# Chapter 10

# Nanobiology

**10.0 Outline of Chapter 10.**
This Chapter focuses role of fluctuations in the small systems that comprise the building blocks of biology. The measurement aspect of nanobiology – single molecule measurements – was dealt with in Chapter 4. We begin with a brief description of natural selection as the central driving force in biology. We go on to describe some of the components of modern molecular biology, getting a glimpse into the role of random processes when we discuss the remarkable phenomenon of gene splicing. We examine the role of structural fluctuations in enzyme catalysis and go on to discuss biological energy and molecular motors, phenomena that rely upon fluctuation-driven electron transfer reactions. Finally we examine the role of fluctuations in gene expression in the development of whole organisms. We use the immune system as an example of the biological process in which random assembly of its basis components is essential. We end that discussion with a brief look at the role of random gene splicing in the development of neural networks, and ultimately possibly, in the development of the mind itself. Readers unfamiliar with molecular biology might find the glossary in Appendix L useful.

**10.1 Natural selection as the driving force for biology**
There is probably no more interesting emergent phenomenon than life. Charles Darwin published "The Origin of Species by Means of Natural Selection" in 1859. It lays theoretical foundation for modern biology and is the first comprehensive description of an emergent phenomenon.

A slightly modernized scheme of natural selection is shown in Figure 10.1. We know today that random variation in the physical manifestation of a particular creature (its *phenotype*) is a consequence of random variation in the sequence of the DNA of the creature or of the placement and repetition of particular genes (its *genotype*). The quest for the connection between phenotype and genotype is a major focus of modern biology. Very few phenotypical features

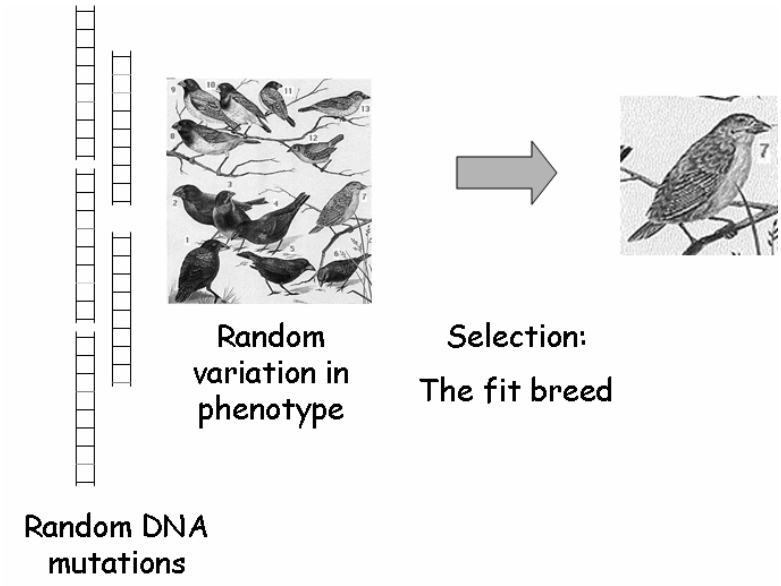


**Figure 10.1** Evolution by natural selection

are associated with just one gene. Darwin, of course, did not know about DNA, but he did observe the large random variations between members of a given species. He also saw how members of the population best adapted to a particular environmental niche formed the most successful breeding cohort, thereby concentrating those phenotypical features best adapted to the particular ecological niche into what would become a new subspecies or species.

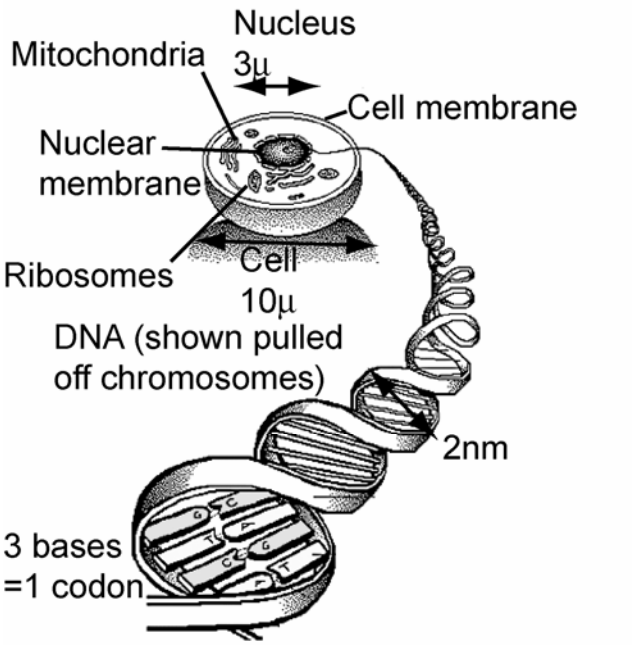


**Figure 10.2:** The eukaryotic cell. The genetic material (DNA) is contained inside an enclosed nuclear membrane with genetic information exported (as an RNA copy of DNA) to the cytoplasm between the nuclear and external membranes where ribsomes (attached to an internal membrane called the endoplasmic reticulum) translate the RNA code to proteins. The entire cell is packaged in a lipid-bilayer membrane.

One of the most striking aspects of biology on the nanoscale is the operation of Darwinian natural selection at the molecular scale, and we will describe processes that mimic evolution by natural selection at the molecular level. At this point it is useful to make a connection with the other emergent phenomena we have studied, which is electron transfer in solution. Recall that a random population of solvent polarizations helped to bring about the degenerate transition state required for electron transfer by Fermi 's Golden rule. We shall see that biological molecules that catalyze reactions (enzymes) use this mechanism, so that natural selection serves as a motif for biological processes all the way from the most elementary chemistry through to cells, whole organisms, and entire ecologies.

**10.2 Introduction to molecular biology**

The smallest independent living systems are composed of single cells, and are called prokaryotes. They consist of a lipid bilayer ‘bag’ that contains DNA and proteins needed to sustain the life of the cell (by synthesizing components using the instructions encoded in the DNA). Multi-celled organisms, called eukaryotes, have a more complex double membrane structure illustrated in Figure 10.2. Because these multicelled organisms (like you and me) have to specialize gene expression according to the particular function of the cell within the organism, the genetic material (DNA folded onto positively charged proteins in a compact structure called chromatin) is contained within an inner membrane in a structure called the cell nucleus. Between the cell nucleus and the outer cell membrane lies the cytoplasm, the region containing all of the machinery needed to generate chemical energy for the cell (the mitochondria) and to translate genetic material

into proteins (the ribosomes). Transport of genetic code from the cell nucleus into the cytoplasm is controlled by a nano scale gatekeeper called the nuclear pore complex. The same complex also regulates the transport of proteins that control gene expression from the cytoplasm back into the nucleus.

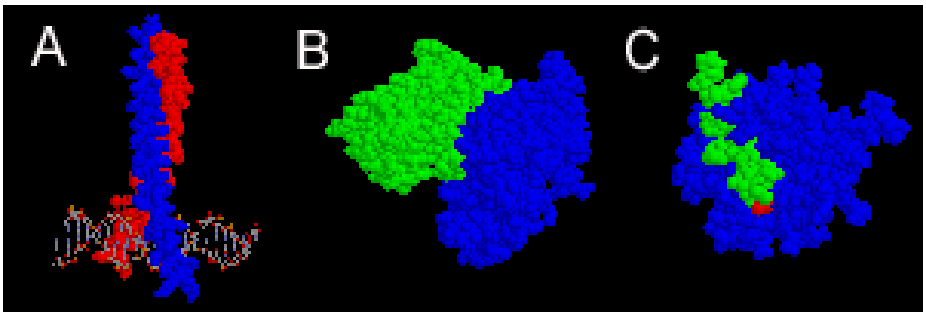


**Figure 10.3:** Examples of protein structures (from http://medweb.bham.ac.uk/bmedsci/bms2/chime/structure/structure.html#anchor259353)

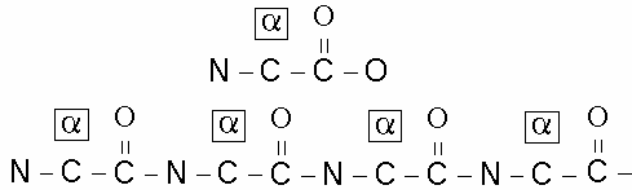


**Figure 10.4:** (Top) Generic structure of a single amino acid where α represents one of the 20 naturally occurring residues illustrated in Figure 10.5. (Bottom) amino acids joined by peptide bonds to form a polypeptide.

DNA is the repository of genetic information and its structure was described in detail in Chapter 6. The nano machines that carry out the chemical processes essential to life are proteins. Proteins come in an amazing variety of shapes and sizes and chemical properties. Some examples are shown in Figure 10.3. Figure 10.3A shows a protein called AP1. It is a transcription factor that binds DNA to regulate the expression of genes. Some idea of the size of the protein is given by the (2 nm diameter) double helix it is shown bound to in the Figure (the DNA is the structure at the bottom). The protein is a dimer formed by the association of two different proteins (one shown in red, the other in blue). The second protein shown in Figure 10.3B is also a dimer composed of two units that associate to form a complex that modifies specific sites on other proteins by adding a phosphate to them. Such enzymes are called protein kinases. The final example shown in Figure 10.3C is a protein that becomes phosphorylated (i.e. a phosphate group is added) at one specific site, shown in red, when the protein binds to a specific target.

This incredible structural and chemical diversity is built from a limited number of quite simple chemical building blocks. The elementary unit of a protein is called an amino acid, and the generic chemical structure of these units is shown at the top of Figure 10.4. Here the variable component is labeled α and it is connected through a carbon atom to an amine at one end and a carboxylate residue at the other end. These units are joined together in a polymer chain where one oxygen is cleaved from the carbon together with two hydrogens from the nitrogen (not shown in the Figure) to generate a water molecule, as the carboxylate joins to the amines to form *peptide bonds*. The resulting chain of amino acids is called a *polypeptide* and it folds into a functional protein. The chemical diversity of proteins is built on the 20 different types of residue symbolically listed as α in Figure 10.4. These 20 amino acids are shown in Figure 10.5. Remarkably all of the chemical diversity of biology is built on just these 20 components, DNA, lipids, sugars and minor chemical modifications to these components. The first five amino acids shown at the top of Figure 10.5 are hydrophobic, so in most proteins, regions of the peptide chain that are rich in these residues tend to be pushed to the interior of the folded protein. The next seven residues shown in the middle of Figure 10.5 are hydrophilic and tend to lie on the outside of the folded protein. Proteins that are embedded in lipid bilayers (i.e. transmembrane proteins) have hydrophobic residues on the surface that contacts the lipid bilayer. The final group of eight amino acids have an intermediate character and can change their properties from hydrophobic to hydrophilic and *vice versa* depending upon

