The global mass and average rate of rubisco

Yinon M. Bar-On[a] and Ron Milo[a,1]

[a]Department of Plant and Environmental Sciences, Weizmann Institute of Science, 76100 Rehovot, Israel

Photosynthetic carbon assimilation enables energy storage in the living world and produces most of the biomass in the biosphere. Rubisco (ᴅ-ribulose 1,5-bisphosphate carboxylase/oxygenase) is responsible for the vast majority of global carbon fixation and has been claimed to be the most abundant protein on Earth. Here we provide an updated and rigorous estimate for the total mass of Rubisco on Earth, concluding it is ≈0.7 Gt, more than an order of magnitude higher than previously thought. We find that >90% of Rubisco enzymes are found in the ≈2×10^{14} m^2 of leaves of terrestrial plants, and that Rubisco accounts for ≈3% of the total mass of leaves, which we estimate at ≈30 Gt dry weight. We use our estimate for the total mass of Rubisco to derive the effective time-averaged catalytic rate of Rubisco of ≈0.03 s^{-1} on land and ≈0.6 s^{-1} in the ocean. Compared with the maximal catalytic rate observed in vitro at 25 °C, the effective rate in the wild is ≈100-fold slower on land and sevenfold slower in the ocean. The lower ambient temperature, and Rubisco not working at night, can explain most of the difference from laboratory conditions in the ocean but not on land, where quantification of many more factors on a global scale is needed. Our analysis helps sharpen the dramatic difference between laboratory and wild environments and between the terrestrial and marine environments.

Rubisco | primary productivity | quantitative biology

T he joint action of the photosynthetic machinery and the Calvin–Benson carbon fixation cycle controls the global carbon cycle and produces the vast majority of the organic carbon present in the biosphere (1). The fixation of atmospheric CO_2 in the Calvin–Benson cycle is enabled by the activity of the Rubisco enzyme, which as such has a pivotal role in the global carbon cycle. Almost 40 y ago, shortly after the discovery of Rubisco, Ellis crowned it the most abundant protein on Earth (2). This statement was derived in only a single paragraph of a much longer paper detailing Rubisco's role in primary productivity, based on carbon fixation in terrestrial environments and the turnover number of the Rubisco enzyme measured in the laboratory.

The brief analysis by Ellis was instrumental in emphasizing the important role of Rubisco in the environment, as well as the power of using back-of-the-envelope calculations as a tool to estimate the abundance of proteins in the biosphere. The actual robustness of this estimate is unclear, however. To demonstrate the uncertainty surrounding the estimate of the total mass of Rubisco, we note that Ellis arrived at approximately 0.04 gigatons (Gt = 10^{15} g) of protein. This can be compared with collagen, the most abundant protein in the human body, which accounts for ≈30% of the ≈10 kg of total protein mass in an adult human (3, 4). Collagen is found not only in humans, but also in livestock. Considering the collagen present in livestock, we arrive at a global mass of collagen of ≈0.05 Gt, higher than the total mass reported for Rubisco. In addition, we note that the original estimate by Ellis did not take into account marine carbon fixation, which supports a similar flux to terrestrial carbon fixation (5).

The aim of the present work was to use an independent methodology to construct a rigorous estimate for the total mass of Rubisco worldwide. We found that the global mass of Rubisco is ≈0.7 Gt, more than an order of magnitude higher than the previous estimate reported by Ellis. We use this independent estimate to probe the average global rate of Rubisco and find that in terrestrial environments, this corresponds to ≈1% of its characteristic k_{cat}.

Results

To estimate the total mass of Rubisco proteins, we estimate the total mass of terrestrial and marine Rubisco separately (Fig. 1). For terrestrial Rubisco, we use a two-step approach. We first estimate the global dry mass of leaves. We then estimate a characteristic mass fraction of Rubisco out of this dry mass. By multiplying these two components, we arrive at an estimate for the total mass of Rubisco, as shown in Fig. 1.

Estimating the Total Mass of Leaves Globally. To estimate the global mass of leaves, we rely on two independent methods. The first method arises from an estimate of the total biomass of terrestrial plants (6). This estimate of ≈450 Gt of carbon translates to ≈900 Gt of dry weight, assuming ≈40–50% carbon content in dry weight (7). To convert this total plant dry weight to the total mass of leaves, we use a meta-analysis of the mass fraction of different plant compartments across different biomes (8, 9). We use the average leaf mass fraction across biomes, weighted by the fraction of plant biomass in each biome (10) (step 1 in Fig. 1). Overall, this approach yields an estimate of ≈50 Gt dry leaf weight, which is ≈6% of the total mass of plants (full calculation at https://bit.ly/2RYH74k).

