A feeling for the numbers in biology

Rob Phillips[a] and Ron Milo[b,1]

[a]*Departments of Applied Physics and Bioengineering, California Institute of Technology, Pasadena, CA 91125; and* [b]*Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel*

Edited by Ken A. Dill, University of California, San Francisco, CA, and approved October 30, 2009 (received for review July 14, 2009)

Although the quantitative description of biological systems has been going on for centuries, recent advances in the measurement of phenomena ranging from metabolism to gene expression to signal transduction have resulted in a new emphasis on biological numeracy. This article describes the confluence of two different approaches to biological numbers. First, an impressive array of quantitative measurements make it possible to develop intuition about biological numbers ranging from how many gigatons of atmospheric carbon are fixed every year in the process of photosynthesis to the number of membrane transporters needed to provide sugars to rapidly dividing *Escherichia coli* cells. As a result of the vast array of such quantitative data, the BioNumbers web site has recently been developed as a repository for biology by the numbers. Second, a complementary and powerful tradition of numerical estimates familiar from the physical sciences and canonized in the so-called ''Fermi problems'' calls for efforts to estimate key biological quantities on the basis of a few foundational facts and simple ideas from physics and chemistry. In this article, we describe these two approaches and illustrate their synergism in several particularly appealing case studies. These case studies reveal the impact that an emphasis on numbers can have on important biological questions.

bionumbers | order of magnitude | physical biology

Although many biological phenomena have been discovered and explained on the basis of qualitative analyses, new insights often follow when they are revisited in quantitative terms. More importantly, in some cases, without a quantitative description, there is no discovery at all. This is perhaps best illustrated by the foundation of genetics, one of the great pillars of modern biological investigation. In a recent biography (1), Mendel's views are paraphrased thus: "… no one has concentrated on the number of different forms that appear among the offspring of hybrids. No one has counted them. But doing all this counting and sorting appears to be the only way by which we can finally solve a question whose importance cannot be overestimated." Mendel's careful tallying of frequencies of occurrence of different traits (2) gave him insights that were impossible to garner on the basis of qualitative observation alone.

The quantitative tradition in genetics was continued in the group of Thomas Hunt Morgan with Alfred Sturtevant's determination of the first chromosomal map, again by counting frequencies, this time of pairs of inherited traits. Sturtevant's characterization of his results, worked out on a night spent examining data from the Morgan lab rather than doing his undergrad homework (or so the story goes) was: "'They [the results] form a new argument in favor of the chromosome view of inheritance, because they strongly indicate that the factors investigated are arranged in a linear series, at least mathematically" (3).

An example of special interest to this article concerns the long history of deriving a properly balanced stochiometric equation for the processes of photosynthesis. This kind of work began at least as early as Van Helmont's oft-cited experiment on the growth of a willow tree in which he carefully measured the mass of the soil before and after the growth of his tree revealing a negligible change in the mass of the soil pointing to the need to look elsewhere for the sources of the material making up the tree. This tradition was carried on through the era of the great "pneumochemists" (4) who set themselves the task of measuring the quantities of gas taken up and liberated by plants during their daily lives. Clearly the long history of the study of photosynthesis has relied on quantitative measurements as a key engine for biological discovery.

However, there is a different way in which biological numeracy can result in conclusions of deep biological significance. In this approach, numbers collected by the scientific community that initially appear unrelated are brought together as a tool of inference to shed light on biological mechanisms. A particularly inspiring example of this idea is revealed in the study of biological fidelity. Protein translation was already well characterized in the 1970s when John Hopfield and Jacques Ninio were struck by its impressive fidelity, after reports of approximately one error for every 10^4 amino acids added onto a nascent polypeptide chain. Inferring the required free energy and considering the even smaller error rates apparent in transcription and DNA replication led them to propose that to get such low error rates an energy-driven proofreading step is necessary. Kinetic proofreading, where an erroneous recognition is detected and rejected trading ATP and its equivalents with accuracy, has been subsequently suggested to exist in other biological systems [e.g., immunology (5), signal transduction and protein degradation (6)]. It is worth noting that no new measurements were needed in this inference; the numbers and basic physical laws held all of the required clues.

Focusing on the present, a longstanding effort that continues to deliver new insights concerns how cells decide where to go. In particular, bacterial chemotaxis is a continuous case study in biological numeracy. Several of the illuminating questions have been: (*i*) can an individual bacterium detect a gradient along its long axis, or instead, does such detection require measurements at different time points (7, 8), (*ii*) what permits bacteria to reveal such an enormous dynamic range in the concentrations that can be detected? That is, the ability of bacteria to discriminate gradients is present over a very wide range of absolute background concentrations and has been interpreted, in part, as resulting from clustering of receptor proteins (9, 10), and (*iii*) how can a robust function be achieved for a sensitive switch experiencing large fluctuations of its molecular components (11, 12)? In all cases, the answers to these questions were obtained primarily through an emphasis on numeracy.

Author contributions: R.P. and R.M. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: ron.milo@weizmann.ac.il.

This article contains supporting information online at www.pnas.org/cgi/content/full/0907732106/DCSupplemental.

Progress of this sort continues at an ever-accelerating rate as a host of new measurement techniques provide quantitative data of all kinds. Quantitative approaches that are even more subtle focus not on mean values of some biological quantity, but rather on variability directly. In a particularly clever example, the imperfect partitioning of proteins upon cell division evident in time-lapse microscopy (13) has been used as a way to count the absolute number of proteins in the mother cell. The probabilistic argument reasons that the decision as to which daughter cell will inherit a particular copy of the protein is effectively the result of a coin flip. In this case, the fluctuations in the difference between the two daughters provides a readout of the unknown calibration factor relating fluorescence and protein number. Probability arguments in general suffuse biology as is evident in yet another example, namely, the ability to differentiate neutral, positive, or purifying selection based on the ratio of synonymous to nonsynonymous mutations.