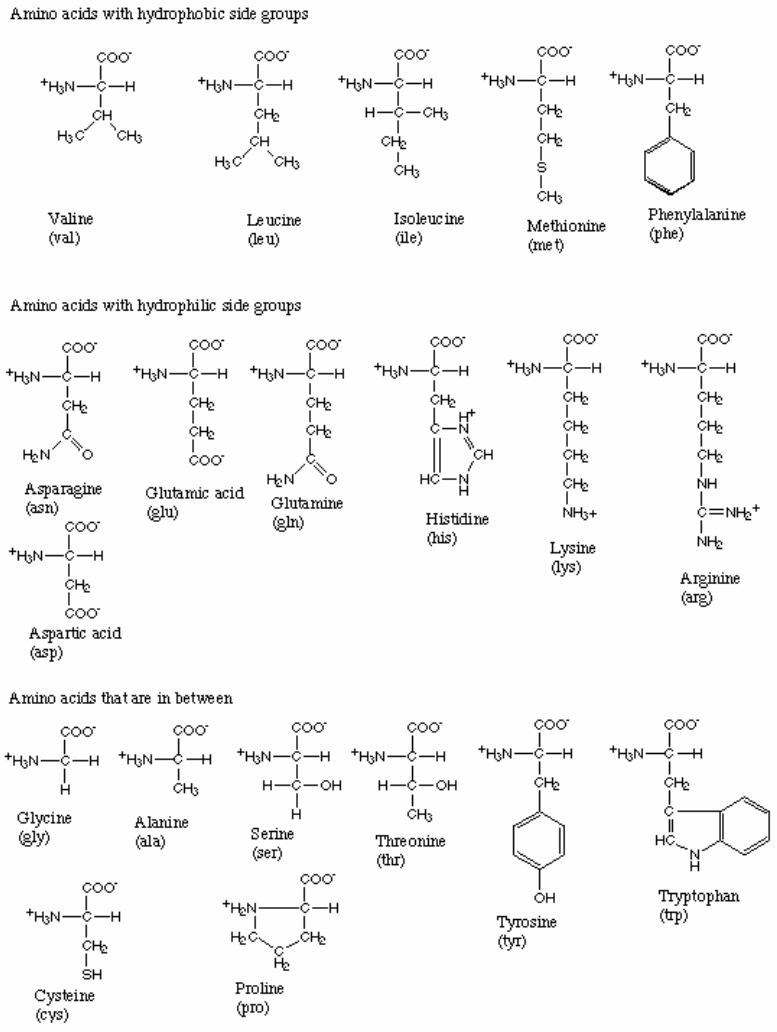


**Figure 10.5:** The 20 naturally occurring amino acids. These are strung together by peptide bonds to form polypeptides that fold into functional proteins. The top 5 are hydrophobic, the middle 7 are hydrophilic and the remaining 8 have intermediate character. Most proteins fold with hydrophobic groups in the interior if this is possible. Membrane proteins (embedded in a lipid bilayer) have hydrophobic exteriors.

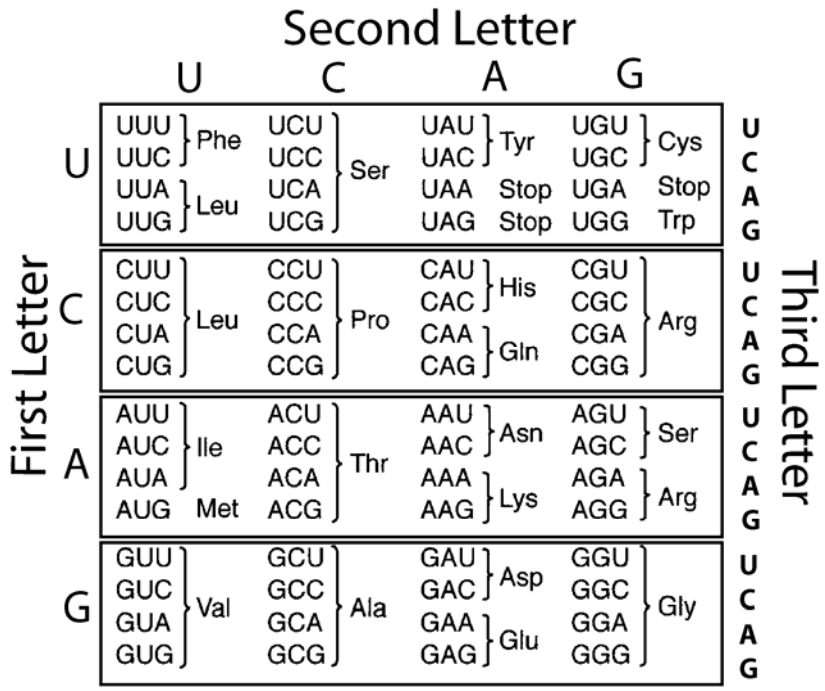


**Figure 10.6:** The genetic code, showing how mRNA bases (complementary copies from DNA, but with U substituting for T) are translated into amino acid residues in a protein (see Figure 10.5). Each group of three bases is called a codon – note how some codons are highly degenerate (though which particular one of the degenerate set is used is often species specific). Three codons – UAA, UAG and UGA are "stop" signals that signify the end of a protein chain.

local chemical conditions such as pH. Accordingly, these residues can act as chemical sensors, driving changes in the structure of the protein in response to changes in the local chemistry of the environment.

Thus we see that these amino acid building blocks can give rise to a fantastic variety of properties of the protein product, depending upon the specific sequence of the residues in the polypeptide backbone. This sequence of residues is specified by the sequence of bases in DNA according to the genetic code. 20 different amino acids require a minimum of three bases (two bases could specify only a maximum of four times four or 16 amino acids). The correspondence between the sequence of bases on the DNA and which particular amino acid they code for was sorted out many years ago and the end result is summarized in Figure 10.6. The DNA remains within the cell nucleus but instructions for the synthesis of specific proteins are passed out from the nucleus to the cytoplasm in the form of RNA. Thus the DNA specifying the sequence of the protein is first translated into a sequence of RNA that corresponds to the complement of the DNA sequence. Thus a T on DNA is translated into an A on RNA, a G on DNA is translated into a C on RNA, and a C on DNA is translated into a G on RNA. The translation of A is handled a little differently, because RNA uses a slightly different version of thymine called uracil (U). Thus an A on DNA is translated into a U on RNA. Figure 10.6 shows how the RNA sequence is translated into the amino acid sequence of the final protein (the abbreviations for the amino acids correspond to the three letter abbreviations given in Figure 10.5). There are 64 combinations that can be made from three sets of the four bases so the genetic code has redundancies. For example four different groups of three bases (called codons) code for arginine: CGU,CGC,CGA, and CGG. Three of the codons, UAA, UAG, and UGA code for a stop signal (i.e. the end of the protein chain). Interestingly this degeneracy in the genetic code is not absolute, because different species use certain codons preferentially. This form of readout requires a universal "start" signal and blocks of DNA code that are upstream of the gene contain

both these start signals and other sequences that are involved in turning transcription of the gene on and off.

A simple schematic pathway for the synthesis of gene products is shown in Figure 10.7. The DNA coding for the "start" signal for a particular protein becomes bound by a large enzyme called RNA polymerase. This translocates along the gene, turning each DNA base into its RNA complement. This RNA is called messenger RNA and it is labeled in Figure 10.7 as mRNA. The messenger RNA is passed from the cell nucleus to the cytoplasm via the nuclear pore complex. In the cytoplasm it binds another piece of protein machinery called the ribosome (ribosomes are generally localized in a folded membrane called the endoplasmic reticulum). The ribosome takes small pieces of RNA linked to amino acids (these are called called transfer RNAs) and uses them to link the amino acids together in the sequence specified by the mRNA, adding the correct residue as each RNA base is read. The result is a peptide chain that spontaneously folds into the desired protein product. Both the ribosome and RNA polymerase are complexes formed by the spontaneous association of small pieces of RNA with protein. The final soup of proteins and small molecules in the cytoplasm feedback into gene expression through factors that are imported by the nuclear pore complex. These factors trigger a cascade of events in the cell nucleus that up- or down-regulate the expression of particular genes.

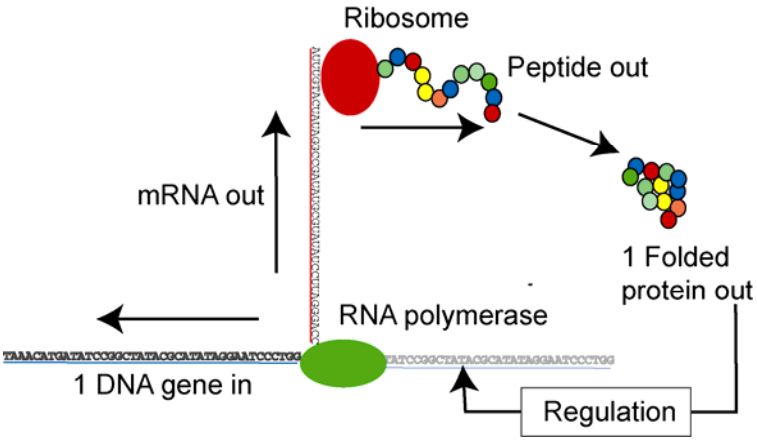


**Figure 10.7:** Molecular biology 1: the simple version. RNA polymererase is a nanomachine that makes an RNA transcript from a DNA strand, substituting uracil in place of the thymine base. The RNA strand leaves the cell nucleus and is translated into a protein according to the genetic code (Figure 10.6) by another nanomachine, the ribosome.

The Human Genome Project has shown that humans have only about 25,000 genes, and yet it is known that there are many millions of different types of human proteins. How can this be? A more sophisticated version of the story just given is shown in Figure 10.8. It turns out that in eukaryotes genes are split into packages of coding DNA (called exons) interrupted by regions of non-coding DNA (called introns). At one point it was supposed that these non-coding regions were useless genetic junk inherited for no good reason. However their sequence is well preserved from generation to generation, implying that they play an important role. Many parts of these introns are completely essential for the correct final assembly of the exons into the correctly spliced code. When RNA polymerase translates this complex DNA sequence into RNA, the regions corresponding to introns are excised. One path way for this excision is the spontaneous folding of parts of the RNA into catalytically active units (called ribozymes) that serve to cut the adjacent RNA. The final assembly of exons into the complete coding

sequence takes place in a nano machine called the spliceosome. Here the RNA is assembled into the final product by putting together the exons in a way that can be regulated by the environment. The various ways of ordering the exons results in various protein products when the different messenger RNAs are a translated into proteins. We shall see that some important processes actually depend upon the random splicing of these exons to form a random library of protein products. Random splicing is part of a

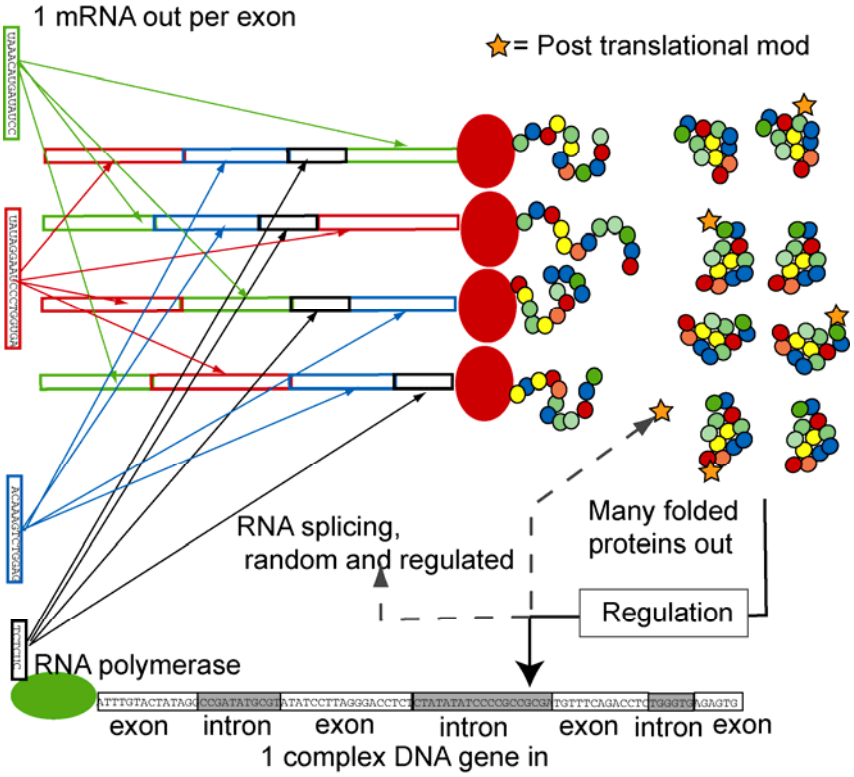


**Figure 10.8:** Molecular Biology 2: some of the complexity. Genes come in packets of coding regions (exons) separated by non-coding regions (introns). The introns are spliced out when the gene is assembled into a continuous transcript as mRNA is assembled in a nanomachine called the spliceosome. The reassembly can be linear, or varied under regulatory control, or even random, with the result that one gene can code for many different proteins. The proteins are further differentiated by chemical modifications made after the proteins are folded, called post-translational modifications.

