Chapter 5: cell – a community of chemical reactions

background reading

Atkins, P. (1994), *The Second Law*, Scientific American Library, W. H. Freeman, New York, paperback edition.

Ball, P. (2023), *How Life Works: a user's guide to the new biology*, Picador, London.

Bray, D. (2009), *Wetware*, Yale University Press, New haven, CT.

Dawkins, R. (2005), *The Ancestor's Tale: a pilgrimage to the dawn of life*, Phoenix, London.

Gee, H. (2021), *A (very) Short History of Life on Earth: 4.6 billion years in 12 chapters*, Picador, London.

Goodsell, D, (2009), *The Machinery of Life*, Copernicus Books, New York, 2nd edition.

Kaplan, M. (2010), *David Attenborough's First Life*, HarperCollins, London.

Lane, N. (2010), *Life Ascending: the ten great inventions of evolution*, Profile books, London.

Lane, N. (2016), *The Vital Question: why is life the way it is?*, Profile books, London.

Margulis, M. and Sagan, D. (1997), *Microcosmos: four billion years of microbial evolution*, University of California Press, Berkeley, CA.

Morowitz, H. (2002), *The Emergence of Everything: how the world became complex*, Oxford University Press, New York.

Pross, A. (2016), *What is Life?: how chemistry becomes biology*, Oxford University Press.

Rutherford, A. (2014a), *Creation: the origin of life*, Penguin.

notes

Constructive comments are welcome.

5.1 from molecules to living organisms

"**The purpose of the first section**", Hoffmann 2012:7, and this is the theme of his book.

5.1.1 an endless flame

"**In this closed system**", Campbell 2015:150, Voet 1999:19.

Figure 5.1 is based on Campbell 2015:figure 6.8. The combustion of methane was shown in figure 4.32. The arrangement shown here is equivalent to using a bunsen burner with the air hole closed, which produces the orange flame, with incomplete combustion, which deposits soot on a cold object put in the flame. The figure shows a simplistic arrangement, which assumes there is complete combustion of the methane. For more on the bunsen burner, maybe to refresh your memory from your time in school, see https://en.wikipedia.org/wiki/Bunsen_burner.

"**This is an open system**", Voet 1999:19, Campbell 2015:150. "*Living things are thermodynamically open systems that tend to maintain a steady state rather than reaching equilibrium*" (Voet 1999:359).

"**The one constant in this system**", fire is an example of something that has a metabolism, that is, it processes energy, but is not alive (Smith 2000:5). One fire can start another, but fires lack heredity and so can't evolve.

5.1.2 "thermodynamic gravity" – surfing the free energy wave

"***the energy that can be dispersed***", Lambert 2002b.

Figure 5.2(a) is based on figure 4.25, and **part (b)** is based on Campbell 2015:figure 6.8.

"**Now the reactant CH_4 and O_2 molecules**", Campbell 2008:148.

"**the chemical system is endlessly falling down**", Peter Atkins writes, "*As food falls down the slope of free energy, and in due course becomes excrement*" (Atkins 1994:172).

5.1.3 "unpicking" the methane molecule

"**If the methane can be burned in small steps**", Cudmore 1977:146.

step by step down the free energy slope

Figure 5.3 is based on Alberts 2008:figure 2-43, Wang 2017:figure 1, and Madigan 2003:figure 17.58. The views of all molecules are from ChemEd DL (http://chemdata.umr.umn.edu/resources/models360/models.php).

There is no scale on the free energy axis, and the steps appear to have the same drop in free energy, but this is not the case. The free energy changes for the reactions can be obtained from Berg 2012:figure 15.9 and Wang 2017:figure 1.

A metabolic chart for methane metabolism is available at https://manet.illinois.edu/pathways.php .

5.1.4 methanotrophs

"**But it is, in fact, the basic metabolism**", Hanson 1996, Madigan 2003:sections 12.6 and 17.24, Margulis 1998:B-3, Wang 2017. The Greek word *trophe* means "food" (Alberts 2008:12).

"**Methanotrophic bacteria**", Hanson1996:439. Hanson1996:figure 1, Wang 2017:figure2, and Madigan 2003:figure 17.58 give the reaction series, and the enzymes catalysing each step. Methanotrophs are a member of a group of methylotrophic bacteria that consume only C_1 organic compounds, that is, molecules that contain only one carbon atom, so they must synthesise all the molecules they need to function from these basic precursors. Methanotrophs appear to be obligate C_1 consumers, unable to utilise molecules that contain carbon−carbon bonds (Madigan 2003:366 and 598).

"***Life is a controlled unwinding of energy***", Atkins 1995:21.

"**There are about 130 known strains**", Hanson 1996:453.

"**A single methanotrophic bacterium**", Hanson 1996:456. One molecule of the MMO enzyme, that catalyses step 1 in figure 5.3, can oxidise one methane molecule per second (Wang 2017:8576), which suggests that a single cell has about 50,000 molecules of that enzyme in it.

"**Methanotrophs are members of a group**", biologists have created terms to describe how organisms sustain themselves.

Organisms have two ways to obtain energy from their surroundings, and two ways to acquire carbon, so there are "*four principal ways in which organisms acquire the energy and materials to grow, function and reproduce*" (Morris 2013:6-1, Purves 1998:536, Garrett 2005:541, Morowitz 2002:74, Kent 2000:367). Bacteria are found in all four of these groups.

Organisms that acquire their energy from sunlight are called ***phototrophs***, with plants as familiar examples, and organisms that acquire their energy from ingesting organic molecules are called ***chemotrophs***, with animals as familiar examples.

Organisms that can convert inorganic carbon in single-carbon molecules, such as CO_2 and CH_4, into the organic form in glucose are called ***autotrophs***, or "self-feeders". Plants can do this, so they are ***photoautotrophs*** ("light self-feeders").

Organisms that can't do this must obtain their carbon from molecules made by other organisms, so they eat other organisms, and are known as ***heterotrophs***, or "other feeders", so animals are ***chemoheterotrophs*** ("chemical other-feeders").

Some writers classify organisms as either ***phototrophs*** or ***lithotrophs***, literally "rock-eaters", and ***chemotrophs*** as ***organotrophs***, because they ingest organic molecules in other organisms (Alberts 2008:12, Madigan 2003:28).

Margulis uses the term "chemolithoautotrophic" by adding the term "litho". This extra term appears to be optional, for her definitions of chemoautotrophs and chemolithoautotrophs are virtually identical (Margulis 1998:456), and other writers don't use the term "litho". I've edited the quotation from Margulis to remove her term "chemolithoautotrophic", without, I trust, any loss of clarity.

"***without sunlight***", "***represent the pinnacle***", and "***are crucial***", all from Margulis 1998:70.

"**Bacteria are metabolically extremely diverse**", Margulis 1998:46.

"***have exploited***", Madigan 2003:547.

"***may not be very big***", Andy Watson, quoted in Walker 2007:279.

5.1.5 a model biological cell

"**The methane is thus transformed**", the direct oxidation of CH_4 to CH_3OH without over-oxidation to H_2CO and HCOOH is extremely difficult (Wang 2017:8575).

"**The molecule C1 acts**", see Atkins 2002:728 and 735 on catalysts.

Figure 5.4 is based on Alberts 2008:figure 2-43 and Madigan 2003:figure 17.58.

This cell is a simple model of a real methanotrophic bacterial cell, and we can ask, along with Capra and Luisi, "*where is cellular life localized? Is there a particular reaction, a particular magical spot, where we can put a tag to say here is life? There is an obvious and very important answer to this question: life is not localized; life is a global property, arising from the collective interactions of the molecular species in the cell*" (Capra 2016:132). They go further to consider complex organisms, so "*where is the life of an elephant, or of a given person, localized? Again, there is no localization; the life of any large mammal is the organized, integrated interaction of heart, kidneys, lungs, brain, arteries and veins … Life, then, is an emergent property – a property that is not present in the parts and originates only when the parts are assembled together*" (Capra 2016:133).

5.1.6 features of all living cells

This chapter is mainly about bacteria. Bacteria are like atoms in that they can't be broken down any further, and so are the smallest unit of biological function (Lane 2016:5). Complex cells, which have a nucleus, are known as eukaryotes, from the Greek, *eu* = "true" and *karyon* = "nut" or "nucleus". Bacteria are known as prokaryotes, because of their lack of a nucleus (Lane 2010:90). While bacterial prokaryotes have their genome in a single loop of DNA, eukaryotes have their genome split into a number of chromosomes, all held within a nucleus. Eukaryotes are much larger then prokaryotes, and have a number of different organelles within them, such as mitochondria, which generate energy in the form of ATP, and chloroplasts, which carry out photosynthesis in plants. Unlike prokaryotes, which divide into daughter cells, and freely transfer DNA between themselves, eukaryotes reproduce sexually (Lane 2010: 91). It's now accepted that mitochondria and chloroplasts were once free-living bacteria that became incorporated into other cells, and thereby became the ancestors of animals and plants, respectively (Lane 2010:73 and 92). Prokaryotic bacteria range from extreme simplicity to limited complexity, but there are no simple eukaryotes (Lane 2010:93). Nick Lane sees the transition from simple bacteria to complex eukaryotes as perhaps the single most important transition in our planet's history (Lane 2010:95).

"**There are four basic features**", Alberts 2008:2-11 lists nine features that are common to all biological cells (their order has been changed here):

a) being enclosed in a selectively permeable membrane,

b) storing hereditary information in the same chemical code (DNA),

c) replicating this hereditary information by templated polymerisation,

d) transcribing parts of this hereditary information into RNA,

e) translating RNA into protein in the same way,

f) having one gene corresponding to one protein,

g) using proteins as catalysts,

h) functioning as biochemical factories dealing with the same molecular building blocks, using ATP as a free energy carrier, and

i) a dynamic chemical system, requiring free energy to function.

These features form four basic themes:

(1) forming a membrane to maintain the chemical environment within the cell (item a),

(2) using protein catalysts to make required chemical interactions happen (items g and h),

(3) requiring free energy, and using ATP as a free energy carrier (items h and i),

(4) storing and reading genetic information (items b-f).

Nick Lane lists 6 basic properties shared by all living cells (Lane 2016:96) …

(1) a continuous supply of reactive carbon, for making new bio-molecules,

(2) a supply of free energy, to drive the metabolic biochemistry,

(3) catalysts to speed up and direct the metabolic reactions,

(4) excretion of waste,

(5) compartmentalisation, in which a cell-like structure separates the inside from the outside,

(6) hereditary material, as RNA and/or DNA, to specify the cell's form and function.

These 6 properties emphasise the thermodynamic aspects of the cell's function, and tally with the 4 basic themes above.

"**Biological life is not possible**", Purves 1998:31. However, living organisms sometimes create micro-environments from which polar water molecules are excluded, so they can't interfere with the required biochemical interactions (Stryer 1995:10).

"***organisms are aqueous systems***", Garrett 2005:49.

"***not so much a thin watery bouillon***", Bray 2009:92.

"**Molecules are endlessly and randomly moving around**", Peter Hoffmann has called this a "*molecular storm*" (Hoffmann 2012:7 and 72).

"***experts at molecular recognition***" and "***a thousand reactions***", both from Goodsell 2009:26.

5.1.7 responding to the external world

bacteria search for food by chemotaxis

"**Bacteria, such as *E. coli***", Madigan 2003:87 and 227, Alberts 2008:942, Bray 2009:6 and 90, Phillips 2013:159 and 873.

"***movement towards chemicals***", Bray 2009:6.

"**Figure 0.1 has a simple sketch of an *E. coli* cell**", the sketch shows the all the flagella gathered at one end of the cell, but they can be scattered over the cell's surface (Campbell 2015:632).

"**The cell has a molecular "motor""**", Bray 2009:6, Alberts 2008:943.

"**The cell has nutrient-detecting molecules**", Bray 2009:90, Alberts 2008:943. The cell typically has about 10,000 receptors of 4 types, that can discriminate about 50 different nutrient attractants and harmful repellents (Bray 2009:90).

"**In the absence of nutrients**", the roughly 1 second cycle is from Alberts 2008:942.

"**However, if the nutrient concentration increases**", Madigan 2003:figure 4.46, Phillips 2013:figure 4.16.

"**By varying the frequency of tumbles**", Madigan 2003:87, Alberts 2008:943, Phillips 2013:161.

"***if conditions are improving***", Bray 2009:94.

"**This behaviour is executed**", Alberts 2008:943, Phillips 2013:873, Bray 2009:89.

"***individual microchips***", Bray 2009:93.

"**These bacteria can detect**", Berg 1977:213, Alberts 2008:043.

the tryptophan repressor

This section is based on Madigan 2003:220, Garrett 2005:956, Alberts 2008:433, Campbell 2015:417. The *lac* operon is a better known example of the regulation of gene expression, but I've used tryptophan here because the mechanism is simpler.

"**The DNA in the bacterium *E. coli***", Alberts 2008:433.

"**One of these amino acids is tryptophan**", Voet 1999:80.

"**The *E. coli* cell requires vast numbers of amino acids**", Phillips 2013:198.

"**The cell's DNA molecule**", Madigan 2003:220, Alberts 2008:433, Campbell 2015:417.

An operon is a set of genes for a particular metabolic pathway, that are grouped together on the DNA molecule, have a single promoter, and are transcribed as a single mRNA molecule. This is sometimes called polycistronic mRNA, because the region of DNA that encodes a protein is sometimes called a cistron, with cistron and gene being essentially equivalent terms.

The step-wise synthesis of tryptophan is shown in Garrett 2005:figure 29.17, and it can be found in Nicholson's metabolic pathways chart (Nicholson 2003).

"**However, the synthesis of tryptophan**" Phillips 2013:198.

"**For example, the *E. coli* cell**", Alberts 2008:433, Campbell 2015:417.

"**In this situation**", Madigan 2003:221, Alberts 2008:433, Campbell 2015:418.

E. coli bacteria show us three things

"***image of the world***", Bray 2009:x and 164.

5.1.8 from what matter is to what matter does

"**We have made the transition**", Adam Rutherford points out that the word "inanimate", as applied to molecules, is "*unsatisfactory as it implies inaction, which chemistry most certainly is not. The opposite of living – dead – is also unhelpful for the same reasons. 'Undead' seems to be a fitting way to describe what is more correctly called 'prebiotic' chemistry: fizzing, active reactions that form the pathway to life*" (Rutherford 2014a:69). Nevertheless, I'll use the familiar term "inanimate" for molecules in Level 4.