In the remainder of the Introduction, we take stock of several different approaches to biological numeracy, subject to the important proviso that quantitative approaches in biology are but one of many distinct ways to come to terms with biological complexity. Indeed, most of our current understanding of the molecular mechanisms of life is the result of qualitative rather than quantitative inferences based on experiments. As a result, it is not surprising that biological training often focuses on these productive and less mathematically demanding approaches. Nonetheless, we try to make the case that there exists a useful niche for quantitative analysis in several pivotal cases and that it is especially timely to increase awareness of this useful addition to the biologist's toolbox as a flood of new quantitative information is becoming available.

First, we discuss the development of a web site (www.BioNumbers.org) that serves as a portal to a vast array of measured numbers that characterize the living world at the molecular and cellular scale. We then describe a complementary approach to these numbers based on calculations with a special emphasis on simple order-of-magnitude estimates. In the rest of the article, we merge these two perspectives to examine several case studies of current interest that show how an insistence on quantitative analysis can sharpen the questions we ask about biological problems and can lead to surprises that contradict either intuition or leading hypotheses in a way that would not be uncovered with strictly verbal descriptions. In this way we find numbers serving as one useful pathway to what Barbara McClintock termed "a feeling for the organism" (14).

BioNumbers: Numbers from Measurements

Even for properties that have been measured numerous times it can often be surprisingly difficult to find their values. Except for a discontinued effort in the 1970s (15), biology does not have the same tradition of developing handbooks of quantitative data that are so common in engineering or physics (16). For example, finding the volume of a nucleus or the ribosome translation rate can result in time consuming and frustrating searches in textbooks or on the internet. To address this need, BioNumbers, the database of key numbers in molecular and cell biology (17), was constructed as a Wikipedia-like community effort (www.BioNumbers.org). More than 4,500 entries are currently available and the all-important pointer to the original literature is supplied. The BioNumbers team and users from the scientific community enter numbers that are deemed useful for other researchers and that have been published in peer-reviewed literature. User comments are taken into account in the curation process by the BioNumbers team, and part of the curation process includes the fact that values that have been superseded by better measurements are replaced. Some quantities of interest appear more than once because the same quantity is often measured by different groups or under different experimental conditions. Users have easy access to the different measurements that have been performed, allowing for simple searches for the most relevant examples for the specific case under study and indicating the range of values reported. In Tables S1–S4, we give some statistics on the most popular searches, and we invite readers to add entries from their fields of study. In the remainder of the article, each time that we invoke some particular BioNumber of interest, we will reference its BioNumbers ID (e.g., BNID 101234) used much like the PubMed ID numbers so familiar from the National Institutes of Health web site. As will be shown through some of the case studies in the remainder of the article, the numbers that emerge from such a search can form the basis for further investigation of particular biological questions. Variety is always evident at the level of the cell and should be kept in mind when discussing the values of biological properties. Our approach in this article is to provide illustrative examples with the proviso that the development of a true feeling for the numbers requires a more thorough investigation of the range of measured values and a detailed discussion of the environmental conditions, as is made possible by the range, measurement method, and other fields in the BioNumbers web site.

BioEstimates: Using Biological Numbers to Make Biological Inferences

In some cases, it is not enough to consider the hard-won measurements of important numbers in biology by themselves. A useful complementary perspective is to see how those numbers make sense on the basis of estimation and detailed calculation. One of the first lessons learned by science students is to check their units and make sure that if one side of an equation has units of kilograms the other side of that equation better not have units of Joules. This is great advice and a key way to see whether calculations make sense. However, a second kind of sanity check that is at once more subtle and has larger scientific reach requires that we put in some numbers and convert abstract expressions into numerical statements. We can then see whether the resulting numbers jibe with our intuition and keep an eye out for clues that suggest relationships that were previously hidden. Two of the greatest episodes in the history of physics resulted from this strategy: (*i*) in the "Principia," Newton compared how far the moon falls in a minute to how far a mass falls at the surface of the Earth in the same time (18, 19) and found the ratio of those distances was "pretty nearly" 1/3,600, the result required by an inverse-square law of gravitational attraction (the ratio of the distances derived from a knowledge of the distance between the Earth and the moon and the radius of the Earth was already known to be ≈60:1). (*ii*) Approximately 200 years later, James Clerk Maxwell derived a wave equation for electromagnetic fields (20, 21) that (in modern notation) involved the peculiar-looking quantity $1/\sqrt{(\mu_0\varepsilon_0)}$, where μ_0 and ε_0 are the magnetic and electrical permeabilities of free space, respectively. Interestingly, the quantity in Maxwell's wave equation has units of velocity and describes the speed with which electromagnetic waves move in free space. When he plugged in the numbers what popped out on the other side was pretty nearly the speed of light, serving as a key theoretical line of evidence that light is an electromagnetic phenomenon.

A biological example of how quantification reveals surprising links between quantities that were previously assumed to be unconnected is given by Stadler's studies of the dependence of mutation rates in maize on the wavelengths of

UV radiation shined on them. He found a mutagenic spectrum with a maximum at ≈260 nm (22). Already knowing that the spectrum of nucleic acids has maximal absorption exactly at that same wavelength region, this brought quantitative support to the Avery experiment (23) and the Hershey–Chase experiment (24) performed later that established DNA as the carrier of genetic information.

