

## 7. THE CELLULAR ENVIRONMENT

14 January 2023

Armed with an appreciation for the variation in the population-genetic environment experienced by different lineages and the principle factors governing evolutionary change, we now consider a few of the most basic chemical and physical constraints dictating the properties of cells. Unlike the population-genetic environment, several aspects of the cellular environment are largely invariant across the Tree of Life. These include the elemental makeups of cells, the diffusion properties of molecules, the effects of temperature on biological processes, and the amounts of energy accessible from various food types. Some lineages have evolved special attributes to cope with such challenges, e.g., increased protein stability in thermophiles, and the use of motors for molecular transport in eukaryotes. Nonetheless, immutable laws of physics and chemistry ultimately dictate what natural selection can and cannot do.

The cellular environment is in large part defined by ancient historical contingencies that established the foundational features of biology. For example, the earliest stages of evolution set the elemental requirements of the biochemical building blocks from which all of today's cell bodies are constructed. Life depends on < 20% of the 119 but aside from carbon, hydrogen, and oxygen, most of these have environmental concentrations thousands to millions of times lower than those found in cellular biomass, highlighting the power of cells to sequester nutrients. Of the myriad forms of organic compounds, life has come to rely on just a handful of fundamental types – amino acids, nucleotides, lipids, carbohydrates, and a few others.

Here, we consider some of the quantitative consequences of biophysical and chemical constraints on cell biology. With an overview of what cells are made of, how many molecules are present per cell, and how much carbon and energy is required for cellular reproduction, the stage will then be set for understanding the breadth of issues covered in subsequent chapters. As introduced here, and further elaborated on in Chapter 8, numerous cellular features scale with cell size in predictable ways that transcend phylogenetic boundaries. Finite numbers of molecules per cell, combined with the physical constraints associated with molecular diffusion and temperature, dictate the possible rates of intracellular biochemical reactions. The energy content of resources constrains the rate at which new biomass can be constructed. These and many other “rules of life” define the ultimate limits of the evolutionary playing field. An excellent overview of many of the points discussed below can be found in Milo and Phillips (2016).

### The Molecular Composition of Cells

Given that life evolved in an aqueous environment, it is not surprising that the primary component of all of today's cells is water, albeit with a much higher solute load than in the surrounding environment. Cell dry weights scale with cell volume in what appears to be a near universal relationship across all phylogenetic groups. Over a range of eleven orders of magnitude in cell volume, there is a smooth power-law relationship of

$$W \simeq 0.57V^{0.92}, \quad (7.1)$$

where cell dry weight  $W$  has units of pg (picograms, or  $10^{-12}$  grams) and cell volume has units of  $\mu\text{m}^3$  (cubic microns, or  $10^{-12}$  ml) (Figure 7.1). The exponent is significantly less than 1.0, indicating that cell density ( $W/V$ ) decreases with the  $\sim 0.08$  power of cell volume. Because 1  $\mu\text{m}^3$  of water weighs 1 pg, these results imply that between one-fifth (large eukaryotic cells) and one-third (small bacterial-sized cells) of total cell weights are comprised of biomolecules and ions. Exceptions occur in diatoms, haptophytes, and foraminiferans, whose cells have hard outer coverings.

**Water.** Because of life's association with water from the start, many of the features of biology have been permanently molded by the unique properties of this simple molecule. Consisting of a bent complex of two hydrogen atoms and one oxygen atom,  $\text{H}_2\text{O}$  molecules have polarity, with a slight negative charge on the oxygen side and a slight positive charge on the hydrogen side. As a result, liquid water naturally forms a three-dimensional network with each molecule being connected to three to four others via hydrogen bonds in a sort of tetrahedral arrangement (Figure 7.2).

These unique organizational features enable water to operate as a highly effective solvent for other polar molecules. Solubility is an essential feature of most biomolecules involved in chemical reactions requiring diffusive encounters with dissolved reactants. On the other hand, the exclusion of nonpolar molecules from the water network provides a pathway for the spontaneous construction of certain cellular features. For example, in water, the hydrophobic tails of lipid molecules naturally aggregate in a highly coordinated fashion (Chapter 15), generating the membranes upon which cells rely.

The hydrogen-bonding ability of water can also present a problem. First, the inner hydrophobic cores that maintain protein structure can be compromised by the intrusion of water molecules. Hydrophobic surface residues also cause proteins to be promiscuously sticky. This imposes strong selective pressure for soluble proteins to achieve their globular structures by populating their outer surfaces with hydrophilic amino acids (Chapter 12). Second, the cohesion within networks of water molecules imposes a drag on large molecules moving through the cytoplasm and on cells moving through aqueous environments, limiting rates of intracellular reactions, and extracellular nutrient uptake and swimming speeds of mobile species (Chapters 16, 18, and 19).

Finally, the thermal properties of water are unique. The viscosity of water declines by nearly 50% from  $4^\circ\text{C}$  to  $40^\circ\text{C}$ , so warm water imposes less resistance to the directed movements of cells but also provides less buoyancy (e.g., imposing higher sinking velocities in aquatic settings). At normal atmospheric pressure, pure water freezes at  $0^\circ\text{C}$ , imposing a lower temperature barrier to single-celled organisms incapable of thermoregulation. However, the fact that water has a maximum density at

4°C provides a buffer against such an extreme, as aquatic environments freeze from the top down, with bottom waters never colder than 4°C. Ball (2008, 2017) provides a comprehensive overview of many additional knowns and unknowns regarding the biological consequences of water.

**Elemental composition.** Of the many dozens of chemical elements found in the natural world, only about 20 are essential to life. Ignoring hydrogen and oxygen, carbon is always predominant in terms of molar composition, followed by nitrogen (Table 7.1). The bulk of the remaining biomass is associated with two other elements, phosphorus and sulfur, incorporated into one or more building blocks of cells (e.g., nucleic acids, amino acids, and lipids), along with five other major ions – sodium, calcium, magnesium, potassium, and chloride. All of these elements generally have intracellular concentrations  $> 1$  mM. Essential trace metals (e.g., iron, manganese, cobalt) that serve as cofactors of individual enzymes are present at  $10\times$  to  $100\times$  lower concentrations. Redfield (1934) first proposed that the ratio of C, N, and P atoms in cells is typically on the order of 106:16:1, and the average of the exemplars in Table 7.1, 100:13:1, is close to this expectation.

A comparison of cellular and environmental elemental concentrations reveals the extent to which cells go to sequester nutrients. There can be considerable variation in the biogeochemistry of different environments, but reliable average estimates exist for the dissolved content of ocean water. As many of the species in Table 7.1 derive from marine environments, molar concentrations in seawater will be used as a reference point. This shows that the degree of cellular enrichment averages  $\sim 5000\times$  for carbon, and 50,000 to 60,000 $\times$  for nitrogen and phosphorus. The remaining major ions range from being nearly isotonic with sea water to enriched by no more than  $25\times$ . On the other hand, several essential trace metals (iron, manganese, and cobalt) are enriched by factors  $> 10^6$ .

To appreciate the challenges imposed by such nutrient acquisition, consider as an example phosphorus, which has an average cellular enrichment of  $\sim 60,000\times$ . Living in an average marine environment, in order to produce an offspring, a bacterial cell with volume  $1\ \mu\text{m}^3$  would need to accomplish the equivalent of fully clearing a surrounding volume of  $\sim 60,000\ \mu\text{m}^3$  of P, and for the trace metals noted above, the equivalent of  $\sim 10^6$  cell volumes would need to be scrubbed clean. For a moderate sized eukaryotic cell,  $100\ \mu\text{m}^3$  in volume, the necessary volumes of environmental clearance are  $100\times$  higher.

When viewed in the context of cell-division times, the impressive rate at which cells harvest nutrients becomes clear. Again, consider a cell with volume  $1\ \mu\text{m}^3$  (equivalent to  $10^{-15}$  liters) at birth. With an average internal concentration of 115 mM for phosphorus (Table 7.1), such a cell would contain  $\sim 7\times 10^7$  P atoms. Cells of this size have a minimum doubling time of  $\sim 0.4$  days at 20°C (Chapter 8), implying an incorporation rate of  $\sim 2000$  P atoms/sec at maximum growth rate. Similar calculations for cells of volume 10, 100, and  $1000\ \mu\text{m}^3$ , growing at maximum rates, indicate incorporation rates of  $\sim 1\times 10^4$ ,  $9\times 10^4$ , and  $6\times 10^5$  P atoms/sec. Given the average 100:13:1 ratio for C:N:P noted above, these incorporation requirements would be 100 and  $13\times$  higher for C and N atoms, respectively. Thus, depending on their size, when growing at maximum rates, cells incorporate on the order of  $10^6$  to  $10^{10}$  atoms per minute.

As can be seen in Table 7.1, there is variation among species in elemental composition, and some of this may relate to cell size. Menden-Deuer and Lessard (2000) summarized the scaling of carbon content with cell volume in a wide variety of unicellular marine eukaryotes. Aside from chrysophytes, which have inexplicably low carbon estimates, the average exponent on the power-law relationship across groups is 0.91 (SE=0.03), so there is a decline in carbon content per cell volume in larger cells. For cells of volume 1, 10, 100, and 1000  $\mu\text{m}^3$ , mean carbon contents are 0.30, 0.23, 0.18, and 0.14  $\text{pg}/\mu\text{m}^3$ , implying a reduction in cell density with increasing cell size, consistent with the results in Figure 7.1. Using Equation 7.1, the average fractional contributions of carbon to dry weight for cells of these sizes are  $\approx 0.53$ , 0.49, 0.46, and 0.44, respectively. Thus, a rough rule of thumb from these and other studies (Ho and Payne 1979; Roels 1980; Finlay and Uhlig 1981; Williams et al. 1987; von Stockar and Marison 1989; de Queiroz et al. 1993) is that  $\sim 50\%$  of average dry weight in both prokaryotic and eukaryotic cells consists of carbon.

**Table 7.1.** Contents of the major elemental constituents (other than hydrogen and oxygen) in a variety of unicellular species. Concentrations in the top half of the table are in units of mM, whereas as those in the bottom half are in units of  $\mu\text{M}$ . Species are in order of increasing cell volume ( $\mu\text{m}^3$ ). The means for Ca and Sr exclude the haptophytes *E. huxleyi* and *Gephyrocapsa oceanica*, which have hard outer shells consisting of these elements. *Prochlorococcus* and *Synechococcus* are cyanobacteria; *Vibrio* and *Escherichia* are heterotrophic bacteria; *Pycnococcus*, *Nannochloris*, *Pyramimonas*, and *Dunaliella* are green algae; *Saccharomyces* is budding yeast; *Nitzschia*, *Amphidinium*, and *Thalassiosira* are diatoms; and *Prorocentrum* and *Thoracosphaera* are dinoflagellates. Seawater concentrations are taken from Nozaki (1997). References: cyanobacteria (Heldal et al. 2003); heterotrophic bacteria (Fagerbakke et al. 1996, 1999); yeast (Lange and Heijnen 2001); and all others (Ho et al. 2003).

Species	Size	C	N	P	S	K	Na	Mg	Ca	Cl
<i>Prochlorococcus</i> sp.	0.16	15323	1682	87	82	49	410	371	25	173
<i>Synechococcus</i> sp.	1.00	14906	1755	122	72	78	248	104	49	120
<i>Vibrio natriegen</i>	3.50	8333	1837	157	116	320	400	73	8	1320
<i>Escherichia coli</i>	3.80	7675	1880	263	74	62	210	61	10	104
<i>Pycnococcus provasoli</i>	10	14000	1900	72	77	89		19	4	
<i>Nannochloris atomus</i>	14	14000	2000	81	29	78		19	2	
<i>Saccharomyces cerevisiae</i>	67	15809	2218	131	27	39	7	26	0	
<i>Nitzschia breviostris</i>	119	11000	1700	250	290	610		150	67	
<i>Emiliana huxleyi</i>	142	10000	1200	130	100	110		18	19000	
<i>Gephyrocapsa oceanica</i>	142	8900	1000	140	140	130		18	18000	
<i>Dunaliella tertiolecta</i>	227	11000	1900	49	14	18		18	1	
<i>Amphidinium carterae</i>	514	1200	160	9	12	1		5	3	
<i>Pyramimonas parkeae</i>	587	6800	570	32	47	27			55	
<i>Prorocentrum minimum</i>	833	22000	1800	16	350	210		160	61	
<i>Thoracosphaera heimii</i>	1353	5100	400	63	82	63		30	2800	
<i>Thalassiosira eccentrica</i>	6627	18000	1900	240	470	790		520	160	
Means		11503	1494	115	124	167	255	106	232	429
Seawater		2.25	0.03	0.002	28	10.2	469	52.7	10.3	546
Cellular enrichment		5,100	50,000	57557	4.4	16.4	0.5	2.0	22.6	0.8

	Sr	Fe	Mn	Zn	Cu	Co
<i>Pycnococcus provasoli</i>	8	910	150	66	38	7
<i>Nannochloris atomus</i>	4	1100	93	140	19	7
<i>Saccharomyces cerevisiae</i>		354	31	642	46	
<i>Nitzschia brevirostris</i>	330	790	590	69	46	14
<i>Emiliana huxleyi</i>	44000	460	940	50	9	39
<i>Gephyrocapsa oceanica</i>	39000	560	990	57	16	50
<i>Dunaliella tertiolecta</i>	4	560	93	74	33	1
<i>Amphidinium carterae</i>	11	120	47	12	5	3
<i>Pyramimonas parkeae</i>	390	500	250	48	20	8
<i>Prorocentrum minimum</i>	470	1100	980	140	440	73
<i>Thoracosphaera heimii</i>	5000	110	79	7	4	6
<i>Thalassiosira eccentrica</i>	950	1600	500	240	68	59
Means	796	680	395	129	62	24
Seawater	89	0.00054	0.00036	0.0054	0.0024	0.000020
Cellular enrichment	8.9	1,260,000	1,086,000	24,000	26,000	1,182,000

**Biomolecules.** The organic fraction of cells consists primarily of macromolecules such as proteins, nucleic acids, lipids, and carbohydrates (as well as their precursor building blocks). Most information on this fundamental issue is confined to quite old literature, sometimes based on methods that are not terribly reliable, and variation is also associated with growth conditions during assays (Chapter 9). The most reliable statement that can be made is that proteins comprise the largest fraction of the organic component of cellular biomass (on a dry weight basis), typically in the range of 40 to 60%, but somewhat lower in eukaryotes than in prokaryotes (Figure 7.3). The other primary contributors are RNA (including messenger, ribosomal, and transfer RNAs), carbohydrates (especially in species with cell walls – most bacteria, and some eukaryotes such as fungi and plants), and lipids (which are more enriched in eukaryotic cells, owing to the presence of internal membranes).