Darwinian mechanism that allows cells to produce the correct product through a process of natural selection.

There is yet another important source of protein diversity based on specific chemical modifications of particular amino acid residues. These generally occur in a protein processing plant within the cell called the Golgi apparatus. This is the place where specific chemical stamps are put onto proteins that are then dispensed to various regions of the cell (to perform the functions that require that particular modification). These post-translational modifications extend the range of protein activities over and above the range produced by splicing variations. In addition to labeling proteins for

dispatch to particular regions of the cell, post-translational modifications also serve a signaling function. Thus a vast range of protein functionality can emerge from a limited number of genetic components. It has been recently discovered that the introns serve many other control functions when translated into RNA, opening a new avenue for RNA-control of cells.[14] A newly discovered class of small RNA fragments (siRNA) act to shut down expression of particular genes, and can be used as 'switches' in biotechnology applications.

**10.3 Some mechanical properties of proteins**

*Density*

A representative value for the density of well packed protein is about $1.4 \times 10^3$ kg/m$^3$,[15] leading to the following value for the radius of a spherical protein (where the mass is expressed in Daltons, *MW*):

$$r = \sqrt[3]{\frac{3}{4\pi}\frac{MW}{822}} = 0.066 \times \sqrt[3]{MW} \text{ nm} \qquad (10.1)$$

Thus a (typical) 100kD protein has a radius of about 3 nm.

*Stiffness*

The stiffness of the material is specified in a geometry-independent way by the Young's modulus. This is the ratio of stress to strain for a given material. Stress is the force per unit area applied to the material and strain is the fractional change in dimension, taken here to be the direction along which the stress acts, for simplicity. Thus if a force *F* acts on the face of a material of area *A*, producing a fractional change in dimension $\frac{\delta\ell}{\ell}$, force and strain are related by the Young's modulus, *E*, according to

$$\frac{F}{A} = E\frac{\delta\ell}{\ell}. \qquad (10.2)$$

It is useful to translate this quantity into an equivalent spring constant for the stretching of a Hookean spring. For the sake of argument we will imagine a force acting on a face of a cube of sides $\ell$. After inserting $A = \ell^2$ into equation 10.2 and then multiplying both sides by $\ell^2$ we arrive at Hooke's Law,

$$F = E\ell\delta\ell \qquad (10.3)$$

where the spring constant, $\kappa$, is given by

$$\kappa = E\ell. \qquad (10.4)$$

Some values for the Young's modulus of various proteins are given in table 10.1 below. These may be used to estimate rough values for a spring constant by inserting the dimension of the protein that is strained.

| Protein | E (GPa) | Function |
|---|---|---|
| Actin | 2 | Skeleton of the cell |
| Tubulin | 2 | Transport 'roadways' in cells |
| Silk | ~5 | Spider webs, cocoons |
| Abductin | 0.004 | Mollusc hinge ligament |
| Elastin | 0.002 | Smooth muscle and ligaments |

**Table 10.1** Young's modulus of various proteins (adapted from Howard[1])

As an example, tubulin has a molecular weight of about 50 kD, so according to equation 10.1 its diameter is about 4.8 nm. Using $\kappa = E\ell$ with $E$ = 2GPa yields a spring constant of about 10 N/m. Elastin has a molecular weight of about 75 kD, yielding a diameter of 5.5 nm and a spring constant of about 0.01 N/m.

*Transport properties*

We will use as an example our 'typical' 100kD protein with a radius of about 3 nm. The drag coefficient in water follows from Stokes' law

$$\gamma = 6\pi r\eta$$

Which, with $\eta = 1\times10^{-3}$ Pa.s, yields 56 x $10^{-12}$ Ns/m.

The diffusion constant can be obtained from equation 3.69, $D = \mu k_B T$ , and using $\mu = (6\pi a\eta)^{-1}$, we obtain for $D$

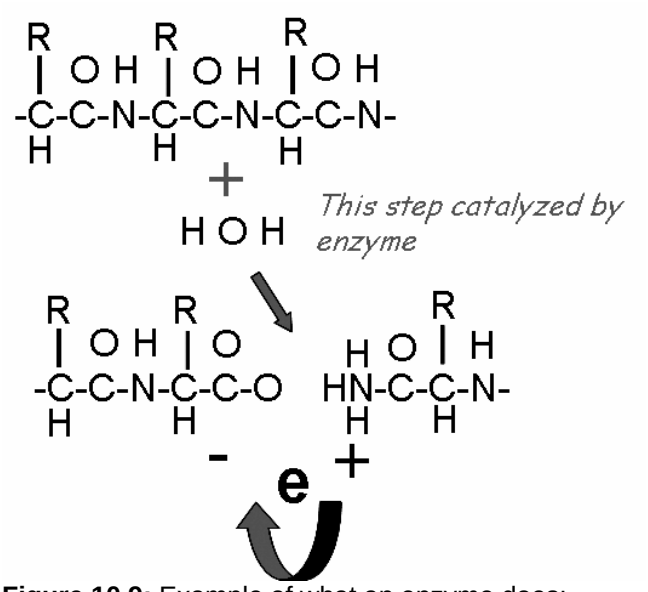


**Figure 10.9:** Example of what an enzyme does: Peptide bond hydrolysis. A protease will bind a protein at a site to be cleaved (specific to a particular bond) and then catalyze an electron transfer reaction involving the peptide backbone and a water molecule. In the absence of the enzyme, the transition state for the reaction is reached once every 500 years – in the presence of the enzyme, it occurs about every ms!

$$D \approx \frac{4.14\times10^{-21}\,J}{56\times10^{-12}\,Ns/m} = 74 \text{ x } 10^{-12}$$

$m^2/s$.

Thus the root mean square displacement owing to diffusion is about 8.6 $\mu$m in one second.

**10.4 What enzymes do**

Enzymes are the functional machines of living systems. They can produce motion using the adenosine triphosphate produced by the mitochondria as a chemical energy source. They can also serve as catalysts, even in the absence of an energy source, and here we focus on just one type of activity, the cleavage of the peptide bond, as an example of enzyme function.

Peptide bonds are broken by means of a chemical reaction with a water molecule, referred to as *hydrolysis*. This reaction is illustrated

in Figure 10.9. It is an electron transfer reaction involving the active participation of a water molecule. Two protons from a water molecule combine with the nitrogen of the peptide bond to form a positively charged amine residue. The oxygen combines with the carbon of the peptide bond to produce a negatively charged carboxalate residue. This reaction can, and does occur in water without any catalysis, but the half life of a peptide bond in pure water varies from 300 to 600 years, depending upon the position of the bond and the nature of the adjacent residues.[16] (The curious reader might like to know that these data were extrapolated from much shorter experiments performed at high temperatures and pressures.) In the presence of an enzyme evolved to cleave the peptide bond, this half life is reduced to times on the order of 1 ms, representing a $10^{13}$-fold speeding up of the kinetics! (The millisecond timescale just quoted is a limiting value based on the assumption that the protein and its target are already assembled - rates will be much smaller in dilute solutions were enzymes and their substrates have to find one another.) Astounding though this speeding up is, this millisecond timescale is actually rather slow when compared with the intrinsic fluctuation-rate of proteins.

The intrinsic fluctuation rate of proteins can be estimated as follows: imagine the protein deformed into a non-equilibrium geometry through the application of a force *F* that suddenly goes to zero at time *t=0*. The strain in the protein will decay according to the result given in Appendix D for a highly damped system with time constant

$$\tau' = \frac{6\pi a \eta}{\kappa}. \qquad (10.5)$$

Using the value 56 x $10^{-12}$ N s/m for $\gamma$ and taking the values for $\kappa$ deduced from *E* (Table 10.1 and equation 10.4) we find values for $\tau$' that range from ca 6 ps (tubulin) to 6 ns (elastin). Another way to estimate these times (see Chapter 3) is to assume that the distance from an equilibrium conformation to the transition state is some fraction of the protein's size – say $\ell_t$ = 1nm – and use

$$\tau \approx \frac{\ell_t^2}{D} \qquad (10.6)$$

for the time between major fluctuations. Using $D \sim 74 \times 10^{-12}$ m$^2$/s (deduced above) yields $\tau \sim 1.4 \times 10^{-8}$s, similar to the time estimated using the elastic properties of elastin.

These estimated fluctuation times are $10^5$ to $10^7$ times faster than the typical catalysis timescale of 1 ms. Thus the fluctuation required to form the transition state for hydrolysis of the peptide bond occurs only in about one in one million of the total long-range conformational fluctuations of the protein. (Very local fluctuations are much more rapid, occurring on the timescale of bond vibrations, i.e. $10^{-14}$ s.) So while enzyme catalyzed peptide bond hydrolysis appears to be almost miraculously fast (1ms vs. 500 years) it involves *extremely rare* configurations of the protein. This slow dynamics presents one of the principal impediments to simulating enzyme catalysis from first principles.

Qualitatively, however, we can describe the process of enzyme catalysis as follows: The enzyme first binds to its target site on the protein to be cleaved. This can be

an extremely sequence-specific event. To take a concrete example, the enzyme HIV protease carries out the task of chopping up a long continuous protein strand that has been synthesized by the host cell of the virus (the virus hijacks the cell's protein production apparatus by transferring its genome into that of the cell). The long protein chain has to be cut at just the right sites to make the collection of smaller proteins that will spontaneously assemble onto the HIV genome. This assembly, of genome plus protein coat, makes the new viruses that will eventually burst the cell to escape and infect other cells. Once the protease is bound at one of the correct sites for cleavage, a water molecule must also become bound at the correct position and in the correct orientation for hydrolysis of the peptide bond to occur. Once all the players are in place, bound in a stable energy minimum, little will happen until a fluctuation of the protein conformation drives the local polarization into the transition state required for the electron transfer (see the discussion of Marcus theory in Chapter 8). The reaction products form in an energetically excited state, but relax rapidly to the ground state of the products. These products are the N-terminus and a C-terminus of two peptide chains made by breaking the original chain.

In this case no energy was required to drive the reaction because the free energy of the reactants - the water molecule and the peptide bond - is higher than that of the products - the two terminal ends of the newly broken chain. What the enzyme did was to lower the kinetic barrier for this reaction to occur. Based on the discussion above, we will assume that the attempt frequency is about $10^8$ Hz. Using $\Delta G = k_B T \ln\left(\frac{f_0}{f}\right)$, a reaction rate of $10^3$ Hz implies a free energy barrier of about 11.5 $k_BT$ (i.e., 0.3eV or 6.9 kcal/mol). If we assume the same attempt frequency in pure water (probably not a good assumption) the kinetic barrier for the uncatalyzed reaction would be about 42 $k_BT$ (i.e, 1eV or 25 kcal/mol).