Beyond the evaluation of expressions like in the famed examples given above, simple order of magnitude estimates have also proven a fruitful avenue for developing intuition. The tradition of simple order-of-magnitude estimates has now been canonized within the physics community as "Fermi problems" (25). This name refers to the penchant of the famed 20th-century physicist Enrico Fermi to carry out order of magnitude estimates for examples ranging from the number of piano tuners in Chicago to the distance a crow can fly to more consequential examples relevant to nuclear physics. This tradition has been carried on in many fronts in different scientific communities (ref. 26 and www.inference.phy.cam.ac.uk/sanjoy/oom/book-a4.pdf). In our opinion, the time is ripe for the emergence of a similar tradition in the biological setting because as we argue throughout the remainder of the article such estimates can reveal gaps in our understanding, relate quantities that were previously not known to be related and serve as the basis for an intuitive understanding of the significance of numbers that result from measurements.

Concomitant with the development of the BioNumbers web site, we have been engaged in trying to develop a systematic set of Fermi problems for biology (i.e., BioEstimates) (28) that have the same reach across scales as are represented in the BioNumbers database. Examples that illustrate this style of thinking in the biological setting are explored in the case studies throughout the remainder of the article.

The Power of Estimates Coupled to Measured Values

Both estimates and measured biological numbers have their place and, in many cases, the most potent insights come from combining them. Order of magnitude estimates provide a useful sanity check, but must be juxtaposed with the measurements that show whether they have merit or not. As a result of this interplay between order of magnitude estimates and measured values of the same quantities, it is possible to reveal fallacies in our understanding of a given problem. However, it is important to note that the accuracy demanded in an estimate should depend on the context the number is being used in. For example, an order of magnitude description for the number of carbons in an *Escherichia coli* bacterium is probably the best we can achieve considering the variability in size and composition of the bacterial cell. However, in thinking about the concentration of carbon dioxide in the atmosphere a factor of two can possibly spell the difference between survival and extinction for some species (29). It is an essential tool of the trade to know what level of accuracy is required for a given problem.

Careful attention to accurate determination of numbers is key in a class of analyses that set limits on biological phenomena. One of the cornerstones of modern science is the second law of thermodynamics that has its foundations in the limits of energy usage as implied by Carnot's study of the maximum efficiency of heat engines. Examples where physical limits can be put on biological processes abound ranging from macroscopic considerations of the strength limits of legs to the largest animals that can walk on water [the Jesus number (30)] to microscopic considerations such as the smallest number of photons that can be detected by a rod cell (31) and the smallest chemical gradients that are detectable by a motile bacterium (7).

In the remainder of the article we flesh out these general concepts with specific case studies. We are constantly trying to find a balance between general statements and the inevitable biological exceptions and between conciseness and accuracy. We begin with one of the most fundamental mysteries of life, namely, how one cell becomes many. In particular, this case study concerns the use of carbohydrates by living organisms to make new cells, with a special emphasis on the growth of microbes such as *E. coli* and yeast. We show how an order of magnitude calculation contrasts with the measured biological number. Our second case study centers on the crowded landscape of the cell surface as it marshals the busy traffic of molecules to and from the cell. One of the messages of this case study will be an impression of the extreme crowding of both the cytoplasm and cell membranes, suggesting experiments to test some of the hypotheses set forth by the quantitative analyses. We next turn to photosynthesis. In this most fundamental of all fueling processes, nuclear reactions in the sun produce photons that are harvested by living organisms on Earth that use them to turn inorganic CO_2 into useful biological substrates. This example gives us the chance to showcase biological numbers at a totally different scale than in the previous examples through an emphasis on numbers of relevance to the entire biosphere. We then build on this analysis through a case study centering on how individual cells carry out the processes of photosynthesis by harvesting materials and energy from the environment to make new cells. These case studies result in interesting biological conclusions and testable hypotheses, thus highlighting the power of a repository of biologically relevant numbers and order of magnitude estimations.

Results

Case Study: Managing the Macromolecules of the Cell. ***The materials to build a cell.*** We choose to begin with one of the "simplest" and most fundamental of biological experiments. In particular, we consider what unfolds if we take a 5-mL tube with sterile growth medium, add a single *E. coli* cell, and then place that tube in a shaker at 37 °C. What factors limit the maximal rate of cell division under such ideal conditions? How many sugar molecules does it take to make a single *E. coli* cell and how does the answer to that question depend on growth conditions? Is the cost mostly for building materials or the energetic investment to put those materials together (i.e., for energetic purposes)? We begin by thinking about the sugar needed to synthesize a new cell. At a representative exponential growth rate of 40 min per division the rod-shaped *E. coli* has a diameter of ≈1 μm and a length of ≈2 μm (note that cell size depends on growth rate). With water content of ≈70% (BNID 100044) the measured dry mass is 0.5 pg (BNID 103892, 103904, 102230, 102242). The elemental composition of *E. coli* is such that approximately half the mass is carbon (BNID 100649) and therefore there are $\approx 10^{10}$ carbon atoms per cell (BNID 103010). Thus, it requires the carbons from $\approx 2 \times 10^9$ glucose molecules to make a new cell when considering only the required carbon building material. How does that compare with the energetic cost?