Although the fractional contributions to biomass from protein, RNA, lipids, and carbohydrates do not obviously scale with cell volume, the data are scant and noisy enough that such patterns cannot be entirely ruled out. However, the matter is readily accessible for DNA, as genomes have been sequenced for a substantial number of species, and  $10^9$  bp of DNA is equivalent to  $\sim 1$  pg dry weight. Here, there is a very strong negative scaling of proportional contribution with cell volume (Figure 7.3). Despite its centrality to all of life, DNA almost never constitutes  $> 10\%$  of the biomass of any cell, and this fraction declines to 0.001% in relatively large eukaryotic cells. Thus, although larger cells tend to have larger genomes (Lynch 2007), scaling as  $\sim V^{0.25}$ , the proportional investment in total cellular biomass is progressively diminished.

### Numbers of Biomolecules per Cell

The preceding results provide a generic view of cellular contents per unit biomass, but finer details (e.g., numbers of molecules per cell volume) are required to under-

stand issues related to the properties of specific gene products, such as reaction rates among colliding particles, cellular stochasticity, random variation in inheritance, etc. High-throughput methods for characterizing and quantifying individual mRNA and protein molecules provide insight into these matters. Although data are only available for a few species, over a range of five orders of magnitude in cell size (including both prokaryotes and eukaryotes), the total number of protein molecules/cell scales nearly isometrically with cell volume ( $V$ , in units of  $\mu\text{m}^3$ ),

$$N_{\text{tot},p} = (2.0 \times 10^6)V^{0.95}, \quad (7.2a)$$

(Figure 7.4). The smallest known bacterial cells harbor  $< 10^5$  total protein molecules, whereas larger eukaryotic cells (like those in metazoans) contain  $> 10^9$ .

To resolve the degree of gene-expression stochasticity, a view at the gene-specific level is necessary. Owing to the fact that large cells often harbor more genes, the average number of proteins within a cell per active gene scales with cell volume more weakly than the total number of proteins per cell,

$$\bar{N}_p = 1820V^{0.68}. \quad (7.2b)$$

Moreover, there is substantial variation in the amount of protein product associated with different genes within a cell around the overall mean  $\bar{N}_p$ . Distributions of the numbers of proteins for individual genes are approximately log-normal (a normal “bell-shaped” distribution on a logarithmic scale), with the mean being considerably larger than the median, owing to the long tail to the right. With such distributions, the smallest known cells are on the edge of having just one (or fewer) proteins per cell for some genes. For a cell the size of *E. coli*,  $\sim 1 \mu\text{m}^3$ , a substantial number of genes are represented by fewer than 100 protein molecules per cell (Figure 7.4). This means that genetically identical offspring resulting from binary fission can vary substantially in their protein contents. If each of the  $n$  copies of a protein in a parental cell is randomly partitioned to daughters, the coefficient of variation (ratio of standard deviation to the mean) among sisters will equal  $\sqrt{1/n}$ . Further complications arise when proteins are aggregated in vesicles (Chapter 9).

What do the preceding numbers mean in terms of cellular concentrations? Focusing on the number of proteins representing an average gene, Equation 7.2b, the concentration on a per  $\mu\text{m}^3$  basis becomes  $1820V^{-0.32}$ . Multiplying this by  $10^{15} \mu\text{m}^3 / \text{liter}$ , and dividing by the number of molecules per mol (Avogadro’s number,  $6.023 \times 10^{23}$ ), yields an average concentration of  $3.0V^{-0.33} \mu\text{M}$  ( $\mu\text{mol}/\text{liter}$ , where  $1 \mu\text{mol} = 10^{-6} \text{mol}$ ). With protein numbers  $10\times$  above and  $10\times$  below the average, this concentration would be multiplied by 10 and 0.1, respectively. Thus, cellular concentrations of proteins are typically in the nM (nanomolar, or  $0.001 \mu\text{M}$ ) to  $\mu\text{M}$  range, with concentrations tending to decline with increasing cell volume.

The situation is much more extreme for messenger RNAs, as even large, well-nourished cells typically harbor orders of magnitude fewer mRNAs than the total number of proteins (Figure 7.4). The total number of transcripts per cell scales more weakly with cell volume than for the case of proteins,

$$N_{\text{tot},t} = 6760V^{0.42}, \quad (7.3a)$$

and the mean number of transcripts per active gene is just

$$\bar{N}_t = 3.2V^{0.26} \quad (7.3b)$$

(Figure 7.4). Typically, there are hundreds to thousands of more copies of proteins than transcripts per gene within cells, and the average gene is represented by fewer than ten mRNAs at any particular time. As a consequence, a substantial number of genes are at least transiently devoid of transcripts in small cells, and this is even true for a small subset of genes in species with the largest of cells.

As will be discussed in Chapter 8, the numbers of ribosomes per cell also scale across the Tree of Life with cell volume in a predictable manner. Here, however, cells are much more guarded against stochastic loss, as the average number of ribosomes per cell is generally  $> 100$  even in the smallest cells, ranging up to  $10^8$  in the largest cells. This should not be too surprising, as complete loss of ribosomes is equivalent to a death sentence.

### Passive Transport of Particles Through the Cytoplasm

To carry out their key functions, biomolecules often have to travel to particular destinations to encounter specific substrate molecules. Except for large complexes and cargoes within vesicles in eukaryotic cells, most molecules spend the majority of their time moving by passive diffusion. Thus, to understand the ultimate biophysical constraints on cellular functions, we require information on how rapidly molecules can diffuse from one location to another. Due to background thermal motion, each molecule within a cell is continuously jostled in random ways (often referred to as Brownian motion), and until encountering an impervious barrier, such as the cell membrane, will diffuse at a roughly constant average rate, depending on the nature of the medium. The average distance moved after  $t$  time units is a function of the diffusion coefficient  $D$ , defined as the average squared distance of molecular movement per unit time (Foundations 7.1).

The reason for focusing on the squared distance is most easily understood in the context of a random one-dimensional diffusion process. In this case, at each time point a particle has an equal probability of moving to the left vs. the right, so the average directional movement of particles is zero. Nonetheless, when molecules move randomly, such that there is no memory in the process, the noise of each incremental move is cumulative. Thus, although the mean location remains constant, with increasing time a diminishing fraction of molecules will remain in the vicinity of their initial location. With respect to the starting position, the probability distribution of locations of individual molecules becomes wider and wider with time ( $t$ ). Taking the square root of the mean-squared distance,  $\sqrt{2Dt}$ , in the case of a one-dimensional process, provides a measure of average absolute dispersion on the original scale.

Of course, not all diffusion processes in biology are one dimensional. Diffusion of individual molecules within a fluid lipid membrane is a two-dimensional process, whereas diffusion through the cytoplasm is three-dimensional. There is, however, a simple algebraic relationship between the expected magnitude of diffusion and the dimensionality of the process. As just noted, under one-dimensional diffusion, a particle can move in only two directions, right vs. left. Adding a dimension increases the magnitude of dispersion, owing to the reduction in the degree of back-tracking (Figure 7.5). For example, considering a two-dimensional grid, a particle can move in four directions (e.g., north, south, east, west), and on a three-dimensional lattice,

there are six possible routes of movement. In these higher-dimensional cases, the dispersion distance is the radial (straight-line) distance from the initial point, and with two and three-dimensional diffusion, the root mean-squared distances after  $t$  time units become  $\sqrt{4Dt}$  and  $\sqrt{6Dt}$ , respectively. Thus, the rate of diffusion relative to an initial location increases with dimensionality, but the scaling with the square root of time is retained. From these expressions, it can be seen that for an  $n$ -dimensional process, the expected time required for a particle to move an absolute distance of  $d$  units is  $d^2/(2nD)$ .

To understand the implications of diffusion limitation for cellular processes, we require information on how the diffusion coefficient depends on the features of a particle and the medium through which it moves. In its most elementary form, a diffusion coefficient is defined as

$$D = \frac{k_B T}{\gamma}, \quad (7.4)$$

where  $k_B$  is the Boltzmann constant ( $1.38 \times 10^{-16} \text{ cm}^2 \cdot \text{g} \cdot \text{sec}^{-2} \cdot \text{K}^{-1}$ ), which relates energy at the particle level to temperature  $T$  in degrees Kelvin, and  $\gamma$  is the friction coefficient, which is a net measure of the resistance imposed on particle movement by the medium (with units of  $\text{g} \cdot \text{sec}^{-1}$ ). The form of this expression is reasonably intuitive – the numerator is a measure of the jostling due to thermal noise, and the denominator is a measure of resistance to such jostling. Because most of life (other than thermophiles) exists in the range of  $T \simeq 280$  to  $315$  K,  $T$  can be approximated as  $300$  K with only slight loss of accuracy. A sampling of diffusion coefficients for small molecules in an aqueous environment is provided in Figure 7.6, where it can be seen that large proteins diffuse up to  $100\times$  more slowly than small ions.

The friction coefficient depends on the medium as well as on the shape and form of the particle, and many expressions have been developed to accommodate such effects (He and Niemeyer 2003; Dill and Bromberg 2011; Soh et al. 2013). For a perfectly spherical particle with radius  $r$  (in units of cm), the Stokes-Einstein equation tells us that

$$\gamma = 6\pi\eta r, \quad (7.5)$$

where  $\pi \simeq 3.142$  is the universal constant (equal to the ratio of a circle's circumference to diameter), and  $\eta$  is the viscosity of the medium (with units  $\text{g} \cdot \text{cm}^{-1} \cdot \text{sec}^{-1}$ ). For water,  $\eta$  is temperature dependent, taking on values of  $0.013$ ,  $0.011$ ,  $0.010$ ,  $0.0089$ , and  $0.0080 \text{ g} \cdot \text{cm}^{-1} \cdot \text{sec}^{-1}$  at  $10$ ,  $15$ ,  $20$ ,  $25$ , and  $30^\circ\text{C}$ . For simplicity, the  $20^\circ\text{C}$  value will be assumed in the following calculations. The diffusion coefficient of a sphere in a typical aqueous environment then becomes

$$D \simeq \frac{22 \times 10^{-6}}{r}, \quad (7.6)$$

where the numerator has units  $\mu\text{m}^3/\text{sec}$ , and  $r$  has units of  $\mu\text{m}$ . If  $r$  is in units of nm,  $D \simeq 22/r \text{ nm}^2/\text{sec}$ .

Biology introduces numerous complications. For example, most biomolecules depart from a perfectly spherical geometry, and cytoplasm is substantially more viscous than water. Here we will focus on proteins, which typically fold into globular structures. The problem of particle shape can then be dealt with by considering the

effective particle radius. For the ideal case of perfectly packed spherical proteins composed of  $N_{AA}$  amino acids, the radius would scale as  $N_{AA}^{1/3}$  (because the volume of a sphere is proportional to the cube of the radius). However, empirical study implies that the average distance of a protein molecule's parts to a central point (often called the radius of gyration) scales (in units of cm) as

$$r_g = (2.2 \times 10^{-8}) N_{AA}^{0.4} \quad (7.7)$$

(Hong and Lei 2009). Subdivision of proteins into domains, less than perfect packing, and various elastic features contribute to this elevated scaling relative to the ideal situation. Tyn and Gusek (1990) find that a protein with radius of gyration  $r_g$  behaves hydrodynamically on average as though the effective radius is  $r \simeq 1.3r_g$ . Applying this correction factor and Equation 7.7 to Equation 7.6, we obtain an expected diffusion coefficient for a protein in an aqueous environment of

$$D \simeq 770 N_{AA}^{-0.4}, \quad (7.8)$$

with units of  $\mu\text{m}^2/\text{sec}$ . Proteins diffuse at rates that are typically <10% of the rates for individual amino acids (Figure 7.6).