"***the essence of life***", Tudge 2006:63.

"***Life is not a substance***", Carroll 2016:2.

"***constantly replacing itself***", Tudge 2006:253.

"**So, we can appreciate one definition**", of the very many definitions of life that have been proposed, none has been found to be completely satisfactory (Pross 2016:39). Kolb 2015 outlines the history and philosophical principles, and Popa 2015 surveys the main ideas and gives a selection of definitions. The number and range of definitions of life is quite remarkable (Popa 2015:327−333), for example, "*Life is matter that makes choices, binds time and breaks gradients*"; "*It's alive if it can die*"; "*Life is what the scientific establishment (probably after some healthy disagreement) will accept as life*"; "*Life is simply a particular state of organized instability*", and very many technical definitions. The activity of defining life has effectively taken on a life of its own, for there were forty definitions of life proposed in the year 2002 (Pross 2016:39). The property most often proposed as distinguishing life is Darwinian evolution (Popa 2015:341).

"***Life is simply a particular state***", Hennet, quoted in Pross 2016:40.

"***Imagine a child playing***", Woese 2004:176.

from what matter is to what matter does

"**Now, things have progressed**", and Philip Ball has observed that for a living organism made of molecules, "*it is not so much a question of what the molecules are, but of what they do, and specifically, of what they do collectively*" (Ball 2023:25).

Peter Hoffmann has put it succinctly - "*Living is programmed molecular dancing*" (Hoffmann 2012:231).

5.1.9 the plan for this chapter

"***constant self-renewal***", Tudge 2006:16.

"I**n this chapter**", this is not a survey of current knowledge, or an account of the origin of life on Earth; for these see Purves 1998:chapter24, Delsemme 1998:chapter 5, Smith 2000, Dawkins 2005:chapter "Canterbury", Mason 1991:chapter 13, Morowitz 2002:chapters 11 and 12, and Gribbin 2008:chapter 9. For more on the biology of single-celled organisms, see Madigan 2003, Purves 1998:chapters 25 and 26, and Margulis 1998.

"**Bacteria live as independent**", free-living bacteria can come together in colonies and biofilms, in which individual cells can have specialist rôles (Campbell 2008:207 and 565, Madigan 2003:636).

"**I'll use the bacterium**", there are a huge number of different strains of *E. coli*, so one must be careful about using it as a "*standard ruler*" (Phillips 2013:1034).

5.2 diffusion

5.2.1 diffusion in biological cells

a molecule's random walk in solution

Figure 5.5(a) is very similar to the computer simulations shown in Berg 1993:figure 1.4 and Halford 2004:figure 2; **part (b)** is very similar to the computer simulation shown in Goodsell 2009:figure 1.4; and **part (c)** is based on Atkins 2006:figure 21.28 and Alberts 2008:75.

Notes on figure 5.5(a) – a molecule's random walk

Part (a) shows a small molecule's random walk in 2−D space, which was generated using EXCEL 2010. The RAND() function was used twice, first to generate a random direction, and again to generate a step length as a random fraction of 1 nm. The SIN and COS functions calculated the components of each step in the *x* and *y* directions, and these were added to the molecule's previous location to find its new location. The small diffusing molecule started at the point (0,0), and a stationary enzyme molecule, with a diameter of 2 nm, was positioned with its centre at (2,0), twice its own radius away from the molecule's starting point.

A total of 50 diffusion runs were executed, each of 1,600 steps, with each step being a random fraction of 1 nm, so the average step length was always very close to 0.5 nm. For every step, the spreadsheet calculated the molecule's distance from its starting point, d_{sim} nm, its path length, *p* nm, and its distance from the centre of the enzyme, so it could indicate if the molecule strayed into the enzyme's space.

Diffusion paths varied greatly with each run of the program, which is to be expected with random behaviour, but over 50 runs, a pattern emerged, as shown below.

number of steps, *n*	100	(x 4→)	400	(x 4→)	1,600
diffusion time, *t* ns	100		400		1,600
path length, *p* nm	50	(x 4→)	200	(x 4 →)	800
simulation diffusion distance, d_{sim} nm	5.4	(x 1.7 →)	9.0	(x 2.2 →)	20.1
ratio d_{sim}^2/t	~0.29		~0.20		~0.25
expected diffusion distance, $d_{expect} = 0.5n^{1/2}$	5.0		10.0		20.0

Over 50 simulation runs, the average length of each step was very close to 0.5 nm, and so the path length was always half the number of steps. If we consider that the time duration of each step is the same, which is quite a long time for a molecule, then we can calculate the diffusion time, *t* ns.

The pattern of the results is that the diffusion distance increases roughly as the square root of the diffusion time, so the ratio d^2/t is roughly constant, and this is consistent with diffusion in the real world (Feynman 1963:41-8, Berg 1993:10, Milo 2016:211). This has been calculated in the table as d_{sim}^2/t, and its value is close to 0.25, and this is a measure of how fast the simulated diffusing molecule moves away from its starting point.

Feynman describes a simple case, in which a diffusing particle's distance from its starting point after *n* random steps, each of length *l*, is on average $n^{1/2}l$ (Feynman 1963:41-8). The table shows how the simulation distances, d_{sim}, are very close to the distances expected, d_{expect}, from Feynman's simple formula.

the hit probability

Over a series of 50 runs, each of 1,600 steps, the diffusing molecule wandered into the enzyme's space in about 80% of the runs, giving a "success rate" of about 0.8. So the probability of the molecule not finding the enzyme, even when it starts so close to it, is about 0.2, or 1 in 5.

The path shown in (a) is a single run, which was chosen as an illustration because the glucose failed to find the enzyme, and so we can see the random walk, and the relation between the diffusion distance and the time. The diffusion distances in this run are larger than the average, but are roughly proportional to the square root of the time.

The EXCEL simulation was to used to look at how the hit probability varied with the distance between the molecule and enzyme at the start of a run. This distance was measured in terms of the enzyme radius, so in the arrangement described above the initial molecule-enzyme distance was 2. The results of 30 runs were as follows.

molecule-enzyme distance:	1.1	1.5	2	3	4	5	6
hit probability:	0.93	0.83	0.84	0.63	0.63	0.30	0.37

This small dataset shows that the hit probability decreases, as one would expect. The only way to be sure that the molecule will find the enzyme is to start the random walk right next to it. The hit probability when the molecule/enzyme distance is 2, the situation shown in figure 5.5, is ~0.8, and once the molecule is further from the enzyme's centre than 4 times the enzyme radius, then it's more likely that the molecule won't find it.

Berg (1977:194) considers a situation where a molecule is released a distance *d* from a completely absorbing sphere of radius *r* in an unbounded medium, and states that the probability, *P*, that the randomly diffusing molecule will be captured by the sphere is given by:

$P = r/d$

so the probability is the inverse of the distance ratio. Capture is more probable if the sphere is larger and if the molecule is closer, as one would expect. However, $P = 1$, meaning the capture is certain, only when the molecule is right next to the sphere, and *P* rapidly decreases with distance. So, if the molecule is twice the sphere's radius away, so $d = 2r$, then $P = 0.5$, and the molecule is as likely to escape as to be caught. The simple simulation described above follows a similar trend, but with hit rates higher than predicted by Berg.

Notes on figure 5.5(b) – molecular "hide-and-seek"

Figure 5.5(b) is a simulation, generated using EXCEL 2010, of the random walks of two molecules, a glucose molecule and an enzyme, as they diffuse from opposite ends of a cell, and follows Goodsell 2009:figure 1.4 and Halford 2004:figures 2 and 3. The "cell" is represented by a simple 2−D square grid of permitted ($\pm x$, $\pm y$) locations, with *x* and *y* having values between 5 and 25. So, the size 5 cell is an 11 x 11 grid with 121 permitted locations, and the size 15 cell is a 31 x 31 grid with 961 locations.

For each step, each molecule is able to move to any one of 8 adjacent locations selected by the RAND function. Each molecule is given a value between 0 and 1 for its mobility, and this set the probability of the molecule making a move at each step. So, with a mobility of 1 a molecule would move to an adjacent location every step, and with a mobility of 0.1 a molecule would move on average once in every 10 steps. This way

one could decide how many times a molecule just rattled about in its location before it actually moved, and also make the big enzyme less mobile than the small glucose. The enzyme's larger size reduces its diffusion speed, and in the simulation it has been given a mobility of 0.25, which means that on average the molecule makes a move on the grid every 4 steps. In this simple, illustrative simulation, the enzyme still occupies one location in the grid.

The cell size was varied from 5 to 15, with the starting points set one diagonal location in from opposite corners, and each diffusion run was 1,000 steps long. The EXCEL program checked each molecule's location, and if a random move took it to the edge of the cell, the spreadsheet reversed this move, so the molecule in effect "rebounded" from the cell's membrane. In this way the molecules were confined to diffuse within the cell for their allotted 1,000 moves. As the molecules' random walks extended, they overlapped, so they were visiting the same locations within the cell. In most cases this was at different times, but occasionally they did meet, and the program detected this, because the molecules had the same coordinates at the same time.

The "shapes" of the molecules' random walks varied greatly from run to run, so no single run is representative, and part (b) illustrates just one successful run in a size 15 cell.

cell size and success rate

One would expect that the probability that the molecules meet at some point in their 1,000 step random walks would depend on cell size, and this idea was tested by doing 30 runs with cell sizes from 5 to 25. Every meeting was counted, and a run was classed as a success if the two molecules met at least once.

The results were as follows:

cell size	5	6	7	8	9	10	11	12	13	14	15	20	25
number of locations	121	169	225	289	361	441	529	625	729	841	961	1681	2601
total number of successes	30	28	30	26	20	17	18	17	10	11	7	3	1
success fraction	1.0	0.93	1.0	0.87	0.67	0.57	0.60	0.57	0.33	0.37	0.23	0.1	0.03
number of meetings/run	9.0	6.7	5.3	4.1	2.3	1.7	2.4	1.0	1.1	0.9	0.6	0.5	0.03
seeking/hiding ratio	8.3	5.9	4.4	3.5	2.8	2.3	1.9	1.6	1.4	1.2	1.0	0.6	0.4

This very simple simulation shows that as the cell is made larger, there is a decreasing probability of the molecules meeting at least once in a single run (success fraction), as one would expect. What is interesting is how quickly this success rate decreases. It seems that with a cell size of 5−7 the molecules are almost certain to meet at least once in a run, but with a cell size bigger than about 25 the probability is approaching zero.

It seems reasonable to surmise that the probability of the two molecules meeting is inversely proportional to the number of available locations in the cell. If there are twice as many locations for the enzyme to be in, then the chance of being found by a randomly diffusing glucose molecule is halved. So, we would expect that doubling the cell size, which increases the number of locations by a factor of 4, should reduce the meeting rate by a factor of 4, and the results in the 5th row roughly follow this pattern. Doubling the cell size from 5 to 10, roughly quarters the meeting rate, from ~9 to ~2, and doubling the size again, from 10 to 20, roughly quarters the rate again, from ~2 to ~0.5.

seeking/hiding ratio

The glucose molecule makes 1,000 moves in its random walk, so in a cell with 200 locations it makes 5 seeking moves for every hiding location, while in a cell with 2,000 locations the seeking/hiding ratio is 0.5. The number of moves per location, the seeking/hiding ratio, is given in the bottom row. The simulation suggests that a seeking/hiding ratio greater than ~4, which occurs with cells up to size 7, ensures that the molecules will meet in any given run.

This simulation is on a tiny scale, and comprises random walks of only 1,000 steps in a cell with fewer than 3,000 molecular locations. However, it suggests that the probability of two particular molecules meeting within a certain time is sensitive to cell size, and that with a seeking/hiding ratio of ~4 or more, the molecules are very likely to meet. These constraints would impose a strict size limit for biologically viable cells.

a neat calculation

Ron Milo raises the question of how long it takes on average for a particular small molecule, such as glucose, to collide with a particular enzyme in a typical bacterial cell, such as *E. coli*, and presents a neat solution (Milo 2016:219).

The rate of substrate-enzyme collisions is dictated by the diffusion limit, which is ~10^9 $s^{-1}M^{-1}$ times the substrate concentration.

A single molecule of glucose in an *E. coli* bacterial cell has a concentration of ~10^{-9} M.

So, the glucose-enzyme collision rate is about 10^9 $s^{-1}M^{-1}$ x 10^{-9} M ≈ 1 s^{-1},

that is, the two molecules will meet within a second on average.

From this we can infer that "*every substrate molecule collides with each and every protein in the cell on average about once per second*" (Milo 2016:220).

Notes on figure 5.5(c) – the distance/time relation for a diffusing glucose molecule

The ratio d^2/t is constant for a freely diffusing particle, and this is the basis for the diffusion coefficient, which is the measure of how far an "average" particle will move from its starting point in a certain time.

If a molecule takes t seconds to diffuse a distance d µm, then to diffuse 10d µm will take 10^2t seconds, so we can generalise and say that to diffuse 10 times further takes 100 times longer. This tells us that "*there is no such thing as a diffusion velocity; displacement is not proportional to time, but rather to the square-root of time*" (Berg 1993:10).

An object travelling at a constant velocity will travel 100 times further in 100 times the time. But a diffusing object travels only 10 times further in 100 times the time, and so its overall diffusion velocity, the rate at which it moves from place to place, is 10 times slower. Diffusion is a very effective means of moving particles over short distances, but its effectiveness decreases rapidly as distances increase.

A particle's diffusion coefficient for its diffusion in a liquid can be calculated using the Stokes-Einstein equation, which, for water simplifies to

$D \approx 200/r$

where *r* nm is the particle's radius (Milo 2016:212, Atkins 2006:775). This tells us that a particle's diffusion coefficient is inversely proportional to its radius, so larger particles diffuse more slowly, which is what we would expect.

A cell depends on nutrient molecules each "finding" their specific enzyme by random diffusion, and in this chapter I take the example of glucose and a small enzyme, with a diameter of 5 nm. A glucose molecule, with a diameter of about 1 nm, has a diffusion coefficient of about 400 $\mu m^2/s$ in water. However, a cell's cytoplasm holds a variety of molecules of all shapes and sizes, and interactions with these slow down small molecules by a factor of about 4 (Milo 2016:214), and so a glucose molecule has a diffusion coefficient in a cell's cytoplasm of about 100 $\mu m^2/s$ (Milo 2016:212 and 218).