Our second approach to estimate the global mass of leaves is based on first estimating the total area of leaves on land and then converting it to mass using an estimate of the mass of leaves per unit leaf area (step 4 in Fig. 1). To estimate the total area of leaves, we rely on both field measurements (11) and remote-sensing (12) of the leaf area index (LAI = total area of leaves per unit land area) across the entire globe. (Details on the construction of the LAI maps are provided in *Methods*.) We thus generate two maps of leaf area, one based on field measurements of LAI and the other based on remote sensing of LAI. We sum the leaf area across the entire terrestrial surface of Earth and arrive at two independent estimates for the global area of leaves. As our best estimate, we use the geometric mean of the two estimates (step 2 in Fig. 1), which is ≈2×10^{14} m^2 (https://bit.ly/2GtzbkN). This is equivalent to 200×10^6 km^2, or approximately twice the

Significance

Rubisco is often claimed to be the most abundant protein on Earth, yet the quantitative evidence to support the estimate of its global mass are scarce. Here we provide a robust and detailed estimate of the global mass of Rubisco, which is an order of magnitude larger than previous estimates. We use this estimate to derive the time-average rate of terrestrial and marine Rubisco and show that they are, respectively, 100-fold and sevenfold lower than the in vitro measured k_{cat} of Rubisco at 25 °C.
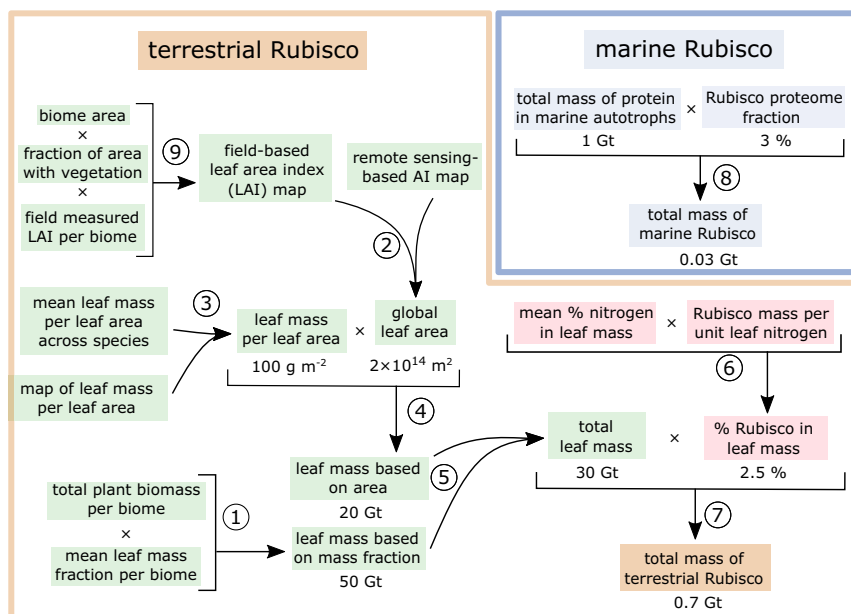
Fig. 1. A schematic representation of our methodology for estimating the mass of Rubisco. Our methodology for estimating the mass of terrestrial Rubisco comprises two parts: estimating the total mass of leaves and estimating the mean mass fraction Rubisco out of the dry leaf mass. Our methodology for estimating the mass of marine Rubisco is based on an estimate of the total biomass of marine autotrophs, as well as on reports on the fraction of Rubisco out of the dry weight of autotrophs.

global ice-free land area. We use our two independent estimates for the global area of leaves (one based on remote sensing and the other based on ground measurements) to evaluate the uncertainty associated with our estimate of the total leaf mass (https://bit.ly/2WnDxzr). We project the uncertainty (akin to a 95% multiplicative confidence interval) to be approximately twofold (*Methods*).

To convert the total area of leaves to the total mass of leaves, we multiply our estimate of the global leaf area by an estimate of the mass of leaves per unit area. We rely on two separate procedures for calculating the mass of leaves per unit area. The first procedure relies on a global database of plant traits (13), while the second relies on a recently generated map of the distribution of plant traits (14). Combining these two data sources (stage 3 in Fig. 1; detailed in *Methods*) yields an estimate of ≈100 g dry weight per square meter of leaf area (https://bit.ly/2Tjx8Dq). By multiplying the average leaf mass per leaf area by the total area of leaves (step 4 in Fig. 1), we arrive at an estimate of ≈20 Gt for the global mass of leaves (https://bit.ly/2Tjx8Dq). This value also includes crops, which are a relatively small fraction (≈2%) of the biomass of plants due to their high turnover rate relative to trees (6).

Our two methods for estimating the mass of leaves are based on independent datasets, each with its own assumptions and caveats. Thus, the relatively modest difference between the leaf mass fraction method (50 Gt) and the leaf area method (20 Gt) for estimating the global mass of leaves suggests a relative robustness of our estimate. As a best estimate for the total mass of leaves, we use the geometric mean of the estimates from both approaches (step 5 in Fig. 1), corresponding to ≈30 Gt. We use our two independent estimates to evaluate the uncertainty associated with our estimate of the total leaf mass. We project the uncertainty (akin to the 95% multiplicative confidence interval) to be approximately twofold (https://bit.ly/2RpopxC).