Intracellular crowding imposes an additional impediment to molecular diffusion. The internal milieu of a cell is hardly the open-water environment assumed in most diffusion theory. Rather, 20 to 40% of the cytoplasmic volume of a typical cell is occupied by proteins and other macromolecules (Zimmerman and Trach 1991; Luby-Phelps 2000; Ellis 2001). As a consequence, the average distance between proteins is on the order of the width of the proteins themselves. This then raises questions as to how the basic composition of cells alters the freedom of movement of the very molecules upon which life depends. On the one hand, molecular crowding reduces the aqueous volume that must be searched to locate a small solute. But on the other hand, transient molecular confinement, aggregation with molecules of opposite charge, and sieving effects can inhibit the free diffusion of proteins. Although the net consequences of these added complications are minor for small metabolites, the diffusion coefficients for proteins are reduced by 10- to 50-fold in *E. coli* (Elowitz et al. 1999; Konopka et al. 2006; Nenninger et al. 2010), and perhaps somewhat less in eukaryotic cells (Luby-Phelps 2000; Dix and Verkman 2008).

For example, in an aqueous environment, green fluorescent protein (GFP), with a chain length of 238 amino acids, has a diffusion coefficient of  $87 \mu\text{m}^2/\text{sec}$ , almost exactly as predicted by Equation 7.8. In contrast, empirical estimates of GFP diffusion within the cytoplasm of multiple bacteria (*Caulobacter crescentus*, *E. coli*, *Lactococcus lactis*, and *Pseudomonas aeruginosa*) yields coefficients in the range of 5 to  $15 \mu\text{m}^2/\text{sec}$  (Konopka et al. 2009; Nenninger et al. 2010; Montero Llopis et al. 2012; Guillon et al. 2013; Mika et al. 2014), and on the order of 25 to  $30 \mu\text{m}^2/\text{sec}$  in the slime mold *Dictyostelium* and mammalian cells (Swaminathan et al. 1997; Potma et al. 2001). Examination of a diversity of proteins in *E. coli* demonstrates that despite the crowdedness of bacterial cytoplasm, diffusion is well-described as a Brownian process once the effective viscosity of the medium is accounted for (Belletto et al. 2022). Large complexes diffuse much more slowly. For example, the estimated rate for a ribosome is  $0.04 \mu\text{m}^2/\text{sec}$  in *E. coli* (Bakshi et al. 2012). Membrane proteins undergoing two-dimensional diffusion through a much more densely

packed lipid milieu have diffusion coefficients in the range of 0.02 to 0.03  $\mu\text{m}^2/\text{sec}$  in bacteria, with the rate declining with the number of transmembrane domains in the protein (Kumar et al. 2010; Mika et al. 2014).

The preceding observations suggest that general diffusion processes may speed up in eukaryotes. On the one hand, the average protein chain length for eukaryotes,  $N_{\text{AA}} = 532$ , is 45 to 60% larger than the means in bacteria (365) and archaea (329) (Wang et al. 2011). On this basis, assuming similar folding architectures, all other things being equal, Equation 7.8 implies that a  $\sim 1.5\times$  increase in total chain length should yield a 15% reduction in the average diffusion coefficient for proteins in eukaryotes. However, given that the density of eukaryotic cytoplasm is lower than in prokaryotes, reduced crowding effects may essentially cancel this particle-size effect. A third effect that merits further consideration is that active processes in eukaryotic cells, such as those created by molecular motors, generate as by-products random diffusion-like forces, thereby enhancing rates of molecular movement throughout the cytoplasm even by non-carrier molecules (Guo et al. 2014).

Whether these cytoplasmic features of eukaryotic cells have evolved to facilitate long-distance diffusion and/or result in relaxed selection against protein stickiness remains a matter of speculation (Soh et al. 2013). There is, however, some evidence that the diffusive properties of proteins coevolve with their proteomic environment. For example, Mu et al. (2017) found that when placed in the cytoplasm of *E. coli*, human proteins tend to stick to their foreign environment, but that modification of a few surface amino-acid residues can yield diffusion rates equivalent to the native *E. coli* proteins.

Finally, to appreciate the time scale of passive molecular diffusion, consider a protein of moderate length with a diffusion coefficient of  $D \simeq 20 \mu\text{m}^2/\text{sec}$ . In a three-dimensional setting (e.g., cytoplasmic diffusion), the root mean-squared distance traveled after  $t$  seconds will be  $d = \sqrt{6Dt} \simeq 11\sqrt{t} \mu\text{m}$ . The expected time to travel  $d \mu\text{m}$  is then  $(d/11)^2 \text{ sec}$ . A spheroid bacterial cell with a  $1 \mu\text{m}^3$  volume has a diameter of  $1.2 \mu\text{m}$ , so it would take  $\sim 0.01$  seconds for the protein to travel the width of the cell. For a moderate-sized eukaryotic cell with volume  $100 \mu\text{m}^3$ , traversing the  $5.8\text{-}\mu\text{m}$  width requires  $\sim 0.28$  seconds. For a large spherical cell with volume  $10^5 \mu\text{m}^3$  (which is attained in some marine diatoms and dinoflagellates), traveling the cell width of  $58 \mu\text{m}$  requires  $\sim 28$  seconds. Thus, molecular delivery across a cell based on diffusion alone is effectively instantaneous in bacteria, and comes with no cost, as it is entirely fueled by background thermal noise. In contrast, diffusion becomes dramatically less efficient in large eukaryotic cells, which often transport material by use of molecular motors, which run on ATP (Chapter 16). One final caveat with respect to all of the above results is that the viscosity of cytoplasm appears to vary significantly with the level of cell nutritional state, increasing in starved cells (Joyner et al. 2016). All of the above issues, and more, are reviewed in Schavemaker et al. (2018).

## Intermolecular Encounter Rates

Proteins do not operate in isolation. More often than not, they aggregate into multimeric complexes, and most engage with particular substrate molecules. Diffusion

theory explains the rates of dispersion of individual particles, but the rate of encounter of interacting particles depends on particle sizes and concentrations. As an entrée into this area, we consider the simple situation in which the two interacting particle types are products of the same genetic locus, as in the case of two monomeric subunits coalescing to form a dimer, a very common situation for proteins. (The more general case of two different particles is derived in Foundations 7.2).

To move forward, we require a measure of the encounter rate per unit concentration,  $k_e$ , which is a function of the particle diffusion rate (Foundations 7.2) and has units of events  $\cdot \text{cm}^3 \cdot \text{sec}^{-1}$  (or some other combination of distance and time units). This must be multiplied by the product of the concentrations of the particles to be joined to account for the fact that both interacting partners are randomly diffusing; in this particular example, each particle has the same concentration  $[C]$ . The resultant rate of encounter per unit volume, which has units of events  $\cdot \text{cm}^{-3} \cdot \text{sec}^{-1}$ , must then be multiplied by the cell volume  $V$  (in units of  $\text{cm}^3$  per cell) to give the total rate of encounter events within the cell. A small modification arises because a particle cannot interact with itself, necessitating a correction factor of  $1 - (1/n)$ , where  $n = [C] \cdot V$  is the expected number of particles per cell. (For  $n > 100$ , this modification can be ignored. The final expression for the total rate of encounter then becomes

$$R_E = k_e \cdot [C]^2 \cdot V[1 - (1/n)] = (11 \times 10^{-12}) \cdot [C]^2 V[1 - (1/n)] \quad (7.9)$$

with units of events/cell/sec, and the constant substituted for  $k_e$  applies to the specific case of two spherical particles of the same size (Foundations 7.2).

To gain some appreciation for the constraints on such encounters, and hence the viability of a strategy to dimerize, consider a cell with a  $1 \mu\text{m}^3$  volume (bacterial sized) and a molecule with a concentration of  $1 \mu\text{M}$ , which as noted above is within the range typically seen for proteins. Using the conversions  $10^6 \mu\text{M}/\text{M}$ ,  $1000 \text{cm}^3/\text{liter}$ , and  $6.02 \times 10^{23}$  molecules/mol, a  $1 \mu\text{M}$  concentration transforms to  $[C] = 6.02 \times 10^{14}$  molecules/ $\text{cm}^3$ . Thus, because there are  $10^{12} \mu\text{m}^3$  in  $1 \text{cm}^3$ , the  $1 \mu\text{m}^3$  cell is expected to contain  $n \simeq 602$  molecules. Application of Equation 7.9 then leads to an encounter rate of  $4 \times 10^6$  events/cell/sec. Increasing the concentration by a factor of  $x$  will increase the encounter rate by a factor of  $x^2$ .

In the preceding example,  $n$  is sufficiently large that the correction factor has essentially no effect. However, decreasing  $[C]$  and/or  $V$  begins to have a nonlinear effect at sufficiently low values. For example, if the concentration is reduced to  $0.01 \mu\text{M}$ , the expected number of molecules/cell is reduced to  $n \simeq 6$ , and the encounter rate is reduced by a factor of  $(0.01)^2(5/6)$  to  $\sim 333$  events/cell/sec. Further reducing the cell volume to  $0.1 \mu\text{m}^3$ , then  $n < 1$ , and a protein would almost always be without partners in a cell. These results demonstrate that constraints on the number of molecules contained within small cells (Figure 7.4) must ultimately limit the reaction rates that can be carried out (Klumpp et al. 2013).

Finally, it should be noted that all of the above considers only the physical encounter rate between particles, assuming an ideal homogeneous setting with no attractive or repulsive forces between particles. In doing so, it also only focuses on translational diffusion across spatial points. Should the surface of each particle contain a restricted reactive patch, this will reduce the effective encounter rate by a factor related to the effective patch size per particle, after also taking into

consideration the process of rotational diffusion (which refers to random movement of a particle on its axes, apart from movement across space). These issues will be taken in up in Chapters 13 and 18, focused on protein multimerization and nutrient uptake.

### Temperature-dependence of Biological Processes

Through its effects on rates of molecular motion, temperature influences virtually all biological processes. For most biochemical interactions, elevated temperature increases the reaction rate, at least up to the point beyond which the stability of the reactants is compromised. Chemical reaction rates depend on the frequency of successful encounters between participating molecules, and most reactions require some amount of energy to go forward. The energetic barrier to a reaction is called the activation energy ( $E_a$ ), with a higher value of  $E_a$  implying a slower response to temperature. A powerful result from statistical mechanics, the Boltzmann distribution, relates the distribution of energy states of molecules to ambient temperature (Foundations 7.3).

This distribution has the useful property of being exponential in form, with the mean energy state of molecules being the familiar  $k_B T$ . For a system in thermodynamic equilibrium, the fraction of molecules with an energy state above the activation energy is simply

$$f_e = e^{-E_a/(k_B T)}. \quad (7.10)$$

As temperature increases, more molecules have high enough energy to overcome the activation barrier, and  $f_e \rightarrow 1$  at a rate that depends on  $E_a$ . The overall reaction rate is the product of the encounter rate between reactants ( $R_E$ ) and the fraction of successful encounters,

$$R_{\text{tot}} = R_E \cdot f_e = R_E \cdot e^{-E_a/(k_B T)} \quad (7.11)$$

Taking the log of this expression demonstrates that a plot of the log of a reaction rate against the inverse of temperature ( $1/T$ ) is expected to yield a straight line

$$\ln(R_{\text{tot}}) = a - b(1/T) \quad (7.12)$$

with the slope ( $b$ ) estimating  $-E_a/k_B$ , and the intercept ( $a$ ) estimating the log of the encounter rate, which is a function of the properties of reactants and their concentrations (Foundations 7.2). Such an inverse relationship between the rate of a molecular reaction and  $1/T$  is known as Arrhenius-rate behavior, after its early advocate (Arrhenius 1889), who derived the expression in a different way than pursued in Foundations 7.3. Because  $k_B$  is a constant, Equation 7.12 provides a simple means for estimating the activation energy of a reaction. (It may be noticed from Equation 7.2.2 that temperature appears in the expression for  $R_E$ , which is inversely related to the viscosity of the fluid, but this is generally ignored under the assumption that the exponential dependence of  $f_e$  on temperature dominates the overall behavior).

Although the Arrhenius equation often provides an excellent description of the temperature-dependence of simple chemical reactions, organisms consist of mixtures of hundreds to thousands of biomolecules. Each biochemical reaction will have its

own activation energy, with the concentrations and stabilities of the interacting partners changing with environmental conditions, including temperature (e.g., Hunter and Rose 1972; Alroy and Tannenbaum 1973; Herendeen et al. 1979). Many of these reactions will operate in parallel (as, for example, independent pathways for uptake of different nutrients), whereas others will operate in series (as in consecutive steps in metabolic pathways). Thus, although there may be one rate-limiting step at any particular temperature, the nature of this step (and its associated activation energy) is likely to change among temperatures. Further complicating matters is the fact that complex biomolecules tend to become increasingly unstable at high temperatures and can have altered properties at low temperatures (Dill et al. 2011).

All of these issues motivate the question as to whether rates of higher-order biological functions scale in accordance with Equation 7.13, and if they do, whether there is any simple mechanistic interpretation of the fitted slopes and intercepts. At best, any estimate of  $E_a$  for a cell-biological process would seem to be a composite “effective” barrier to activation of the process. Nonetheless, it is often argued that processes such as metabolic and developmental rates, and the “rate of living” (inverse of life span), scale in close accordance with the Arrhenius equation, at least below temperatures at which key molecular/cellular processes begin to break down (Gillooly et al. 2001; Savage and West 2006). It has been argued that in *E. coli*, most biochemical reaction rates have similar responses to temperature under nonextreme conditions, leading to an overall adherence to Arrhenius-rate behavior (Mairet et al. 2021).