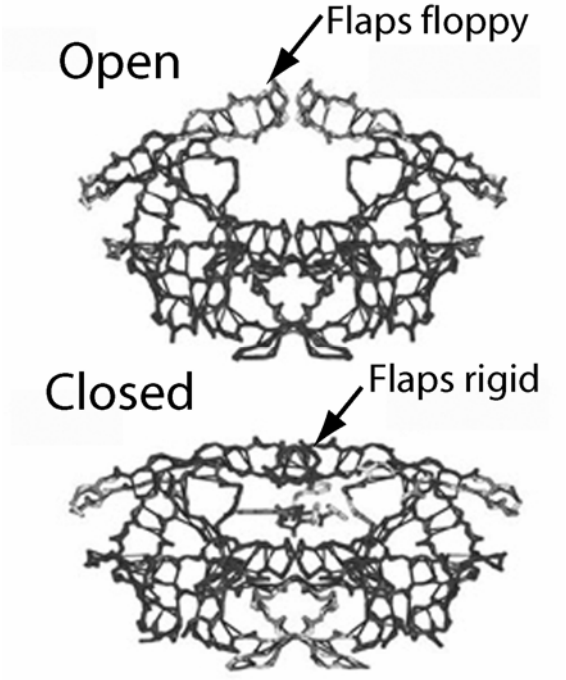


**Figure 10.10:** Large scale motion of a protease occurs at the site where the peptide to be cleaved passes through a cavity in the protease. Presumably this structure puts Brownian motion where it is needed to catalyze the electron transfer reaction. (Images of HIV protease from Michael Thorpe)

If catalysis is such a rare event, it might appear to be impossible to locate the protein confirmations responsible for it. Certainly conventional molecular dynamics is not up to the task. A geometric approach to the flexibility of complex structures like proteins has been developed[17] and this identifies the large-scale cooperative motions that are possible. Interestingly they are generally few in number and their nature can be quite evocative of the function of a protein. Figure 10.10 shows two frames from a movie of the random fluctuations of HIV protease

made with this geometric approach. The hole in the structure at the top of the Figure surrounds the catalytic site where the protein backbone to be cleaved passes through the enzyme. Clearly the large-scale fluctuations of this enzyme occur at the active site. This is because the bonding in this protein makes other degrees of freedom much more rigid. The protein might be thought of as a 'mechanical focusing system' for Brownian motion. The fluctuations are still quite random, but the internal mechanics of the protein have evolved so as to ensure that the fluctuations are largest where they need to be, in order for the protein to carry out its function.

We turn next to biomolecular machines that require an energy input in order to direct their motion.

**10.5 Powering bio-nano machines: where biological energy comes from.**

Living organisms do useful directed work which requires a source of fuel. The universal biological energy source is a molecule called adenosine triphosphate (ATP) and all cells contain machinery for ATP synthesis. In animals, this machinery is contained in the mitochondria (reviewed in Chapter 6). In the mitochondria, electrons (released from a pyruvate molecule in the course of degradation to carbon dioxide) are eventually transferred to oxygen, reducing it to form water. The energy released in this process is used to develop a proton gradient across a membrane. Proton flow, driven by the chemical potential gradient, powers an enzyme called *ATP synthase* which synthesizes adenosine triphosphate from adenosine diphosphate and available phosphate ions. Animals obtain carbohydrate food molecules by eating plants, and these carbohydrates are degraded to pyruvate by machinery outside the mitochondria. Plant cells obtain their energy directly from sunlight using organelles called chloroplasts. The chloroplasts contain a photoreaction center that uses the reaction of photo generated energetic chemical species to produce a proton gradient across the membrane of the chloroplast. This drives the same molecular machine used in mitochondria – ATP synthase - to synthesize ATP from ADP. In plants, the ATP is used to drive the energetically unfavorable synthesis of carbohydrates. Nature has evolved a common pathway for making a common fuel (adenosine triphosphate) from a proton gradient across the cell membrane in both plants and animals. Thus, in nature, the source of power is a *protomotive* force, in contrast to the electromotive force that drives the electrical appliances that we humans make. To summarize: the energy chain starts with sunlight driving the synthesis of adenosine triphosphate in plants which synthesize carbohydrates that are then eaten by animals who use the carbohydrate to synthesize adenosine triphosphate in mitochondria.

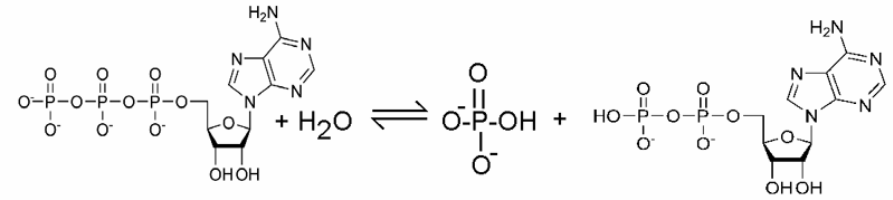

**Figure 10.11:** Hydrolysis of adenosinetriphosphate (ATP) to adenosinediphosphate (ADP) as the source of energy for some enzymes and all molecular motors. The ATP is made in the mitchondria as described in Chapter 6.

**10.6 Adenosine triphosphate - the gasoline of biology**

The hydrolysis of adenosine triphosphate (illustrated in Figure 10.11) is a universal source of biochemical energy for ATP dependent enzymes and molecular motors. It is worth taking a close look at the structure of these molecules. Hopefully they are familiar! The linked adenine, sugar and phosphate are almost identical to some of the components of the DNA molecule we discussed in Chapter 6. This is a wonderful illustration of the way in which nature re-uses molecular components, and an encouragement to those who fear that biochemistry requires unlimited memorization of structures.

Referring to Figure 10.11, the hydrolysis reaction proceeds by the water mediated cleavage of the bond between the second and third phosphate groups (reading outwards from the adenosine these groups are referred to as the $\alpha$-, $\beta$- and $\gamma$-phosphates). This reaction releases a phosphate ion plus adenosine diphosphate. The forward reaction, in which adenosine diphosphate is produced, usually proceeds in concert with some desired motor function. That is to say the hydrolysis of adenosine triphosphate is directly coupled to some conformational change in a motor protein. The equilibrium constant for this reaction (it is dimensionless – Chapter 3.16) is a given by

$$K_{eq} = \frac{[ADP][P_i]}{[ATP]} = 4.9x10^5. \qquad (10.7)$$

Here, the various concentrations represent the sum of all the ionic species that contain an ATP, ADP or phosphate, for example, $MgATP^{2-}$, $ATP^{4-}$, etc. Water is not included in the calculation of the rate constant because the reaction is assumed to take place at constant pH. By solving $K_{eq} = \exp -\frac{\Delta G}{k_B T}$ for $\Delta G$, we obtain the available free energy from the hydrolysis of a single adenosine triphosphate as 12 kcal/mol. This energy is summarized in various units in Table 10.2.

The reverse reaction, the phosphorylation of adenosine diphosphate, is carried out by the molecular motor, ATP synthase. In this case the reaction requires energy and this is supplied by a proton gradient across the membrane containing the ATP synthase. As protons defuse through the motor from the low pH aside of the membrane to the high pH side of the membrane, they turn a 'turbine' inside the motor. The rotation occurs in discrete steps of 120° and a bound adenosine diphosphate is phosphorylated to form adenosine triphosphate at each step.

12kcal/mole
20kT at 300K
0.52eV/molecule
$8.3x10^{-20}$J/molecule
82pN.nm

**Table 10.2** Energy from ATP hydrolysis in various units.

**10.7 The thermal ratchet mechanism.**

One mechanism for molecular motors is called the 'thermal ratchet,[1] outlined in Figure 10.12. Here we will consider a 'downhill' process in which the hydrolysis of ATP directs the motion of a molecular motor by formation of the (lower energy) ADP. The thermal ratchet model has similarities to the Marcus model for electron transfer, discussed in Chapter 8.

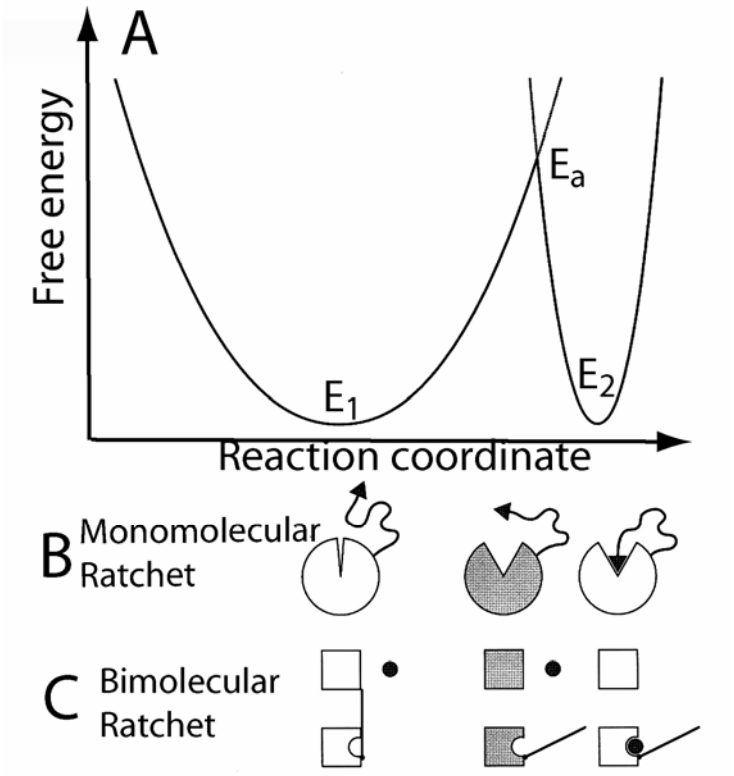


**Figure 10.12:** The thermal ratchet mechanism for molecular motors (adapted from Figure 5.7 of Howard[1]).

The reactants at $E_1$ on the left of the Figure are the motor protein bound with a water molecule in the presence of a significant concentration of ATP. The products at $E_2$ on the right of the Figure are a phosphate ion, ADP and a motor that has advanced one step. Binding of ATP and protein fluctuations (on a timescale given by equation 10.5) eventually result in the transition state ($E_a$ in the Figure) being reached. At this point, the electron transfer reaction involving the water molecule and ATP occurs, and the ATP is hydrolyzed to ADP. This is no longer stably bound, so it is released, stabilizing the (one-step-advanced) motor against backward fluctuations. This principle of operation is illustrated schematically for a monomolecular reaction in Figure 10.12b by a fluctuating cleft in the molecule which, when opened, allows the flexible ratchet to bind in the cleft, preventing it from closing in the face of future fluctuations. This is not a very good model for the multi-particle reaction just described, and a schematic bimolecular ratchet is shown in Figure 10.12b. Here binding of the small molecule prevents the 'lid' from closing over the binding pocket. The real operation of molecular motors is considerably more complex involving as it does several reactants and several products, but the 'ratchet' is, apparently, the release of ADP after hydrolysis of ATP.