We begin with an order of magnitude estimate of the energetic cost and then use experimental values to assess how good this estimate is. Approximately half of the carbons used to make up a cell are tied up in amino acids [lipids and nucleic acids being the next major constituents (32); BNID 101436]. There are on average approximately five carbons per amino acid, implying $\approx 0.5 \times 10^{10}/5 \approx 10^9$ amino acids per cell. This compares well with an experimental value of 1.3×10^9 amino acids per cell under a 40-min cell cycle division time (BNID 100089). We can explore the

energetic consequences of all of this protein synthesis by noting that it requires four ATP molecules to add an amino acid to a nascent polypeptide chain (BNID 101442). We thus find an energetic cost of $\approx 4 \times 10^9$ ATPs per cell for protein polymerization that is known to be the main energetic cost for cell biosynthesis [energy costs for synthesis of DNA, cell wall, lipids, etc. are much smaller (33)].

How do these rough estimates compare with measured values? Experimentally, one measures the decrease in sugar concentration in the medium per unit of biomass produced. From knowing how many of the sugars are used as cell building blocks (2×10^9; see above) and the number of ATP produced from each sugar in either aerobic (≈30 ATP/glucose) or anaerobic (≈3 ATP/glucose; BNID 105011) conditions, the experiments imply that *E. coli* growing on glucose requires $\approx 10–20 \times 10^9$ ATPs (BNID 101981, 101983; for dependence on growth rate, temperature, etc. see http://openwetware.org/wiki/Ecoli_ATP_requirement). A large part of the difference between this value and the energetic cost of protein polymerization ($\approx 4 \times 10^9$ ATPs) is suggested to arise from the cost of keeping the membrane in an energized state (34). Although the simple estimate gave us the correct scale of total ATP consumption to build a new cell, BioNumbers enabled us to assess the accuracy of this estimate and, more importantly, infer the existence of other major costs. Going back to the cost in terms of sugars, under aerobic conditions glucose can be maximally used to make ≈30 ATP molecules (BNID 101778) so the energetic requirement is $\approx 3–6 \times 10^8$ glucose molecules on top of the $\approx 2 \times 10^9$ needed for the fundamental building blocks. Thus, in this case the work cost (energy) is somewhat cheaper than the building material cost (carbon source). Under anaerobic conditions, only approximately three ATPs are produced in mixed acid fermentation of glucose [BNID 105011, 103350, versus two ATP for alcohol (ethanol) fermentation in yeast or homolactate fermentation in our muscles]. The cell then needs another $\approx 3–6 \times 10^9$ glucose molecules. So under these conditions the work costs more than the building materials. In addition to giving us insight into how the energy budget is spent, these numbers teach us that if 10^{10} ATPs are used in ≈2,000 s of generation time then the standing pool of ≈3 mM of ATP in *E. coli* (BNID 101181; corresponding to $\approx 3 \times 10^6$ ATP per cell) is turned over approximately every second.

Similar estimates can be carried out for any of a number of other cellular constituents in a growing bacterium as highlighted elsewhere (28, 32, 35). The key point here is to illustrate how a few simple facts (cell size and density) can help us construct a meaningful census of the vast array of different mechanisms that have to work in concert to turn growth media into living matter.

Delivering the materials to build a cell. As shown above, for cells growing with only glucose as their carbon source, a steady stream of sugar molecules must make their way from the external environment to the cellular interior. What fraction of the *E. coli* membrane has to be covered by carbon source transporters when growing at maximal rate? This question forces us to think about physical limits to biological phenomena like those described in the Introduction, but this time with special reference to supplying the cell with the necessary ingredients for doubling. *E. coli* under ideal conditions, in media containing preformed amino acids, can divide every 20 min (≈1,200 s; BNID 103514), whereas in the previous example where glucose is the sole carbon source, and amino acids need to be synthesized from scratch, we analyzed a characteristic rate of ≈40 min. Approximately 10^{10} carbon atoms (see previous case study) have to be transported into the cell in a generation time. For simplicity we do not include the sugars that should be transported for energy production and that will be lost in the form of CO_2 or fermentation byproducts in glycolysis and the tricarboxylic acid cycle.

For calculating transport rates, assume that the carbon source is provided exclusively in the form of glucose or glucose equivalents. Is the maximal division rate dictated by the limited real estate on the surface of the cell membrane to locate glucose carbon transporters? From the rate of the glucose transporter in *E. coli* [BNID 102931 with similar values for glucose transporters in yeast (BNID 101737, 101738, 101739) and the lactose transporter in *E. coli* (BNID 103159)] we have an estimate of ≈100 sugar molecules per s as the saturated turnover rate. The surface area of the membrane is $\approx 6\ \mu m^2$ (BNID 103339 and 105026). The LacY lactose transporter has an oval shape normal to the membrane with dimensions of 6×3 nm (BNID 102929), assuming a similar size for the glucose transporter, the area it occupies on the membrane is $\approx 14\ nm^2$. For importing $\approx 2 \times 10^9$ sugar molecules into the cell (each consisting of six carbon atoms) within a cell cycle, the fraction of the area required is ≈0.04, or 4% of the membrane (see Table S5). Thus, a substantial part of the membrane has to be occupied just to provide the necessary carbon source. Can it be that faster growth is constrained by the ability to transport the carbon source? Dedicated experiments, motivated by this analysis, can clarify if there is a limitation on increasing this value further (say to 10%). We also note that detailed quantitative studies found that ribosome concentration grows linearly with growth rate (35) and that the rate of translation may dictate the limits on maximal growth rate. Indeed, it is clear that there is more to the determination of maximal growth rates than the transport of nutrients across the cell membrane, although at the same time, these estimates clearly demonstrate the need for careful thought about the management of membrane real estate.