Herein lies the problem. Although the range of temperatures consistent with Arrhenius rate behavior are often referred to as being “biologically relevant,” this is usually little more than a matter of convenience, with the edges of such regions often being quite arbitrary. When taken to even moderately extreme temperatures, the responses of cellular growth rates to temperature are virtually always curvilinear, in contrast to the expectations from Equation 7.12, with the optimal temperature and the form of the response curve often varying substantially among species (Figure 7.7). Even within the range of “meaningful” temperatures, not all biological rates scale exponentially with temperature, with the response of growth rate to temperature approaching linearity in various unicellular eukaryotes (Montagnes et al. 2003). Nor do all biological features respond in a positive way to a thermal increase. For example, the cell sizes of unicellular eukaryotes often decline by  $\sim 2$  to 4% for each  $1^\circ\text{C}$  increase in temperature (Montagnes and Franklin 2001; Atkinson et al. 2003).

Some have suggested that these kinds of variations in temperature response curves can be accommodated by relatively simple modifications of the Arrhenius equation, such as by subtracting or dividing one exponential expression by another to account for contrasting responses of cell features to temperature (Mohr and Krawiec 1980; Ratkowsky et al. 1983, 2005; Corkrey et al. 2014; Arroyo et al. 2022). For example, Dill et al. (2011) show how deviations from ideal Arrhenius behavior can be accommodated by multiplying Equation 7.10 by a function that accounts for increasing protein denaturation with temperature. Although the fits of such mathematical relationships to biological features are often quite good over a substantial temperature range, and the underlying models are frequently viewed as first-principles sorts of derivations, caution is warranted in attaching too much biological meaning to them. With four or more parameters, a wide variety of non-

linear functions can yield essentially identical fits to the same data. Indeed, more than 24 alternative mathematical functions have been proposed for the relationship between reaction rates and temperature (Noll et al. 2020). Nonetheless, such general statistical fits are of great interest, as they suggest the operation of universal scaling features of biological traits, begging the question as to the underlying mechanisms that apply across the Tree of Life.

One of these alternatives is a common rule-of-thumb in biology, the so-called  $Q_{10}$  rule, which states that biological rates typically increase by a factor of 2 to 3 with a  $10^\circ\text{C}$  increase in temperature (Raven and Geider 1988; Hansen et al. 1997), again with a presumed focus on a “biologically relevant” temperature range. The idea was first raised by Arrhenius’ Ph. D. advisor, Van’t Hoff, and can be crudely related to the Arrhenius equation. For example, considering two commonly used temperatures, 12 and  $22^\circ\text{C}$  (i.e.,  $T = 285$  and  $295$ ), then from Equation 7.10 the ratio of Arrhenius rates at the high vs. low temperature is  $\sim e^{0.00012(E_a/k_B)}$ . If  $E_a/k_B \simeq 8333$ , then  $e^1 \simeq 2.72$ , which is within the range of commonly observed  $Q_{10}$  estimates. This implies that  $E_a/k_B$  must typically be on the order of 8333. However, somewhat different results will be obtained with different limits on the temperature range. For example, applying temperatures of 22 and  $32^\circ\text{C}$  yields a ratio of  $e^{0.00011(E_a/k_B)}$ . Assuming  $E_a/k_B = 8333$  still holds, this would imply  $Q_{10} = 2.52$ . Thus, the  $Q_{10}$  approach is an approximation, albeit a fairly good one, if the system behaves in accordance with the Arrhenius equation. However, there is little justification for claiming the superiority of one approach over the other, as both are phenomenological with respect to describing biological functions.

## Energy, Carbon Skeletons, and Cell Yield

Heterotrophic organisms, incapable of fixing  $\text{CO}_2$ , are reliant on the uptake and assimilation of organic compounds for the production of new cellular biomass. The key materials consist of reduced carbon compounds containing hydrogen, usually with some oxygen, nitrogen, phosphorus, and/or sulfur atoms also present. In today’s organisms, these substances are almost always ultimately derived from cellular materials or excretory products of photoautotrophs, with many undergoing secondary modification in herbivores and detritivores before again being ingested by carnivores. Food materials provide both the carbon skeletons necessary for biosynthesis of the monomeric building blocks of the cell, e.g., amino acids, nucleotides, and lipids. They are also the source of energy for subsequent transformation into the cell’s energetic currency, ATP.

The specific organic composition of food ultimately dictates the rate at which a heterotroph can invest biomass and energy into self-maintenance, growth, and reproduction. In the organism, as in the furnace, the oxidation of organic substrates releases energy, and some organic substances have higher energy contents than others. The maximum amount of extractable energy of a substance is equivalent to its heat of combustion,  $\Delta H_C$ , with the absolute limit to biological energetics being set by the product of the latter and the consumption rate (ignoring the costs of building and maintaining the metabolic machinery itself).

A deeper understanding of the biological relevance of heats of combustion can

be achieved by considering the chemical composition of a substrate and the fates of carbon-associated electrons upon combustion. Kharasch and Sher (1925) classified organic compounds on the basis of the number of electrons that experience a transition from a methane-type bond (C-H) to a carbon dioxide-type bond (C=O) upon combustion,

$$N_E = 4N_C + N_H - 2N_O \quad (7.13)$$

for a molecule containing  $N_C$  carbon,  $N_H$  hydrogen, and  $N_O$  oxygen atoms. The structure of this expression follows from the fact that each carbon atom has four outer-shell (valence) electrons of its own, sharing one additional electron with each bonded hydrogen atom and two with each bonded oxygen. The electrons shared with each hydrogen atom are free to move upon combustion, whereas the two associated with each oxygen are already in the position expected after oxidation. Complete combustion reconfigures hydrogen atoms into water, and oxygen atoms into  $\text{CO}_2$ , which from Equation 7.13 has  $N_E = 0$ . For glucose,  $\text{C}_6\text{H}_{12}\text{O}_6$ ,  $N_E = 24$ .

For carbon substrates commonly employed in laboratory growth experiments, this composite measure of the degree of electron movement upon transformation to  $\text{CO}_2$  and water is nearly perfectly correlated with known heats of combustion determined in chemistry labs (Figure 7.8), with  $\Delta H_C$  (in units of kcal/mol) being closely approximated by  $27N_E$ . Extensions to organic substrates containing nitrogen and/or sulfur can be found in Kharasch and Sher (1925) and Williams et al. (1987).

These purely physico-chemical descriptors of substrate molecules are informative with respect to growth rates of pure cultures of unicellular organisms raised in chemostats (Chapters 8 and 17). For situations in which a single substrate is the sole source of carbon and energy (and all other nutrients in excess supply), a compilation of data from studies involving alternative carbon sources indicates that the growth yield (g cell dry weight/g carbon consumed) increases with the heat of combustion per carbon atom in the substrate, with no obvious differences between bacteria and eukaryotes (Figure 7.8). However, beyond the point at which the caloric content of the substrate exceeds 10 kcal/g carbon, the cell yield levels off at  $\sim 1.4$  g dry weight/g carbon consumed. Because the values in Figure 7.8 are derived from cultures growing at maximum rates, only a small fraction of the cell's energy budget is allocated to maintenance (Chapter 8), so the estimates provided are close to the maximum biomass yields associated with the substrate.

This overall pattern, first suggested by Linton and Stephenson (1978), with many fewer data than contained in Figure 7.8, implies that for low-energy substrates (heats of combustion  $< 10$  kcal/g carbon), heterotrophic cells are intrinsically energy limited, i.e., they are incapable of experiencing the maximum possible yield of  $\sim 1.4$  g dry weight/g carbon consumed. Above a substrate heat of combustion of 10 kcal/g carbon, the constant cell growth yield per unit carbon implies a progressive decline in the efficiency of energy extraction with increasing energetic content of the substrate. Thus, an energy content of  $\sim 10$  kcal/g carbon appears to separate a lower domain in which the substrate provides insufficient energy to assimilate the available carbon from an upper domain where energy is in excess of the requirements for carbon assimilation. Foundations 2.3 takes these kinds of observations a step further to estimate the amount of energy in units of ATP hydrolyses needed to produce a unit of biomass.

Finally, recalling from above that the average fractional carbon mass per cell

dry weight is  $\sim 0.5$ , the cell yields in Figure 7.8 can be rescaled to units of g cellular carbon/g substrate carbon, providing a measure of assimilation efficiency for carbon. With dry-weight cell yields / g substrate carbon being in the range of 0.8 to 1.6 for nearly all common substrates (Figure 7.8), this implies typical carbon assimilation efficiencies in the range of 0.4 to 0.8. After nearly four billion years, this is the best that natural selection has been able to achieve. 100% conversion of substrate carbon into biomass is unobtainable, as energy must be extracted from some of the substrate to carry out cellular functions, and some carbon is lost as  $\text{CO}_2$ .

### Summary

- Between 65 and 80% of the wet weight of cells consists of  $\text{H}_2\text{O}$ , eukaryotic cells being more watery than those of bacteria. Across the Tree of Life  $\sim 50\%$  of cell dry weight is comprised of carbon atoms, and one- to two-thirds of the dry weights of most cells consist of protein.
- The unique physical properties of water govern almost every aspect of biology, as they dictate the folding stability of proteins, the ability of lipid molecules to aggregate into membranes, the diffusion rates of molecules, and the challenges to motility.
- Of the 20 chemical elements essential to life, many have intracellular concentrations enriched by factors of  $10^3$  to  $10^6$  relative to environmental levels. Such factors are equivalent to the volume of the environment relative to that of the cell that needs to be fully harvested to produce an offspring cell.
- Despite its centrality to life, the fractional contribution of genomic DNA to cellular biomass scales negatively with cell volume, declining from  $\sim 10\%$  in the smallest bacterial cells to  $< 0.001\%$  in the largest eukaryotic cells.
- The total number of protein molecules per cell and the average number per gene increase sublinearly with cell volume, consistent with larger cells being less dense with biomaterials. Messenger RNAs are typically 100- to 10,000-fold less abundant per cell than their cognate proteins, with the mean number per gene often being in the range of 1 to 10. With the distributions of both mRNAs and protein molecules per gene per cell being approximately log-normal in form, there can be significant stochastic variation in gene expression among genetically uniform cells. Moreover, there must be a lower bound to cell size below which adequate numbers of molecules cannot be harbored to reliably sustain key biochemical reaction rates.
- Many molecules travel through cells by passive diffusion processes. Fueled by

background thermal noise, such transport imposes no costs to the host cell. For small bacterial-sized cells, an average protein can diffuse across a cell diameter in several milliseconds, whereas such a sojourn can require up to half a minute in some of the larger eukaryotic cells. Thus, diffusion limits to intracellular transactions can ultimately constrain the rates of biological processes in eukaryotic cells.

- Through its influence on the motion of all molecules, temperature plays a governing role in all reaction rates. A number of mathematical expressions have been proposed as summary descriptors for the response of biological processes to temperature, although the mechanistic interpretation of the fitted parameters is open to debate.
- Life ultimately depends on the acquisition of energy. For aerobic heterotrophs (most organisms other than photosynthesizers), food comes in the form of reduced carbon compounds, which provide both carbon skeletons for constructing biomass and energy for carrying out cellular functions. The heats of combustion of substrates provide reliable measures of the energy that organisms can extract from such compounds. More reduced carbon compounds provide more energy, but there is an intermediate level of substrate reduction (approximately equivalent to that in glucose) above which carbon starts to be limiting.
- The upper limit of evolved assimilation efficiency of carbon compounds (fraction of ingested carbon atoms incorporated into biomass) is  $\sim 0.8$ , and the construction of 10 grams dry weight of cellular biomass requires the energy released from the hydrolysis of  $\sim 1$  mol of ATP.

**Foundations 7.1. Intracellular diffusion.** In a homogeneous medium, small particles are subject to random walks as a consequence of background thermal perturbations. This leads to diffusive particle movement from a starting location in a symmetric fashion. To minimize the mathematical details, the focus here will be on a one-dimensional diffusion process, with a summary of the general results for two and three dimensions following the initial details.

Consider a particle moving randomly to the right and left with equal probabilities of 0.5 and fixed jump lengths, independent of prior motion at each time unit. Let  $t$  be the total number of jumps, with  $t_+$  being the number to the right and  $t_-$  the number to the left, so that

$$t = t_+ + t_-.$$

The net displacement relative to a starting point at position 0 is then

$$x = t_+ - t_-.$$

Given  $t$  jostling episodes, the probability of  $t_+$  draws in the positive direction is given by the binomial distribution

$$P(x) = \frac{t!}{t_+!t_-!} \left(\frac{1}{2}\right)^{t_+} \left(\frac{1}{2}\right)^{t-t_+} = \frac{t!}{t_+!t_-!} \left(\frac{1}{2}\right)^t, \quad (7.1.1a)$$

where  $y! = y \cdot (y-1) \cdot (y-2) \cdots 1$  is the factorial function.