Estimating the Mass Fraction of Rubisco Proteins Out of Leaf Dry Mass. We next estimate the average fraction of Rubisco out of the total leaf mass (highlighted in red in Fig. 1). We rely on a recent meta-analysis that characterized several physiological parameters across a wide variety of plant species (15). We supplement this dataset with data on C4 plant species (13, 16–19). The first parameter we use for our analysis is the amount of nitrogen in Rubisco per unit leaf nitrogen. We convert the amount of nitrogen in Rubisco per unit leaf nitrogen to the total mass of Rubisco per unit leaf nitrogen by using of the fact that, similar to other proteins, nitrogen accounts for ≈15% of the total

mass of Rubisco (20). The second parameter that we use is the concentration of nitrogen in leaves. By multiplying these two numbers, we obtain an estimate for the mass of Rubisco per unit of dry leaf mass (step 6 in Fig. 1). Our dataset contains measurements for woody plants as well as herbaceous C3 and C4 plants. For each, we calculate the geometric mean of the mass fraction of Rubisco out of the dry leaf mass. We estimate that Rubisco accounts for ≈2% of the dry leaf mass in woody plants, ≈5% of the dry leaf mass in herbaceous C3 plants, and ≈1% of the dry leaf mass in herbaceous C4 plants. We estimate that leaves of woody plants account for ≈70%, leaves of C3 herbaceous plants account for ≈20%, and leaves of C4 herbaceous plants account for ≈10% of the total mass of leaves (*Methods*). We apply the characteristic fraction for each growth form on its fraction of the total leaf mass. Overall, we estimate that Rubisco accounts for ≈2.5% of the global mass of leaves (Fig. 2) (which, on a soluble protein basis, would be in line with the several tens of percent measured in the literature). Combining our estimates for the global leaf mass and the mass fraction of Rubisco in leaves (step 7 in Fig. 1), we estimate that the total mass of terrestrial Rubisco is ≈0.7 Gt (https://bit.ly/2UssG5A). We propagate our uncertainties for each parameter used to estimate the total mass of Rubisco to evaluate the uncertainty associated with our best estimate of the mass of terrestrial Rubisco. We project an uncertainty of approximately threefold associated with our estimate of the global mass of Rubisco (https://bit.ly/2Ust9ES).

One potential caveat in our analysis is that measurements of leaf mass can also include leaf tissues that are not photosynthetic, such as the petiole and midrib. These tissues account for ≈20% of the total leaf nitrogen (21), meaning that even if the inclusion of these tissues causes an overestimate of the mass of Rubisco, this is small in relation to the uncertainty that we project for our estimate.

Estimating the Mass of Marine Rubisco Proteins. Approximately one-half of global net primary productivity occurs in the oceans (5), and thus one would expect the global mass of Rubisco proteins in the marine environment to be significant. We estimate the total mass of Rubisco proteins in the marine environment by combining an estimate for the total biomass of marine autotrophs with estimates of the Rubisco content of marine autotrophs (highlighted in blue in Fig. 1). We recently estimated the total marine autotrophic biomass at ≈1 Gt C (6). Assuming that carbon constitutes ≈50% of the dry biomass, we estimate the total biomass of marine autotrophs as ≈2 Gt. We focus on microalgae, which are likely to
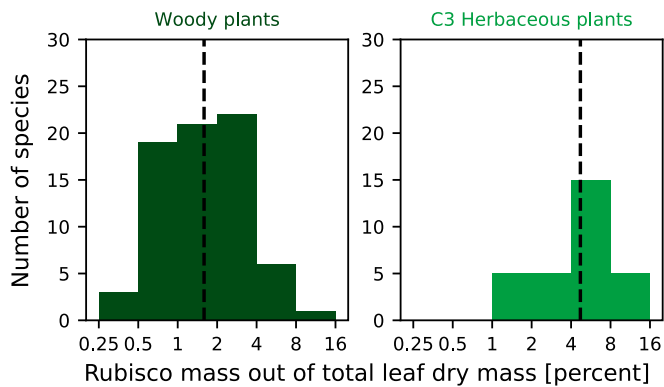
Fig. 2. The distribution of the fraction of Rubisco out of the dry mass of leaves. Each color denotes a different category of plant growth form, with dark-green values representing woody plants and the lighter-green values representing C3 herbaceous plants. The vertical line indicates the average fraction for each growth form used for the analysis.

dominate over macroalgae and seagrass as ocean autotrophs (6). Microalgae usually have a protein content of ≈50% of dry mass (22). Therefore, we estimate ≈1 Gt of proteins in marine autotrophs. We next estimate the mass fraction of Rubisco proteins out of the protein mass of marine autotrophs. We rely on previous reports (23–33) of values of 0.1–20% for different species of microalgae and cyanobacteria. We use the geometric mean of the measured proteome fraction of Rubisco for each group of phytoplankton. We use data on the relative biomass of each taxonomic group (6) to calculate the weighted global mean proteome fraction of Rubisco. We estimate that in marine phytoplankton, Rubisco accounts for ≈3% of the total protein mass. Multiplying our estimate of ≈1 Gt proteins in marine autotrophs by our estimate that Rubisco accounts for ≈3% of the total cellular protein (step 8 in Fig. 1), we estimate a total mass of marine Rubisco proteins of ≈0.03 Gt, which is <10% of the total mass of Rubisco (https://bit.ly/2RWZkiV). We use different estimates for the biomass of marine autotrophs, as well as the

variability in measurements of the proteome fraction of Rubisco in marine autotrophs, to evaluate the uncertainty associated with our estimate of the total mass of marine Rubisco. We project an uncertainty of approximately fourfold associated with our estimate (https://bit.ly/2G8ni3t).