For a molecule diffusing within the finite 3−D volume of a cell, the diffusion coefficient is

$D = d^2/6t$

where D $\mu m^2/s$ is the diffusion coefficient, d μm is the "average" diffusion distance and t seconds is the diffusion time (Milo 2016:213, Berg 1993:equation 1.14).

With this we can estimate that a glucose molecule can diffuse across an *E.coli* cell, a distance of about 1 μm, in a time of

$t = d^2/6D$ = 1/600 seconds = 1.7 ms ≈ 2 ms

Alberts states that a small organic molecule, such as glucose, "*takes only about one-fifth of a second on average to diffuse a distance of 10 μm*" in a cell's cytoplasm (Alberts 2008:75). At this rate the molecule would take $1/100^{th}$ of this time, or 2 ms, to diffuse 1 μm, and so this is consistent with the calculation above.

The larger enzyme molecule diffuses more slowly, and has a a diffusion coefficient of around 10 $\mu m^2/s$, and might take ~10 ms to cross an *E. coli* cell (Milo 2016:213:213 and table 4-1).

The distance/time scale in part (c) has been constructed for glucose, starting with the fact that it take 2 ms to diffuse 1 μm, and the relation that it takes 100 times longer to diffuse 10 times further. The distances and times are scaled in decades in order to cover the big size range, but these are not proper log scales, because they start at zero.

Part (c) includes values of the glucose molecule's effective diffusion speeds over different distances. It's clear that the diffusion speed drops dramatically as the diffusion time increases, and this is due to the molecule's path having more random steps in it. A molecule diffusing for 100 times longer duration will only move $100^{1/2}$ = 10 times further in distance, and so its overall speed is reduced by a factor of 10. Consequently, diffusion speed is inversely proportional to distance; when the molecule diffuses 10 times further, its average speed is 10 times slower.

the "true" speed of a glucose molecule

We saw in chapter 4 (see the notes to section 4.2.5) that the average energy of a water molecule in liquid water is given by the expression $3kT/2$ Joules, and so its average speed is given by the expression $v_{av} = (3kT/m)^{1/2}$, which gives a value of about 640 m/s at a temperature of 300 K.

We're concerned with reactions between biomolecules in a cell's watery cytoplasm, and so we're interested in the energies of large moving molecules, such as glucose. Richard Feynman imagines a situation where a small particle is suspended in a gas or a liquid, and states that its mean kinetic energy "*will be 3kT/2, even though it is very heavy compared with a molecule*" (Feynman 1963:41-1).

A mole of glucose molecules, $C_6H_{12}O_6$, has a mass of 180 g, and comprises 6×10^{23} molecules, so the mass of one molecule is $180 \times 10^{-3}/6 \times 10^{23} = 3.0 \times 10^{-25}$ kg.

So, the average speed of a glucose molecule at 300 K is:

$v_{glucose} = (3kT/m)^{1/2} = (3 \times 1.38 \times 10^{-23} \times 300/3.0 \times 10^{-25})^{1/2} \approx 200$ m/s (~460 miles/hour).

This is the speed given at the start of the distance/time scale in part (c), before the glucose molecule has any collisions. This is a considerable speed, but the molecule can go only a tiny distance before it collides with a water molecule, and many such collisions lead to the random walk, shown in figure 5.5.

"***no tendency to move***", Berg 1993:12.

a human analogue of molecular diffusion

"**It's tempting to try to find some human analogue**", David Goodsell gives this example: "*You enter an airport terminal and must reach a certain window. If the room is empty, the distance is traversed in a matter of seconds. But imagine that the room is extremely crowded … it takes you fifteen minutes to get across the room, a thousand times longer than the one or two seconds it would take if empty. You would bump into many different people and would definitely not travel in a direct line to your destination*" (Goodsell 1991). But this is misleading, because you are trying to reach a particular destination, and molecules are not – they have no sense of purpose.

molecular "hide-and-seek"

cell size and success rate

diffusion – twice the distance, four times the time

5.2.2 why are cells so small?

"**Single-celled bacteria are very small**", typical size ranges are given in Kent 2000:59, Campbell 2015:figure 7.2, and Purves 1998:figure 4.1.

diffusion in and out of the cell

"**Figure 5.4 shows that a cell**", a cell relies on diffusion to bring nutrient molecules in and take waste molecules out.

It might seem counterintuitive that the random motions of individual molecules can result in the directional movement of many. But imagine that you can shrink and "stand" on a molecule in solution, with 2 dissolved molecules to your left and 1 on your right. All these molecules move randomly in any direction, and so there is a chance that they will pass in front of you. Because there are twice as many molecules on your left as on your right, there is twice the chance that you will see a molecule moving from left to right as the other way. In other words, it is likely that there is a net movement of dissolved molecules away from the region of high concentration, and down the concentration gradient (illustrated in Campbell 2015:figure 8.10).

From this we can think of a general case, and surmise that the net rate of movement of diffusing molecules is proportional to the concentration gradient. This is the case and is known as Fick's first law of diffusion (Kent 2000:68, Atkins 2006:757).

Once the dissolved molecules are evenly spread out, the concentration gradient is zero, and net movement ceases. This doesn't mean that diffusion stops, for the molecules will always be randomly moving about, but there is no net movement in any particular direction.

In a system of randomly interacting molecules, a localised cluster is an improbable configuration, and the most probable arrangement is one in which the molecules are dispersed as widely as possible. So, diffusion is a spontaneous process, because it increases the entropy of a system of

molecules (section 4.3.5). The way that molecules diffuse away from a region of high concentration is analogous to the way a gas expands into a bigger volume (figure 4.29).

"**The cell's metabolism is constrained**", the slowness of diffusion limits organisms to a size of about 1 mm in their smallest dimension, so that no cell is more than about 0.5 mm from the organism's surroundings (Voet 1999:31). Some of the earliest and simplest invertebrate animals have structures of thin sheets of cells, with large surface areas. For example, sponges have thin bodies through which water circulates, and flatworms have bodies that are no more than a few cells thick (Morris 2013:39-2).

Organisms get bigger, not by growing bigger cells, but by growing more small cells, and then actively transporting substances between them, and enabling cells to pass signals to each other. This links to chapter 6, which deals with multicellular organisms, and particularly neurons, the cell specialised for transmitting signals.

"**To illustrate this**", Gee 2021:109 and 269, and Purves 1998:figure 4.2. I've used a cube because the surface area and volume are easy to calculate, and the pattern is the same for other cell shapes. For example, a spherical shape has a surface area of $4\pi r^2$ and a volume of $4\pi r^3/3$, and so shows the same change in area/volume ratio with size (Madigan 2003:figure 4.14).

"***limits the exchange of nutrients***", Purves 1998:66.

"**There comes a point**", many bacterial cells are elongated shapes that facilitate diffusion to their interiors (Morris 2013:26-3).

it may be that bacterial cells operate fairly close to the area/volume limit (Szenk 2017). At low growth rates cells obtain their energy wholly by aerobic respiration, but at high growth rates they switch to anaerobic fermentation, even when oxygen is plentiful. Aerobic respiration produces more ATP molecules per glucose molecule, but requires a bigger area of membrane. Fermentation produces fewer ATPs per glucose, but requires a much smaller membrane area. Fermentation produces about 4 times more ATPs than respiration for the same area of membrane. So, to grow quickly, cells adjust their metabolisms to make more efficient use of their limited membrane area.

diffusion within the cell

"**In the past**", Peter Hoffmann observes that "*Without the shaking and rattling of the atoms, life's molecules would be frozen in place, unable to move*. … *Without randomness, there would be no universe, no life, no humans, and no thought*" and so he concludes that "*Chaos* ***is*** *the life force*" (Hoffmann 2012:7).

"**Every molecule possesses**", "*The random motions of the atoms in our bodies are an afterglow of the creation of the universe, the big bang. The big bang created a universe full of energy, and, eventually, it created stars like our sun. With the sun as an intermediary, the energy of the big bang shakes the atoms of our cells – making life on Earth possible*" (Hoffmann 2012:7).

"***simply bumps around***", Goodsell 2009:6.

"***bump around***" and "***it will only take about a second***", both from Goodsell 2009:6.

"**So, a glucose molecule**", this is 500 separate random walks, each of 2 ms duration and each with a new starting point. The scale in figure 5.5(c) tells that a single journey lasting 1 second would take a glucose molecule only about 22 µm from its original starting point.

"**4,000 different types of protein**", the DNA of a typical *E. coli* bacterium codes for about 4,300 different kinds of protein (Alberts 2008:25).

"**about 250,000 protein molecules**", Goodsell 1991.

"***collides with each and every protein***", Milo 2016:220.

"***a few molecules of DNA***", Dyson 1987.

"***random motions are essential***", Dyson 1987.

"***a transport mechanism***", Purves 1998:105.

5.3 proteins

"***by far the most structurally complex***", Alberts 2008:125 and 6.

"***hardly anything happens***", Morris 2013:4-1.

"**Many proteins are enzymes**", Morris 2013:4-1, Stryer 1995:17, Voet 1999:94.

"***all life is built from or by proteins***", Rutherford, 2014b:17.

5.3.1 amino acids

"***letters differ in their shapes and sounds***", Morris 2013:4-2.

"**The "meaning" of an enzyme protein**", whereas all words are highly selective, and no word uses all the letters of the alphabet, most proteins use most or all of the full set of 20 amino acids (Voet 1999:78). Some amino acids are used more than others: leucine is "popular" at ~9 % average occurrence, arginine is "average" at ~5 % (100/20), and tryptophan is the least used, at ~1.4 % (Voet 1999:table 4-1).

the amino acid alphabet

"**There is a vast number**", there are hundreds of known amino acids, and it's not clear why the particular 20 have been chosen, for they are not "*the most naturally abundant or the most easily synthesized or even the most chemically versatile*" (Voet 1999:89). There also seems to be no clear reason why only the L-isomers have been selected, rather than the D-isomers. However, we can understand that individual amino acids must have only one isomeric form, because the different isomers will have different effects on the final shape of the protein. Since a protein's biochemical action depends critically on its shape, the amino acids must be either the L- or the D-form.

"**at least 2 billion years old**", Stryer 1995:19.

"**These 20 amino acids**", Berg 2012:section 2.1, Garrett 2005:80, Alberts 2008:panel 3-1.

"**Both types attract polar water molecules**", Campbell 2008:80.

"**Glutamine and glutamic acid**", Berg 2012:28.

Figure 5.6(a-i) is based on Morris 2013:figure 4.1 and Garrett 2005:figure 4.1, and **part (a-ii)** is based on Campbell 2008:figure 5.18; **part (b)** is based on Berg 2.8-2.11, Alberts 2008:panel 3-1, and Campbell 2008:figure 5.17.

The five amino acids shown in **part (b)** represent the four types of amino acid (Berg 2012:section 2.1, Garrett 2005:section 4.1, Campbell 2008:78, Alberts 2008:125). Glycine is sometimes classified as a polar amino acid (Garrett 2005:80). The carboxyl group in glutamic acid exists as COO^- at pH7, and so has a net negative charge.

Histidine can be uncharged or positively charged, depending on its local environment, and it's often found at the active sites of enzymes, where it can bind or release protons in the course of catalysed reactions (Berg 2012:31 and figure 2.10). Histidine has been included here because it plays a part in the function of myoglobin, shown in figure 5.10. Glutamine and glutamic acid are included because they feature in figure 5.17.

The molecular views in parts (a-i) and (b) are from ChemEd DL, at http://chemdata.umr.umn.edu/resources/models360/models.php. All amino acids shown are the L-isomers, and their ChemSpider IDs are as follows: alanine (5735), glycine (730), valine (6050), glutamine (5746), glutamic acid (30572), and histidine (6038). The invariant part shown is from alanine, orientated to show the tetrahedral bonding of the α-carbon and cropped to remove the side group. The side groups have been orientated to show their 3−D structure as clearly as possible, and cropped to remove the invariant part. The two parts of each amino acid are shown separated to emphasise the modular basis of amino acids.

I've used "R" to represent a side group, following the convention that calls a monomer a "residue" when it is part of a polymer (Purves 1998:52, Morris 2013:44).

At neutral pH7 amino acid molecules are doubly ionised zwitterions, with the amino group as NH_3^+ and the carboxyl group as COO^- (Voet 1999:79, Garrett 2005:83). To keep things simple, I've shown the amino acids un-ionised.

The alpha-carbon atom bonding is tetrahedral, so a protein backbone is a zig-zag, with the hydrogen atoms and side-groups lying above and below the visual plane, and this makes it hard to see the simple structural motifs (see, for example, Berg 2012:figure 2.7 and Stryer 1995:figure 2-28). The 2−D schematic representation in (a-ii) is based on Campbell 2008:figures 5.17−18, and shows the amino and carboxyl groups as collinear, running left to right, and the side groups set above the backbone. Molecular structures of the amino acids are given by Garrett 2005:figure 4.3, Stryer 1995:figures 2−19, and Berg 2012:figures 2.8−11.

The amino acid cysteine is one of two amino acids with a sulphur atom in its side chain (the other being methionine), and is sometimes classed as hydrophobic (Alberts 2008:127), and sometimes as polar (Berg 2012:figure 2.8 and Campbell 2008:figure 5.17).

the stereochemistry of amino acids (Garrett 2005:section 4.4, Voet 1999:section 4-2)

With the exception of glycine, all the amino acids have asymmetrical molecules, because the central alpha-carbon atom is bonded to four different atomic groups, so it is called chiral (from the Greek *cheir* = "hand"). This means that these amino acids exist in two molecular forms that are mirror images of each other, analogous to our left and right hands. These two forms are called optical isomers, or enantiomers.

Molecules containing a chiral atom are structurally asymmetrical, and their solutions rotate the plane of polarised light, so they are described as optically active. One isomer rotates the plane clockwise, in the + direction, and is called dextrorotatory (Greek, *dextro* = right), while the other isomer rotates the plane anti-clockwise, in the – direction, and is called levororotatory (Greek, *levo* = left).

This is the basis of the Fischer convention for discriminating enantiomers, and uses the single-carbon glyceraldehyde molecule as its reference. This has two enantiomers, one known as L-glyceraldehyde, because it rotates the plane of polarised light to the left, and the other known as D-glyceraldehyde, because it rotates the plane to the right.

