here to make some comparisons and think about what an organism will actually do with the light energy it absorbs. In this part of the problem, we will compare the energy of a photon to each of the following processes:

- One of the consequences of oxygenic photosynthesis is that $H_2O$ is split into protons and the waste product $O_2$ through the overall chemical reaction

$$2H_2O \rightarrow O_2 + 4e^- + 4H^+. \tag{1}$$

  This reaction requires about 125 $k_BT$ to complete.

- Photosynthesis allows the cell to produce ATP, which is a significant source of energy in biochemical reactions. When ATP is converted to ADP + phosphate via ATP hydrolysis, approximately 20 $k_BT$ of energy is liberated.

- Light energy will eventually be stored in the form of glucose. Combustion of 1 g of glucose releases about 16 kJ of energy, which tells us how much energy is stored in glucose. (Note that we will convert the energy into units of $k_BT$ per molecule.)

Question 1f

**Question 1f:** Using the conversion factor 1 $k_BT = 4.1 \times 10^{-21}$ J, how many photons are required to provide enough energy to split water into protons and $O_2$? How many photons' worth of energy is released during ATP hydrolysis? How many photons' worth of energy is stored in a single molecule of glucose?

## 2. Converting Light Into Energy

After a photon is absorbed, the energy must be converted into a form that can be stored and then used as needed by the cell. In this problem, we will focus on energy storage in the form of a transmembrane hydrogen ion gradient. The establishment of a proton gradient allows the cell to store free energy as a *proton-motive force*. This free energy is then used to power the synthesis of ATP, which occurs when protons pass through the impressive molecular machine known as ATP synthase.

The free energy of the proton-motive force consists of two components: the energy due to the electric potential across the membrane, and the free energy stored in the concentration gradient across the membrane, so that we have $\Delta G_{tot} = \Delta G_{elec} + \Delta G_{chem}$. Here, $\Delta G_{tot}$ gives us the free energy per proton, and $\Delta G_{elec}$ and $\Delta G_{chem}$ are likewise determined on a per-proton basis. The electrical component is given by $\Delta G_{elec} = q\Delta V$, where $q$ is the charge of a proton and $\Delta V$ is the voltage across the membrane, typically of order 100 mV.

Question 2a

To get a sense of energy scales, what is the value of $\Delta G_{elec}$ for a proton in a membrane potential of 100 mV? Note that the charge of a proton is about $1.6 \times 10^{-19}$ C and 1 V = 1 $\frac{\text{J}}{\text{C}}$. Write your answer in units of $k_BT$.

Now we need to find an expression for the chemical component, $\Delta G_{chem}$. We will think of $\Delta G_{chem}$ as the free energy change when we take one proton from the outside of the membrane and transfer it to the inside of the membrane. When a proton is transferred across the membrane in this manner the free

energy on either side of the membrane is changed, so that

$$\Delta G_{chem} = \Delta G_{out} + \Delta G_{in}. \tag{2}$$

We note that $G = H - TS$, and since the contribution of enthalpy in this system is negligible, we can rewrite this as $G \approx -TS$. When $T$ is constant, $\Delta G = -T\Delta S$. $\Delta S$ for each side of the membrane can be written as $\Delta S = S_{final} - S_{initial}$, where *initial* indicates an initial state where there are $M$ protons on the inside of the membrane and $N$ protons on the outside, and $final$ indicates a state where there are $M + 1$ protons on the inside of the membrane and $N - 1$ protons on the outside.

**Question 2b**

Using the above definitions for $G$ and $\Delta S$, rewrite equation 2 in terms of $T$ and $S$.

We can now move forward by using the Boltzmann definition of entropy, $S = k_B \ln W$, where $W$ is the number of states available to the system. To find an expression for $W$, we will use the lattice model as shown in Figure 4.

**Question 2c**

Using a lattice model as discussed in class, find an expression for $W$ in terms of $N$ and $\Omega$, where $\Omega$ is the number of lattice positions available in the system. Note that $\Omega >> N$, which should ultimately allow you to derive the expression $W(N, \Omega) = \frac{\Omega^N}{N!}$. Carefully explain each of the steps you use to arrive at this answer, and provide an intuitive description of the functional form of the answer.

**Question 2d**

Using your answer to **2c** and the Boltzmann definition of entropy, find expressions for $\Delta S_{out}$ and $\Delta S_{in}$ in terms of $\Omega$, $N$, and $M$. Note that for $M >> 1$, $M + 1 \approx M$.