The frequency of crossing, $k$, is determined by the approach rate multiplied by the exponential of the free energy corresponding to the transition state ($E_a$):

$$k \approx \frac{\kappa}{6\pi a \eta} \exp\left[\frac{-E_a}{kT}\right]. \qquad (10.8)$$

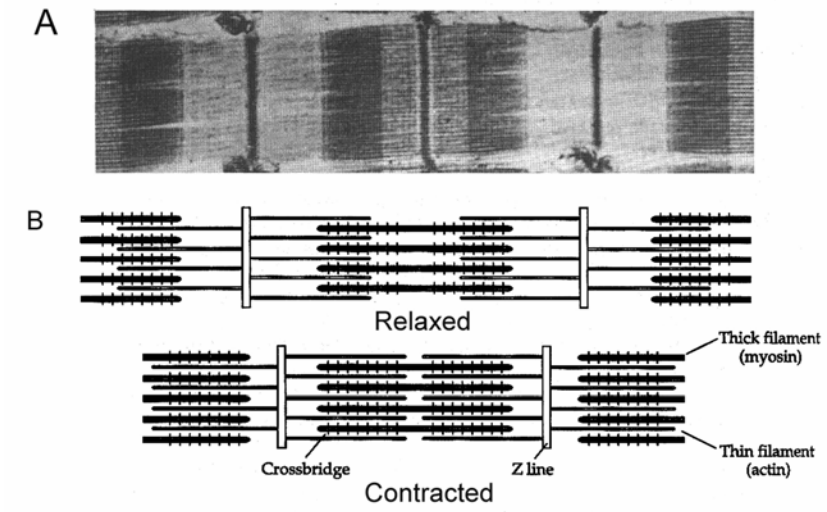


**Figure 10.13:** Molecular motors in muscle sarcomeres. A TEM image of sarcomeres is showm in A. The crossbridge model is illustrated in B. The myosin 'walks' along the interleaved actin filaments to draw the crossbridges together, resulting in muscle contraction. (adapted from Figure 1.1 of Howard[1])

In this expression, $E_a$ is a compound quantity that incorporates free energy donated by ATP hydrolysis together with a work expended in the power stroke of the motor.

**10.8 Types of molecular motor.** Muscles are probably the biological 'motors' most familiar to us. The motor function is carried out by an actin-myosin complex found in muscle tissue. The structure of the active component (the sarcomeres) of smooth muscle tissue is shown in Figure 10.13. It consists of a thick filament to which are attached many myosin molecules. These thick filaments are interdigitated with thin filaments, fibers composed of bundles of the protein actin. During muscle contraction, the thick filaments move against the thin filaments to pull the fixed components of the sarcomere together. In some animals, muscle tissue can contract by more than 20% in length in a period of tens of milliseconds.

The active motor component of the muscle tissue is found in the cross bridges that connect the thin filaments to the thick filaments. These cross bridges are made from myosin molecules which consist of a long stalk that is permanently attached to the thick filament, and a pair of head units that transiently connect the actin filaments. An electron micrograph of myosin molecules is shown in Figure 10.14A. The arrows point to the head units at the end of the long stalks. A view of the crystal structure of the mysosin head, looking in from the two domains (lower right) that bind actin, is shown in 10.14B. A schematic view of the motion along the actin filament is shown in 10.14C. ATP binding (first two steps) causes disengagement from the actin with motion down one step which is stabilized by release of the phosphate removed in the degradation of ATP to ADP, followed by release of the ADP. 10.14D shows a modern view of the details of the walking motion. One of the heads becomes unbound from the actin filament on binding of ATP. The second head remains bound, leaving the first free to diffuse until it finds a binding site downstream of the bound leg. Strain in the bound leg biases the diffusion towards the downstream direction. The new configuration is stabilized by release of phosphate and ADP, leaving the myosin translocated by one repeat of the actin filament and ready for the next ATP binding event.

Myosin motors consume one ATP molecule per power stroke and contact the actin filament only transiently. That is to say they have a small duty cycle, with only a

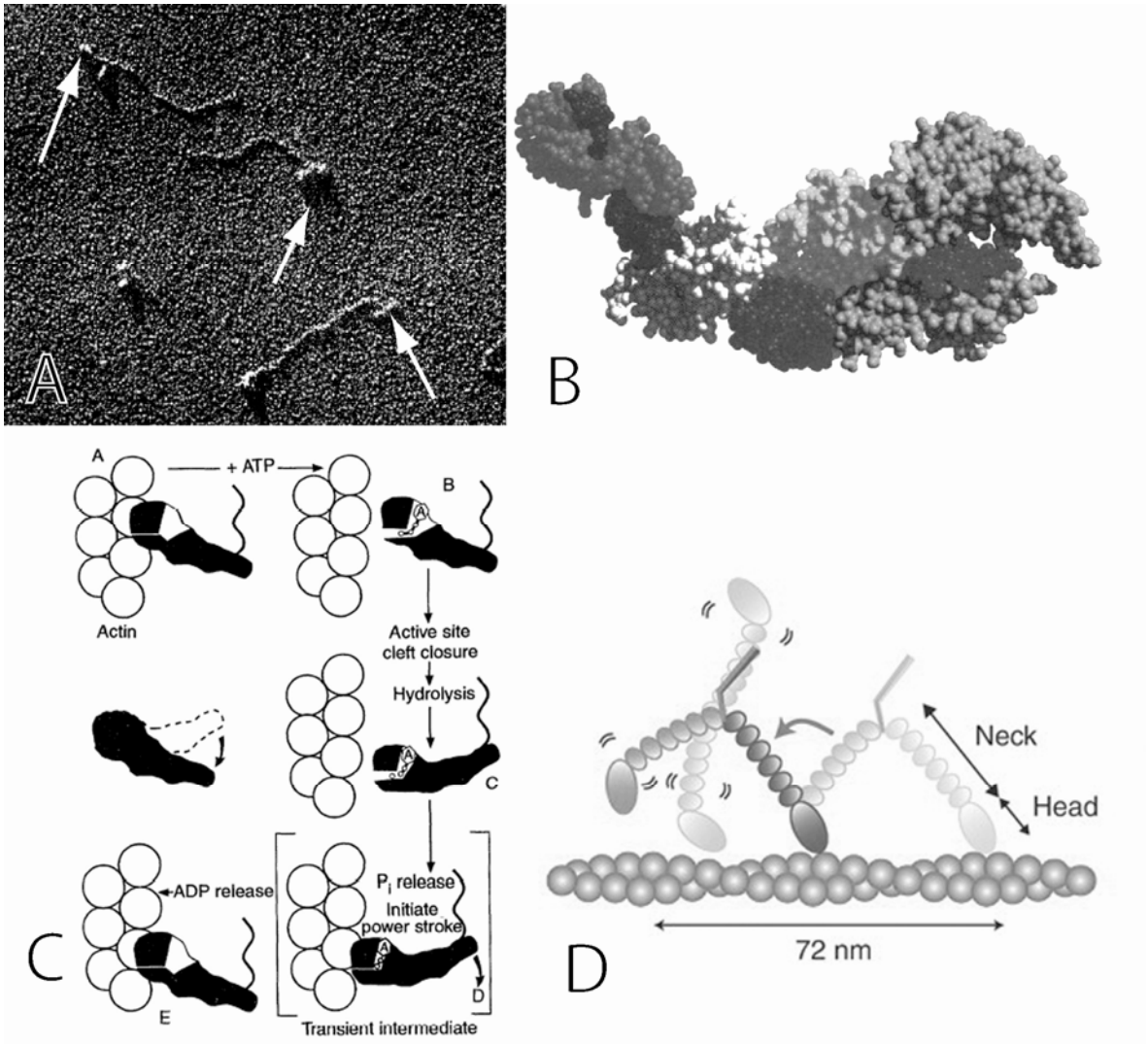


**Figure 10.14:** The actin-myosin molecular motor complex. (A) Electron micrograph of myosin molecules. The arrows point to the heads that engage and walk on the actin filaments. (From Figure 1 of Elliot et al.[5]). (B) Crystal structure of the myosin head unit. The two actin binding domains are in the lower right of the figure. (From Figure 3 of Rayment et al.[11]). (C) ATP hydrolysis and the power stroke in the actin-myosin complex. The figure reads from A to E showing motion down one repeat of the actin filament. (From Figure 5 of Rayment et al.[12]). (D) Detail of the walking motion of myosin on actin showing a step from right to left on the figure. (From Figure 1 of Shiroguchi and Kinosita.[13])

few of the myosin motors in contact with the actin filament at any time, shuffling the actin filament along. ATP binding and hydrolysis is believed to cause strain only when the myosin head is attached to the actin filament. The power stroke distance per binding event per myosin head is about 5 nm and the measured force generated is about 1.5 pN. A lever action is generated by the long attached stalk being moved by the rotation of the head, and this results in a motion of the actin filament of about 36 nm. Thus the work

done per ATP is about 1.5 pN x 36 nm or 54 pN.nm (an efficiency of about 65% - see table 10.2).[1]

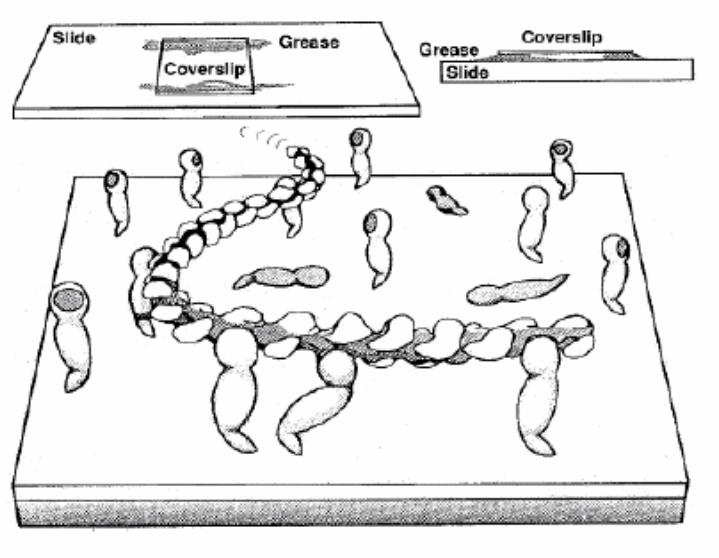


**Figure 10.15:** Experimental setup of an in vitro motility assay. (Top) Front (left) and side (right) view of a flow cell. Most in vitro motility assays are performed using flow cells, i.e., open systems that allow solutions to be exchanged by flowing them through the cell. In its simplest design, a flow cell consists of a coverslip that is mounted on a glass slide using silicon grease as a spacer (i.e., to achieve a separation of ~0.5-1.0mm). (Bottom) In this cartoon, myosin heads are depicted adsorbed to a nitrocellulose-coated coverslip and are in the process of moving an actin filament. The coverslip is shown upside down, i.e., the myosin molecules and the actin filaments face the slide. (adapted from Figure 9 of Warrick & Spudich.[10])

In-vitro visualization of the action of myosin molecules has been carried out by functionalizing a microscope slide with myosin molecules so that their stalks are attached to the slide (Figure 10.15). Dye-loaded actin molecules were then placed on the slide, on top of the myosin layer. When the system was fed ATP, the actin filaments were observed to start to move on the slide (using fluorescence microscopy to visualize the actin filaments). The myosin moves on actin filaments in one direction only (i.e. the actin molecules are polarized) but the actin filaments will move in random directions (Figure 10.16) because the system is randomly assembled. Small ATP driven linear motors have been made by orienting the protein components appropriately.[18]