A similar calculation can be performed for the yeast *Saccharomyces cerevisiae*. The volume and thus the number of carbons required is ≈50 times (BNID 100427) larger than in *E. coli*, whereas the surface area is ≈10 times larger and the fastest generation time is ≈5 times longer (BNID 100270). Thus, the areal fraction required for carbon building blocks is suggested to be similar. Notice though that under maximal growth rate conditions yeast performs fermentation to supply its energy needs, which dictates a significant additional transport of sugars. A measurement shows that under growth rates up to one division per 140 min, approximately half the carbon is lost in fermentation (with an even higher proportion at faster growth rates) (BNID 102324). Thus, the required surface fraction covered by transporters is suggested to be double that found in the bacterial setting, resulting in ≈8% areal coverage. We found this case study so tantalizing that R.M. is considering experimentally testing whether the expression of a membrane protein not related to transport will decrease the maximal growth rate of yeast and *E. coli* more than a control cytosolic protein overexpression as a result of limiting the available area for transporters.

This same kind of estimate can be played out again and again for other membrane occupants as well. For example, one can perform similar numerical sanity checks to see what fraction of membranes need to be occupied by the machinery of ATP synthesis to serve the energy needs of a rapidly growing cell. The result is a picture of the cell membrane that is riddled with hosts of different membrane proteins, each serving some different function. In a series of impressive recent measurements, it has been possible to perform a census (36). For examples of other census measure-

ments see refs. 36–38 of both the lipid and protein content of different types of membranes, resulting in a picture leading the authors of ref. 36 to assert: "A picture is emerging in which the membrane resembles a cobblestone pavement, with the proteins organized in patches that are surrounded by lipidic rims, rather than icebergs floating in a sea of lipids." Our calculation points at the necessity for such a constellation of membrane proteins and at rough quantitative predictions that could be tested experimentally.

As is evident from the variability and condition dependence of the assumptions used in the calculations given above we do not expect better than factor-of-two accuracy in the calculation, but would expect better than a factor of 10. In the next case study, the estimates involve much larger numbers and our resolution is thereby reduced, resulting in the fact that we will then expect to only get approximately the correct order of magnitude (that is to within a 10-fold or so accuracy). With each estimate, it is crucial to bear the uncertainties in mind.

Case Study: "Eating the Sun." Our first case study focused on a range of quantitative descriptions that tell us something about the management of "natural resources" by growing cells. Similarly interesting biological numbers arise in contemplating the origins of these resources in the fundamental process of photosynthesis, the process in which photons are harnessed to synthesize the sugar molecules that sustain humans and their animal friends. The numbers characterizing how living organisms eat the sun (39) are intriguing because they allow us to address questions at the level of the entire biosphere and at the level of the individual molecules and cells that power this planetwide process.

Photosynthetic efficiency and carbon balance at the global level. Questions of energy and the environment are at the center of current scientific and political discourse worldwide. In <10,000 years, humanity has gone from being but one of many occupants of the planet with a negligible footprint to the curators of the chemistry of Earth's atmosphere. The overall carbon budget of the atmosphere, of living matter, and humanity's impact on that budget is a useful starting point for scientific and political discussions alike. In recent years, many different experimental threads have come together to shed light on these global issues ranging from satellite missions that measure the color of the ocean water [revealing the quantity of chlorophyll (40)] to measurements of atmospheric CO_2 on distant volcanic mountaintops (41) to cell counts of cyanobacteria in a milliliter of sea water (42, 43). How can we think more deeply about the numbers that emerge from these studies? What do they tell us about photosynthesis and the redistribution of carbon on the planet? To respond to these questions, we explore some of the relevant orders of magnitude and contrast them with their corresponding BioNumbers.

We begin by asking what fraction of the energy arriving at the Earth from the sun is converted by living cells into chemical energy? The energy flux in full sunlight is ≈ 1 kW/m^2 (BNID 103709). Multiplying by the approximate overall cross-section of the Earth results in $\approx \pi \times (6.4 \times 10^6)^2$ m^2 $\times$ 10^3 kW/m^2 $\approx 10^{17}$ W (BNID 100943). How does this power compare with the current demands of humanity (44)? For an accurate estimate we would need to tally up a variety of different human energy demands, but for order of magnitude estimates it is sufficient to remember the rule of thumb that a person in the developed world uses ≈ 1 kW of electric power ($\approx$700 kW h per month; check your electricity bill). Earth's human population of ≈ 7 billion consumes with a total rate equivalent to ≈ 2 billion developed world energy consumers. Given that electricity is produced at an efficiency of $\approx$30% we arrive at an energy requirement of $10^3 \times 2 \times 10^9/0.3 \approx 10^{13}$ W = 10 TW. This is a "stick in the sand" method to estimate humanity's overall energy consumption that is currently $\approx$15 TW (BNID 101694). This simple estimate reveals a four order of magnitude "excess" of energy impinging on Earth each second compared with that used by humanity. From this perspective, Earth is actually an energy-rich planet, not an energy-poor planet. Because there are $\approx$8,000 h per year it can be said that the solar energy impinging on Earth in 1 h is equivalent to all of humanity's needs over a year (45). This overly bright result is clouded by several obstacles that we will discuss in the context of photosynthesis.

The theoretical efficiency of photosynthesis (the energy content in energy currency products ATP and NADPH divided by the energy in the incoming solar radiation) is limited to $\approx$10% because of the physics and photochemistry in play (46). This arises from the limited wavelengths that can be used (below a rough threshold of 700 nm), from the fact that wavelengths with more energy are only partially used (only the equivalent of 700-nm photon energy excites the reaction center) and from the electron chain stoichiometry relating electrons to ATP and NADPH. Changes in any of these factors would require a fundamental alteration in how photosynthesis is performed.