At this point, we are nearly done with our derivation of a usable expression

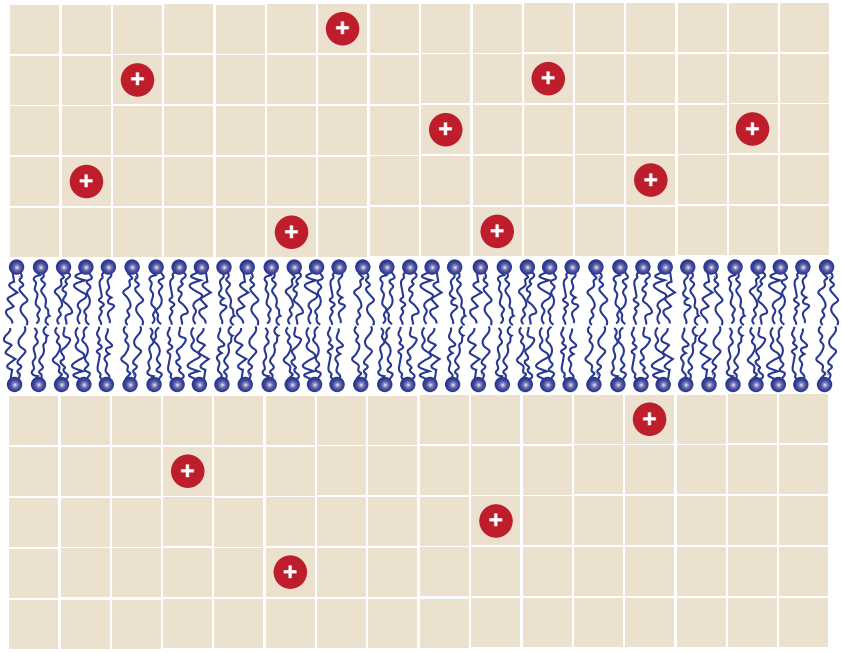

Figure 4: **The lattice model of a solution.** This schematic shows the lattice model that will be used to determine the number of states available on either side of the membrane. The available lattice sites on each side of the membrane, $\Omega$, are shown as tan squares. We think of the protons (red circles) as occupying distinct lattice sites. There are $N$ protons on the outside of the membrane (top) and $M$ protons on the inside of the membrane (bottom), each of which occupies one of the $\Omega$ lattice sites.

for $\Delta G_{chem}$. All that remains is to rewrite our expression in terms of more useful parameters, namely concentration and pH.

**Question 2e**

The concentration of protons on the outside of the membrane can be written as $c_{out} = \frac{N}{\Omega}$ and the concentration on the inside of the membrane can likewise be written as $c_{in} = \frac{M}{\Omega}$. Using these expressions for concentration and your answer to **2c**, rewrite your expression for $\Delta G_{chem}$ in terms of $c_{out}$ and $c_{in}$.

Question 2f

Using the relationship $c = 10^{-pH}$ for proton concentrations, write a final expression for $\Delta G_{chem}$ in terms of $\Delta pH$. Finally, recall that $\Delta G_{elec} = q\Delta V$ and finish writing the expression for $\Delta G_{tot}$ associated with the proton-motive force.

Question 2g

Given the model you just derived, what is the value of $\Delta G_{chem}$ given a physiologically typical $\Delta pH$ of 1.0, where $\Delta pH = pH_{out} - pH_{in}$? Given your answer to **2a**, what is the value of $\Delta G_{tot}$ associated with a proton moving across the membrane? Write your answers in units of $k_BT$.

## 3. Turning Protons into ATP

The electrochemical potential established by the proton gradient is used to power ATP production. ATP synthase (see Figure 5) is a membrane protein complex that can either harvest the energy stored in the proton gradient to synthesize ATP, or it can perform the reverse process of using the energy released by ATP hydrolysis to establish a proton gradient. During ATP synthesis, a lower-energy ADP molecule is combined with a phosphate to form ATP, a process that requires about 20 $k_BT$. During ATP hydrolysis, 20 $k_BT$ of energy is released by splitting an ATP into ADP and phosphate.

ATP synthase is composed of two primary subunits, the $F_0$ subunit, which spans the membrane, and the large $F_1$ subunit, which protrudes from the membrane (see Figure 5). ATP synthase is an example of a rotary motor, a remarkable class of molecular motors that acts by rotating to generate torque. Figure 6 outlines the single-molecule experiment that was used to observe the rotation of the $F_1$ motor. In this experiment, an $F_1$ motor was tethered to a slide, and a fluorescently labeled actin filament was attached to the opposite end of the $F_1$ motor. As the motor rotated, the orientation of the actin changed, and thus it was possible to observe that $F_1$ rotates through three positions, each separated by 120°. By comparison, $F_0$ only moves through 30° each time it rotates.