Myosin is an example of a molecular motor that works only in concert with many other molecules. By itself, it will detach from an actin filament, because its duty cycle (the fraction of the power stroke time for which the molecule is attached) is small. An example of molecules that remain attached throughout the motor cycle are those that transport material by walking along microtubules. Their function requires that they remain attached. Microtubules are the self assembled 'freeways' of the cell. Composed of proteins called tubulins, they spontaneously assemble and disassemble as needed to move materials into various regions of the cell. Kinesin, and dynein are two different types of motor protein that 'walk' along microtubules. Both have been studied by single molecule techniques and a fairly detailed model of kinesin motion is now available.[19] Kinesin appears to walk in a hand over hand fashion whereby one the head will attach, the second let go, passing the first and attaching at the next tubulin subunit along the microtubule, whereupon the first headed detaches, and the cycle is repeated. In this case the duty cycle of the motor is at least 50% for each head so that the kinesin molecule always remains attached to the microtubule. Like myosin, kinesin hydrolyzes

one ATP molecule a step. Kinesin can generate a force of up to three pN and moves 8 to 16 nm a step (an efficiency of 30 to 60%).[1]

The third kind of molecular motor we will consider is a *rotary* motor, an analog of the familiar electric motor. We will focus on ATP synthase, already discussed as the molecular machine that synthesizes ATP from ADP using a proton gradient. Remarkably, the motor can work backwards as well. It uses ATP, converting it to ADP, to generate rotary motion which pumps protons. Rotary motors are found elsewhere in nature, such as those that drive the flagellae of swimming bacteria. The structure of the ATP synthase motor is shown schematically in Figure 10.17. The

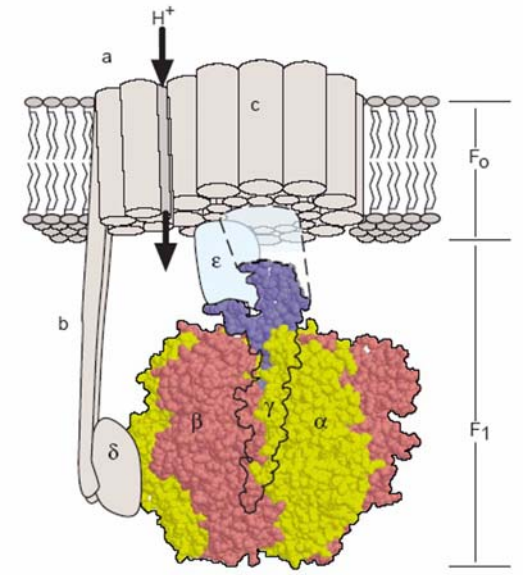


**Figure 10.17:** ATP synthase. The F1 unit can be driven into rotation by a flow of protons through the enzyme driven by the proton gradient across the cell membrane. The rotation is accompanied by the conversion of ADP to ATP. Conversely, a high concentration of ATP will drive the motor into rotation in the other direction as the ATP is hydrolyzed to ADP. (From Figure 1 of Wang and Oster.[6])

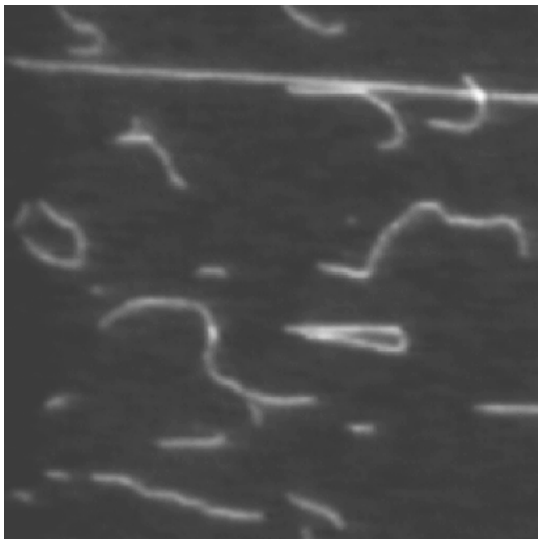

**Figure 10.16:** Fluorescently labeled actin filaments imaged with an optical microscope. They are attached to myosin heads (as shown in Figure 10.15) and move over the surface when ATP is introduced into the system. (From http://www.mih.unibas.ch/Booklet/Chapter3/Fig2large.gif)

rotating head unit sticks out into the region of higher pH where it phosphorylates ADP. A base unit (which translocates protons from the low pH side of the membrane to the high pH side) is embedded in the membrane. The coupling of proton translocation to motor rotation and ADP phosphorylation is shown in Figure 10.18.

The rotary motion of individual ATP synthase molecules has been visualized directly by attaching gold beads to one part of the motor (a fragment corresponding to the moving part of the proton translocation apparatus) while the other part (in this case the ATP hydrolysis subunit) was fixed to a glass slide. Movement of the gold bead was detected by laser dark-field optical imaging (where the illumination is aimed so that it can only enter the collection optics by scattering). The experimental arrangement is illustrated in Figure 10.19. Experiments like this have shown that the motor takes three 120° steps to complete one rotation,

hydrolyzing one ATP per step. The steps are subdivided into 80° and 40° sub-steps, corresponding to the binding of ATP (80°) and its subsequent hydrolysis (the remaining 40°).[7, 20, 21]

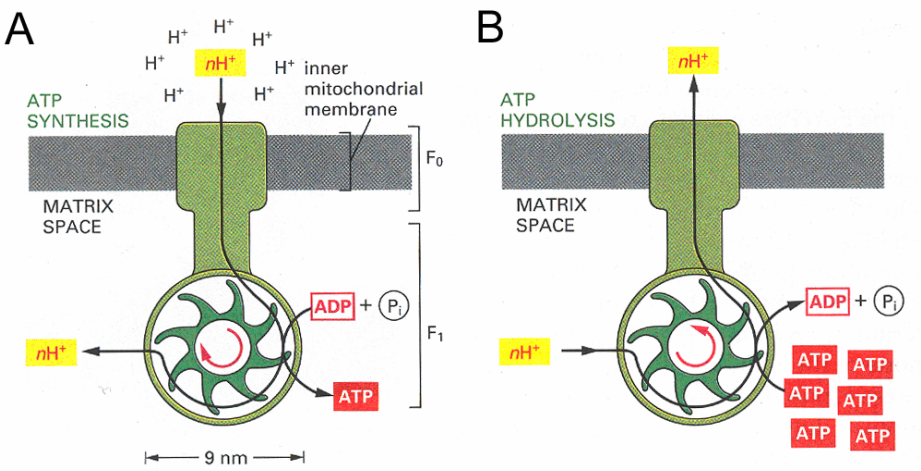


**Figure 10.18:** (A) showing operation of ATP synthase as (A) a motor driven by a proton chemical potential gradient that synthesizes ATP by phosphorylation of the less energetic ADP and (B) a motor, operated in the reverse direction in the presence of a large excess of ATP which is hydrolyzed to ADP to turn the motor. (From Figure 14.27 of The Molecular Biology of the Cell".[4]

**10.9 The central role of fluctuations in biology.** Fluctuations matter at every level of biology. The death of a single dinosaur is a critical event to the species if that dinosaur is a member of the last breeding pair in existence. Our modern understanding of molecular biology can appear so deterministic that it becomes easy to overlook the role of fluctuations and randomness in molecular biology. The central dogma, that genes code for the proteins that self assemble into the organism, leaves out the role of random fluctuations. The basic randomness of biology is

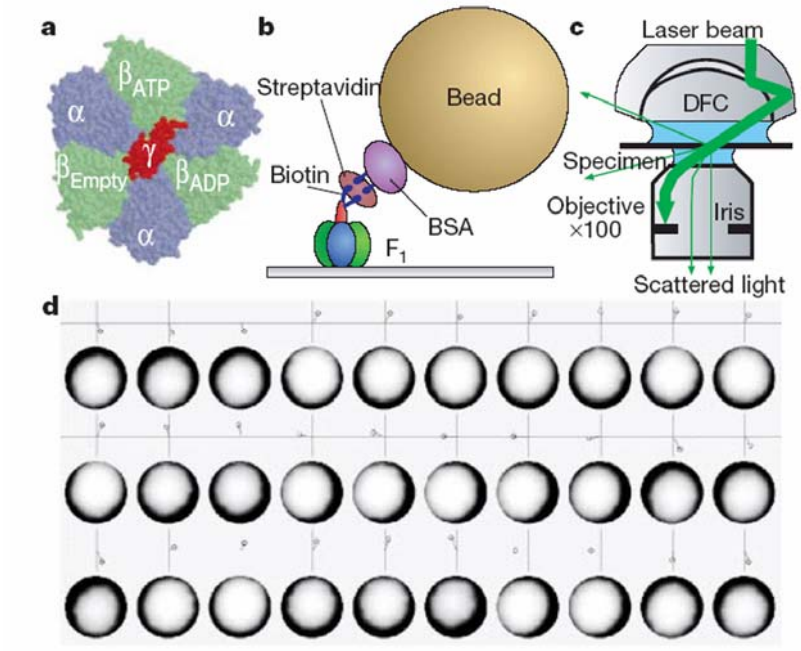


**Figure 10.19:** Experiment to visualize the rotation of the F1 portion of ATP synthase. The γ subunit (a) is mutated to incorporate biotin that binds to a streptavidin that also binds biotinylated BSA attached to a 40 nm diameter gold bead (b). Rotation of the bead is observed directly by laser-dark-field microcopy (c) yielding a series of scattered light images like those shown in (d). (From Figure 1 of Yasuda et al.[7])

readily illustrated by comparing identical twins or clones, animals with identical genomes but, as it turns out, quite distinct phenotypes. Figure 10.20 shows of the thumbprints of identical twins (A). They are clearly quite distinct. Figure 10.20B shows (right) CC, the first cloned cat, with her genetically identical offspring (left). Mother and daughter are similar in marking, but clearly not identical. Thus, the fluctuations so essential to chemical reactions and enzyme function must play a role in the development also.

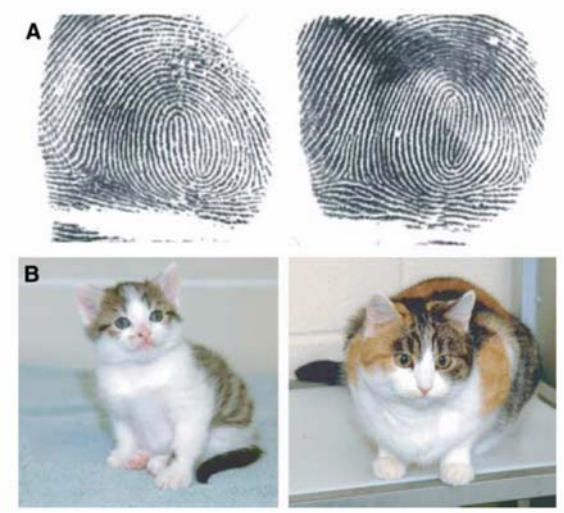


**Figure 10.20:** Manifestation of fluctuations in gene expression in biology: (A) The thumbprints of identical twins (B) CC (right) and her daughter, the first cloned cat. (From Figure 1 of Raser and O'Shea.[3])

One source of fluctuations in developmental processes lies simply in a statistical fluctuation in the number of molecules that regulate gene expression inside a single cell. If the number of molecules is small, then the relative amount of fluctuation in the number is large (on the order of $\sqrt{N^{-1}}$). A beautiful experiment that illustrates this point was conducted by exploiting the fact that bacterial genes are often read from a circular genome, with RNA polymerases moving both clockwise and anti-clockwise to translate the DNA to messenger RNA. Elowitz et al.[9] modified the circular genome of the gut bacterium *E. Coli* to insert a pair of genes symmetrically with respect to an important control region called the origin of replication. Each of the genes contained the same promoter sequence upstream of the gene, and would therefore be expressed at the same rate if the factors that turn gene expression on (the transcription factors) acted in the same way on each gene. One gene contained the DNA sequence corresponding to a fluorescent protein called yellow fluorescent protein (YFP) and the other contained in the DNA sequence corresponding to a fluorescent protein called cyan fluorescent protein (CFP). Thus, if the rates at which both these genes were expressed was a constant from bacterium to bacterium, each of the bacteria would appear the same color (some mix of yellow and cyan) under a fluorescence microscope. The results of this experiment are shown for several different gene promoter sequences in Figure 10.21. Sequences that correspond to proteins that are produced at a high copy number in the bacterium do indeed give rise to a reasonably uniform population of colors in images of the bacteria (Figures 10.21 B and E). However, promoter sequences associated with genes that are expressed at low copy numbers give rise to very different colors in the individual bacteria (Figures 10.21 A, D, C and F). These fluctuations are intrinsic (i.e. they do not result from variations from bacterium to bacterium, because these would affect both genes equally). The enhancement of fluctuations in gene expression associated with the production of small numbers of proteins (and therefore, presumably, small numbers of transcription factors) indicates that these are simply statistical fluctuations as the quantity $\frac{\sqrt{N}}{N}$ approaches unity.