Humanity's ability to siphon off the energy available from sunlight is actually even more limited. In modern agriculture, even under favorable conditions of irrigation and fertilization the efficiency for conversion to biomass is usually about an order of magnitude lower at $\sim$1% on a yearly basis (46, 54). This is partially because of respiration losses and the limited ability to cope with high levels of illumination that result in saturation of the photosynthetic machinery. On a global basis the conversion of solar energy to biomass has an effective efficiency an order of magnitude lower at $\sim$0.1–0.3% (BNID 100761, including oceanic areas that suffer from nutrient limitations). This is because of seasonal changes, the existence of large areas of land on our planet that do not sustain vegetation and that, in natural ecosystems, nutrients, water, pests, and pathogens can be limiting factors. This global value is the most difficult to assess and is based on a combination of satellite-based information on the concentration of chlorophyll around the globe tied to local measurements of the relation of chlorophyll to productivity (27). Therefore, of the four orders of magnitude of excess energy impinging on Earth, the biosphere is able to harvest $\approx 10^{14}$ W, an order of magnitude more than our electricity needs. Currently, for purposes of growing our food [in large part because of the increasing demand for meat that requires feeding of livestock (47)], it is estimated that humanity is already appropriating $\approx$1/4 of the terrestrial photosynthetic primary productivity (48), a value that should serve as an alarming warning shot across the bow concerning our increasing effect on the planet.

Once the photons that are the carriers of all of this energy have been absorbed, how does this translate into carbon fixation of atmospheric CO_2 into carbohydrates? To answer this question, we perform a simple sanity check calculation. The energy content of dry biomass is $\approx$4 kcal/g biomass (BNID 103499; equal to $\approx$16 kJ/g biomass). Thus, the estimate in the preceding paragraph $\sim 10^{14}$ W is equivalent to $\approx 10^{14}/(4 \times 10^3$ cal/g $\times$ 4 J/cal) = 10^{10} g/s = 10^4 ton/s. On a yearly basis that is approximately (10^4 ton/s) $\times$ (3×10^7 s/year) = 300 Gt biomass/year (Gt = gigaton). Because carbon is approximately half of the dry biomass this yields an order of magnitude estimate that the total carbon fixation is $\approx$150 Gt carbon/year. Evidence from satellites coupled to calibrated models estimate $\approx$50 Gt carbon/year of terres-

trial net primary productivity (BNID 102934) and a similar quantity fixed in oceanic environments (BNID 102936), although most of this latter material is respired and returned to the atmosphere within several days (BNID 102947). We thus see that on the basis of a relatively meager investment of facts our back of the envelope calculations are in reasonable accord with global estimates.

As we have argued throughout the article, often one of the most useful outcomes of playing with the numbers is that one can relate seemingly disparate phenomena and observations. The Keeling curve, which shows the atmospheric concentration of CO_2, has become one of the most iconic images of modern science (http://scrippsco2.ucsd.edu/images/graphics_gallery/original/mlo_record.pdf). Before discovering the overall increase in CO_2, Keeling (41) first observed striking annual periodic variations in atmospheric CO_2 concentrations. These annual variations correspond to differences in photosynthesis in the Northern and Southern Hemispheres as a result of the 2-fold ratio in land mass in these two hemispheres. Interestingly, by once again making relatively meager assumptions about the processes that are in play, the amplitude of the annual variation in photosynthesis can be used to estimate the net primary productivity of photosynthesis on Earth and the results are in good accord (approximately a factor of 10) with the numbers explored above.

The molecules and cells of photosynthesis. Despite the staggeringly large numbers associated with terrestrial photosynthesis, at the end of the day, the entirety of this carbon fixation is taking place on a cell-by-cell basis. At the cellular level, what does it take for an organism to be able to absorb solar energy? How many layers of pigments does it take to absorb half of the available photons? Further, given our molecular understanding of the carbon fixation process, what do these numbers tell us about the number of cells and molecules mediating all of this carbon chemistry?

The solar flux at wavelengths that can be used by photosynthesis (400–700 nm) is ≈2,000 μEinstein (that is μmole of photons per m^2 per s; BNID 100329). The photosynthetic reaction center has a diameter in the membrane of ≈10–20 nm (ref. 49; BNID 100904, 103907, 103908). The flux of photons onto that area is $\approx (2 \times 10^{-3}$ moles photons/$m^2) \times (6 \times 10^{23}$ photons/ mole) $\times$ $(10^{-16}$ $m^2) \approx 10^5$ photons per s. This contrasts with the reaction center maximal turnover rate of 100–10,000 s^{-1} (BNID 100349). We conclude that many layers are required to absorb half of the incident photons. In practice one indeed finds several layers of membrane in every chloroplast or photosynthetic bacterial cell. Moreover, the photosynthetic capacity in plant leaves saturates at 10–20% of maximal solar flux. This is an example where we can rationalize structural and functional features based on knowing the relevant numbers.