All the amino acids used in proteins are L-isomers, as referenced to the glyceraldehyde molecule (Berg 2012:27). However, the presence of the side chains influences their effects on polarised light and some amino acids rotate the plane to the left, while others rotate it to the right (Grisham 2005:table 4.2, Voet 1999:87).

Most biomolecules are chiral and "*biosynthetic processes almost invariably produce pure stereoisomers*" (Voet 1999:88). One optical isomer's reaction with another optical isomer can depend on their relative handedness. A synthetic protein made only of D-residues will react only with the corresponding D-amino acid. When animals are fed with both L- and D-isomers of an amino acid, they metabolise only the L-isomer, which matches the L-amino acids in their proteins, and excrete the D-isomer (Voet 1999:89). Amino acids with the D-configuration are found in some organisms, in some antibiotics and in the cell walls of some micro-organisms.

5.3.2 the peptide bond

"**This is known**", Purves 1998:47, Alberts 2008:85 and panel 3-1. Condensation reactions are also known as dehydration reactions (Campbell 2008:68 and figure 5.18, Purves 1998:520).

"**First, we can see**", Garrett 2005:103, Purves 1998:54.

"**Third, the peptide bond is rigid**", the C–N peptide bond has some degree of double bond character, and so does not allow rotation (Berg 2012:36). This means that the four atoms that contain the peptide bond, which are $O{=}C_O{-}N{-}H$, all lie in the same plane (Berg 2012:figure 2.18, Alberts 2008:panel 3-1).

"**By convention**", Berg 2012:34.

Figure 5.7(a) is based on Campbell 2008:figure 5.18 Alberts 2008, panel 3-1, and Stryer 1995:figure 2-30. The peptide bond has some double bond character, and so it won't rotate, but the bonds on either side will rotate (Garrett 2005:figure 5.3, Berg 2012:figure 2.22 and Stryer 1995:27); **part (b)** shows two views of the canonical leu-enkephalin molecule as given by PubChem (ID 461776), using their SMILES code, CC(C)CC(C(=O)O)NC(=O)C(CC1=CC=CC=C1)NC(=O)CNC(=O)CNC(=O)C(CC2=CC=C(C=C2)O)N, which can be visualised on the St. Olafs site, at https://chemapps.stolaf.edu/jmol/jmol.php?model= [SMILES code]. I've oriented the molecule to give the optimal view of the zig-zag backbone and the side-groups.

Wikipedia has an entry for this molecule, at https://en.wikipedia.org/wiki/Leu-enkephalin. This gives a different SMILES code, O=C(O)[C@@H](NC(=O)[C@@H](NC(=O)CNC(=O)CNC(=O)[C@@H](N)Cc1ccc(O)cc1)Cc2ccccc2)CC(C)C, which gives the same molecule, but in a different pose, with side chains oriented so they can't all be seen clearly. ChemSpider (ID 406229) shows the same molecule in a third pose. Berg 2012:figure 2.14 gives the 2−D structure of the leu-enkephalin molecule.

An animation of the thermal motion of a short alpha helix peptide is available at https://proteopedia.org/w/Thermal_motion_of_peptide .

The ball-and-stick and space-filling views are standard conventions for representing molecules (Berg 2012:figure 1.21 and Alberts 2008:figure −12).

an astronomical number of possible proteins

"***vastly greater***", Voet 1999:95, and Purves 1998:54.

"**From this vast number**", Alberts 2008:137.

5.3.3 a very short polypeptide

"**The peptide in the figure**", Berg 2012:34.

5.3.4 three levels of protein structure

"**Finally, the particular sequence of amino acids**", the three levels of protein structure are all shown together in Morris 2013:figure 4.4.

5.3.5 secondary structure – alpha helix and beta sheet

"**Every amino acid has an –N–H group**", Berg 2012:figure 2.15 and Campbell 2015:figure 5.18.

the alpha helix

"**The tight spiral winding**", Berg 2012:38, Morris 2013:4-6, Garrett 2005:157. Both right- and left-handed helices are allowed, but there is much less steric (spatial) conflict between the side groups and the backbone with right-handed helices, and just about all helices found in natural proteins are right-handed (Berg 2012:40).

the beta sheet

"**Two types of pairing are possible**", Berg 2012:40, Garrett 2005:161, Alberts 2008:131.

"**Most natural polypeptides**", Berg 2012:34, Alberts 2008:131.

"**The directions of the side groups**", Berg 2012:41 and figure 2.30.

"**In such proteins**", Berg 2012:39−41. The effective length of an amino acid in a straight polypeptide strand, such as in a beta sheet, is ~0.35 nm. The pitch of an alpha helix is 0.54 nm, with 3.6 amino acids per complete turn, and single helices are usually less than ~4.5 nm long, so they contain ~8 turns and 30 amino acids (Berg 2012:39−41, Alberts 2008:figure 3-7).

Figure 5.8(b) is based on Berg 2012:figure 2.15 and Campbell 2015:figure 5.18. All the peptide bonds have the trans form, and the adjacent side-groups lie on either side of the backbone. The alpha helix in **part (c)** is based on Garrett 2005:figures 6.6 and 6.9, Morris 2013:figure 4.6, which number the amino acids, and also Berg 2012:figure 2.24. It's very difficult to follow hydrogen bonding in even well-drawn alpha helixes, as in Berg, for example, and I've gone for the simple scheme in Garrett:figure 6.9. The beta sheet in **part(d)** is based on Berg 2012:figures 2.31−32 and Garrett 2005:figure 6.11. The figure shows a mixed beta sheet, as is shown in Berg 2012:figure 2.33. The polypeptide chain in part (b) has been used to make the chains in the beta sheet.

5.3.6 real proteins

determining protein structure

"**However, this approach**", Voet 1999:140, Goodsell 2016, chapter 2. The wavelengths of the X-rays can be set to match the distances between atoms of 0.1 – 0.2 nm, and so produce interference patterns. Since every spot in the interference pattern is the sum of the contributions from all the atoms in the molecule, then from the intensity and position of each spot one can reconstruct the entire molecule (Voet 1999:140).

In fact, it's possible to recreate an optical image from its interference pattern, and this has been done with a photograph of the Parthenon in Athens (Stryer 1995:figure 3-39, and also available at http://www.sci.sdsu.edu/TFrey/Bio750/FourierTransforms.html).

"***a three-dimensional***", Berg 2012:99.

"**With this information**", Goodsell 2016, chapter 3.

a protein for transporting fatty acids

"**The desert locust**", Haunerland 1994.

"**A fatty acid**", Campbell 2015:122, Alberts 2008:58.

"**Within the tertiary structure**", the beta barrel is a common secondary structure in proteins (Alberts 2008:632, Garrett 2005:180).

Figure 5.9 is taken from Berg 2012:figure 2.35, which is presented as an example of a protein constructed largely from beta sheets, and which also has two short alpha helices. This is a fatty acid-binding protein taken from the desert locust (Haunerland 1994).

The images of the protein (ID 1FTP) are from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) at https://www.rcsb.org/, and made available by PDB-101 (PDB101.rcsb.org) under a CC BY 4.0 license.

The protein has two molecules, chains A and B and the figure shows only chain A, which consists of a strand of 133 amino acids, configured in ten strands of anti-parallel beta sheet that wrap around to form a configuration known as a beta barrel, and two small alpha helixes. **Part (a)** shows the whole of chain A. **Part (b**) shows the alpha helix formed by amino acids 28−34, showing the hydrogen bonds in the region of number 31, arginine. The upper image shows the whole alpha helix, and the lower image is a cut-away section, showing the hydrogen bonds within the helix. **Part (c**) shows the three strands in the anti-parallel beta sheet, containing residues 40 and 66, showing the hydrogen bonds in the region of amino acid 54, threonine.

An internal slice of the 3−D molecule can be viewed by setting "clipping planes" to fade out the structure above and below this slice, using SHIFT + the mouse scroll wheel. To see the whole protein: select "3D View: structure"; to view only one chain (say A or B): select "1D-3D view", then de-select the chain you don't want to see.

Guidance on viewing is available at https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/introduction .

myoglobin – a protein for holding oxygen

This section on myoglobin is based on Garrett 2005:493, Berg 2012:45 and 195, Voet 1999:162, Morris 39-12, and Vojtěchovský 1999.

"**Our bodies need to be able to transport oxygen**", Morris 2013:39-10, Garrett 2005:493, Berg 2012:195.

"**The first protein, hemoglobin**", Berg 2012:195.

"**The iron atom is itself held**", the group of carbon-nitrogen rings is known as a porphyrin ring (Phillips 2013:145). I've followed a number of texts in referring to the iron as an atom, but strictly speaking it's an ion, since it's bonded to other atoms (Garrett 2005:492).

"**However, the iron atom**", when isolated heme is exposed to oxygen the iron atoms are irreversibly oxidised, but this can't happen when the heme group is held within a myoglobin molecule (Voet 1999:163).

"**Consequently, a myoglobin molecule**", Morris 2013:29-12, and also see Berg 2012:198 on how myoglobin prevents the iron atom being irreversibly oxidised.

"***molecular bucket brigade*** " and "***efficiently relays oxygen***", Voet 1999:164.

"**But the molecular view**", Vojtěchovský 1999:2155.

"***constantly in motion***", Goodsell, 2000.

"***flexible and rapidly fluctuating molecules***", and "***conformational flexibility, or breathing***", Voet 1999:157, and figure 6-41.

The protein molecule shown in **figure 5.10** is the oxygenated form of sperm-whale myoglobin, a single-strand protein with 151 amino acids and 1,584 atoms (Vojtěchovský 1999).

The images of the protein (ID 1A6M) are from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) at https://www.rcsb.org/, and made available by PDB-101 (PDB101.rcsb.org) under a CC BY 4.0 license.

Illustrations of myoglobin and heme are given by Berg 2012:figures 2.43-4 and 7.2, Voet 1999:figures 7-1 to 7-3, and Garrett 2005:figures 15.23 to 15.26. Depictions of the myoglobin molecule showing the positions of just about all the amino acids are given in Garrett 2005:figure 15.23 and in Voet 1999:figure 7-1.

Myoglobin requires iron to store oxygen, and this is common, for more than one third of all enzymes make use of metals for their activity (Berg 2012:267).

protein folding

"**The way the polypeptide chain**", Berg 2012:45.

"**The only polar amino acids**", Berg 2012:45 and figure 2.44.

"**Myoglobin illustrates**", Alberts 2008:figure 3-5, Berg 2012:45.

"***folds spontaneously***", Stryer 1995:34.

"**The secret of burying**", Berg 2012:46.

"**The different sizes**", Stryer 1995:20.

"**Generally, protein interiors**", Erickson 2009. One consequence is that all proteins have about the same density, 1.37 g/cm3.

"**The final tertiary structure**", Alberts 2008:130, Garrett 2005:166.

There is a final quaternary level of protein structure in which two or more polypeptide chains, each folded into its tertiary structure, come together in one functional macromolecule (Berg 2012:48, Campbell 2015:131, Garrett 2005:table 5.1). For example, hemoglobin, the oxygen-transport protein in blood, has four sub-units, each virtually identical to myoglobin (Garrett 2005:495). However, we can understand a protein molecule's function largely from its tertiary structure, and so I've omitted the quaternary level of structure.

why are protein molecules so large?

"**So, a long chain of many amino acids**", Morris 2013:6-11.

"**The vast majority of proteins**", Voet 1999:95.

the precision of protein structures

"***the term polypeptide***", Campbell 2015:128.

"**The tertiary structure is held together**", Garrett 2005:110, Alberts 2008:126 and figure 3-4.

"***precise three-dimensional shape***" and "***the bonds holding the tertiary structure***", both from Kent 2000:35.

"***all the information necessary***", Garrett 2005:110.

"***the change of even a few atoms***", Alberts 2008:137.

5.4 enzymes – protein catalysts

5.4.1 a simple chemical reaction

"***Now we are at the heart of chemistry***", Atkins 2006:869.

"**We can start with the simple ideas**", Atkins 2006:869, Hoffmann 2012:148.

"**To illustrate this, I'll consider a very simple system**", Campbell gives an example, using A, B, C and D to represent portions of molecules, which undergo a reaction whereby A–B + C–D → A–C + B–D (Campbell 2008:figure 8.14), and I've simplified this to make them atoms rather than parts of molecules. This is highly simplistic, and maybe chemically impossible, but I just want a system that is simple, both conceptually and visually, in order to illustrate transition states and multiple reaction pathways.

Figure 5.11 is based on Campbell 2015:figure 6.13.

The images of water molecules are from ChemSpider, CSID:937, http://www.chemspider.com/Chemical-Structure.937.html (accessed 07:39, Jul 3, 2023).

"**The water molecules**", "*in solution, the reactant molecules may simply diffuse together and then acquire energy from their immediate surroundings while they are in contact*" (Atkins 2006:870).

"**The bonds in molecules A–B and C–D**", the nature of the molecular potential energy curve, in which the equilibrium bond length corresponds to the energy minimum, at which point repulsive and attractive forces are in balance, means that bonds behave like springs (Atkins 2006:figure 11.1).

"**If these molecules remain together**", the collision rate times the time, $10^{13} \times 10^{-9} = 10^{4}$ collisions.

"***transitory molecular structure***", Berg 2012:225.

"**It is very short-lived**", transition states have very short lifetimes, around 10^{-14}–10^{-13} seconds (Voet 1999:288, Garrett 2005:447, Zewail 1999:figure 1).

a choice of paths in reaction-space

"***the energy required***", and "***the amount of energy***", Campbell 2008:152.

5.4.2 controlling the reaction

"**One way of doing this is to raise the temperature**", Garrett 2005:412.

"**Most of the biochemical reactions**", the temperature sensitivity of a reaction can be quantified using the quotient Q_{10}, which is the ratio of reaction speeds at two temperatures 10°C apart (Purves 1998:816). A Q_{10} value of 1 means that a reaction's rate does not change with temperature. Most biological Q_{10} values are in the range 2–3, meaning that if the temperature rises by 10°C their reactions proceed between 2 and 3 times faster.

Reactions that follow the Arrhenius equation have rate constants that are temperature-dependent (Garrett 2005:412, Atkins 2002:707). The higher the activation energy, the stronger the temperature dependence. A reaction with a moderately high activation energy of about 50 kJ/mol (or ~0.5 eV/molecule) doubles its rate constant for a 10°C temperature increase (Atkins 2002:710, Milo 2016:220).

"**For example, we see the effect**", Milo 2016:220-221.

"**Third, all reactions are sped up**", Alberts 2008, caption to figure 2-45, Campbell 2008:153.