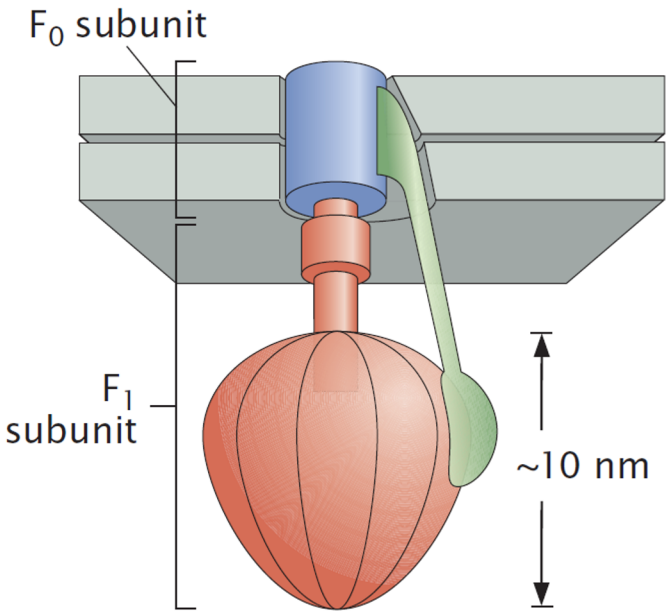


Figure 5: **Schematic of ATP synthase.** ATP synthase is a ubiquitous protein complex that is embedded in the membrane. Protons diffuse down the proton gradient through the $F_0$ subunit, powering ATP synthesis. (Phillips, R., Kondev, J., Theriot, J., Garcia, G. (2012) Physical Biology of the Cell, 2nd edition)

The $F_0$ motor uses the energy of the proton gradient to rotate, which forces the $F_1$ motor to rotate in the same direction. When the $F_1$ subunit rotates this way, it generates ATP by combining ADP and phosphate. This process will occur whenever the transmembrane gradient is sufficiently strong. However, when the transmembrane gradient is weak, the $F_1$ motor will rotate in the opposite direction to perform ATP hydrolysis and use the resulting energy to establish a proton gradient.

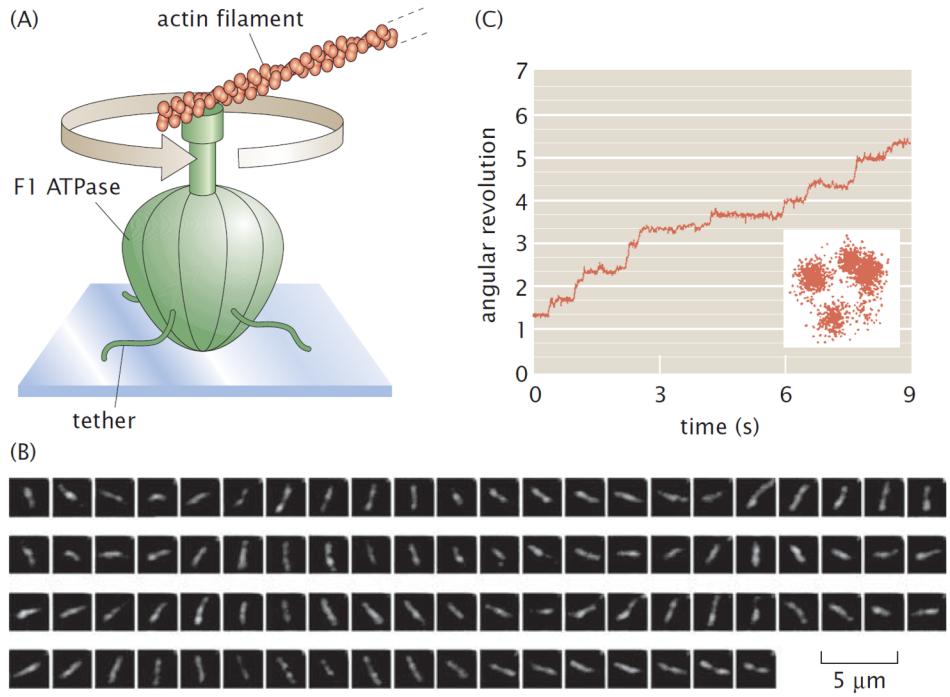


Figure 6: **An elegant single-molecule experiment for observing the rotation of the $F_1$ motor.** (A) The $F_1$ motor was tethered to a slide, and a fluorescently-labeled actin filament was attached to the other end of the motor. (B) Fluorscence microscopy was used to observe the actin molecule as it rotated. (C) Plotting the revolutions over time made it clear that $F_1$ rotated in discrete steps, of 120° each (see inset). (Phillips, R., Kondev, J., Theriot, J., Garcia, G. (2012) Physical Biology of the Cell, 2nd edition)