The process of storing harvested solar light in the form of stable chemical energy requires carbon fixation using the enzyme Rubisco, claimed to be the most abundant protein on Earth (50). However, how abundant must a protein like Rubisco be to qualify for status as the most abundant protein on Earth? To see how much Rubisco there is, consider the net fixation of ≈100 Gt carbon per year, and an average effective rate of carbon fixation of ≈1 carbon per s. Using these numbers, we arrive at a global need of 4×10^{10} kg of functional Rubisco or ≈5 kg per person on Earth (see Table S6). Assuming that ≈10% of global photosynthesis is carried out by cyanobacteria, we can use this estimate for the overall quantity of Rubisco either to figure out how many cyanobacteria there are in the world's oceans by using the known number of Rubiscos per cyanobacterium [e.g., ≈200 Rubisco octamers/carboxysome (BNID 101678) and 6 carboxysomes/cell (BNID 102623) (51), leading to an order of magnitude estimate of 10^{29} cyanobacteria worldwide], or alternatively, using an independent estimate of the number of cyanobacteria, we can compute the approximate number of Rubiscos per cyanobacterium. The value of ≈5 kg per person transforms an intangible and astronomical number given in exponential notation into a quantity that conveys some sense for the prevalence of Rubisco by reporting it on a per-human basis.

These biological numbers and associated estimates concerning photosynthesis leave us with the impression that many claims about the energy and carbon balance of the biosphere and of cells can be assessed by knowing some basic numbers and performing order of magnitude estimates. Numbers like those described in this section are clearly an important part of the equipment needed to reason about any sort of energy policy relevant to issues as diverse as the reasonableness of subsidies for corn-derived ethanol or the promise of roof-top solar heating (52).

Discussion

In this article, we have shown how order of magnitude estimates in conjunction with the accessibility of measured numbers of biological significance provide a useful picture of a vast array of biological problems, although this approach is only one of many and we are not advocating it as the unique or "right" way to study living systems. Through a series of illustrative (rather than comprehensive) case studies including: (*i*) one of the great mysteries of cell biology, namely, how from one cell come many, (*ii*) the mechanisms governing the regulated flow of materials in and out of living cells, and (*iii*) a study of the carbon budget in photosynthesis both at the scale of biosphere and individual cells, we see that biological numeracy can be a powerful tool for understanding the living world that complements the powerful tools based on qualitative reasoning that have given rise to modern biology.

It is fair to wonder whether this emphasis on quantification really brings anything new and compelling to the analysis of biological phenomena. We are persuaded that the answer to this question is yes and that this numerical spin on biological analysis carries with it a number of interesting consequences. First, a quantitative emphasis makes it possible to decipher the dominant forces in play in a given biological process (e.g., demand for energy or demand for carbon skeletons). Second, order of magnitude BioEstimates merged with BioNumbers help reveal limits on biological processes (minimal generation time or human-appropriated global net primary productivity) or lack thereof (available solar energy impinging on Earth versus humanity's demands). Finally, numbers can be enlightening by sharpening the questions we ask about a given biological problem. Many biological experiments report their data in quantitative form and in some cases, as long as the models are verbal rather than quantitative, the theory will lag behind the experiments. For example, if considering the input–output relation in a gene-regulatory network or a signal-transduction network, it is one thing to say that the output goes up or down, it is quite another to say by how much (53).

Given the flood of data emanating from new molecular techniques there is every reason to believe that more and more quantitative hints will be available for ever more sophisticated inferences about the mechanisms of biological action. We hope that readers of this article will be inspired to join us in our enthusiasm for the quantitative approach advocated here and make their own submissions to the BioNumbers database and similarly, will use simple order of magnitude estimates as a way to discover previously uncovered

linkages or call attention to paradoxes and conundrums in their own research areas.

ACKNOWLEDGMENTS. We thank Niv Antonovsky, Danny Ben-Zvi, Maja Bialecka, Lacra Bintu, Jed Buchwald, David Cahen, Trek Changeux, Sidney Cox, Yuval Eshed, Nir Friedman, Hernan Garcia, Bill Gelbart, Peter Goldreich, Shura Grosberg, Adrian Jinich, Stephanie Johnson, Paul Jorgensen, Jane Kondev, Avi Levy, Sanjoy Mahan, Debbie Marks, Simon Mawer, Elliot Meyerowitz, Uri Moran, Elad Noor, Steve Quake, Chris Sander, Dave Savage, Moselio Schaechter, Eran Segal, Richard Sever, Guy Shinar, Mike Springer, Tim Skype, Wilfred (Zeev) Stein, Bodo Stern, Arbel Tadmor, Julie Theriot, Jon Widom, and Mike Widom for insights on biological numeracy and comments on the manuscript. This work is supported by a National Institutes of Health Director's Pioneer Award (to R.P.) and a grant from the Israel Science Foundation (to R.M.). R.M. is incumbent of the Anna and Maurice Boukstein career development chair.