"**Finally, high temperatures**", Purves 1998:816.

"***accelerates reactions***", Atkins 2002:707.

5.4.3 enzymes – protein catalysts

steric specificity and induced fit

Figure 5.12 is based on Alberts 2008:figures 2-16, 3-37 and 3-42, Berg 2012:figure 8.9, and Voet 1999:figure 11-1.

"***clasping handshake***", Campbell 2008:154.

"***dynamic recognition***", Berg 2012:228.

"***stressing and bending***", Campbell 2008:154.

"**The substrate molecule can be dismantled**", Alberts 2008:73 and figure 2-46.

"***is due to the precise interaction***", Berg 2012:221.

5.4.4 an enzyme-catalysed reaction

Figure 5.13(a) is based on Alberts 2008:figure 3-52 and Campbell 2015:figures 6.13, 6.14, and 6.16; part (b) is based on Campbell 2008:figures 8.14−15, Garrett 2005:figure 13.5, Berg 2012:figure 8.3, and Voet 1999:figure 11.5. I've used the simple version of the free energy diagram given by these references, and not used the more complicated version, given by Alberts 2008:figure 3-46, Voet 1999:figure 11-13, and Garrett 2005:figure 14.1. These show the binding energy of the enzyme-substrate complex, which we don't need to consider here.

The full and proper treatment of enzyme-catalysed reaction kinetics is given in Berg 2012:section 8.2, Garrett 2005:section 13.2, and Alberts 2008:72 and 158.

"**Part (a-i) shows the enzyme molecule**", Campbell 2015:figure 6.16.

"**The new product molecules**", Kent 2000:43.

"***more molecules have the energy***", Berg 2012:226.

enzymes enable a dynamic metabolism

"**At any moment**", Garrett 2005:405.

"**Without enzymes**", Milo 2016:215

"**they typically make reactions**", Garrett 2005:444.

"***roughly equivalent***", Kent 2000:42.

"***direct each of the many different molecules***", Alberts 2008:73.

"***dynamic metabolism***", Campbell 2008:153.

5.4.5 a snapshot of a working enzyme

"**A typical bacterial cell**", the figure for the number of proteins is for *E*. *coli* (Alberts 2008:25).

"**The cell is constantly recycling**", there appear to be no figures for *E*. *coli*, and these figures are for yeast (Berg 2012:676). The typical lifetime of 40 minutes is from Milo 2016:247 and figure 4-22.

"**The number of substrate molecules**", Alberts 2008:160, Garrett 2005:416, Berg 2012:234 and Table 8.5. Kent gives an average enzyme a turnover number of 1,000 (Kent 2000:42).

The turnover number is defined more strictly as "*the number of substrate molecules converted into product per enzyme molecule per unit time when the enzyme is saturated with substrate*" (Garrett 2005:416).

"**The cell contains around 60,000 nutrient substrate molecules**", we can get an idea of these figures using basic maths with representative values of relevant parameters.

A 1 M solution is a concentration of 1 mole per litre, so it holds 6×10^{23} molecules/1,000 cm^3, or 6×10^8 molecules/μm^3.

An *E*. *coli* bacterium is approximately cylindrical, with rounded ends and a size of ~1 x 2 μm, with a volume of ~1.3 μm^3 (Milo 2016:xxxi and 10).

I'll take a representative bacterial cell to have volume of 1 μm^3, so this cell will hold 6×10^8 = 600 million molecules of a 1 M solution, and 0.6, or ~1 molecule of a 10^{-9} M solution. Milo gives a "rule of thumb", that 1 molecule in a bacterium constitutes a 1 nanomolar solution, or 10^{-9} M (Milo 2016:xxxi).

The substrate molecules acted on by enzymes are present in a range of concentrations, and a typical concentration is around 0.1 mM, or 10^{-4} M (Milo 2016:217). This is in accord with Alberts, who states that "*some abundant substrates are present at a concentration of 0.5 mM*" (Alberts 2008:75). I'll take a representative concentration of our substrate molecule as ~10^{-4} M, so there are ~60,000 substrate molecules in the bacterial cell.

What is the concentration of a representative enzyme protein? Alberts states that most enzyme concentrations are much lower than 10^{-6} M (Alberts 2008:169), so for this simple example, I've set the enzyme concentration at 10^{-7} M, which means there are ~60 molecules in the cell.

"**All these molecules are diffusing around**", Peter Hoffmann writes, "*It has been estimated that the average space between proteins in living cells is less than ten nanometers. Since proteins are between ten and one hundred nanometers in size, this is equivalent to a crowded parking lot with just a foot or less between each car. When things are this tight, it becomes tricky to maneuver past each other. In addition to this crowding, every space between proteins is filled with water, ions, sugars, and other assorted small molecules. You now have an idea of what a crowded mess a cell is*" (Hoffmann 2012: 107).

Figure 5.14(b) is based on Goodsell 1991:figures 1(a) and (b) and Halford 2004:figure 3. Alberts 2008:figure 2-49 is a colour version of Goodsell's figure 1(a). I'm taking the general substrate molecule as similar to glucose.

"**The substrate has already had two unsuccessful collisions**", there will be many random substrate-enzyme collisions before the substrate is successfully bound. With a substrate concentration of 0.5 x 10^{-3} M the active site on its enzyme has about 500,000 collisions from the substrate in 1 second (Alberts 2008:75). In the example in figure 5.14, with a substrate concentration of 10^{-4} M, we can estimate there will be ~100 collisions in 1 millisecond, only one of which is successful and leads to a reaction.

"**The water molecules are shown as stationary**", Hoffmann 2012:145.

Also, see the notes to section 4.2.5.

"**These collisions have set the enzyme molecule spinning**", Alberts states that "*a large globular protein is constantly tumbling, rotating about its axis about a million times per second*", so it spins 1,000 times in 1 ms (Alberts 2008:74). This surprising figure is supported by computer simulations of bacterial cytoplasm, which show that rotation speed decreases with increasing molecular size, and that macromolecules with diameters around 8−9 nm rotate around 1,000,000 times per second (Yu 2016:figure 5A).

"**A lot happens in the next 1/1000th of a second**", Peter Hoffmann briefly outlines such a scenario (Hoffmann 2012:150).

5.5 respiration – the combustion of glucose

This section is based on Alberts 1998a:chapters 4 and 13, Alberts 2008:chapters 2 and 14, McMurry 2000:chapter 29, Purves 1998:chapter 7, Campbell 2008:chapter 9, Morris 2013:chapters 6 and 7, Berg 2012:chapters 15−18.

"**This section will cover**", Alberts 1998a:108.

5.5.1 overview of respiration

"The complete oxidation", Purves 1998:142. The standard figure for the free energy released by the complete oxidation of glucose is −2,870 kJ/mol (Purves 1998:142, Campbell 2015:240). Some texts give energies in kcal/mol, and 1 kcal = 4.18 kJ. Reaction energies are conventionally given in kilojoules/mole (kJ/mol), which is the energy change for one mole, that is 6 x 10^{23} molecules.

It will be simpler to deal with one molecule, and with energies in eV rather than Joules, where 1 J = 6.3 x 10^{18} eV.

An energy of 1 kJ/mol is 1,000 J per 6 x 10^{23} molecules or 1.67 x 10^{-21} J/molecule, and since 1 J = 6.3 x 10^{18} eV, this is ~0.01 eV/molecule.

So, the free energy released by the complete combustion of 1 glucose molecule is 2,870 x 0.01 = 28.7 ≈ 30 eV/molecule.

Figure 5.15 is based on Alberts 2008:figures 2-69 and 14-12, and Campbell 2015:figure 10.5.

The glucose molecule is alpha-D-glucopyranose (ChemSpider ID 71358), or just alpha-D-glucose, and referred to here as just glucose (Alberts 2008:panel 2-8, Garrett fig. 7.5, Purves 1998:figure 7.6, and Campbell 2015:figure 10.9). The glucose molecule shown is from https://commons.wikimedia.org/wiki/File:Alpha-D-Glucopyranose.svg , made publicly available by NEUROtiker.

"***the cell could not use it***", Cudmore 1977:146.

"**Every reaction step releases**", ATP is the most important and versatile of the energy carrier molecules (Alberts 2008:80).

"**The molecule that is used as an energy carrier**", Morowitz 2002:73.

"**The views in parts (a) and (b)**", Purves 1998:142, Berg 2012:222.

"**The first is an explosive release**", Alberts 1998a:108.

"***controlled unwinding of energy***", Atkins 1995:21.

"***cashes in the large denomination***", Campbell 2015:243.

"***falls down the slope of free energy***", Atkins 1994:172.

"**In this way, a bacterial cell**", a biological cell has to do 3 types of work: (1) chemical work, such as driving endergonic reactions – pushing molecules up the free energy "hill"; (2) transport work – pumping substances across membranes against their concentration gradients; (3) mechanical work – moving itself about against the resistance of the physical world (Campbell 2015:150). Here we'll just look at item (1).

5.5.2 ATP – the energy currency of the cell

Figure 5.16 is based on Purves 1998:figure 6.7, Alberts 2008:figure 2-57, and Campbell 2015:figure 6.9.

The 2−D structure of ATP is from https://commons.wikimedia.org/wiki/File:ATP_chemical_structure.svg, made available by user Hbf878, under the CC0 1.0 Public Domain Declaration.

A molecule of water is involved in the hydrolyis of ATP, but we don't need it to follow the equation in part (b), and so I've left it out.

"***the triphosphate tail of ATP***", Campbell 2015:151, and also see Alberts 2008:80. The phosphate bonds in ATP are sometimes called "high energy" bonds, but this is misleading because the energy come from hydrolysis, and is not intrinsic to the bonds themselves (Purves 1998:126, Campbell 2008:149).

"**The third phosphate group**", a general description is "*the cleavage of a covalent bond with the accompanying addition of water*", so we can write this as a general formula, A—B + H—OH → A—H + B—OH (Alberts 2008:Glossary-18).

Under cellular conditions, ATP hydrolysis yields around 50 kJ/mol, equivalent to about 0.5 eV/molecule (Garrett 2005:72, Campbell 2008:149, Purves 1998:125, Alberts 2008:80 and 825, Milo 2016:182).

The hydrolysis of ADP to AMP yields slightly more free energy than does ATP (36 kJ/mol under standard conditions, Garrett 2005:table 3.3), but this reaction does not occur often in cells, and hydrolysis of AMP, removing the last phosphate group, yields little free energy, about 9 kJ/mol (Garrett 2005:table 3.3, Purves 1998:126). Consequently, I've left these energies out of figure 5.16.

"***the minute-to-minute existence***", Garrett 2005:60.

"**ATP molecules package free energy**", Purves 1998:144.

5.5.3 how ATP couples catabolic and anabolic reactions

"**The combustion of glucose**", definitions of catabolism and anabolism are from glossaries in Purves 1998 and Campbell 2015.

"**So, catabolic reactions**", Campbell 2015:figure 6.12, Garrett 2005:543.

"**The cell uses the energy storage molecule ATP**", Purves 1998:figure 6.9, Campbell 2015:figure 6.12.

The coupled reactions, shown in **figure 5.17(a)** are based on Purves 1998:figure 6.9, Campbell 2015:figure 6.12, Morris 2013:figure 6.2, Garrett 2005:figure 17.4, and Alberts 2008:figure 2-55. The gear wheel symbol that couples a pair of reactions is based on Alberts 2008:figure 2-51 and Atkins 2007:figure 17. The structure of the enzyme glutamine synthetase (2GLS) is shown in Voet 1999:figure 20-26, and at https://www.rcsb.org/.

The synthesis of glutamine, shown in part (b) is based on Alberts 2008:figure 2-59, Campbell 2015:figure 6.10, Berg 2012:710, and Garrett 2005:figure 25.10. Both reaction steps take place on the surface of a molecule of glutamine synthetase (Alberts 2008:figure 2-59).

All the images of molecules are from ChemSpider: L-glutamic acid, CSID:30572, http://www.chemspider.com/Chemical-Structure.30572.html (accessed 08:15, Jul 3, 2023); L-glutamyl phosphate, CSID:167893, http://www.chemspider.com/Chemical-Structure.167893.html (accessed 08:16, Jul 3, 2023); ammonia, CSID:217, http://www.chemspider.com/Chemical-Structure.217.html (accessed 08:18, Jul 3, 2023); L-glutamine, CSID:5746, http://www.chemspider.com/Chemical-Structure.5746.html (accessed 08:19, Jul 3, 2023); and water, CSID:937, http://www.chemspider.com/Chemical-Structure.937.html (accessed 08:20, Jul 3, 2023).

The $–NH_2$ group was cropped from the view of the glutamine molecule, and the –OH and –COOH groups added to show the glutamic acid and glutamyl phosphate molecules. This is a simplistic view of the reaction, and doesn't show the molecules in their ionised states, as they would be inside the cell.

The standard free energy of formation of glutamine from glutamic acid is ~0.14 eV/molecule (Campbell 2015:152).

"***leading to the formation***", Atkins 1994:173.

5.5.4 the cell's ATP economy

ATP in bacteria

"***the circulating currency***", Purves 1998:127.

"**A typical E. coli cell**", Milo gives an ATP concentration of ~10^{-3} M for *E. coli* growing on glucose (Milo 2016:96:184:table 3-2), which means there are 10 x 10^{-3} x 6 x 10^{23} = 6 x 10^{21} molecules/litre, or 6 x $10^{21}/10^{15}$ = 6 x 10^{6} ATP molecules/µm^3 of cytoplasm, since 1 litre = 1,000 cm^3 = 10^{15} µm^3.

An *E. coli* cell is roughly cylindrical, ~1 x 2 µm, with a volume of ~1.3 µm^3, which is 1.3 x 10^{-15} litres, and a surface area of ~6 µm^2 (Milo 2016:10 and 131). So, the number of ATP molecules in a cell at a typical moment is 6 x 10^{6} x 1.3 ≈ 8 x 10^{6} = 8 million.

"**This might seem a vast number**", an *E. coli* cell's rate of oxygen consumption tells us that it consumes about 10^{7} ATP molecules/second (Milo 2016:202), so the average lifetime is 8 x $10^{6}/10^{7}$ ≈ 1 second, which is consistent with Milo's figure of 2 seconds (Milo 2016:230:table 4-3).