For large  $t$ , this discrete-state formula can be simplified to a continuous distribution by first noting that  $t_+ = (t+x)/2$  and  $t_- = (t-x)/2$ , substituting into Equation 7.1.1a, and then logarithmically transforming to obtain

$$\ln[P(x)] = \ln(t!) - \ln\left\{\left[\frac{t}{2}\left(1+\frac{x}{t}\right)\right]!\right\} - \ln\left\{\left[\frac{t}{2}\left(1-\frac{x}{t}\right)\right]!\right\} - t \ln 2. \quad (7.1.1b)$$

Factorial functions can be unwieldy, but large  $t$  allows the use of Stirling's approximation for the logarithm of large factorials,

$$\ln(y!) \simeq \frac{\ln(2\pi y)}{2} + y \ln(y) - y, \quad (7.1.2)$$

application of which simplifies Equation 7.1.1b to

$$\ln[P(x)] = \ln[(2/\pi t)^{0.5}] - \left(\frac{t+x+1}{2}\right) \ln\left(1+\frac{x}{t}\right) - \left(\frac{t-x+1}{2}\right) \ln\left(1-\frac{x}{t}\right).$$

Further simplification is accomplished by noting that for  $y < 0.5$ ,

$$\ln(1+y) \simeq y - (y^2/2), \quad (7.1.3a)$$

$$\ln(1-y) \simeq -y - (y^2/2). \quad (7.1.3b)$$

Applying these approximations to the preceding expression, followed by exponentiation to return to the original scale eventually leads to

$$P(x) \simeq \left(\frac{2}{\pi t}\right)^{1/2} \exp\left(-\frac{x^2}{2t}\right). \quad (7.1.4a)$$

We now modify Equation 7.1.4a to a more familiar and general form. First, we note that the variance in the number of jumps to the right follows from the properties of the binomial distribution. When the probability of each type of event is 0.5, the binomial variance associated with each event is  $0.5 \cdot 0.5$ , and summing over  $t$  independent events leads to variance  $\sigma^2 = t/4$ . Second, the disparity between numbers of right and left jumps,  $x$ , can be rewritten as  $(t_+ - t_-) = (2t_+ - t)$ , and because the expected number of jumps to the right (the mean) can be written as  $\mu = t/2$ , this further reduces to  $x = 2(t_+ - \mu)$ . Substituting the latter expression and  $t = 4\sigma^2$  into Equation (7.1.4a), we obtain

$$P(t_+) = \left( \frac{1}{2\pi\sigma^2} \right)^{1/2} \exp \left( -\frac{(t_+ - \mu)^2}{2\sigma^2} \right). \quad (7.1.4b)$$

This is the widely used normal (or Gaussian) distribution of a variable (in this case  $t_+$ ) with mean  $\mu$  and variance  $\sigma^2$ .

In the current case, diffusion results in movement from the initial point, but with no net bias, so we can rescale to a mean of zero. The variance  $\sigma^2$  can also be written as the mean-squared deviation  $2Dt$ , where  $D$  is the diffusion coefficient, with units of length<sup>2</sup>/time (see main text). The one-dimensional diffusion distance  $d$  then has probability distribution

$$P(d) = \left( \frac{1}{4\pi Dt} \right)^{1/2} \exp \left( -\frac{d^2}{4Dt} \right). \quad (7.1.4c)$$

Note that in the one-dimensional case, the diffusion variance is proportional to  $2D$ , and increases linearly with time. In two dimensions,  $2Dt$  becomes  $4Dt$ , and with three dimensions, it becomes  $6Dt$ . The standard deviation is the root mean-squared distance that a particle is expected to have traveled (with equal probability in all directions) after  $t$  time units. Thus, regardless of the dimensionality, the expected distance traveled increases with the square root of time. Berg (1993) provides a useful compendium of results and biological applications of diffusion theory.

**Foundations 7.2. Rates of encounter by molecular diffusion.** A purely physical limit to the encounter rate between two molecules can be derived from diffusion theory developed by Smoluchowski (1915), who independently of Einstein outlined a number of the general principles of Brownian motion. We start by considering the random diffusion of two spherical molecules, with respective radii  $r_a$  and  $r_b$ , moving randomly through an otherwise homogeneous environment. A collision between these two molecules will occur whenever their centers come within a distance  $r_c = r_a + r_b$  from each other. To simplify the overall analysis, one may then consider an imaginary sphere around the center of either particle, with radius  $r_c$ , whose overall surface area  $4\pi r_c^2$  represents the entire boundary across which a flux of one particle or the other constitutes a collision (Figure 7.9).

To proceed further, we require the total rate of particle movement, which is determined by the sum of the diffusion coefficients associated with each particle type. From Equations 7.4 and 7.5,

$$D = D_a + D_b = \frac{k_B T (r_a + r_b)}{6\pi\eta(r_a r_b)}, \quad (7.2.1)$$

with units of cm<sup>2</sup>/sec, where  $k_B$  is Boltzmann's constant,  $T$  is the temperature (in Kelvins), and  $\eta$  is the viscosity of the medium (see main text for the assumed values of these parameters).

To complete the derivation of the encounter rate, we require an expression for the rate of diffusion across a planar surface. This is given by Fick's first law, which states that the flux rate of a diffusing substrate across a point is equal to the product of the concentration gradient at that point and the diffusion coefficient. Here, the concentration gradient can be approximated by treating the concentration inside the sphere of radius  $r_c$  as zero and denoting the bulk concentration (outside the sphere) as  $[C]$ , implying a concentration gradient of  $([C] - 0)/r_c$  and flux rate  $[C]D/r_c$ . After multiplying by the total surface area ( $4\pi r_c^2$ ) and dividing by the concentration, this scales up to a flux rate per unit concentration of  $(4\pi r_c^2)(D/r_c) = 4\pi r_c D$ . Substituting Equation 7.2.1 for  $D$ , we then obtain an expression for the encounter-rate coefficient,

$$k_e = 4\pi r_c D = \left( \frac{2k_B T}{3\eta} \right) \left( \frac{(r_a + r_b)^2}{r_a r_b} \right). \quad (7.2.2)$$

After substituting for the average temperature of life (see main text) and the viscosity of water at 20° C, this reduces further to

$$k_e = (2.8 \times 10^{-12}) \left( \frac{(r_a + r_b)^2}{r_a r_b} \right), \quad (7.2.3)$$

with units  $\text{cm}^3 \cdot \text{sec}^{-1}$ . (Note that if  $[C]$  is expressed in molar-concentration units, then  $k_e$  needs to be divided by 1000 to convert to liters and multiplied by Avogadro's number to convert to molecules, making the prefix  $1.7 \times 10^9$ ). The product of this encounter-rate coefficient and the concentrations of both particle types (each in units of molecules/ $\text{cm}^3$ ) yields the expected number of collisions between the two particle types in a 1  $\text{cm}^3$  volume per second,

$$R_e = k_e [C_a][C_b]. \quad (7.2.4)$$

For two spherical particles identical in size ( $r_a = r_b$ ), as in the case of two monomeric subunits forming a homodimeric protein, Equation 7.2.3 reduces to

$$k_e \simeq 11 \times 10^{-12}. \quad (7.2.5a)$$

In this case, the rate coefficient is independent of the particle size because any increase in target size is perfectly balanced by a reduction in the rate of diffusion. On the other hand, if one particle type is much larger than the other,  $r_b \ll r_a$ ,

$$k_e \simeq (2.8 \times 10^{-12}) \left( \frac{r_a}{r_b} \right), \quad (7.2.5b)$$

showing that the encounter rate depends only on the ratio of particle sizes, not on their absolute sizes.

The encounter rates denoted by these expressions are sometimes referred to as the Smoluchowski limits. They denote the encounter rate in the ideal situation in which there are no attractive or repulsive forces between colliding particles, and otherwise no barriers for diffusion through a homogeneous medium, any of which can become important in various biological contexts.

**Foundations 7.3. The Boltzmann probability distribution for alternative molecular states.** Numerous situations are encountered in cell biology where it is

necessary to know the distribution of alternative states of the individual members of a population of molecules, as these often determine the average rates and stochasticities of cellular processes. Theoretical results in this area are generally derived from the field of statistical mechanics, which takes a microscopic view of particle states within a closed system assumed to be in thermodynamic equilibrium. There are numerous ways to achieve the final result (e.g., Schroeder 2000; Phillips et al. 2012). The route taken here uses the properties of combinatorics, along with a few mathematical approximations.

The starting assumption is a system containing  $n$  molecules, which together harbor a fixed amount of energy,  $\Sigma$ . We assume discrete energy states, taking on values of  $0, \epsilon, 2\epsilon, \dots, k\epsilon$ , so  $N = \Sigma/\epsilon$  represents the total number of discrete energy packets available to the system. Individual particles are free to change energy states, but the overall probability distribution of alternative states remains constant under the assumption of equilibrium. It is this equilibrium probability distribution that we wish to determine, i.e., the probability that a random particle is in energy states  $i = 0, 1, \dots, k$ . To accomplish this, we must account for the full distribution of the alternative states that a set of  $n$  molecules can take on, conditional on their sum equaling  $N$ . Given the large number of particles typically involved, this can be a dauntingly complex problem, but a few mathematical tricks simplify the overall derivation.

We first note that the total number of ways that  $N$  packets of energy can be partitioned among  $n$  molecules is given by

$$T(N, n) = \frac{(N + n - 1)!}{N!(n - 1)!}, \quad (7.3.1a)$$

where  $!$  again denotes the factorial function. To obtain this general result, note that there are  $n$  bins within which  $N$  energy packets must be partitioned. The numerator is the total number of ordered ways that  $N$  distinct packets can be randomly assigned to  $n$  bins. But because the energy packets are all identical in content, the ordering in which they are assigned is irrelevant, and the two terms in the denominator discount the numerator to account for the redundancy associated with ordering of packets and bins.

Now consider the situation where one specific molecule has energy  $i\epsilon$ , so there are a remaining  $(N - i)$  packets to partition among  $(n - 1)$  molecules. Modification of the previous expression then leads to

$$T(N - i, n - 1) = \frac{(N + n - 2 - i)!}{(N - i)!(n - 2)!}. \quad (7.3.1b)$$

Thus, the probability of a particle having energy content  $i\epsilon$  is

$$p(i) = \frac{T(N - i, n - 1)}{T(N, n)} = (n - 1) \cdot \frac{N!(N + n - 2 - i)!}{(N - i)!(N + n - 1)!}. \quad (7.3.2)$$

Further simplification is possible if it is assumed that the energy in the system is substantial enough that  $N \gg n$ , which makes reasonable the approximations  $N!/(N - x)! \simeq N^x$ , and  $n/(N + n) \simeq n/N$ . Noting as well that the number of molecules is large, so that  $n - 1 \simeq n$ ,

$$p(i) \simeq nN^i \cdot (N + n)^{-(i+1)} \simeq \frac{n}{N} \cdot [1 + (n/N)]^{-i}, \quad (7.3.3a)$$

which further reduces to

$$p(i) \simeq \frac{n}{N} \cdot e^{-in/N}, \quad (7.3.3b)$$

using  $e^x \simeq (1+x)$  for  $x \ll 1$ . Thus, what started as a complex problem reduces to a relatively simple expression (a negative exponential distribution) under the assumption of large numbers.

Letting  $\bar{E} = N\epsilon/n$  denote the average energy per particle, the preceding expression implies that the probability of a particle having energy state  $E_i = \epsilon_i$  is

$$p(E_i) = C \cdot e^{-E_i/\bar{E}}, \quad (7.3.4)$$

where  $C$  is a normalization constant that ensures that the total probability distribution sums to 1.0, satisfied in this case by  $C = 1/\bar{E}$ . Letting  $\bar{E} = k_B T$  be the average energy per particle yields the Boltzmann distribution,

$$p(E_i) = C \cdot e^{-E_i/(k_B T)}. \quad (7.3.5)$$

Note that the cumulative function for this exponential distribution, which defines the probability of being in a state below  $E_i$  is  $1 - e^{-E_i/\bar{E}}$ .

**Foundations 7.4. The yield of cellular biomass per ATP usage.** Observations in the final section of the main text allow for a crude estimate of the amount of energy required to build new cellular material (in terms of  $\text{ATP} \rightarrow \text{ADP}$  hydrolyses), an issue that will be addressed in more detail in Chapter 17. Here we will assume a relatively high-energy carbon substrate with a heat of combustion of 9.3 kcal/g C (carbon), the approximate value for most six-carbon sugars (including glucose). From Figure 7.8, such a substrate leads to an expected 1.3 g DW (dry weight) produced/g C consumed, or inversely  $(1/1.3) = 0.77$  g C consumed/g DW produced. Multiplying by 9.3 kcal/g C leads to an estimated cellular energy-intake requirement of 7.2 kcal/g DW produced.

How much of this required consumption is diverted to energy production for cell functions? Surveys of multiple bacterial and eukaryotic species suggest average caloric contents of 5.41 (0.05) and 5.13 (0.04) kcal/g cell DW (Supplemental Table 8.1), respectively. Assuming an average value of 5.3, this implies that of the 7.2 kcal consumed/g DW produced,  $7.2 - 5.3 = 1.9$  kcal (26%) must be used in cellular processes required to produce new cellular material (with the rest of the substrate providing carbon skeletons used in the construction of the monomeric building blocks of the cell). Thus,  $\sim 0.77 \times 0.26 = 0.20$  g C of substrate must be converted to energy in order to produce 1 g of cell DW.

What does this energetic investment mean in units of ATP, the cellular currency of bioenergetics? One mole of glucose contains 72 g C, and assuming complete aerobic metabolism, observations from biochemistry tell us that each mole of metabolized glucose generates  $\sim 32$  mole of ATP (Chapter 17). This suggests that, in units of ATP, the energetic requirement for the production of 1 g DW of cells is  $\sim 0.20$  g C consumption  $\times$  (1 mole glucose/72 g C)  $\times$  (32 mol ATP/mole glucose) = 0.089 mole ATP. It then follows that the yield of cells is  $\sim 1/0.089 = 11.2$  g DW/mole ATP. This rough estimate is quite close to the average value of 10.5 for more direct estimates found in a wide variety of organisms raised on a diversity of carbon substrates (Payne 1970).