Fluctuations have biological utility when they control the process of gene splicing described in section 10.2 of this chapter. One of the first and most vivid examples of this is the molecular operation of the immune system. In this case, the fluctuations operate at two levels. One is in the *site specific recombination* whereby gene segments are shuffled during differentiation of the B-cells that make antibodies. The other lies with the splicing process itself. The immune response is a remarkable biological phenomenon. Antibodies are capable of recognizing a foreign invader - any foreign invader - even if it has never been encountered before (though training does help and this is the basis of vaccination). The immune system has a second amazing aspect to it which is a built-in recognition of the elements that belong in the organism, a feature we will not discuss further here. How does an antibody bind to a foreign invader if it does not "know" what to look for? The answer lies in the diversity of the antibody population. The structure of a typical antibody is illustrated in Figure 10.22A (a crystal structure) and 10.22B (a schematic layout). It is a dimer, consisting of two proteins, each one of which is terminated in two separate chains linked together by a disulfide bonds. Part of each of the four chains comprising the antibody has a variable sequence. Each of the arms has a long variable region combined with a constant region called the heavy chain, and a shorter variable region (with a constant component also) called the light chain, Any one antibody carries two identical copies of the heavy and light chains, and the sequence of these chains is (almost) unique for each antibody. The genetic coding process behind this variability is outlined in Figure 10.23. The variability comes about as a consequence of the random splicing of DNA as the various exons are shuffled during production of the cells that make antibodies (B-cells). Specifically, the variable parts of each chain are encoded for

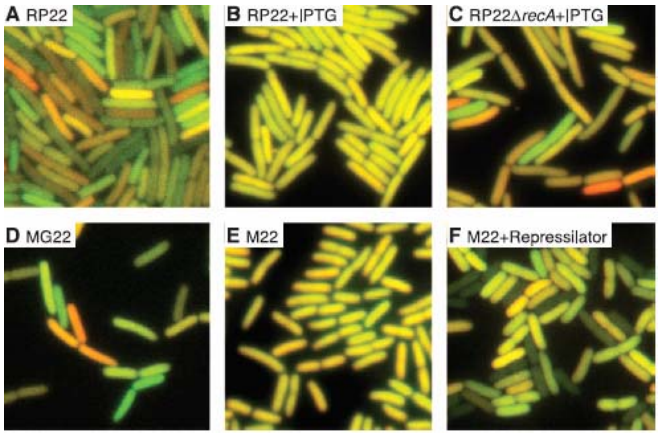


**Figure 10.21:** Stochastic gene expression in single cells. Genes were labeled with sequences that coded for green fluorescent proteins and cyan fluorescent proteins in such a way that constant expression of the genes would produce equal amounts of fluorescent protein, resulting in an average yellow color. This is what is seen in genes expressed at high levels (B and E) but not the case for genes expressed at lower levels (A, C, D and F). The colors reflect stochastic fluctuations in gene expression that results from the small number of molecules in each cell involved in gene regulation of low expression levels. (from Figure 2 of Elowitz et al.[9]).

| Segment | κ-Light Chain | Heavy Chain |
|---|---|---|
| V | 300 | 500 |
| J | 4 | 4 |
| D | 0 | 12 |

**Table 10.3** Approximate number of gene components for a light chain (V = variable, J = joining) and a heavy chain (D = diversity) resulting in 300x4x500x4x12 possible variants from these components alone.

by blocks labeled "V" (for variable). These are joined to the constant region (C) by another block (labeled "J" for joining). In addition, heavy chains have a further set of gene blocks called "D" (for diversity). To add to the mix, there is more than one type of light chain 'constant' region. The first step occurs by means of a process called site directed recombination. The exons have adjoining sequences that trigger shuffling (but just within antibody genes) by proteins that direct recombination. The process occurs as B cells develop from germ line cells. The result is that each B-cell has a different ordering of the various V (or D) regions in its genome. Thus, when translated into

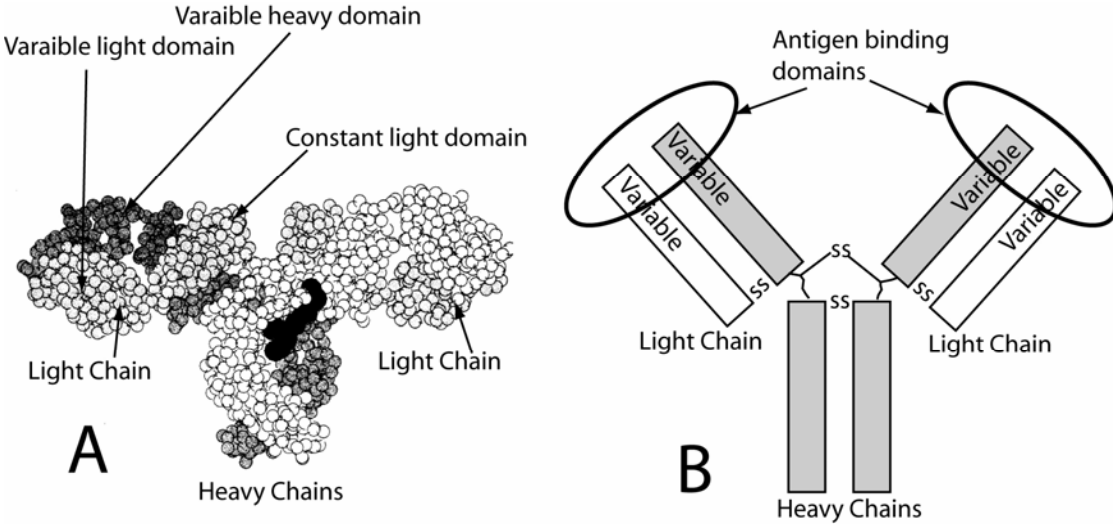


**Figure 10.22: (**A) Crystal structure of a human Immunoglobulin (antibody). (Structure from Figure 3 of Silverton et al.[2]) (B) Schematic layout of the chains in the antibody. The 'ss' show disulfide bonds that hold the chains together.

messenger RNA, the particular V block adjacent to a J block is random (as is the choice of J block itself). Extra J regions in the mRNA transcript are removed by splicing to produce a final V-J-C mRNA to be translated into protein. The resulting diversity for just one type of light chain (the so called $\kappa$ light chain) is shown in Table 10.3 (these data are for the well-studied mouse immune system). This one combination of chains produces about $3 \times 10^7$ variants. Added to this, the splicing process itself is somewhat variable, so the actual diversity exceeds $10^8$. This army of random peptides is almost guaranteed to contain a variant that will bind an invading particle (called an antigen). The binding of an antibody to an antigen produces a cascade of signals that results in the rapid *cloning* of the B-cell that made the right antibody. Antibodies are attached to the surface of B-cells (one cell for each sequence of antibody) and binding of the antigen to the antibody causes the cell (containing the unique DNA sequence for its particular antibody) to reproduce rapidly, producing a large population of antibody clones with identical sequence. These will then bind other foreign invaders that are identical to the one that triggered the first reaction, which, when combined with other components of the immune system, results in destruction of the invading population. Random splicing of the genetic codes for the antibodies is therefore an essential component of the immune system and it is combined with a 'natural selection' based on cloning of the successful B-cell. Darwinian evolution operates inside each one of us every time we fight off a cold.

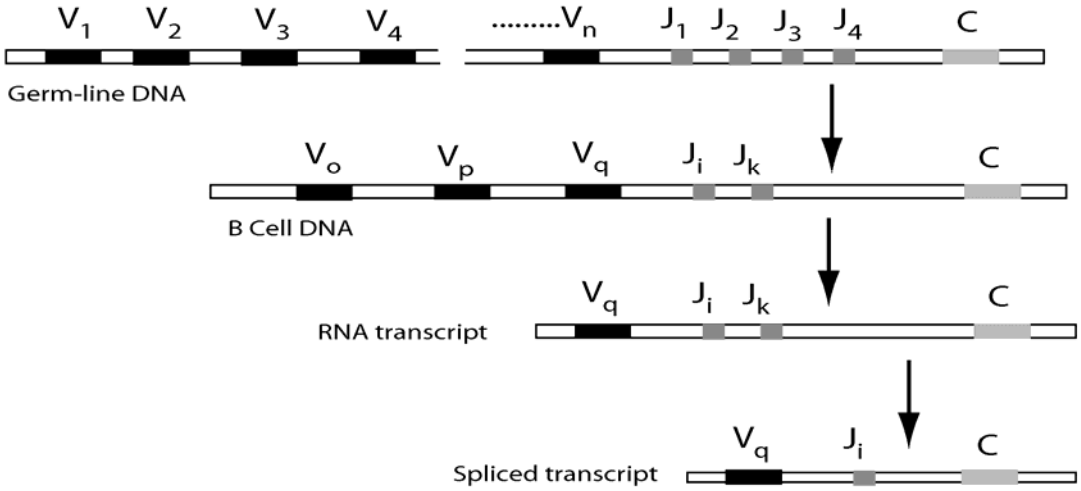


**Figure 10.23:** Assembly of mRNA for the light-chain of an antibody. Exons are shown as

A process somewhat like this appears to operate in the development of physiologically complex structures like neural networks and we end with some recent work that demonstrates the crucial role of molecular diversity in neural development. Cells in complex organisms find the right place to go to partly as a result of specifically sticky molecules that are expressed on their surfaces. These molecules, called cell adhesion molecules, bond to targeted receptors on other cells, helping to form the complex three-dimensional arrangement required for the development of an organ. One particular cell adhesion molecule of importance in gluing neurons together is called Dscam. It was reported recently that the gene for this protein has 38,016 splicing variants, implying that there are up to 38,016 different types of Dscam involved in 'gluing' neural networks together (at least in the fruit fly with this study was carried out). Is this diversity essential or is it just a complication? The same group that reported this diversity found a way of modifying the gene for Dscam so that its essential components were retained but the number of splicing variants that could be made was reduced (to 22,176).[8] Fruit fly embryos with the modified Dscam gene could not develop proper neural networks as evidenced in studies of the networks that are responsible for controlling bristles. A summary of these results is given in figure 10.24. The conclusion is that a molecularly diverse population of cell adhesion molecules is *essential* for the successful development of the gross physiology of the organ. Presumably while each normal fruit fly bristle controlling network looks essentially identical under an optical microscope, they differ dramatically one from the other in the molecular composition. In other words, there are probably many 'right ways' to develop successfully, given *adequate population diversity* in the first place. Apparently, reduced population diversity does not allow for successful in evolution of an organ in the face of environmental pressures on the developing embryo.