Estimating the Effective Rate of Terrestrial and Marine Rubisco. In contrast to the method used by Ellis, our approach for estimating the total mass of Rubisco is not based on the rate of Rubisco. Therefore, we can use our estimate for the total mass of Rubisco to estimate its average effective rate of carbon fixation. In this section, we calculate the effective rate of both terrestrial and marine Rubisco. Our methodology is similar to recent efforts to quantify in vivo rates of enzymes (34). Namely, we estimate the total flux (reactions per unit time) that is supported by the combined action of all Rubisco proteins in a given environment (terrestrial or marine). We then divide the total flux by an estimate of the total number of Rubisco active sites, which we derive from our estimates of the total mass of Rubisco. By dividing the total flux by the number of protein active sites that support this flux, we get an estimate for the average catalytic rate of a single Rubisco enzyme (Fig. 3A).

For the terrestrial environment, gross primary productivity (GPP), which incorporates all carbon fixation, including the amount respired by the organism, is estimated at ≈120 Gt C y^{-1} (35). This value represents the total flux of carbon fixed on land each year, and although the exact value remains a matter of debate, it is estimated to be accurate to better than twofold (36), which is sufficient for the purposes of our analysis. We use the terrestrial GPP as a measure of the total flux that is supported by the combined action of all terrestrial Rubisco proteins. To calculate the average effective rate of Rubisco, which is measured in reactions per second, we convert the estimate of terrestrial GPP to units of molecules of CO_2 fixed per second. Because each CO_2 molecule contains one carbon atom, which has a molecular weight of 12 Da, we can express the global GPP flux in units of carbon atoms fixed per second. Because each year has ≈3 × 10^7 s, the total flux of terrestrial carbon fixation is ≈2 × 10^{32} carbon atoms (and thus CO_2 molecules) per second, as derived in Fig. 3B.

We convert our estimate for the total mass of terrestrial Rubisco proteins into an estimate of the total number of Rubisco active sites by using the molecular weight of a Rubisco active site,
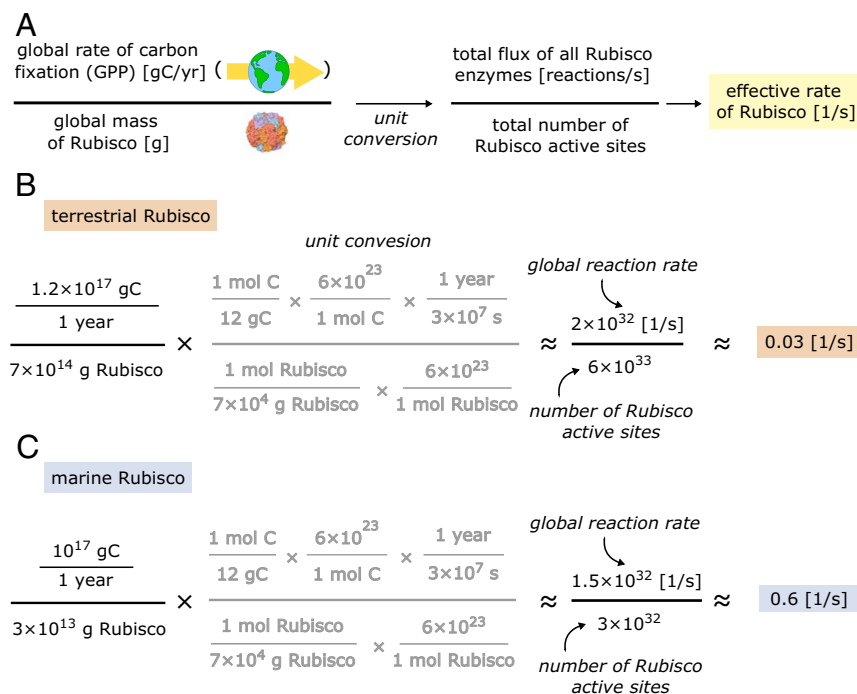

Fig. 3. Estimating the effective rate of an average Rubisco. (A) We use estimates for the total annual rate of carbon fixation—the gross primary productivity—in the terrestrial and marine environments, in conjunction with our estimates of the total mass of Rubisco in those environments. We convert the units of the total rate of carbon fixation to reactions per second using the fact that each reaction of Rubisco fixes one carbon atom, which has a mass of 12 Da. We also convert our estimates for the total mass of Rubisco into the number of active sites by dividing the total mass by the molecular weight of an active site. By dividing the total amount of reactions that all the active sites are performing each second by the total number of active sites, we get an estimate of the average rate of a single active site. B and C show the calculations for terrestrial and marine Rubisco, respectively.

which is $\approx$70 kDa [one large and one small subunit in type I Rubisco (37)]. We calculate that the total number of Rubisco active sites is $\approx$6 × 10^{33} (or $\approx$$10^{33}$ Rubisco L8S8 octamers). Dividing the total rate of all Rubisco enzymes by the total number of Rubisco enzymes, we calculate that the average catalytic rate of a single Rubisco is $\approx$0.03 s^{-1}, as depicted in Fig. 3C (https://bit.ly/2DG4GpY). We propagate our projections for the estimate of the total mass of Rubisco, as well as the uncertainty associated with the estimate of the terrestrial GPP, to evaluate the uncertainty associated with the estimate of the time-average rate of terrestrial Rubisco. Overall, we project an uncertainty of approximately fourfold associated with our estimate of the time-average rate of terrestrial Rubisco (https://bit.ly/2MJauly).