**Literature Cited**

- Alroy, Y., and S. R. Tannenbaum. 1973. The influence of environmental conditions on the macromolecular composition of *Candida utilis*. *Biotechnol. Bioeng.* 15: 239-256.
- Arrhenius, S. 1889. Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Saeuren. *Z. Physik. Chem.* 4: 226-248.
- Arroyo, J. I., B. Díez, C. P. Kempes, G. B. West, and P. A. Marquet. 2022. A general theory for temperature dependence in biology. *Proc. Natl. Acad. Sci. USA* 119: e2119872119.
- Atkinson, D., B. J. Ciotti, and D. J. Montagnes. 2003. Protists decrease in size linearly with temperature: ca. 2.5% °C<sup>-1</sup>. *Proc. Biol. Sci.* 270: 2605-2611.
- Bakshi, S., A. Siryaporn, M. Goulian, and J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol. Microbiol.* 85: 21-38.
- Ball, P. 2008. Water as an active constituent in cell biology. *Chem. Rev.* 108: 74-108.
- Ball, P. 2017. Water is an active matrix of life for cell and molecular biology. *Proc. Natl. Acad. Sci. USA* 114: 13327-13335.
- Bellotto, N., J. Agudo-Canalejo, R. Colin, R. Golestanian, G. Malengo, and V. Sourjik. 2022. Dependence of diffusion in *Escherichia coli* cytoplasm on protein size, environmental conditions, and cell growth. *eLife* 11: e82654.
- Berg, H. C. 1993. *Random Walks in Biology*. Princeton Univ. Press, Princeton, NJ.
- Corkrey, R., T. A. McMeekin, J. P. Bowman, D. A. Ratkowsky, J. Olley, and T. Ross. 2014. Protein thermodynamics can be predicted directly from biological growth rates. *PLoS One* 9: e96100.
- de Queiroz, J. H., J.-L. Uribelarra, and A. Pareilleux. 1993. Estimation of the energetic biomass yield and efficiency of oxidative phosphorylation in cell-recycle cultures of *Schizosaccharomyces pombe*. *Applied Microbiol. Biotech.* 39: 609-614.
- Dias-Lopes, G., J. R. Wiśniewski, N. P. de Souza, V. E. Vidal, G. Padrón, C. Britto, P. Cuervo, and J. B. De Jesus. 2018. In-depth quantitative proteomic analysis of trophozoites and pseudocysts of *Trichomonas vaginalis*. *J. Proteome Res.* 17: 3704-3718.
- Dill, K. A., and S. Bromberg. 2011. *Molecular Driving Forces*, 2nd Ed. Garland Science, New York, NY.
- Dill, K. A., K. Ghosh, and J. D. Schmit. 2011. Physical limits of cells and proteomes. *Proc. Natl. Acad. Sci. USA* 108: 17876-17882.
- Dix, J. A., and A. S. Verkman. 2008. Crowding effects on diffusion in solutions and cells. *Annu. Rev. Biophys.* 37: 247-263.
- Ellis, R. J. 2001. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* 26: 597-604.
- Elowitz, M. B., M. G. Surette, P. E. Wolf, J. B. Stock, and S. Leibler. 1999. Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* 181: 197-203.
- Fagerbakke, K. M., M. Heldal, and S. Norland. 1996. Content of carbon, nitrogen, oxygen, sulfur and phosphorus in native aquatic and cultured bacteria. *Aquatic Microb. Ecol.* 10: 15-27.

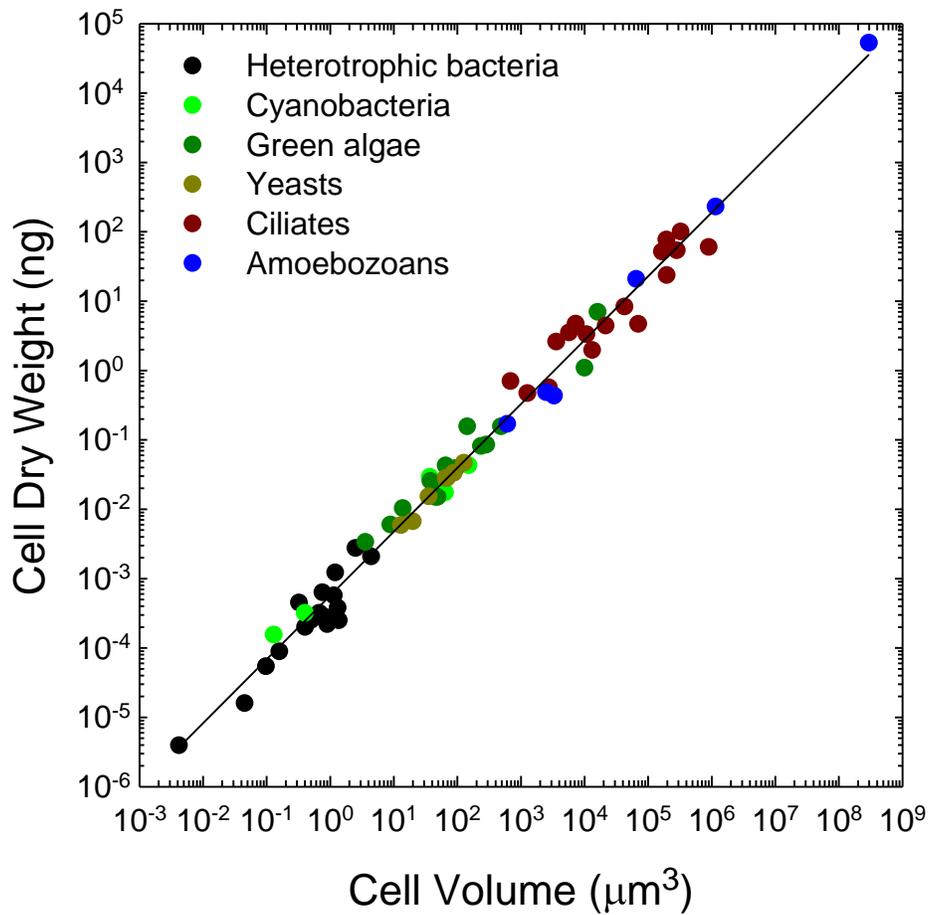
- Fagerbakke, K. M., S. Norland, and M. Heldal. 1999. The inorganic ion content of native aquatic bacteria. *Can. J. Microbiology* 45: 304-311.
- Finlay, B. J., and G. Uhlig. 1981. Calorific and carbon values of marine and freshwater Protozoa. *Helgoländer Meeresuntersuchungen* 34: 401-412.
- Gifford, S. M., J. W. Becker, O. A. Sosa, D. J. Repeta, and E. F. DeLong. 2016. Quantitative transcriptomics reveals the growth- and nutrient-dependent response of a streamlined marine methylophore to methanol and naturally occurring dissolved organic matter. *mBio* 7: e01279-16.
- Gillooly, J. F., J. H. Brown, G. B. West, V. M. Savage, and E. L. Charnov. 2001. Effects of size and temperature on metabolic rate. *Science* 293: 2248-2251.
- Guillon, L., S. Altenburger, P. L. Graumann, and I. J. Schalk. 2013. Deciphering protein dynamics of the siderophore pyoverdine pathway in *Pseudomonas aeruginosa*. *PLoS One* 8: e79111.
- Guo, M., A. J. Ehrlicher, M. H. Jensen, M. Renz, J. R. Moore, R. D. Goldman, J. Lippincott-Schwartz, F. C. Mackintosh, and D. A. Weitz. 2014. Probing the stochastic, motor-driven properties of the cytoplasm using force spectrum microscopy. *Cell* 158: 822-832.
- Hansen, P. J., P. K. Bjørnsen, and B. W. Hansen. 1997. Zooplankton grazing and growth: scaling within the 22,000 $\mu\text{m}$  body size range. *Limnol. Oceanogr.* 42: 687-704.
- He, L., and B. Niemeyer. 2003. A novel correlation for protein diffusion coefficients based on molecular weight and radius of gyration. *Biotechnol. Prog.* 19: 544-548.
- Heldal, M., D. J. Scanlan, S. Norland, F. Thingstad, and N. H. Mann. 2003. Elemental composition of single cells of various strains of marine *Prochlorococcus* and *Synechococcus* using X-ray microanalysis. *Limnol. Oceanogr.* 48: 1732-1743.
- Herendeen, S. L., R. A. VanBogelen, and F. C. Neidhardt. 1979. Levels of major proteins of *Escherichia coli* during growth at different temperatures. *J. Bacteriol.* 139: 185-194.
- Ho, K. P., and W. J. Payne. 1979. Assimilation efficiency and energy contents of prototrophic bacteria. *Biotech. Bioeng.* 21: 787-802.
- Ho, T.Y., A. Quigg, Z. V. Finkel, A. J. Milligan, K. Wyman, P. G. Falkowski, and F. M. M. Morel. 2003. The elemental composition of some marine phytoplankton. *J. Phycol.* 39: 1145-1159.
- Hong, L., and J. Lei. 2009. Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity. *J. Polymer Sci. Part B: Polymer Physics* 47: 207-214.
- Huggett, M. J., D. H. Hayakawa, and M. S. Rappé. 2012. Genome sequence of strain HIMB624, a cultured representative from the OM43 clade of marine Betaproteobacteria. *Standards Genomic Sci.* 6: 11-20.
- Hunter, K., and A. H. Rose. 1972. Influence of growth temperature on the composition and physiology of micro-organisms. *J. Appl. Chem. Biotechnol.* 22: 527-540.
- Joyner, R. P., J. H. Tang, J. Helenius, E. Dultz, C. Brune, L. J. Holt, S. Huet, D. J. Müller, and K. Weis. 2016. A glucose-starvation response regulates the diffusion of macromolecules. *Elife* 5: e09376.
- Kharasch, M. S., and B. Sher. 1925. The electronic conception of valence and heats of combustion of organic compounds. *J. Phys. Chem.* 29: 625-658.
- Klumpp, S., M. Scott, S. Pedersen, and T. Hwa. 2013. Molecular crowding limits translation and

- cell growth. Proc. Natl. Acad. Sci. USA 110: 16754-16759.
- Konopka, M. C., I. A. Shkel, S. Cayley, M. T. Record, and J. C. Weisshaar. 2006. Crowding and confinement effects on protein diffusion *in vivo*. J. Bacteriol. 188: 6115-6123.
- Konopka, M. C., K. A. Sochacki, B. P. Bratton, I. A. Shkel, M. T. Record, and J. C. Weisshaar. 2009. Cytoplasmic protein mobility in osmotically stressed *Escherichia coli*. J. Bacteriol. 191: 231-237.
- Kumar, M., M. S. Mommer, and V. Sourjik. 2010. Mobility of cytoplasmic, membrane, and DNA-binding proteins in *Escherichia coli*. Biophys. J. 98: 552-559.
- Lange, H. C., and J. J. Heijnen. 2001. Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*. Biotechnol. Bioeng. 75: 334-344.
- Linton, J. D., and R. J. Stephenson. 1978. A preliminary study on growth yields in relation to the carbon and energy content of various organic growth substrates. FEMS Microbiol. Letters 3: 95-98.
- Luby-Phelps, K. 2000. Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area. Int. Rev. Cytol. 192: 189-221.
- Lynch, M., and G. K. Marinov. 2015. The bioenergetic costs of a gene. Proc. Natl. Acad. Sci. USA 112: 15690-15695.
- Lynch, M., and B. Trickovic. 2020. A theoretical framework for evolutionary cell biology. J. Mol. Biol. 432: 1861-1879.
- Mairet, F., J. L. Gouzé, and H. de Jong. 2021. Optimal proteome allocation and the temperature dependence of microbial growth laws. NPJ Syst. Biol. Appl. 7: 14.
- Menden-Deuer, S., and E. J. Lessard. 2000. Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. Limnol. Oceanogr. 45: 569-579.
- Mika, J. T., P. E. Schavemaker, V. Krasnikov, and B. Poolman. 2014. Impact of osmotic stress on protein diffusion in *Lactococcus lactis*. Mol. Microbiol. 94: 857-870.
- Milo, R., and R. Phillips. 2016. Cell Biology by the Numbers. Garland Science, Taylor & Francis Group, New York, NY.
- Mohr, P. W., and S. Krawiec. 1980. Temperature characteristics and Arrhenius plots for nominal psychrophiles, mesophiles and thermophiles. J. Gen. Microbiol. 121: 311-317.
- Montagnes, D. J. S., and D. J. Franklin. 2001. Effect of temperature on diatom volume, growth rate, and carbon and nitrogen content: reconsidering some paradigms. Limnol. Oceanogr. 46: 2008-2018.
- Montagnes, D. J. S., S. A. Kimmance, and D. Atkinson. 2003. Using  $Q_{10}$ : can growth rates increase linearly with temperature? Aquatic Microb. Ecol. 32: 307-313.
- Montero Llopis, P., O. Sliusarenko, J. Heinritz, and C. Jacobs-Wagner. 2012. *In vivo* biochemistry in bacterial cells using FRAP: insight into the translation cycle. Biophys. J. 103: 1848-1859.
- Mu, X., S. Choi, L. Lang, D. Mowray, N. V. Dokholyan, J. Danielsson, and M. Oliveberg. 2017. Physicochemical code for quinary protein interactions in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 114: E4556-E4563.
- Nenninger, A., G. Mastroianni, and C. W. Mullineaux. 2010. Size dependence of protein diffusion