"**This huge turnover**", the molar mass of ATP is about 507 Daltons (Chemspider ID 5742), which is 507 x 1.66 x 10^{-27} kg ≈ 8.4 x 10^{-25} kg , so the cell's consumption rate is 10^{7} x 8.4 x 10^{-25} ≈ 8 x 10^{-18} kg ATP/second.

An *E. coli* cell has a mass of about 10^{-15} kg (Milo 2016:10), so a cell consumes its own weight of ATP in 10^{-15}/8 x 10^{-18} ≈ 120 seconds.

"**A typical adult human**", the following notes explain how the figures in this paragraph have been estimated.

ATP in humans

Because human body cells use ATP to store and transfer the energy released by the oxidation of food, we can estimate the rate of ATP consumption in an average human body, in two ways: (1) from the breathing rate, and (2) from the daily energy intake. We assume that the body is in a steady state, so the ATP content is stable, and the rates of ATP production and consumption are equal.

(a) ATP consumption rate

(1) breathing (based on Flamholz 2014)

A resting adult human breathes in ~6 litres/minute of air containing ~20% O_2, and breathes out air with ~15% O_2, and so consumes ~0.3 l/min of O_2.

A volume of 1 l of O_2 has a mass of ~1.3 g, so this oxygen has a mass of 0.3 x 1.3 ≈ 0.4 g.

The molar mass of O_2 is 32 g, so this is ≈ 0.012 moles, and since there are 6 x 10^{23} molecules in 1 mole, this is ≈ 7.5 x 10^{21} O_2 molecules consumed each minute.

Human body cells oxidise a molecule of glucose, $C_6H_{12}O_6$, with the consumption of 6 molecules of O_2, to produce ~30 molecules of ATP (Flamholz 2014:3499), so the conversion rate is ~5 ATPs per O_2 molecule. ATP has the formula $C_{10}H_{16}N_5O_{13}P_3$, and a molar mass of 507 g.

So, the rate of ATP consumption is 7.5 x 10^{21} x 5 ≈ 3.8 x 10^{22} molecules/minute, which is 3.8 x $10^{22}/6$ x 10^{23} ≈ 0.062 moles, which is 0.062 x 507 ≈ 32 g/minute, or ≈ 45 kg/day.

(2) Energy intake (based on Garrett 2005:73)

The average energy intake for a fairly sedentary human is ~12,000 kJ/day, and ~50% of this is stored in ATP.

The energy yield from ATP hydrolysis under cellular conditions is ~50 kJ/mol (Milo 2016:182), so this amount of energy requires 6 x $10^{6}/50$ x 10^{3} = 120 moles of ATP, which has a mass of 120 x 507 ≈ 60 kg.

So, these two approaches give similar figures for the amount of ATP turned over in an average adult human body of ~50 kg/day. This agrees with published figures of 65 kg/day (Garrett 2005:73), 40 kg/day and 83 kg/day (Berg 2012:435 and 525 repectively), 40 kg/day (Phillips 2013:575), and 36 kg/day (Voet 1999:366), and 60–100 kg/day (Lane 2016:63).

(b) cellular ATP turnover rate

The adult human body contains a total of ~3.7 x 10^{13} cells (Milo 2016:314, table 6-1), but ~70% of these are red blood cells, so the number of cells that can oxidise glucose is ~10^{13}.

So, if we take the ATP turnover by a human body as 50 kg/day, then the rate of ATP consumption by an "average" body cell is 50 x $10^{3}/10^{13}$ = 5 x 10^{-9} g/cell/day, and since 1 molecule of ATP has a mass of ~8.4 x 10^{-22} g, this is 5 x $10^{-9}/8.4$ x 10^{-22} ≈ 6 x 10^{12} molecules/cell/day, or ~70 million ATP molecules/cell/second.

(c) human body's ATP content

How much ATP does the average adult human body contain at any moment? There are a range of estimates for this: 100 g and 250 g (Berg 2012:435 and 525 respectively), 50 g (Garrett 2005:73), and less than 50 g (Voet 1999:519). Wolpert and Alberts state that there are about 1 billion ATP molecules in a typical animal cell, which are used and replaced in 1-2 minutes (Wolpert 2009:36, Alberts 1998a:110). Nick Lane says that the human body contains ~60 grams of ATP, and each molecule is recharged once or twice a minute (Lane 2016:63).

However, a figure of 10^{9} ATP molecules/cell in a human body would give a total ATP content of 10^{9} x 10^{13} = 10^{22} molecules, with a mass of 10^{22} x 8.4 x 10^{-22} ≈ 8 grams, which seems very low compared to these figures.

Voet states that the lifetime of an ATP molecule varies from seconds to minute depending on the cell type and its level of metabolic activity (Voet 1999:366). For a body content of only 8 grams to sustain a daily turnover of 50 kg means that the 8 gram must be recycled 50,000/8 ≈ 6,000 times a day, so the average ATP lifetime is ~14 seconds, which seems very short.

Milo gives the ATP concentration in human muscle cells as 8 x 10^{-3} M (Milo 2016:184:table 3-2), which means there are 8 x 10^{-3} x 6 x 10^{23} = 4.8 x 10^{21} molecules/litre.

The linear dimensions of mammalian cells vary widely, but their volumes are in the range 10^{3} – 10^{4} µm^3 (Milo 2016:315), with the majority being around 10^{3} µm^3, which is 10^{-12} litres, since 1 litre = 10^{15} µm^3 (Milo 2016:16).

So, the number of ATP molecules in an "average" mammalian cell is 4.8 x 10^{21} x 10^{-12} ≈ 5 x 10^{9}, or ~5 billion. This is about 5 times the figure given by Alberts and Wolpert.

This gives a total body content of 5 x 10^{9} x 10^{13} = 5 x 10^{22} molecules, which have a mass of 5 x 10^{22} x 8.4 x 10^{-22} ≈ 40 grams. This is still low, but closer to the estimates given above.

This amount of ATP will be recycled 50,000/40 = 1,250 times a day, so the average ATP lifetime is ≈ 70 seconds. This is short, but fairly consistent with the times given by Wolpert, Alberts, and Lane.

To summarise, we can take an average human body cell as having a complement of ~5 billion ATP molecules, with an average lifetime of ~1 minute, giving a total body complement of ~40 grams, which is recycled 1,250 times each day, to give a daily turnover of ~50 kg. These figures are self-consistent and consistent with the figure given above.

"**It costs about four ATP molecules**", Alberts states at least 4 ATPs (Alberts 2008:385), while Lane gives a figure of 5 ATPs (Lane 2016:172).

"**If we assume that each significant reaction**", the cell size is from Milo 2016:16.

"**Blink in surprise**", a blink lasts about one third of a second (Kwon 2013).

5.5.5 cellular respiration – from glucose to ATP

This section is based on Purves 1998:chapter 7, Campbell 2015:chapter 10, Garrett 2005:chapters 18-20, Berg 2012:chapters 16-18, Alberts 2008:88-102 and 817-840, and Morris 2013:chapter 7.

Figure 5.18(a) is based on Purves 1998:figure 7.11; Morris 2013:figures 7.3 and 7.12; Alberts 2008:panels 2-8 and 2-9; Campbell 2015:figures 10.5, 10.9-10.11, and 10.16. The glucose molecule is as in figure 5.15.

Part (b) is based on Purves 1998:figure 7.14; Morris 2013:figures 7.9 and 7.10; Alberts 2008:figures 14-12, 14-14, 14-15, 14-26, 14-29, and 14-33; Campbell 2015:figures 10.5 and 10.13-10.15; and Berg 2012:figures 18.6 and 18.24. Free energy profiles are given in Alberts 2008:figure 14-29, Campbell 2008:figure 9.13, and Berg 2012:figure 18.6. Alberts gives the free energy per electron in the electron transport chain, and I've based the scale in **part (b)** on this. *E. coli* bacteria have a double membrane, with the inner membrane being the plasma membrane (Alberts 2008:665). In these bacteria the proteins of the electron transport chain are embedded in the inner, plasma membrane (Alberts 2008:figure 14-33).

Notes on figure 5.18.

(a) dismantling the glucose molecule

Part (a), depicting the 1st stage, includes glycolysis, pyruvate oxidation and the citric acid cycle. The free energy profile is based on Purves 1998:figure 7.11, and Morris 2013:figure 7.3, which give the loss of free energy at the end of the citric acid cycle as ~29 eV.

(b) the electron transport chain (also called the respiratory chain)

The proteins in the electron transport chain are grouped into four major complexes, usually indicated as *I* to *IV* (Morris 2013:figure 7.9, Berg 2012:figures 18.6 and 18.17, Campbell 2015:248). There are two electron carrier molecules: the major carrier is NADH, and $FADH_2$ is the minor carrier. Electrons are delivered by NADH to complex *I*, and by $FADH_2$ to complex *II*, and transferred from these complexes to complexes *III* and *IV*. Complexes *I*, *III* and *IV* pump H^+ ions across the cell membrane. **Figure 5.15(b)** shows electrons being delivered by a generic carrier molecule to a single generic enzyme complex, which pumps H^+ ions across the membrane.

Electrons are shuttled to the electron transport chain by two carriers, NADH and $FADH_2$. The molecule flavine adenine dinucleotide, better known as FAD, will accept 2 H atoms to become $FADH_2$, with an increase in free energy, and will deliver these to where they are needed (Purves 1998:131).

In respiration the great majority of electrons are transported by nicotinamide adenine dinucleotide, better known as NAD, which has an oxidised form, NAD^+, and a reduced form, NADH (Campbell 2015:241). In some steps in the breakdown of glucose, 2 hydrogen atoms – 2 protons (H^+) + 2 electrons (e^-) – are removed and transferred to NAD^+, which accepts the 2 electrons and 1 proton, and releases the other into the surrounding solution, as follows …

$2H^+ + 2e^- + NAD^+ \rightarrow NADH + H^+$ (Campbell 2015:figure 10.4)

The oxidation of NADH by oxygen gas from the air releases a large amount of energy …

$NADH + H^+ + ½O_2 \rightarrow NAD^+ + H_2O + 2.2$ eV (Purves 1998:144 and 155, Alberts 1998a:figure 13-5, Campbell 2015:242)

In the same way that ATP transfers energy in 0.5 eV bundles, NAD stores energy in 2.2 eV bundles (Purves 1998:144).

Each NADH molecule delivers its pair of electrons to the electron transport chain, and once these electrons have given up all their energy, they are released into the cytoplasm to combine with hydrogen ions there and oxygen atoms from the air to become water …

$2e^- + 2H^+ + ½O_2 \rightarrow H_2O$ (Campbell 2015:249 and figure 10.15, Alberts 2008:figure 14-12)

In the text, I've simplified things to show a generic carrier molecule (CM), which can pick up a pair of H atoms, becoming $CM–H_2$, so CM represents the oxidised form NAD^+ and $CM–H_2$ represents the reduced form NADH. The sharp-eyed reader may notice that the 12 molecules of $CM–H_2$ will transport 24 H atoms, and there are only 12 in the glucose molecule, and this is because some of the H atoms have come from the breakdown of water molecules in the cytoplasm.

$CM–H_2$ delivers the pair of H atoms to the electron transport chain, and I've simplified the NADH oxidation reaction given above, so that in the text it is given as …

$CM–H_2 + ½O_2 \rightarrow CM + H_2O + 2.2$ eV equation 5.2

"**The process is executed in two main stages**", cellular respiration involves dozens of highly intricate chemical reactions, and their complementary enzymes. Here I've simplified it into just 2 main stages; the 1st stage covers glycolysis, pyruvate oxidation, and the citric acid cycle, and the 2nd covers the processes of the electron transport chain, also known as the respiratory chain. I think this gives an outline of the major processes and principles underlying respiration, without "dumbing down".

"**In the 1st stage**", pyruvate oxidation and the reactions of the citric acid cycle are the sources of the CO_2 we breathe out (Morris 2013:7-9).

"**The two stages of cellular respiration**", a well-fed adult human exhales about 1 kilogram of CO_2 each day (Garrett 2005:47). We can see the water produced by the 2nd stage in the patch of condensation that appears when we breathe on a cold window pane.

the 1st stage – dismantling the glucose molecule

"**The cell starts breaking down**", Purves 1998:147, Allberts 2008:panel 2-8.

"**There is a set of six steps, marked R_2**", the O atoms needed to make CO_2 do not come from molecular O_2 in the air, but from splitting H_2O molecules in the cytoplasm (Alberts 1998a:121 and panel 4-2).

"**Every single step is executed**", Alberts 2008:panels 2-8 and 2-9.

Proteins don't work alone, and "*instead of a cell dominated by randomly colliding individual protein molecules, we now know that nearly every major process in a cell is carried out by assemblies of 10 or more protein molecules. And, as it carries out its biological functions, each of these protein assemblies interacts with several other large complexes of proteins. Indeed, the entire cell can be viewed as a factory that contains an elaborate network of interlocking assembly lines, each of which is composed of a set of large protein machines*" (Alberts 1998b). The enzymes of the metabolic pathways are sometimes physically separate, with substrates diffusing from one enzyme to the next. But enzymes are often physically linked in a complex, either in the cytoplasm or within the cell membrane, so the metabolites can't diffuse away (Garrett 2005:545 and figure 17.5). These multi-enzyme complexes are sometimes called metabolons, meaning "units of metabolism".

the 2nd stage – the electron transport chain

The academic texts describe the operation of the electron transport chain in the mitochondria of eukaryotic animal cells (Purves 1998:154, Alberts 2008:819, Morris 2013:7-10, Campbell 2015:248). Here I describe the electron transport chain in *E. coli* bacteria, which is similar to the process in mitochondria, and is located in the cell's inner plasma membrane (Purves 1998:145, Campbell 2015:248).

The electron transport chain is an example of a chemiosmotic process (Lane 2010:31 and 2016:83). Osmosis is the name given to the movement of water molecules through a membrane, driven by a concentration gradient, and so, in broad terms, chemiosmosis is the movement of protons through a membrane (Lane 2010:31). In respiration, the energy of the electrons stripped from nutrient molecules is used to pump protons across a membrane. As the protons flow through the membrane, driven by their concentration gradient, they cause an ATP-generating protein to spin, which re-generates ATP from ADP (Lane 2010:31).

Chemiosmosis enables a cell to store energy in the "currency" of ATP molecules, from the small change of electrons stripped from nutrients. All biological cells use chemiosmotic processes to generate ATP; it is as universal as the genetic code (Lane 2010:30 and 2016:83).