For the marine environment, two independent lines of evidence suggest that GPP is approximately twice the net primary productivity. First, measurements of autotrophic respiration, photorespiration, and dissolved organic carbon secretion imply that roughly 50% of the carbon fixed by photosynthesis is lost by these processes (38, 39). Second, measurements of the gross oxygen production (the total mass of oxygen produced by photosynthesis) are approximately 2.7-fold higher than the measured net primary productivity (38). Not all oxygen produced by photosynthesis is coupled to carbon fixation; some processes, such as the Mehler reaction and additional terminal oxidases (40), use the electrons produced in water splitting for other purposes. These processes usually account for 20–25% of the gross oxygen production (41). This means that the remaining 75–80% is coupled to carbon production, which dictates that the gross carbon production (the GPP) is approximately twofold higher than the net primary productivity. The global net marine primary productivity is estimated as $\approx$50 Gt C y^{-1} (5), and thus the global gross marine primary productivity is $\approx$100 Gt C y^{-1}. As with the terrestrial environment, we convert the GPP into units of reactions per second and arrive at an estimate of $\approx$1.5 × 10^{32} carbon atoms (and thus CO_2 molecules) fixed per second. Our estimate of the total mass of marine Rubisco is $\approx$0.03 Gt, which corresponds to $\approx$3 × 10^{32} marine Rubisco active sites. Dividing the total rate of all marine Rubisco enzymes by the total number of Rubisco active sites, we calculate that the average catalytic rate of a single Rubisco in the marine environment is $\approx$0.6 s^{-1}, roughly an order of magnitude higher than Rubisco in the terrestrial environment (https://bit.ly/2DG4GpY). We follow the same error propagation procedure as for terrestrial Rubisco and project an uncertainty of approximately fourfold associated with our estimate of the time-averaged rate of marine Rubisco (https://bit.ly/2RSVyXA).

To validate our results, we compare our global estimates for the rate of Rubisco with measurements of productivity and producer biomass at several different locations on land and in the ocean. Calculating the rate of Rubisco across 20 locations yields numbers well within the uncertainty we report for our global estimate, which increases our confidence in the validity of the approach (*SI Appendix*).

Discussion

Our work provides a methodology for estimating the total global mass of Rubisco. Whereas Ellis used the in vitro catalytic rate of Rubisco to estimate the total amount of Rubisco, we rely on mass fractions of the total autotrophic biomass. We estimate that the total mass of Rubisco enzymes is $\approx$0.7 Gt in the terrestrial environment and $\approx$0.03 Gt in the marine environment. Our estimate is more than an order of magnitude higher than the longstanding estimate of $\approx$0.04 Gt (2). Relying on measured mass fractions allows for much better constraints on the parameters used to estimate the total mass of Rubisco, resulting in the large difference from previous values. The large difference between our estimate and the estimate of Ellis demonstrates that his original claim stating that Rubisco is the most abundant protein on Earth was not well established. Even with our much higher estimate, it is not clear that Rubisco is indeed the most abundant protein in the biosphere. A comprehensive comparison of the mass of Rubisco with the mass of other ubiquitous proteins is required to substantiate

this claim. This is a research direction beyond the scope of this paper, but one that we are currently investigating.

One additional benefit of our methodology is that because it is not based on the catalytic rate of Rubisco, we can use our estimate to infer the effective time-averaged rate of Rubisco. We find that the effective rates of terrestrial and marine Rubisco are $\approx$0.03 s^{-1} and $\approx$0.6 s^{-1}, respectively. How does this rate compare with the maximal catalytic rate of Rubisco in plants? Using a collection of kinetic parameters of Rubiscos (42), we estimate the characteristic k_{cat} of terrestrial plants is $\approx$3 s^{-1} at 25 °C (https://bit.ly/2CNq6j2). Thus, the effective catalytic rate of Rubisco in wild terrestrial environments is $\approx$1% of its maximal rate. The characteristic k_{cat} of marine autotrophs is not significantly different from that of terrestrial plants, even though that of cyanobacterial Rubisco is usually faster (42), and we estimate it at $\approx$4 s^{-1} (https://bit.ly/2CNq6j2). As such, the effective catalytic rate of Rubisco in the marine environment is $\approx$15% of its maximal rate at 25 °C.

There are two trivial factors that help explain part of the difference between our estimates and the in vitro measured k_{cat} values. The first factor is that the flow of solar energy that drives carbon fixation is limited to daytime, so we expect our annual mean rate to be twofold lower than the rate of fixation in daytime (or, more accurately, $\approx$1.9-fold for the marine environment due to productivity being mostly in summer at high latitudes with long daytime; https://bit.ly/2CSY5qB). The second factor is the temperature at which carbon fixation occurs. Usually, the maximal rate of Rubisco is measured in vitro at 25 °C, but the fixation rate is dependent on the temperature at which the enzyme is working, so if in nature Rubisco is working at lower temperatures, we would expect its maximal rate to be lower than the in vitro measured rate of $\approx$3–4 s^{-1}. We use global maps of mean temperatures and primary productivity to estimate the average temperatures at which carbon fixation is occurring on land and in the ocean. We estimate that Rubisco is operating on average at $\approx$24 °C on land and at $\approx$10 °C in the ocean (https://bit.ly/2HG9AXV). Using data on the temperature dependence of Rubisco k_{cat} (32), we estimate a maximal Rubisco rate of $\approx$3 s^{-1} on land and $\approx$1 s^{-1} in the ocean at the ambient average temperatures. For the marine environment, these two factors explain most of the difference between the time-averaged rate of Rubisco and the maximal rate measured in vitro, implying that Rubisco is working near its k_{cat} in phytoplankton (32). On land, however, even when these factors are taken into account, Rubisco is still more than an order of magnitude beneath its k_{cat}.