**Question 3a**

Given your answers to part **2g**, in which you found the free energy associated with moving a proton across the membrane under typical physiological conditions where $\Delta pH \approx 1.0$, about how many protons are required to power the production of 1 ATP?

## 4. Street-fighting the Ribosome

One of the most important molecular assemblies in the cell is the ribosome. This giant complex is responsible for protein synthesis by moving along a messenger RNA and adding the next amino acid in the polypeptide chain that will eventually be a full protein. The number of ribosomes per cell dictates how fast cells can grow. *E. coli* growing with a division time of 24 minutes have 72,000 ribosomes per cell, and slow growing *E. coli* with a division time of 100 minutes have a factor of ten fewer ribosomes with a count of $\approx$ 6800 ribosomes.

**Question 4a**

In this part of the problem, we will use our street fighting skills to explore the ribosomal density in another organism as shown in Figure 7, and then see how well our results from the electron microscopy study square with the numbers quoted above. By examining the figure, make an estimate of the number of ribosomes per $\mu m^3$ and compare that result to the numbers quoted for *E. coli* above.

**Question 4b**

In a beautiful turn of the millennium paper by Tania Baker and Stephen Bell whose abstract is shown in Figure 8, they imagined a world in which DNA polymerase was the size of a FedEx truck and explored what copying DNA would look like. Write a one-paragraph abstract of your own which carries out a similar analysis, but this time for the ribosome which has an error rate of roughly 1 incorrect amino acid incorporation per 10,000 amino acids added.

The typical speed of translation of bacterial ribosomes is between 10 and 20 amino acids per second. In this part of the problem, we perform a sanity check to see whether $f \times 10^4$ ribosomes can perform all the polymerization reactions needed to double the protein content of a cell.

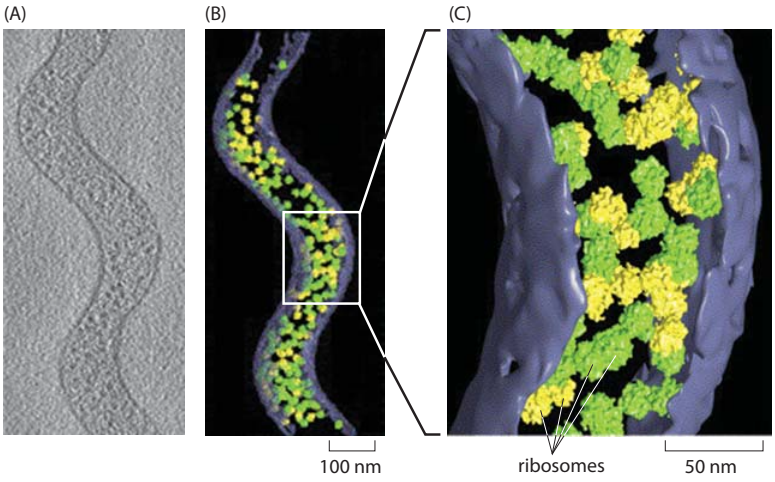


Figure 7: Cryo EM study of a bacterial cell. These images are of the tiny bacterium, *Spiroplasma melliferum*. Using algorithms for pattern recognition and classification, components of the cell such as ribosomes were localized and counted. (A) Single cryo-electron microscopy image. (B) 3D reconstruction showing the ribosomes that were identified. Ribosomes labeled in green were identified with high fidelity while those labeled in yellow were identified with intermediate fidelity. (C) Close up view that you should use to make your count. Adapted from JO Ortiz *et al.*, J. Struct. Biol. 156, 334-341 (2006).