1. Mawer S (2006) *Gregor Mendel: Planting the Seeds of Genetics* (Harry N. Abrams, New York).
2. Mendel G (1866) Versuche über plflanzen-hybriden. *Verh Naturforschenden Ver-eines Brünn* 4:3–47.
3. Sturtevant AH (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Zool* 14:43–59.
4. Rabinowitch E, Govindjee (1969) *Photosynthesis* (Wiley, New York).
5. McKeithan TW (1995) Kinetic proofreading in T cell receptor signal transduction. *Proc Natl Acad Sci USA* 92:5042–5046.
6. Rape M, Reddy SK, Kirschner MW (2006) The processivity of multiubiquitination by the APC determines the order of substrate degradation. *Cell* 124:89–103.
7. Berg HC, Purcell EM (1977) Physics of chemoreception. *Biophys J* 20:193–219.
8. Berg HC (1993) *Random Walks in Biology* (Princeton Univ Press, Princeton).
9. Sourjik V, Berg HC (2002) Receptor sensitivity in bacterial chemotaxis. *Proc Natl Acad Sci USA* 99:123–127.
10. Sourjik V, Berg HC (2004) Functional interactions between receptors in bacterial chemotaxis. *Nature* 428:437–441.
11. Alon U, Surette MG, Barkai N, Leibler S (1999) Robustness in bacterial chemotaxis. *Nature* 397:168–171.
12. Barkai N, Leibler S (1997) Robustness in simple biochemical networks. *Nature* 387:913–917.
13. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307:1962–1965.
14. Keller EF (1983) *A Feeling for the Organism: The Life and Work of Barbara McClintock* (Freeman, San Francisco).
15. Altman PL, Katz DSD (1978) *Biology Data Book* (Wiley, New York), 2nd Ed.
16. Lide DR (2008) *CRC Handbook of Chemistry and Physics* (CRC, Boca Raton, FL), 89th Ed.
17. Milo R, Jorgensen P, Moran U, Weber G, Springer M (2009) BioNumbers: The database of key numbers in molecular and cell biology. *Nucleic Acids Res*, in press.
18. Chandrasekhar S (1995) *Newton's Principia for the Common Reader* (Clarendon, Oxford).
19. Newton I, Cajori F, Crawford RT, Motte A (1934) *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World* (Univ California Press, Berkeley).
20. Maxwell JC (1892) *A Treatise on Electricity and Magnetism* (Clarendon, Oxford).
21. Maxwell JC (1954) *A Treatise on Electricity and Magnetism* (Dover, New York), unabridged 3rd Ed.
22. Stadler LJ, Uber FM (1942) Genetic effects of ultraviolet radiation in maize. IV. Comparison of monochromatic radiations. *Genetics* 27:84–118.
23. Avery OT, MacLeod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* 79:137–158.
24. Hershey AD, Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* 36:39–56.
25. Morrison P (1963) Fermi questions. *Am J Phys* 31:626–627.
26. Harte J (1988) *Consider a Spherical Cow: A Course in Environmental Problem Solving* (University Science Books, Mill Valley, CA).
27. Field CB, et al. (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281:237.
28. Phillips R, Kondev J, Theriot J (2008) *Physical Biology of the Cell* (Garland, New York).
29. Erwin DH (2006) *Extinction: How Life on Earth Nearly Ended 250 Million Years Ago* (Princeton Univ Press, Princeton).
30. Vogel S (2003) *Comparative Biomechanics: Life's Physical World* (Princeton Univ Press, Princeton).
31. Bialek W (1987) Physical limits to sensation and perception. *Annu Rev Biophys Biophys Chem* 16:455–478.
32. Neidhardt FC, Ingraham JL, Schaechter M (1990) *Physiology of the Bacterial Cell: A Molecular Approach* (Sinauer, Sunderland, MA).
33. Stouthamer AH (1973) A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie van Leeuwenhoek* 39:545–565.
34. Stouthamer AH, Bettenhaussen CW (1977) A continuous culture study of an ATPase-negative mutant of *Escherichia coli*. *Arch Microbiol* 113:185–189.
35. Bremer H, Dennis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed Neidhardt EA (Sinauer, Sunderland, MA), 2nd Ed, p 421.
36. Takamori S, et al. (2006) Molecular anatomy of a trafficking organelle. *Cell* 127:831–846.
37. Wu JQ, Pollard TD (2005) Counting cytokinesis proteins globally and locally in fission yeast. *Science* 310:310–314.
38. Ghaemmaghami S, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
39. Morton O (2008) *Eating the Sun: How Plants Power the Planet* (Harper, New York).
40. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281:237–240.
41. Keeling CD (1998) Rewards and penalties of monitoring the Earth. *Annu Rev Energy Environ* 23:25–82.
42. Vaulot D, Marie D, Olson RJ, Chisholm SW (1995) Growth of prochlorococcus, a photosynthetic prokaryote, in the Equatorial Pacific Ocean. *Science* 268:1480–1482.
43. Waterbury JB, Watson SW, Guillard RRL, Brand LE (1979) Widespread occurrence of a unicellular, marine, planktonic cyanobacterium. *Nature* 277:293–294.
44. Smil V (2003) *The Earth's Biosphere: Evolution, Dynamics, and Change* (MIT Press, Cambridge, MA).
45. Lewis NS, Nocera DG (2006) Powering the planet: Chemical challenges in solar energy utilization. *Proc Natl Acad Sci USA* 103:15729–15735.
46. Long SP, Zhu XG, Naidu SL, Ort DR (2006) Can improvement in photosynthesis increase crop yields? *Plant Cell Environ* 29:315–330.
47. Steinfeld H, et al. (2006) *Livestock's Long Shadow: Environmental Issues and Options* (Food and Agriculture Organization of the United Nations, Rome).
48. Haberl H, et al. (2007) Quantifying and mapping the human appropriation of net primary production in Earth's terrestrial ecosystems. *Proc Natl Acad Sci USA* 104:12942–12947.
49. Scheuring S, Lévy D, Rigaud J-L (2005) Watching the components of photosynthetic bacterial membranes and their in situ organization by atomic force microscopy. *Biochim Biophys Acta* 1712:109–127.
50. Ellis RJ (1979) The most abundant protein in the world. *Trends Biochem Sci* 4:241–244.
51. Iancu CV, et al. (2007) The structure of isolated *Synechococcus* strain WH8102 carboxysomes as revealed by electron cryotomography. *J Mol Biol* 372:764–773.
52. Muller R (2009) *Physics for Future Presidents: The Science Behind the Headlines* (Norton & Co., New York).
53. Ellis T, Wang X, Collins JJ (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol* 27:465–471.
54. Transeau EN (1926) The accumulation of energy by plants. *Ohio J Sci* XXVI:1–10.