Many factors could explain the lower effective catalytic rate of Rubisco. The rate of Rubisco may be limited by abiotic factors, such as the availability of solar radiation (e.g., when leaves are shaded in the canopy), CO_2 concentration to which Rubisco enzymes are exposed, water supply, nutrient supply, temperature, and others. It could also be caused by physiological processes like photorespiration, regeneration of RuBP, activation state of Rubisco, and others. This is only a partial list of factors that should be explored quantitatively in the future and compared between the terrestrial and marine environments.

Another way of phrasing the question of inefficiency is to consider why we see a large number of Rubisco enzymes operating at a submaximal rate as opposed to a smaller amount of Rubisco working faster. There are several possible explanations for this conundrum, which we touch on briefly. One line of argument suggests that excess Rubisco enables plants to respond more quickly to changing environmental conditions, such as alterations in illumination conditions (e.g., sun flecks). This is akin to the suggested excess ribosomal pool in carbon-limited bacteria (43). Another possible hypothesis is that Rubisco has a role in storage of nitrogen in plant tissues. In terms of elemental stoichiometry, plants have an abundant supply of carbon from the atmosphere but are limited by the supply of other crucial elements, such as nitrogen and phosphorus. Proteins have elemental stoichiometry suited for storing nitrogen when carbon is abundant without the requirement of phosphorus, which would be required for storage in nucleic acids. Thus, plants

can use protein as reservoirs of nitrogen, and as an abundant protein within plants, Rubisco could fulfill this role.

Our analysis of the effective rates of Rubisco enzymes in the terrestrial and marine environments exposes a strong difference between the two environments, with marine Rubisco enzymes operating at an order of magnitude greater rate than that on land. Why are marine Rubisco enzymes so much faster in the ocean? A partial explanation for this difference is CO_2 undersaturation on land. Most of the photosynthesizing organisms in the marine environment—cyanobacteria and eukaryotic phytoplankton—are equipped with carbon-concentrating mechanisms that increase the local concentration of CO_2 in the vicinity of Rubisco, helping reduce CO_2 subsaturation and its associated limitations (44). We believe that the results presented here motivate a dedicated analysis using detailed measurements to rigorously analyze the drivers for the overall quantitative difference between the marine and terrestrial environments.

Overall, our analysis sheds light on the distribution of Rubisco in the natural environment and provides a didactic framework for evaluating the effects of different plant traits on the abundance of Rubisco. We use the estimated total mass of Rubisco to show that on average, Rubisco is operating far below its maximal rate. Further studies will reveal the relative importance of factors that contribute to limiting the rate of Rubisco and explore the extent to which Rubisco might play additional roles besides its catalytic function.

Methods

A full description of our analysis, including the data sources and the code used to generate our results, can be found at https://github.com/milo-lab/rubisco_mass/.

Calculating the Total Area of Leaves. To estimate the total area of leaves, we construct two maps of the distribution of leaf area across the globe. The first map is the GLASS LAI product map (12), which is based on remote sensing. Since the number of leaves changes throughout the year due to deciduous plants, we use monthly composite maps and calculate the annual average of the total leaf area. We chose to use the composite map with a total leaf area closest to the annual mean.

As remote sensing of LAI can become saturated at high LAI values (45), we use ground-measured values of LAI in different biomes as an independent source (11). The average ground-measured LAI values for each biome represents the amount of leaf area per vegetated land surface, but in many biomes (e.g., deserts), most of the land surface is not vegetated. We use a recent study that produced a global map of vegetation coverage (46). For each location, we multiply the fraction of land that is vegetated (either by trees or by short vegetation) by the average LAI measured in the specific biome in which the location resides (step 9 in Fig. 1). Ground-based measurements of LAI are likely to overestimate the annual mean LAI in deciduous biomes, as LAI values of 0 are not usually reported. By combining remote-sensing–based estimates, which likely underestimate the actual leaf area, as well as ground-based measurements, which are likely overestimates, we make our estimate of the total leaf area more robust.

Estimating the Mass of Leaf per Unit Leaf Area. To estimate the characteristic mass of leaves per unit area, we rely on two methodologies. Our first methodology is based on a global database of plant traits. This database includes measurements of leaf dry weight per unit leaf area for ≈2,000 plant species (13). We calculate the geometric mean of leaf mass per unit leaf area across all species and arrive at an estimate of ≈100 g m^{-2}.

Our second methodology relies on a recently generated map of the distribution of plant traits (14). This map details the mass of leaves per unit leaf area in each location. We calculate the average of all the pixels in this map weighted by the area of leaves in each pixel. When estimating the total area of leaves, we generate two maps of leaf area, one based on remote sensing and the other based on field measurements. This generates two estimates for the average mass of leaves per unit leaf area, one average weighted by the remote sensing-based leaf area map and the other weighted by the field-measured leaf area map. We use the geometric mean of these two estimates, which is ≈100 g m^{-2}, as our best estimate for the mass of leaves per unit leaf area based on the trait map reported by Butler et al. (14).