"**This stage requires gaseous oxygen**", Alberts 1998a:121.

"**The overall action**", this equation is based on Purves 1998:144 and 155 and Alberts 1998a:figure 13-5. The free energy released is 220 kJ/mol, which is equivalent to ~2.2 eV per molecule (Campbell 2008:165). Purves shows the equation for $NADH/NAD^+$, and I've simplified this to show the generic carrier molecule releasing two H atoms in a single high-energy oxidation step.

"***an untamable reaction***", Purves 1998:155.

"**Instead, the cell breaks down the oxidation reaction**", Purves 1998;155.

an electron cascade

"**An electron loses potential energy**", Campbell 2015:240.

"**We can visualise this situation**", Campbell 2015:241.

"**Biological cells have evolved**" Milo 2016:191. Campbell 2015:section 10.4, Morris 2013:section 7.5.

"***the fall of electrons***", Campbell 2015:249.

"***oxygen pulls electrons***", Campbell 2015:242

"**This is the reaction**", Campbell 2015:249 and figure 10.15, Alberts 2008:figure 14-12.

protons and ATP synthase

"**ATP synthase**", Campbell 2015:figure 10.14, Alberts 2008:figure 14-15, Morris 2013:figure 7.10, Berg 2012:figure 18.24, Garrett 2005:figure 20.25, and Hoffmann 2012:195–200.

Nick Lane gives some simple figures for mitochondria in eukaryotic cells (Lane 2016:71–73). The energy of two electrons passing through the electron transport chain is enough to pump 10 protons across a membrane. These 10 protons then "spin" the ATP synthase protein molecule once and this re-generates 3 ATP molecules. The ATP synthase "rotor" can spin at up to 100 revolutions/second.

However, these figures are only representative, for there are different ATP synthase molecules, and the "gear ratio" – the number of protons that need to pass through the protein molecule to generate 1 ATP – is generally a non-integral number between 3 and 5 (Alberts 2008:822, and see also Grisham 2005:666).

ATP synthase is universal to all biological cells that use respiration to generate energy, excepting a minority of species that rely on fermentation. It is "*as universal as the genetic code itself*" (Lane 2016:73).

"***two sets of proteins***", Alberts 2008:822.

"**In this way**", Alberts 2008:822, Berg 2012:549.

"**The imbalance of protons**", Morris 2013:7-12.

"***a turbine***", Alberts 1998a:409.

"**We've seen that a typical *E. coli* cell**", the figures for ATP synthase have been estimated as follows.

An *E. coli* cell consumes ~10^7 ATPs/second. It can generate ~36 ATPs/glucose molecule, and 32 of these are produced by ATP synthase, at the end of the electron transport chain, so ATP synthase provides $10^7 \times 32/36 \approx 9 \times 10^6$ ATPs/second.

An E. coli cell has a membrane surface area of ~6 $\mu m^2 = 6 \times 10^6$ nm^2, since 1 µm = 1,000 nm (see note to section 5.5.4)

This membrane can produce ~4 ATPs/second/nm^2 (Szenk 2017:box 1), so the production of 9×10^6 ATPs/second requires a membrane area of $9 \times 10^6/4 \approx 2 \times 10^6$ nm^2, or about ⅓ of the total membrane area.

An ATP synthase complex occupies ~35 nm^2 area of membrane (Szenk 2017:box 1), so this area is occupied by $2 \times 10^6 /35 \approx$ 60,000 ATP synthase complexes.

There are ~60,000 ATP synthase complexes producing ~9×10^6 ATPs/second, so each complex produces $9 \times 10^6/60{,}000 \approx$ 150 ATPs/second.

This is consistent with measured ATP production rates of ~160/second (Tomashek 2004), and a maximum rate of ~270/second (Etzold 1997, Szenk 2017), and rates exceeding 100 ATPs/second (Alberts 2008:822).

The situation is comparable for animal cells, in which the processes of the electron transport chain take place in the mitochondrion, and ATPs have a lifetime of about 1 minute. In these cells a typical ATP molecule "*shuttles out of a mitochondrion and back into it (as ADP) for recharging more than once per minute*" (Alberts 2008:824).

respiration is now complete

"**Respiration is now complete**", these figures are for a bacterial cell like *E. coli*. There are two electron carrier molecules, NADH, and $FADH_2$, and the yield from the oxidation of 1 glucose molecule is 10 NADH and 2 $FADH_2$ (Campbell 2015:figure 10.2).

The energy carried by each NADH is enough to generate a maximum of about 3 ATPs, and for $FADH_2$ the figure is 2 ATPs (Campbell 2015:252, Garrett 2005:650), so the potential ATP yield from the electron transport chain is about (10x3) + (2x2) = 34 ATPs. If we add this figure to the 4 ATPs generated in the 1st stage we get a total of up to 38 ATPs generated from a single glucose molecule (Garrett 2005:670 and 882). This represents an energy efficiency of about (38 x 0.5)/30 ≈ 60%.

In eukaryotic animal cells, the reactions of the electron transport chain take place within organelles called mitochondria. Some energy must be used to transport ATP and NADH through the membranes of these mitochondria, and Garrett estimates that about ¼ of the energy generated by the electron transport chain is used to do this (Garrett 2005:666). In eukaryotic cells, 1 NADH molecule generates 2.5 ATPs, and 1 $FADH_2$ generates 1.5 ATPs, and so the maximum ATP yield is about 25 + 3 = 28 (Campbell 2015:252). This limits the cell's total ATP yield from 1 glucose

molecule to between 30 and 32 (Morris 2013:figure 7.12, Berg 2012:555, Campbell 2015:252, Alberts 2008:823), so the energy efficiency of eukaryotic cells is ~50%.

5.6 metabolism

"**So, we come to what is enduring**", metabolism is derived from the Greek word *metabole* meaning "change" (Campbell 2008:142).

This section is based on Voet 1999:chapter 13, Stryer 1995:chapter 17, Berg 2012:chapter 15, Purves 1998:chapters 6 and 7, Morris 2013:chapters 6, 7 and 40, Campbell 2008:chapters 8 and 9, Alberts 2008: 65 and 101, and Garrett 2005:chapter 17.

5.6.1 outline and principles

Figure 5.19 is based on Alberts 1998a:figure 2-14, Alberts 2008:figures 2-36 and 2-55, and Garrett 2005:figure 17.4. The nutrient molecules and the cell's biomolecules are made of the same small molecules, represented by the geometric shapes, but arranged in different ways. For example, the cell will consume another organism, and break its DNA down into nucleotides with the bases G, C, A and T, but then assemble these into the sequence of its own DNA.

"**These two sets of reactions**", Berg 2012:428.

"**Second, they break down**", Garrett 2005:8.

"***we eat proteins***", Purves 1998:42.

"***the final energy-transfer molecule***", and "***heat our bodies***", Morowitz 2002:73.

"**Virtually every metabolic reaction**", Garrett 2005:19, Alberts 2008:101.

"***the vast world***", Morowitz 2002:73.

5.6.2 a cell's metabolism increases the universe's entropy

"***the total chemical activity***", Purves 1998:120.

"***the overall process***", Voet 1999:354.

"**The totality**", Alberts 2008:67.

5.7 from chemistry to biology

5.7.1 from a finite event to a continuous process

Figure 5.20(a-i) has been drawn by Hilary McNeil; **part (a-ii)** is after Hokusai.

mechanical systems

chemical systems

"***the most stable form***", Pross 2016:172.

"***directed to what is called***", Pross 2016:172.

"**Because there are always unoxidised glucose molecules**", Addy Pross presents a bird hovering by flapping its wings as "*a metaphor for all living things*" (Pross 2016:25). The bird is in an unstable state because if it stops flapping its wings it will fall to the ground. The bird is able to maintain its unstable state by the continual expenditure of energy, and so overcome the Earth's gravitational pull. A biological cell, "*from a thermodynamic point of view, is also unstable and exists in what is termed a far-from-equilibrium state in that it also must continuously expend energy to maintain that state*" (Pross 2016:25).

From this point of view, the biological cell is seen as a purposeful entity, which is flapping its "chemical wings" in continually taking in nutrients and expelling wastes, in order to maintain its non-equilibrium state. However, the surfer's position on the wave is not sustained by paddling uphill, but by the wave's steady advance. Similarly, the cell's position on the free energy "wave" is not sustained by flapping any chemical "wings", but by the wave's constant advance. From this viewpoint the cell's cycle of metabolic reactions go round like a turbine in a hydro-electric power station, which spins as the water falls through it. This is consistent with Carl Woese's view of biological organisms as "*patterns in an energy flow*" (Woese 2004:176).

"***controlled unwinding of energy***", Atkins 1995:21.

"… **the cell can create a variety of metabolic reaction paths** …", Morowitz has given a simple argument to show that the flow of energy through a chemical system generates complexity, in the form of larger molecules (Morowitz 1992:82 and figure 3).

a biological cell is a conduit for the dispersal of free energy

"***Chemistry requires the flow of energy***", Goodenough 2000:18.

"**The biological cell**", Auletta observes that "*biological systems download entropy into the environment so as to acquire order*" (Auletta 2011:208). Adam Rutherford summarises: "*To be alive is to struggle against entropy*" (Rutherford 2014a:77). The "surfing" cell is not struggling against entropy, but rather going with the flow of dispersing free energy.

"**So, rather than the biological cell**", Peter Hoffmann has written, "*Life does not exist despite the second law of thermodynamics; instead, life has evolved to take full advantage of the second law wherever it can*" (Hoffmann 2012:87). This viewpoint seems to see life in terms of entities with purpose and agency, which exist apart from the universe and the Second Law. I rather see biological life as the natural consequence of a universe whose entropy must always be increasing. From this viewpoint, biological organisms are local swirls in the universe's molecular continuum, part of the bigger pattern of dispersing energy.

We've seen how molecules randomly diffuse around inside the cell, driven by their thermal energy, so they come together in every possible combination (section 5.2.2). But the requirement that the cell's metabolism must disperse free energy, and add to the universe's entropy, sets strict limits on the chemical reactions that are possible.

Here we see the partnership of the first and second laws of thermodynamics. The universe possesses an inexhaustible thermal energy (the First Law), which drives the endless chaotic motion of molecules. But their range of their interactions is limited by the requirement that the universe's entropy must always be increasing (the Second Law). Thus, the cell's existence is a "*creative collaboration between chaos and necessity*" (Hoffmann 2012:8).

The First and Second Laws, operating at Level 4, describe how the outcomes of chance molecular interactions are constrained by necessity. These two thermodynamic principles perhaps pre-figure Darwinian natural selection, operating at Level 5, in which random genetic variations produce a range of organisms, but the necessity of surviving in a competitive world eliminates all but the fittest (see section 5.7.5).

"***have exploited every conceivable means***", Madigan 2003:547.

"***Organisms are resilient patterns***", Woese 2004:176.

the spontaneous evolution of biological life

"**And so this leads**", Freedman 2002:section 29-2. An alternative statement is "*whatever is possible is compulsory*" (Gilmore 2001:124).

"**From this viewpoint**", Charles Darwin was equivocal on this point, caught between the scientific argument and the cultural pressure of the time. At the end of the 1st (1859) edition of "The Origin of Species", he wrote, "*There is grandeur in this view of life … having been originally breathed into a few forms or into one; …*" (Darwin 1985:459). However, in the 2nd (1860) edition, he changed this to, "*There is grandeur in this view of life … having been originally breathed by the Creator into a few forms or into one; …*" (https://darwin-online.org.uk/contents.html). Darwin later "*expressed regret for bowing to pressure to put a Creator back into the scheme of life's origins*" (Carroll 2009:281). He perhaps expressed his true opinion in one of his last letters, in which he speculated that "*the principle of life will hereafter be shown to be a part, or consequence of some general law*" ("Some unpublished letters of Charles Darwin", edited by Sir Gavin de Beer, letter 41, page 59, available at https://darwin-online.org.uk/content/frameset?itemID=F1595&viewtype=text&pageseq=1). See also section 9.3.

The processes by which biological "life" has spontaneously emerged from geochemical processes are becoming increasingly clear. For very readable accounts, see Lane 2010:chapter 1, Lane 2016:chapters 3 and 4, and Rutherford 2014a. Morowitz has outlined a scenario for the emergence of biological life, based on the spontaneous formation of membrane-bounded protocells (Morowitz 1992:chapter 13).

For a concise discussion of Darwinian evolution and "intelligent design", see Kenneth Miller, at http://www.millerandlevine.com/km/evol/design2/article.html.

"***The conditions of the infant earth***", Rutherford 2014a:117.

5.7.2 going beyond chemical reactions

"**The accounting of atoms**", Garrett 2005:section 27.2.

"***reveals a fundamental difference***", Garrett 2005:881.

5.7.3 long-range order in crystals and cells

"**A crystal is an inanimate aggregate**", Cotterill 2008:chapter 4.

"**So, in a crystal, we find particles in fixed positions**", Atkins 1994:180.

"***A living cell is not in equilibrium***", Campbell 2015:150.

5.7.4 a musical analogy

"**If we extend the metaphor to a living cell**", Goodenough sees music in cellular functions, so that "*patterns of gene expression are to organisms as melodies and harmonies are to sonatas. It's all about which sets of proteins appear in a cell at the same time (the chords) and which sets come before or after other sets (the themes) and at what rate they appear (the tempos) and how they modulate one another (the developments and transitions)*" (Goodenough 2000:58).

5.7.5 the world becomes Darwinian

metabolism and replication

"***A non-living thing is passive***", Tudge 2006:16.

"***it is not vital to reproduce***", Tudge 2006:16.

"***only those that reproduce endure***", Tudge 2006:17.

"**were both crucial elements**", Pross 2016:159. "*In fact, the moment some non-metabolic (downhill) replicator acquired an energy-gathering capability, could be thought of as the moment that life began*" (Pross 2016:158).

The argument over what came first, metabolism or replication, is "*a barren debate. Replication is doubling, which consumes building blocks in an exponential fashion. Unless those building blocks are replenished at a similar rate, replication swiftly ceases*" (Lane 2016:96).

However, Nick Lane argues from thermodynamic principles that "*to make a cell from scratch requires a continuous flow of reactive carbon and chemical energy across rudimentary catalysts in a constrained through-flow system*" (Lane 2016:120). Deep-sea alkaline thermal vents provide these conditions, and he envisages geochemical reactions taking place within the vent's porous structure, but "*without any genetic basis or real complexity*" (Lane 2016:134). Subsequently, the genetic code evolved, "*enabling protocells to make more or less exact copies of themselves. The earliest forms of selection, based on relative rates of synthesis and degradation, gave way to proper natural selection, in which populations of protocells with genes and proteins began to compete for survival within vent pores*" (Lane 2016:135).