**10.10 Do nanoscale fluctuations play a role in the evolution of the mind?**

The three themes of this book have been the incredible density of information encoded at the nanoscale, the role of quantum phenomena on this scale, and the importance of

fluctuations in small systems, with the consequent emergence of new phenomena. As simple a process as electron transfer solution can now be seen as an exemplar for emergent phenomena. But it appears that it is biology that has produced one of the first empirical tests of the quantitative degree of diversity required to underpin an emergent phenomenon (22,176 Dscam molecules were too few for neural networks to emerge but 38,016 different molecules was adequate). These experiments suggest that complex organs like neural networks develop as a result of environmental selection pressures acting on a randomly formed network. Is this true also for the formation of the mammalian brain? And if it is, could it be that the mind itself emerges through a process of Darwinian selection? If so, it is a special type of selection, depending not upon diversity in the members of a species, but rather upon molecular diversity arising from the fluctuations inherent in the nanomachines that put the molecular components of biological systems together. It is a process of Darwinian selection that operates over the timescale of the development of individual animals as opposed to the original Darwinian selection that operates over a timescale of many generations.

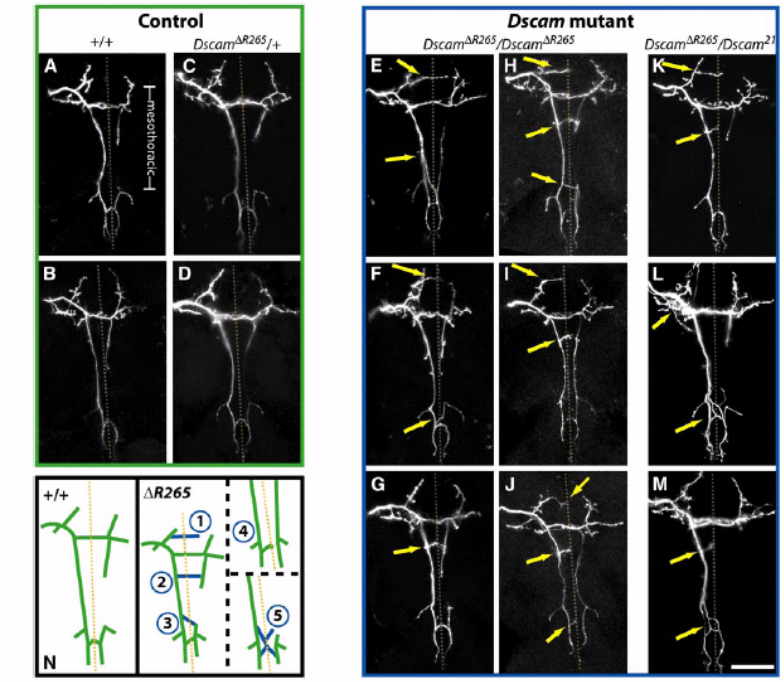


**Figure 10.24:** Stochastic splicing as the basis for brain development? A gene that controls fruit fly neuron adhesion through the cell adhesion molecule Dscam has 38,016 splicing variants. Nonetheless, this diverse protein results in the development of normal robust neural networks A-D (though each must differ enormously at the molecular level). When the splicing diversity is reducing by genetic engineering (to 22,176 variants) defective networks are formed (E-M, defects highlighted by arrows) (From Chen et al.[8]).

One of the most lucid exponents of this molecular Darwinism as the driving force behind the origin of the thinking mind is Gerald Edelman, the brilliant scientist who first conceived of the gene shuffling mechanism responsible for the development of the random library of antibodies. He was the recipient of the 1972 Nobel Prize in physiology or medicine for this discovery. Edelman has proposed that the mind develops as the environment acts upon a complex multiply connected assembly of neurons in the human brain, causing certain groups of connections to stabilize and others to wither. Central to the generation of an adequately diverse neural network is a random population of cell

adhesion molecules, substrate adhesion molecules (molecules that link cells to intercellular material) and cell junctional molecules (molecules that fold cells linked by cell adhesion molecules into sheets). In Edelman's view, the mind emerges as the environment acts upon an interconnected and *adequately random* network.[22]

The importance of fluctuations in small systems has been one of the dominant themes of this book. Here we see that fluctuations in nanomachines – the spliceosomes responsible for assembling the final messenger RNA – produce the molecular diversity needed for the correct assembly of neural networks, and perhaps for the evolution of the mind. Thus perhaps it is true that the same mind that builds simplistic, reductionist models of the physical world has evolved from the chaotic processes in small systems where the scale of energy fluctuations is comparable to the energy that holds components of the system together. Einstein's famous rejection of quantum mechanics ("God does not play dice with the universe") may have come from a mind that required randomness in order to develop the ability to make this famous critique!

## 10.11 Bibliography

*B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts and J.D. Watson, .Molecular Biology of the Cell. 1994, New York: Garland Press.*
The classic comprehensive beginner's text in molecular biology. You'll find everything here explained from the ground up. With this book, and regular updates from review articles, you can become an expert in molecular biology.

*J. Howard, Mechanics of motor proteins and the cytoskeleton. 2001, Sunderland, MA: Sinauer Associates.*
This book is much broader than its title. A masterful and easy-to-follow treatment of the biophysics of molecular machines.

## 10.12 Exercises for Chapter 10

1. Taking the volume of each basepair to be a disk of 2 nm diameter and 0.34 nm height, how many bases could fit into a nucleus of 3 μm diameter?

2. Translate the following RNA sequences (reading from left to right) into amino acids sequences, and classify the resulting peptides as hydrophillic, hydrophobic or intermediate:

a) GUCGUCCUAAUG

b) AACCACAAA

c) GCCACAUGG

3. Which codon is a stop sequence in GUCGUCUAGCUAAUG ?

4. Which amino acids have only one codon? What properties do these amino acids have?

5. A particular gene has 5 exons, two of which are identical. How many distinct ways can the mRNA be spliced?

6. A 10 kD globular protein has a Young's modulus of 1 GPa. What is its drag coefficient in water at 25 °C? What is its diffusion constant? What is its viscous relaxation time (use equation 10.5). Compare this to the relaxation time calculated using equation 10.6.

7. Using the relaxation time calculated using equation 10.5 for the protein in question 6, calculate its turnover rate as an enzyme if the transition state for catalysis is $10\,k_BT$ .

8. What is the spontaneous equilibrium concentration of ATP in a solution of 1 mM ATP and 1 mM phosphate ions?

6. How far does the head of a myosin molecule move in one step? What is the efficiency of this step? What is the force generated at the *head*? (The force quoted in the Chapter is as multiplied by the lever action of the stem).

7. How much does ATP synthase rotate on ATP binding? And how much more after release of ADP and phosphate.

8. If a force is applied 2 nm off the axis of ATP synthase, how much force would be required to stall it?

9. Given that the largest color fluctuations in Figure 10.21 correspond to a 20% variation in gene expression,, how many molecules are involved n regulating the most variable gene expression?

10. Suppose that 50 of the V regions involved in light chain diversity (Table 10.3) were identical. What would the final diversity of the combination of light and heavy chains be?

**References for Chapter 10**


1. J. Howard, Mechanics of motor proteins and the cytoskeleton. 2001, Sunderland, MA: Sinauer Associates.
2. E.W. Silverton, M.A. Navia and D.R. Davies, Three-dimensional structure of an intact human immunoglobulin, Proc. Nat. Acad. Sci., 1977. **74**: 5140-5144.
3. J.M. Raser and E.K. O'Shea, Noise in Gene Expression: Origins, Consequences and Control, Science, 2005. **309**: 2010-2013.
4. B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts and J.D. Watson, Molecular Biology of the Cell Third ed. 1994, New York: Garland Press.
5. A. Elliott, G. Offer and K. Burridge, Electron Microscopy of Myosin Molecules from Muscle and Non-Muscle Sources, Proceedings of the Royal Society of London. Series B, Biological Sciences, 1976. **193**: 45-53.
6. H. Wang and G. Oster, Energy transduction in the F1 motor ofATP synthase, Nature, 1998. **296**: 279-282.
7. R. Yasuda, H. Noji, M. Yoshida, K. Kinosita and H. Itoh, Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase, Nature, 2001. **410**: 898-904.
8. B.E. Chen, M. Kondo, A. Garnier, F.L. Watson, R. Püettmann-Holgado, D.R. Lamar and D. Schmucker, The Molecular Diversity of Dscam Is Functionally Required for Neuronal Wiring Specificity in Drosophila, Cell, 2006. **125**: 607-620.
9. M.B. Elowitz, A.J. Levine, E.D. Siggia and P.S. Swain, Stochastic Gene Expression in a Single Cell, Science, 2002. **297**: 1183-1186.
10. H.M. Warrick and J.A. Spudich, Myosin structure and function in cell motility, Annual Reiews of Cell Biology, 1987. **3**: 379-421.
11. I. Rayment, W.R. Rypniewski, K. Schmidt-Base, R. Smith, D.R. Tomchick, M.M. Benning, D.A. Winkelmann, G. Wesenberg and H.M. Holden, Three-Dimensional Structure of Myosin Subfragment-1: A Molecular Motor, Science, 1993. **261**: 50-58.
12. I. Rayment, H.l.M. Holden, M. Whittaker, C.B. Yohn, M. Lorenz, K.C. Holmes and R.A. Milligan, Structure of the Actin-Myosin Complex and Its Implications for Muscle Contraction, Science, 1993. **261**: 58-65.
13. K. Shiroguchi and K. Kinosita, Myosin V Walks by Lever Action and Brownian Motion Science, 2007. **316**: 1208-1212.
14. M.T. McManus and P.A. Sharp, Gene Silencing in mammals by small interfereing RNAs, Nature Reviews Genetics, 2002. **3**: 737 - 747.
15. J. Tsai, R. Taylor, C. Chothia and M. Gerstein, The packing density of proteins, J. Mol. Biol., 1999. **290**: 253-266.
16. A. Radzicka and R. Wolfenden, Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases, J. Am. Chem. Soc., 1996. **118**: 6105 - 6109.
17. D.J. Jacobs, A.J. Rader, L.A. Kuhn and M.F. Thorpe, Protein flexibilty predicitions using graph theory, Proteins: Structure, Function, and Genetics, 2001. **44**: 150-165.
18. J.J. Schmidt and C.D. Montemagno, Bionanomechanical systems, Annual Review of Materials Research, 2004. **34**: 315-337.

19. R.D. Vale and R.A. Milligan, The Way Things Move: Looking Under the Hood of Molecular Motor Proteins, Science, 2000. **288**: 88-95.
20. M. Diez, B. Zimmermann, M. Börsch, M. König, E. Schweinberge, S. Steigmiller, R. Reuter, S. Felekyan, V. Kudryavtsev, C.A.M. Seidel and P. Gräber, Proton-powered subunit rotation in single membrane bound F0F1-ATP synthase, Nature Structural & Molecular Biology, 2004. **11**: 135-141.
21. R.M. Berry, ATP Synthesis: The World's Smallest Wind-Up Toy, Current Biology, 2005. **15**: R385-R387.
22. G.M. Edelman, Bright air, brilliant fire. 1992: BasicBooks.