Cell, Vol. 92, 295–305, February 6, 1998, Copyright ©1998 by Cell Press

# Polymerases and the Replisome: Machines within Machines

**Review**

**Tania A. Baker* and Stephen P. Bell**
Department of Biology
*Howard Hughes Medical Institute
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Synthesis of all genomic DNA involves the highly coordinated action of multiple polypeptides. These proteins assemble two new DNA chains at a remarkable pace, approaching 1000 nucleotides (nt) per second in *E. coli*. If the DNA duplex were 1 m in diameter, then the following statements would roughly describe *E. coli* replication. The fork would move at approximately 600 km/hr (375 mph), and the replication machinery would be about the size of a FedEx delivery truck. Replicating the *E. coli* genome would be a 40 min, 400 km (250 mile) trip for two such machines, which would, on average make an error only once every 170 km (106 miles). The mechanical prowess of this complex is even more impressive given that it synthesizes two chains simultaneously as it moves. Although one strand is synthesized in the same direction as the fork is moving, the other chain (the lagging strand) is synthesized in a piecemeal fashion (as Okazaki fragments) and in the opposite direction of overall fork movement. As a result, about once a second one delivery person (i.e., polymerase active site) associated with the truck must take a detour, coming off and then rejoining its template DNA strand, to synthesize the 0.2 km (0.13 mile) fragments.

In this review we describe our current understanding of the organization and function of the proteins of the replication fork and how these complexes are assembled at origins of replication. Understanding the architecture of DNA polymerases is relevant to RNA polymerases as well, as the core of the polynucleotide polymerization machine appears to be similar for all such enzymes. In the discussion of the replisome, we particularly focus on features shared by the machinery from different organisms.

### *Polymerases: Template-Directed Phosphoryl Transfer Machines*

Synthesis of the new DNA strands occurs as a result of a collaboration between the synthetic capacities of multiple polymerases. Two types of polymerases are required: primases, which start chains, and replicative polymerases, which synthesize the majority of the DNA (Kornberg and Baker, 1992). The replication fork, however, contains at least three distinct polymerase activities: a primase and a replicative polymerase for each of the two template strands. In *E. coli*, primase is a single polypeptide, and the replicative polymerase is a dimer of DNA polymerase (pol) III core and several accessory proteins that together form the pol III holoenzyme (reviewed in Marians, 1992; Kelman and O'Donnell, 1995). Similarly, phage T4 has one primase and one replicative polymerase that appears to function as a dimer (Alberts, 1987; Munn and Alberts, 1991). The situation in eukaryotic cells is slightly different (Stillman, 1994). The primase is in a tight complex with a DNA polymerase (pol $\alpha$) and eukaryotic cells have two distinct replicative polymerases: polymerase $\delta$ (pol $\delta$) and polymerase $\epsilon$ (pol $\epsilon$).

All the replicative polymerases have one large subunit that contains the polymerase active site and, with the exception of pol $\alpha$–primase, the same subunit or an associated polypeptide carries a proofreading 3′→5′ exonuclease. The polymerase subunits also interact with proteins that dramatically influence their association with DNA. In *E. coli*, the replicative polymerase is found in a complex with proteins that control polymerase processivity; this holoenzyme, consists of 10 distinct polypeptides (Kelman and O'Donnell, 1995). In contrast, neither the T4 nor the eukaryotic polymerases copurify in a complex with the processivity factors (Alberts, 1987; Stillman, 1994). Therefore, these proteins are called accessory proteins rather than subunits (see Table 1).

*Polymerase Architecture.* The central feature of all the known polymerase structures is the existence of a large cleft comprised of three subdomains referred to as the fingers, palm, and thumb by virtue of the similarity of the structures to a half-opened right hand (Figure 1; polymerase structures are reviewed in Joyce and Steitz, 1994, 1995; Sousa, 1996). A diverse set of polymerases—

Figure 8: Abstract of a paper from Tania Baker where she maps the action of DNA polymerase onto human length scales to give a sense of its amazing properties. This parable is the basis of your own analysis of the ribosome. Adapted from Baker TA and Bell, SP Cell, Vol. 92, 295Ð305, February 6, (1998).

Question 4c

Given that the volume of a bacterial cell is roughly 1 $\mu m^3$ and that 70% of the mass of the cell is water, estimate the total number of proteins per cell. To justify your estimate, give an approximate argument as to what fraction of the dry mass of a cell is protein. To work out this estimate, you will need to estimate the mass per protein by describing the number of amino acids per typical protein and the average mass per amino acid.

Question 4d

Given the number of ribosomes per cell, the characteristic speed of translation of ribosomes and the number of proteins per cell, estimate how long it will take to double the protein content of a bacterium in preparation for cell division.