As our best estimate for the characteristic leaf mass per unit leaf area, we use the geometric mean of the estimate based on the species database and

the estimate based on the plant trait maps (step 3 in Fig. 1). Our best estimate for the mass of leaves per unit leaf area is ≈100 g m^{-2}.

Estimating the Fraction of the Total Leaf Area in Woody and Herbaceous Plants. Our analysis generated an estimate of the total mass of leaves. Because different plant types, such as woody plants and C3 and C4 herbaceous plants, contain different characteristic amounts of Rubisco per leaf dry weight, we need to estimate the fraction of each plant type out of the global mass of leaves. To arrive at our estimate, we follow two steps: (*i*) estimating the fraction of the global leaf mass in woody plants and (*ii*) dividing the herbaceous leaf mass between C3 and C4 plants.

To estimate the fraction of the global leaf mass in woody plants, we use the same two methodologies that we use to estimate the total mass of leaves: one based on leaf mass fractions out of plant mass and the other based on leaf area. In each method, we divide the leaf mass estimate to woody or herbaceous leaves based on the biome in which the leaf is located. We define leaves as belonging to herbaceous plants if they are located in grasslands of croplands and consider the remaining biomes to belong to woody plants. Each methodology yields estimated fractions of the global leaf mass in woody and herbaceous plants. We use the geometric mean of the two estimates as our best estimate of the fraction of the global leaf mass in woody and herbaceous plants: 70% and 30%, respectively (https://bit.ly/2HQJ7qX).

We next estimate the fraction of the herbaceous leaf mass in C3 and C4 plants. To generate our estimate, we estimate the fraction of leaf mass in C4 plants in croplands and in nonagricultural herbaceous biomes separately . For croplands, we rely on crop distribution maps for ubiquitous C4 crops (sugarcane, maize, and sorghum) from published sources (47). We combine the distribution maps for all these crops and generate a map of the fraction of land that contains C4 crops. We then overlay this map with our maps estimating the global distribution of leaf mass and integrate across the entire plant to estimate the total mass of leaves in C4 crops. For nonagricultural herbaceous plants, we rely on a map of the global distribution of C4 plants (48). This map quantifies the fraction of land dominated by C4 plants. We exclude C4 plants in croplands from this map, because we calculated their mass separately in the previous section. We overlay the natural C4 plant distribution map with our maps estimating the global distribution of leaves and integrate across the entire planet to estimate the total mass of leaves in natural C4 herbaceous plants. The global distribution map of C4 plants indicates that C4 plants are also dominant in such biomes as savanna and shrubland, which contain woody plants. This is because C4 grasses can be found in the understory of those biomes.

Because it is not clear how much of the mass of leaves in these biomes is herbaceous, we generate two estimates based on the distribution of natural C4 plants, one including only grasslands and the other including also savannas and shrublands. In each case, we estimate the total mass of leaves of nonagricultural C4 plants. We use the geometric mean of the two estimates as our best estimate for the mass of leaves of natural C4 plants. We sum our estimates of the total mass of leaves of C4 crops and nonagricultural C4 plants to estimate the total mass of leaves of C4 plants. We divide the total leaf mass of C4 plants by our best estimate of the total leaf mass based on leaf area estimates to generate an estimate of the fraction of the total leaf mass that is in C4 plants. Overall, we estimate the ≈10% of the global leaf mass is in C4 plants, which leaves ≈20% of the total leaf mass in C3 herbaceous plants (https://bit.ly/2FWZgsL).

Uncertainty Analysis. Along with describing the procedures leading to the estimate of parameters used to derive the global mass and rate of Rubisco, we quantitatively survey the main sources of uncertainty associated with each parameter and calculate an uncertainty range for each. We follow the same methodology described by Bar-On et al. (6). We choose to report uncertainties as representing, to the best of our ability given the many constraints, what is equivalent to a 95% confidence interval for the estimate of the mean. Uncertainties reported in our analysis are multiplicative (fold change from the mean) and not additive (± change of the estimate). We chose to use multiplicative uncertainty because it is more robust to possible outliers in the underlying data and because it is a natural way to report uncertainty associated with the geometric mean of a sample. To estimate the total mass of terrestrial and marine Rubisco, we first estimate several quantities, such as the total mass of leaves and the Rubisco content in leaf mass (Fig. 1). Each of those quantities is calculated as a geometric mean of several data sources or of estimates from independent methods. We rely on the difference between independent methods for estimating the same quantity, or of the variability in the data on which we base our estimate, as the source for evaluating the uncertainty associated with our estimate. We then propagate the

uncertainty in each quantity to our final estimate of the total mass of terrestrial and marine Rubisco and their time average rates.