John Maynard Smith and Eörs Szathmáry observe, succinctly, that "*Before there can be heredity there must be reproduction, and before that there must be growth*" (Smith 2000:6).

natural selection

"**The emergence of metabolising, self-replicating cells**", Morowitz 2002:78 and 85. Morowitz's list is almost identical to the one summarised by Darwin in 1859: "*These laws, taken in the greatest sense, being Growth with Reproduction; Inheritance which is almost implied by reproduction; Variability from the indirect and direct action of the external conditions of life, and from use and disuse; a Ratio of Increase so high as to lead to a Struggle for Life, and as a consequence to Natural Selection, entailing Divergence of Character and the Extinction of less-improved forms*" (Darwin 1985:459).

Mayr has summarised Darwin's model of Natural Selection (Mayr 2002:128, box 6.1).

"***novelty piles upon novelty***", Morowitz 2002:78.

"***At the first step***", Mayr 2002:133.

"**So, the first step**", and Sean Carroll summarises this as "*chance and necessity*" (Carroll 2007:163). He also adds a third step, which is time (Carroll 2007:155), so evolutionary change accumulates over time, like compound interest (Carroll 2007:43). Darwin's deep understanding of geology had enabled him to comprehend the "*cumulative power of gradual change over the enormous spans of time available*" (Carroll 2007:212).

Genetic change in a competitive environment leads to selection, whose efficacy can be quantified as a selection coefficient, which is analogous to an interest rate (Carroll 2007:49). Thus, a beneficial trait that results in the production of 101 offspring, where those without it produce 100 offspring, will be like a 1% compound interest rate. Selection coefficients can be positive or negative, and indicate fitness. However, evolution acts only in the present, within a given environment, and acts only on what a species needs at the time, not on what is not yet needed. So, "*fitness is a relative, transient status, not an absolute or permanent state*" (Carroll 2007:61).

Sean Carroll points out that there are two or more alternative forms of a gene present in substantial numbers within a species or a population of organisms. So, evolution is not "*a matter of "waiting" for a new mutation*", but is "*the preferential increase or decrease of alternative forms in responses to changes in conditions*" (Carroll 2007:163).

We may think of natural selection as only generating novelty and change in the form of new beneficial traits, but Sean Carroll points out that it also acts to remove "injurious change", and thereby preserve existing traits (Carroll 2007:72). For example, there are about 500 genes that are found in the genomes of Archaea, bacteria, fungi, plants and animals (Carroll 2007:79). These genes are vital to fundamental processes in cells, and have remained the same for more than 2 billion years – they are effectively immortal. These genes are as vulnerable to chance mutations as all other genes, but have survived because natural selection has acted to purify them of any changes that would diminish their function (Carroll 2007:81).

The two steps in natural selection lead to the competitive exclusion principle, which can be stated very succinctly, and alliteratively, as "*complete competitors cannot coexist*" (Hardin 1960:1292). So, if two non-interbreeding populations, *A* and *B*, occupy (1) the same ecological niche, and (2) the same geographic territory, and (3) if population *A* multiplies even the least bit faster than population *B*, then ultimately *A* will replace *B*, which will become extinct.

This is analogous to the Pauli exclusion principle, which states, for example, that "*no two electrons in an atom can be in the same quantum state*", that is, no two electrons can have the same set of values for the 4 quantum numbers (Tipler 1999:833 and 1185). The Pauli exclusion principle was described in section 3.7.1.

"**This evolutionary "two-step"**", Jastrow 2008, chapter 12, Campbell 2015, chapter 25, Kaplan 2010.

"**If we think of the cell as a ship**", "*Natural selection is a rudder, not a driving force*" Pross 2016:149.

5.7.6 self and non-self

"**Inside its membrane**", in the sense that an autonomous system "*can specify its own laws, what is proper to it*" (Maturana 1992:48).

"***builds an image of the world***", Bray 2009:*x* and 164.

"***No life without a membrane***", Margulis 1997:55. And Harold Morowitz writes, "*To be an entity, distinguished from the environment, requires a barrier to free diffusion*", so that the formation of isolated regions enclosed in membranes "*represents a discrete transition from nonlife to life*" (Morowitz 1992:8 and 9).

"***semipermeable boundaries***", Margulis 1997:54.

"***by integrating in a new way***", Auletta 2011:271.

5.8 a review of Levels 1–5

"***endless forms most beautiful***", Darwin 1985:460.

references

Alberts, B., *et al*. (1998a), *Essential Cell Biology*, Garland Publishing, New York.

Alberts, B. (1998b), "The cell as a collection of protein machines: preparing the next generation of molecular biologists", *Cell*, volume 92, p. 291–294.

Alberts, B., et al. (2008), *Molecular Biology of the Cell*, Garland Science, New York, 5th edition.

Atkins, P. (1994), *The Second Law*, Scientific American Library, W. H. Freeman, New York, paperback edition.

Atkins, P. (1995), *The Periodic Kingdom: a journey into the land of the chemical elements*, Weidenfeld and Nicolson, London.

Atkins, P. and Jones, L. (2002), *Chemical Principles: the quest for insight*, W.H.Freeman, New York, 2nd edition.

Atkins, P. and de Paula, J. (2006*), Atkins' Physical Chemistry*, Oxford University Press, 8th edition.

Atkins, P. (2007), *Four Laws that Drive the Universe*, Oxford University Press.

Auletta, G. (2011), *Cognitive Biology: dealing with information from bacteria to minds*, Oxford University Press.

Ball, P. (2023), *How Life Works: a user's guide to the new biology*, Picador, London.

Berg, H. and Purcell, E. (1977), "Physics of chemoreception", *Biophysical Journal*, volume 20, p. 193–219.

Berg, H. (1993), *Random Walks in Biology*, Princeton University Press, expanded paperback edition.

Berg, T., Tymoczko, J. and Stryer, L. (2012), *Biochemistry*, W. H. Freeman, New York, 7th edition.

Bray, D. (2009), *Wetware*, Yale University Press, New haven, CT.

Campbell, N. *et al*. (2008), *Biology*, Benjamin Cummings, San Francisco, CA, 8th edition.

Campbell, N. *et al*. (2015), *Biology: a global approach*, Pearson, Harlow, Essex, Global 10th edition.

Capra, F. and Luisi, P. L. (2016), *The Systems View of Life: a unifying vision*, Cambridge University Press, Cambridge.

Carroll, S. (2016), *The Big Picture: on the origins of life, meaning and the universe itself*, Oneworld Publications, London.

Carroll, S. B. (2007), *The Making of the Fittest: DNA and the ultimate forensic record of evolution*, W. W. Norton, New York.

Cotterill, R. (2008), *The Material World*, Cambridge University Press.

Cudmore, L. (1977), *The Centre of Life: a natural history of the cell*, David and Charles, Newton Abbot, Devon.

Darwin, C. (1985), *The Origin of Species*, Penguin Classics, edited by J. Burrow, 1st edition, 1859.

Dawkins, R. (2005), *The Ancestor's Tale: a pilgrimage to the dawn of life*, Phoenix, London.

Delsemme, A. (1998), *Our Cosmic origins: from the big bang to the emergence of life and intelligence*, Cambridge University Press, Cambridge.

Dyson, F. (1987), "Random walks in Biology", *Physics Today*, volume 40, p.74; a review of Berg's 1984 book.

Erickson, H. (2009), "Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy", *Biological Procedures Online*, volume 11, Number 1, available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3055910/.

Feynman, R., Leighton, R. and Sands, M. (1963), *Lectures on Physics: volume I*, Addison-Wesley, Reading, Massachusetts, 6th printing.

Flamholz, A., Phillips, R. and Milo, R. (2014), "The quantified cell", *Molecular Biology of the Cell*, volume 25, p. 3497–3500.

Gee, H. (2021*), A (very) Short History of Life on Earth: 4.6 billion years in 12 chapters*, Picador, London.

Garrett, R. and Grisham, C. (2005), *Biochemistry*, Thomson Brooks/Cole, Belmont, CA, 3rd edition.

Gilmore, R. (2001), *The Wizard of Quarks*, Copernicus Books, New York.

Goodenough, U. (2000), *The Sacred Depths of Nature*, Oxford University Press.

Goodsell, D. (1991), "Inside a living cell", *Trends in Biochemical Sciences*, volume 16, p. 203−206.

Goodsell, D, (2009), *The Machinery of Life*, Copernicus Books, New York, 2nd edition.

Goodsell, D. (2016), *Atomic Evidence: seeing the molecular basis of life*, Copernicus Books, Springer Nature.

Gribbin, J. (2008), *The Universe: a biography*, Penguin, London.

Halford, S. and Marko, J. (2004), "How do site-specific DNA-binding proteins find their targets?", *Nucleic Acids Research*, volume 32, p. 3040−3052.

Hanson, R. and Hanson, T. (1996), "Methanotrophic bacteria", *Microbiological Reviews*, volume 60, p. 439−471.

Hardin, G. (1960), "The competitive exclusion principle", *Science*, volume 131, p. 1292−1297.

Haunerland, N. *et al*., (1994), "Three-dimensional structure of the muscle fatty-acid-binding protein isolated from the desert locust *Schistocerca gregaria*", *Biochemistry*, volume 33, p. 12378−12385.

Hoffmann, P. (2012), *Life's Ratchet: how molecular machines extract order from chaos*, Basic Books, New York.

Jastrow, R. and Rampino, M. (2008), *Origins of Life in the Universe*, Cambridge University Press.

Kaplan, M. (2010), *David Attenborough's First Life*, HarperCollins, London.

Kent, M. (2000), *Advanced Biology*, Oxford University Press.

Kolb, V. (2015), "Origins of Life: chemical and philosophical approaches", *Evolutionary Biology*, DOI:10.1007/s11692-015-9361-4.

Kwon, K-A, *et al*. (2013), "High-speed camera characterization of voluntary eye blinking kinematics", *Journal of the Royal Society Interface*, volume 10:20130227.

Lambert, F. (2002b), "Entropy is simple, qualitatively", *Journal of Chemical Education*, vol.79, p. 1241−1246, and an updated (2005) version is available at https://web.archive.org/web/20140110153956/http://entropysite.oxy.edu/entropy_is_simple/index.html

Lane, N. (2010), *Life Ascending: the ten great inventions of evolution*, Profile books, London.

Lane, N. (2016), *The Vital Question: why is life the way it is?*, Profile books, London.

Madigan, M., Martinko, J. and Parker, J. (2003), *Brock Biology of Microorganisms*, Pearson, Upper Saddle River, NJ, 10th edition.

Margulis, M. and Sagan, D. (1997), *Microcosmos: four billion years of microbial evolution*, University of California Press, Berkely, CA.

Margulis, L. and Schwartz, K. (1998), *Five Kingdoms: an illustrated guide to the phyla of life on Earth*, W. H. Freeman, New York, 3rd edition.

Mason, S. (1991), *Chemical Evolution: origins of the elements, molecules and living systems*, Clarendon Press, Oxford.

Maturana, H. and Varela, F. (1992), *The Tree of Knowledge: the biological roots of human understanding*, Shambhala, Boston, MA, revised edition.

Mayr, E. (2002), *What Evolution Is*, Phoenix, London.

McMurry, J. (2000), *Organic Chemistry*, Brooks/Cole, Pacific Grove, CA, 5th edition.

Milo, R. and Phillips, R. (2016), *Cell Biology by the Numbers*, Garland Science, New York.

Morowitz, H. (1992), *The beginnings of Cellular Life: metabolism recapitulates biogenesis*, Yale University Press, New Haven and London.

Morowitz, H. (2002), *The Emergence of Everything: how the world became complex*, Oxford University Press, New York.

Morris, J., et al., (2013), *Biology: how life works*, W. H. Freeman, New York.

Nicholson, D. (2003), *Metabolic Pathways Chart*, 22nd edition, available at http://www.iubmb-nicholson.org/chart.html.

Phillips, R. et al. (2013), *Physical Biology of the Cell*, Garland Science, New York, 2nd edition.

Popa, R. (2015), "Elusive definition of life", chapter 15, p. 325−348, in *Astrobiology: an evolutionary approach*, edited by Vera Kolb, CRC Press, Taylor and Francis Group, Boca Raton, FL.

Pross, A. (2016), *What is Life?: how chemistry becomes biology*, Oxford University Press.

Purves, W., et al., (1998), *Life: the science of Biology*, Sinauer Associates, Sunderland, MA, 5th edition.

Rutherford, A. (2014a), *Creation: the origin of life*, Penguin.

Rutherford, A. (2014b), *Creation: the future of life*, Penguin.

Smith, J. M. and Szathmáry, E. (2000), *The Origins of Life: from the birth of life to the origins of language*, Oxford University Press.

Stryer, L. (1995), *Biochemistry*, W. H. Freeman, New York, 4th edition.

Szenk, M., Dill, K. and de Graff, M. (2017), "Why do fast-growing bacteria enter overflow metabolism? Testing the membrane real estate hypothesis", *Cell Systems*, volume 5, p. 95−104.

Tudge, C. (2006), *The Secret Life of Trees: how they live and why they matter*, Penguin, London.

Voet, D., Voet, J. G., and Pratt, C. W. (1999), *Fundamentals of Biochemistry*, John Wiley, New York.

Vojtěchovský, J. et al. (1999), "Crystal structures of myoglobin-ligand complexes at near-atomic resolution", *Biophysical Journal*, volume 77, p. 2153−2174.

Walker, G. (2007), *An Ocean of Air: a natural history of the atmosphere*, Bloomsbury, London.

Wang, V. et al. (2017), "Alkane oxidation: methane monooxygenases, related enzymes, and their biomimetics", *Chemical Reviews*, volume 117, p. 8574−8621.

Woese, C. (2004), "A New Biology for a New century", *Microbiology and Molecular biology Reviews*, volume 68, p. 173−186.

Wolpert, L. (2009), *How We Live and Why We Die*, Faber and Faber, London.

Yu, I. et al. (2016), "Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm", *Elife*, volume 5, e19274.

Zewail, A. (1999), "Femtochemistry. Past, present, and future", *Pure and Applied Chemistry*, volume 72, p. 2219−2231.