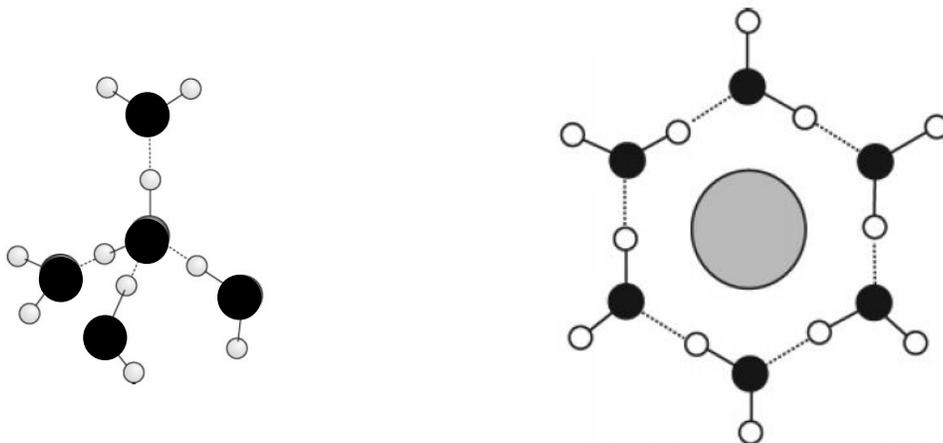
- in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* 192: 4535-4540.
- Noll, P., L. Lilge, R. Hausmann, and M. Henkel. 2020. Modeling and exploiting microbial temperature response. *Processes* 8: 121.
- Nozaki, Y. 1997. A fresh look at element distribution in the North Pacific. *EOS* 78: 221-223.
- Payne, W. J. 1970. Energy yields and growth of heterotrophs. *Annu. Rev. Microbiol.* 24: 17-52.
- Phillips, R., J. Kondev, J. Theriot, and H. Garcia. 2012. *Physical Biology of the Cell*, 2nd Ed. Garland Science, New York, NY.
- Pinho, N, J. R. Wiśniewski, G. Dias-Lopes, L. Saboia-Vahia, A. C. S. Bombaa, C. Mesquita-Rodrigues, R. Menna-Barreto, E. Cupolillo, J. B. de Jesus, G. Padrón, et al. 2020. In-depth quantitative proteomics uncovers species-specific metabolic programs in *Leishmania* (Viannia) species. *PLoS Negl. Trop. Dis.* 14: e0008509.
- Potma, E. O., W. P. de Boeij, L. Bosgraaf, J. Roelofs, P. J. van Haastert, and D. A. Wiersma. 2001. Reduced protein diffusion rate by cytoskeleton in vegetative and polarized *Dictyostelium* cells. *Biophys. J.* 81: 2010-2019.
- Ratkowsky, D. A., R. K. Lowry, T. A. McMeekin, A. N. Stokes, and R. E. Chandler. 1983. Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *J. Bacteriol.* 154: 1222-1226.
- Ratkowsky, D. A., J. Olley, and T. Ross. 2005. Unifying temperature effects on the growth rate of bacteria and the stability of globular proteins. *J. Theor. Biol.* 233: 351-362.
- Raven, J. A., and R. J. Geider. 1988. Temperature and algal growth. *New Phytol.* 110: 441-461.
- Redfield, A. C. 1934. The proportions of organic derivatives in sea water and their relation to the composition of plankton. *James Johnstone Memorial Volume* 176-192.
- Roels, J. A. 1980. Application of macroscopic principles to microbial metabolism. *Biotech. Bioeng.* 22: 2457-2514.
- Savage, V. M., and G. B. West. 2006. Biological scaling and physiological time: biomedical applications, pp. 141-163. In T. S. Deisboeck and J. Y. Kresh (eds.) *Complex Systems Science in Biomedicine*. Springer Inc., New York, NY.
- Shavemaker, P. E., A. J. Boersma, and B. Poolman. 2018. How important is protein diffusion in prokaryotes? *Front. Mol. Biosci.* 5: 93.
- Schroeder, D. V. 2000. *An Introduction to Thermal Physics*. Addison Wesley Publ. Co., San Francisco, CA.
- Smoluchowski, M. 1915. Über Brownsche Molekularbewegung unter Einwirkung äußerer Kräfte und den Zusammenhang mit der verallgemeinerten Diffusionsgleichung. *Ann. Phys.* 353: 1103-1112.
- Soh, S., M. Banaszak, K. Kandere-Grzybowska, and B. A. Grzybowski. 2013. Why cells are microscopic: a transport-time perspective. *J. Phys. Chem. Lett.* 4: 861-865.
- Swaminathan, R., C. P. Hoang, and A. S. Verkman. 1997. Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophys. J.* 72: 1900-1907.
- Tyn, M. T., and T. W. Gusek. 1990. Prediction of diffusion coefficients of proteins. *Biotechnol. Bioeng.* 35: 327-338.

- von Stockar, U., and I. W. Marison. 2005. The use of calorimetry in biotechnology bioprocesses and engineering. *Adv. Biochem. Eng. Biotech. Book Series* 40: 93-136.
- Wang, M., C. G. Kurland, and G. Caetano-Anollés. 2011. Reductive evolution of proteomes and protein structures. *Proc. Natl. Acad. Sci. USA* 108: 11954-11958.
- Williams, K., F. Percival, J. Merino, and H. A. Mooney. 1987. Estimation of tissue construction cost from heat of combustion and organic nitrogen content. *Plant Cell Environ.* 10: 725-734.
- Zimmerman, S. B., and S. O. Trach. 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.* 222: 599-620.

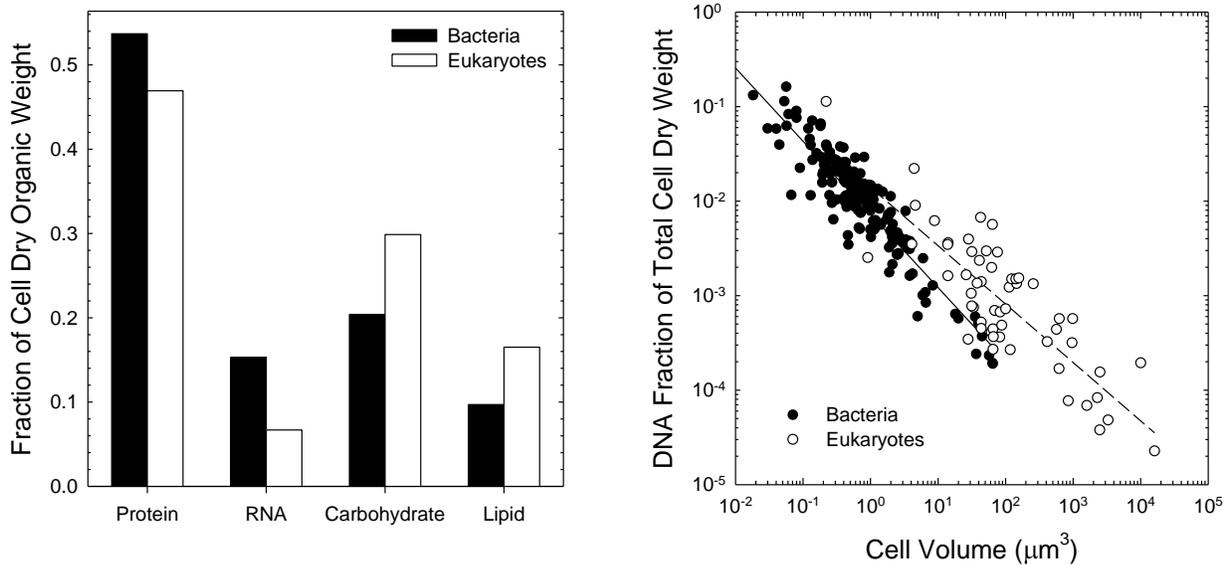
**Figure 7.1.** Relationship between dry weights and volumes of individual cells. The regression line is applied to all groups simultaneously;  $\log_{10}(\text{DW}) = -3.244(0.040) + 0.920(0.013)\log_{10}(\text{CV})$ ; standard errors of the parameters are in parentheses;  $r^2 = 0.99$ ,  $n = 68$ . Data taken from various sources in the literature are recorded in Supplemental Table 7.1.



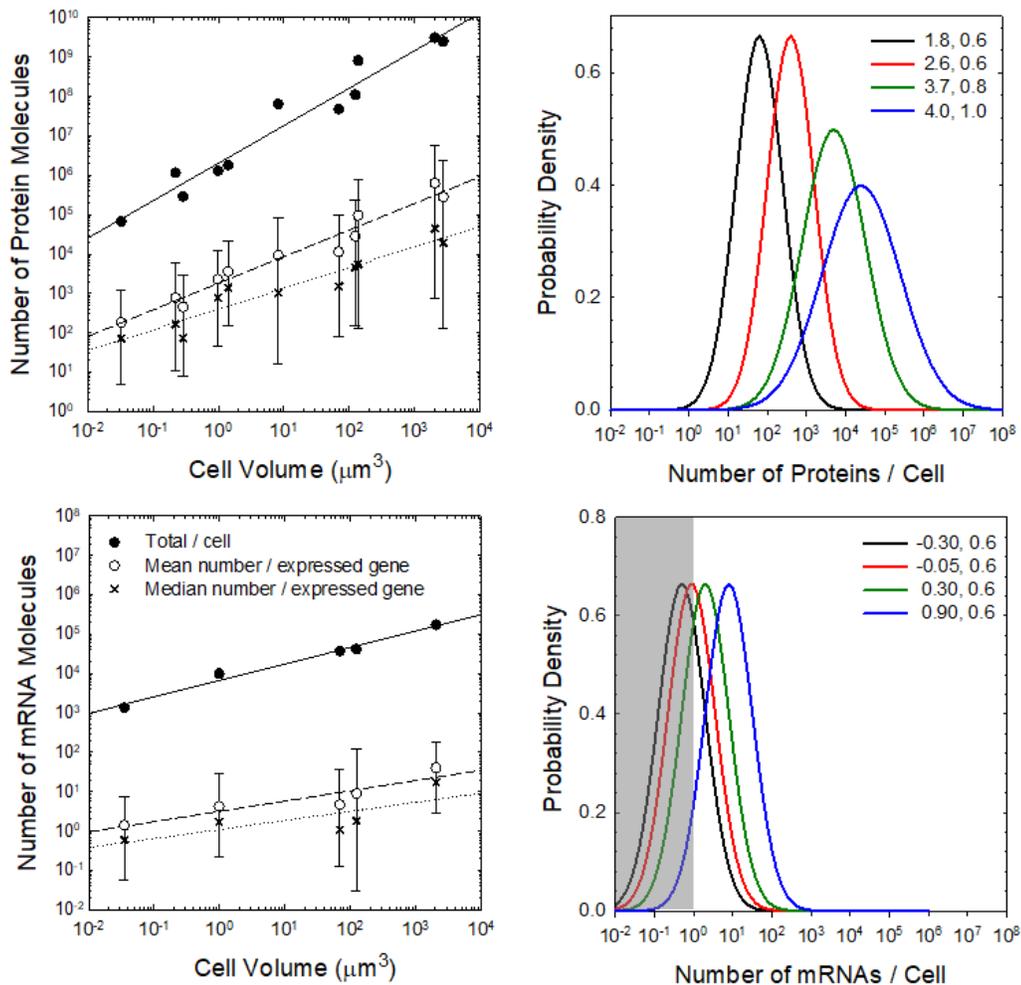
**Figure 7.2.** The organizational features of water molecules. **Left)** Free water,  $\text{H}_2\text{O}$ , consists of a fluid network of hydrogen bonds wherein each oxygen atom (black ball) is conjoined with an average of  $\sim 3.5$  hydrogens (white balls). **Right)** The general view is that water builds a cage around soluble particles (interior gray ball). Further details can be found in Ball (2008).



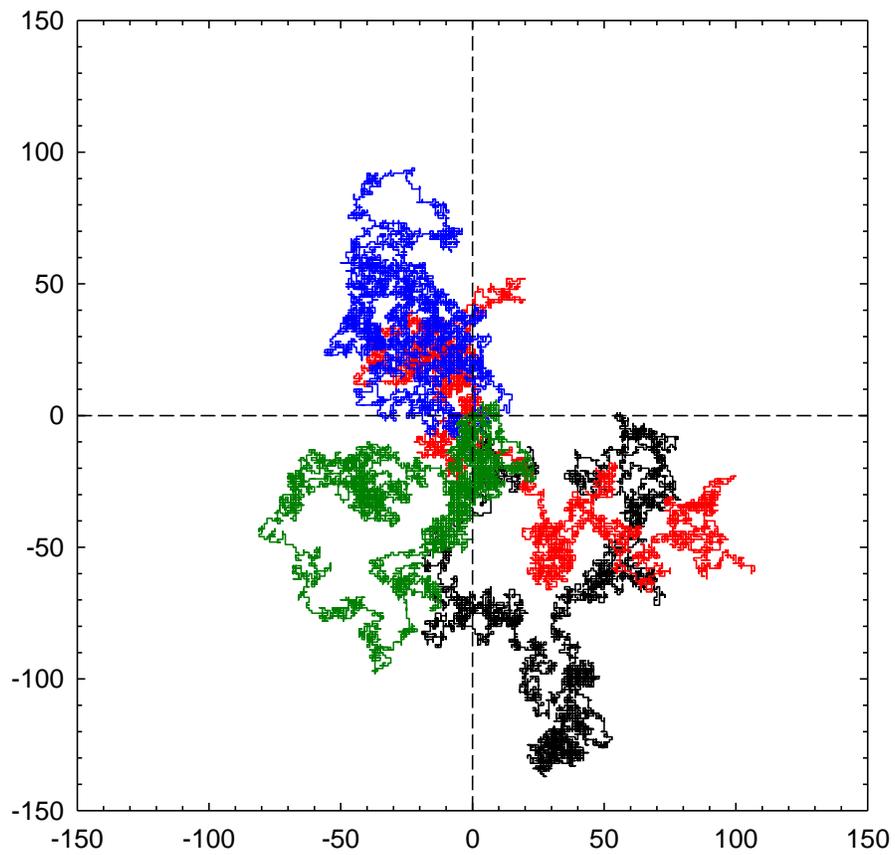
**Figure 7.3. Left)** The fractional contributions of major biomass components to the total dry weights of cells, determined by averaging over 18 species of bacteria and 15 species of eukaryotes (nine of which are photosynthesizers). The plotted fractions for each group sum to 1.0, and exclude some biomass such as chlorophyll in plants. Data are provided in Supplemental Table 7.2. **Right)** The negative scaling between the fractional contribution of DNA to cellular dry weight and cell volume. DNA dry weight per cell was obtained from the number of nucleotides in the total genome sequences of species, and assumes a single haploid genome per cell; cell dry weights were obtained by applying the function in Figure 7.1 to species with known cell volumes. The fitted power functions for the fractional dry weight of DNA are  $0.0072V^{-0.77}$  for bacteria and  $0.014V^{-0.62}$  for eukaryotes, where  $V$  is the cell volume in units of  $\mu\text{m}^3$ . Too much credence should not be attached to the apparent scaling differences between bacteria and eukaryotes, as both sets of data assume cells with haploid genomes, whereas a number of the eukaryotes may be diploid, and bacterial cells in active growth phases often contain several genomes. As a consequence, the plotted estimates may be somewhat downwardly biased (although by no more than a factor of two in eukaryotes). Data are provided in Supplemental Table 7.3.



**Figure 7.4.** Numbers of proteins and messenger RNAs per cell. **Left panels)** Upper solid lines denote the total number of molecules per cell, summed over all genes. The brackets for numbers of molecules per gene denote the lower 2.5% and upper 97.5% cutoffs in the overall distributions; and the dashed and dotted lines are the regressions involving the means and medians. **Right panels)** Approximate distributions are given for the numbers per cell for different proteins in four different sized cells, ranging from (left to right) the bacteria *Mycoplasma* ( $0.05 \mu\text{m}^3$ ) and *E. coli* ( $1 \mu\text{m}^3$ ) to an approximate yeast ( $100 \mu\text{m}^3$ ) to a mammalian cell ( $2400 \mu\text{m}^3$ ). These distributions are based on the empirical values given in the left panels; means and standard deviations on the  $\log_{10}$  scale are given in the insets. The gray region for the mRNA plot denotes the fraction of cells expected to harbor zero transcripts at any point in time. Data are from the previous survey in Lynch and Marinov (2015), a more recent quantitative transcriptomic study on an unclassified marine picobacterium OM43 (Huggett et al. 2012; Gifford et al. 2016), and proteomics studies of the unicellular eukaryotes *Trichomonas* (Dias-Lopes et al. 2018) and *Leishmania* (Pinho et al. 2020).



**Figure 7.5.** Four realizations (each given by a different color) of a two-dimensional random walk, all starting at the same point (the intersection of the vertical and horizontal lines). In each case, 10,000 steps of unit length were recorded with the movement (right, left, up, down) being directionally random.



**Figure 7.6.** Some diffusion coefficients for simple substances in water at 20°C, as a function of molecular weight; a few of these are specifically annotated. The solid line is obtained with Equation 7.8 by multiplying the numbers of amino acids in the chain by the average molecular weight of an amino acid, 137 g/mol. The units may be changed by noting that 1 cm = 10<sup>4</sup> μm = 10<sup>7</sup> nm.

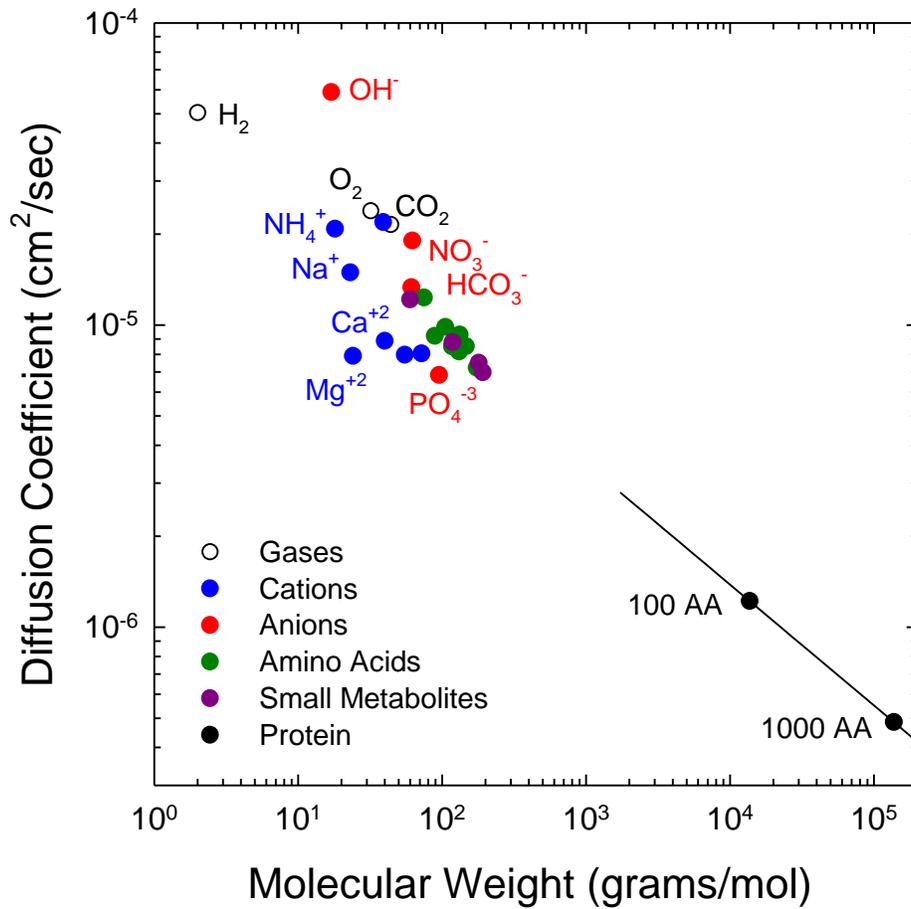
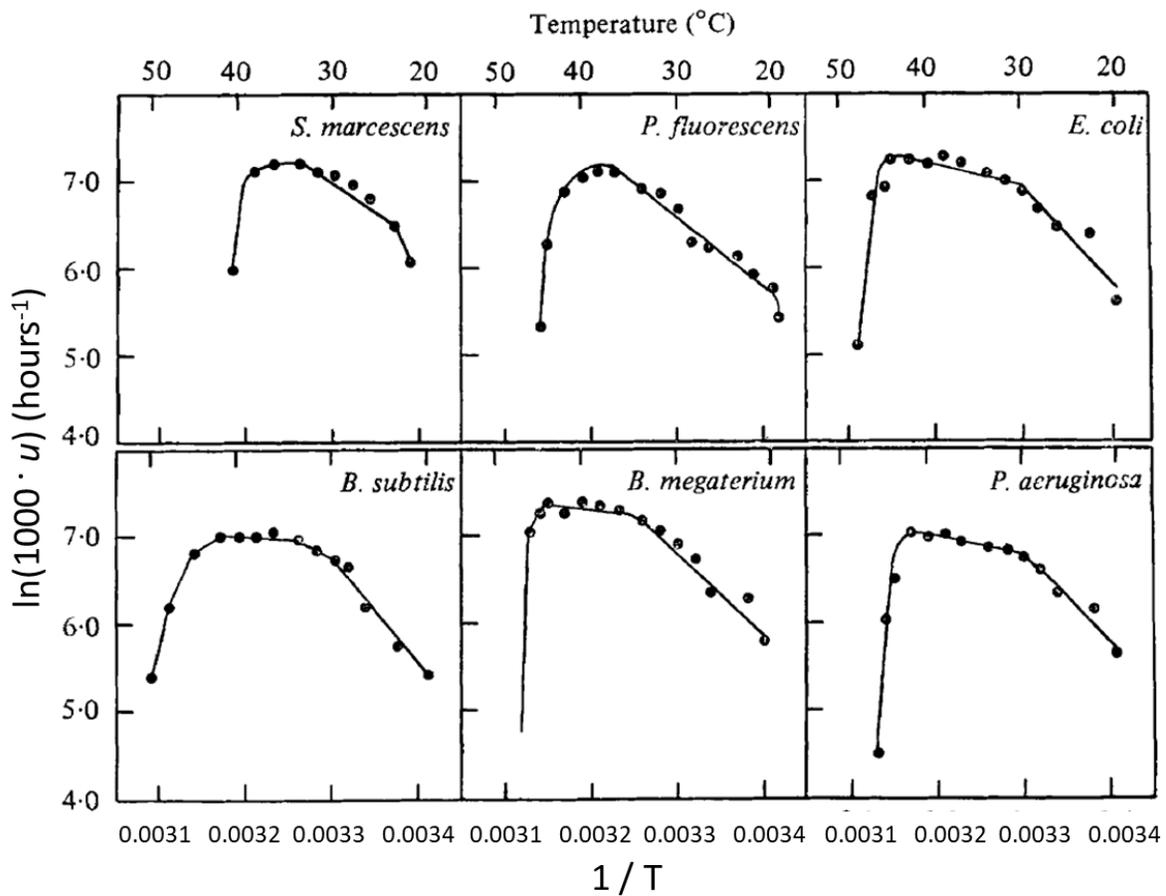
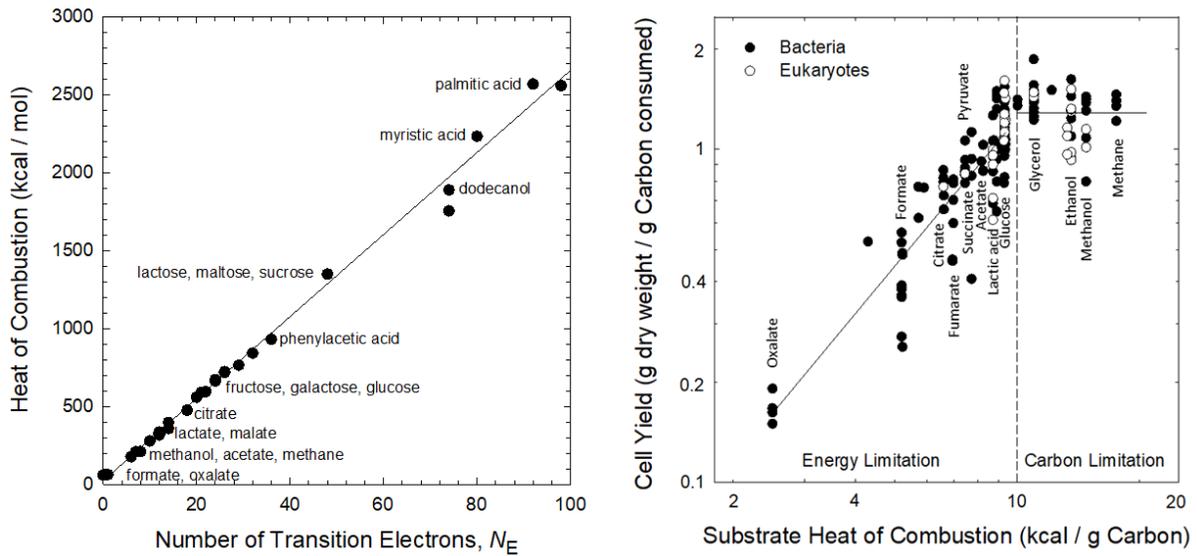


Figure 7.7. Examples of the response of cellular growth rate to temperature in six bacterial species. Growth rate ( $u$ ) is plotted on a log scale as a function of the inverse of temperature in degrees Kelvin ( $1/T$ ). Temperatures in Celsius are given along the top axis. Only in the central regions of the plots do the growth-rate responses follow the linear decline with  $1/T$  expected under Arrhenius rate behavior, and even then this is often just an approximation to a broader curvilinear pattern. From Mohr and Krawiec (1980).



**Figure 7.8.** **Left)** The heats of combustion of organic substrates commonly applied in growth experiments of microbes. The fitted regression is  $\Delta H_C = 26.2N_E + 25.8$ . **Right)** The yield of cell dry weight per gram of carbon consumed as a function of the heat of combustion of the substrate in units of carbon follows a power-law relationship,  $Y = 0.042H_C^{1.47}$  ( $r^2 = 0.81$ ), where  $H_C$  is the normalized heat of combustion, up to a carbon-specific heat of combustion  $\simeq 10$ , and thereafter levels off with constant value  $\simeq 1.3$  (units are defined in the axis labels). Note that the heats of combustion plotted here are equivalent to the plotted values on the  $y$  axis in the left panel divided by the number grams of carbon/mol of substrate. From Lynch and Trickovic (2020).



**Figure 7.9.** Two spheres with respective radii  $r_a$  and  $r_b$  will contact each other whenever the distance between their centers is  $r_c < r_a + r_b$ .