We calculate the uncertainty of each quantity around the geometric mean of the data used to estimate it (the data sources or estimates based on independent methods) by taking the logarithm of the values reported either within studies or from different studies. Taking the logarithm moves the values to log-space, where the SE is calculated (by dividing the SD by the square root of the number of values). We then multiply the SE by a factor of 1.96, which will give the 95% confidence interval if the transformed data are normally distributed. Finally, we exponentiate the result to get the multiplicative factor in linear space that represents the confidence interval (akin to a 95% confidence interval if the data are lognormally distributed). When data are ample, the uncertainty around the geometric mean will be low (as we base our uncertainty of SE in log-space). Nevertheless, this type of uncertainty does not consider the possibility that the distribution of values in the sample data does not represent the natural environment faithfully. To take this into account in our uncertainty projection, we generate an additional multiplicative uncertainty based on the SD and not on the SE in log-space. We consider the SE-based multiplicative uncertainty as an underestimate of the actual uncertainty and the SD-based multiplicative uncertainty as an overestimate of the actual uncertainty (because it does not include the decrease in uncertainty due to averaging). As our measure of uncertainty, we use the geometric mean of the SE-based multiplicative uncertainty and the SD-based multiplicative uncertainty. While this is not a standard statistical procedure, we consider it to be a reasonable compromise for deriving a robust uncertainty estimate.

1. Raven JA (2009) Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat Microb Ecol* 56:177–192.
2. Ellis RJ (1979) The most abundant protein in the world. *Trends Biochem Sci* 4:241–244.
3. Verzár F (1964) Aging of the collagen fiber. *International Review of Connective Tissue Research*, ed Hall DA (Elsevier, Amsterdam), pp 243–300.
4. Wang Z, et al. (2003) Total body protein: A new cellular level mass and distribution prediction model. *Am J Clin Nutr* 78:979–984.
5. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281:237–240.
6. Bar-On YM, Phillips R, Milo R (2018) The biomass distribution on Earth. *Proc Natl Acad Sci USA* 115:6506–6511.
7. Tang Z, et al. (2018) Patterns of plant carbon, nitrogen, and phosphorus concentration in relation to productivity in China's terrestrial ecosystems. *Proc Natl Acad Sci USA* 115:4033–4038.
8. Poorter H, et al. (2012) Biomass allocation to leaves, stems and roots: Meta-analyses of interspecific variation and environmental control. *New Phytol* 193:30–50.
9. Poorter H, et al. (2015) How does biomass distribution change with size and differ among species? An analysis for 1200 plant species from five continents. *New Phytol* 208:736–749.
10. Erb K-H, et al. (2018) Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature* 553:73–76.
11. Asner GP, Scurlock JMO, Hicke JA (2003) Global synthesis of leaf area index observations: Implications for ecological and remote sensing studies. *Glob Ecol Biogeogr* 12:191–205.
12. Liang S, Xiao Z (2012) Global land surface products: Leaf area index product data collection (1985–2010). Available at glcf.umd.edu/data/lai/. Accessed August 16, 2018.
13. Wright IJ, et al. (2004) The worldwide leaf economics spectrum. *Nature* 428:821–827.
14. Butler EE, et al. (2017) Mapping local and global variability in plant trait distributions. *Proc Natl Acad Sci USA* 114:E10937–E10946.
15. Onoda Y, et al. (2017) Physiological and structural tradeoffs underlying the leaf economics spectrum. *New Phytol* 214:1447–1463.
16. Sage RF, Pearcy RW, Seemann JR (1987) The nitrogen use efficiency of C(3) and C(4) plants, III: Leaf nitrogen effects on the activity of carboxylating enzymes in *Chenopodium album* (L.) and *Amaranthus retroflexus* (L.). *Plant Physiol* 85:355–359.
17. Tazoe Y, Noguchi K, Terashima I (2006) Effects of growth light and nitrogen nutrition on the organization of the photosynthetic apparatus in leaves of a C4 plant, *Amaranthus cruentus*. *Plant Cell Environ* 29:691–700.
18. Ghannoum O, et al. (2005) Faster Rubisco is the key to superior nitrogen-use efficiency in NADP-malic enzyme relative to NAD-malic enzyme C4 grasses. *Plant Physiol* 137:638–650.
19. Makino A, Sakuma H, Sudo E, Mae T (2003) Differences between maize and rice in N-use efficiency for photosynthesis and protein allocation. *Plant Cell Physiol* 44:952–956.
20. Kuehn GD, McFadden BA (1969) Ribulose 1,5-diphosphate carboxylase from *Hydrogenomonas eutropha* and *Hydrogenomonas facilis*, I: Purification, metallic ion requirements, inhibition, and kinetic constants. *Biochemistry* 8:2394–2402.
21. Niinemets U, et al. (2007) Do we underestimate the importance of leaf size in plant economics? Disproportional scaling of support costs within the spectrum of leaf physiognomy. *Ann Bot* 100:283–303.
22. Bleakley S, Hayes M (2017) Algal proteins: Extraction, application, and challenges concerning production. *Foods* 6:E33.
23. Raven JA (1991) Physiology of inorganic C acquisition and implications for resource use efficiency by marine phytoplankton: Relation to increased CO_2 and temperature. *Plant Cell Environ* 14:779–794.
24. Losh JL, Young JN, Morel FMM (2013) Rubisco is a small fraction of total protein in marine phytoplankton. *New Phytol* 198:52–58.
25. Zorz JK, et al. (2015) The RUBISCO to photosystem II ratio limits the maximum photosynthetic rate in picocyanobacteria. *Life (Basel)* 5:403–417.
26. McCarthy A, Rogers SP, Duffy SJ, Campbell DA (2012) Elevated carbon dioxide differentially alters the photophysiology of *Thalassiosira pseudonana* (Bacillariophyceae) and *Emiliania huxleyi* (Haptophyta). *J Phycol* 48:635–646.
27. Levitan O, et al. (2010) Combined effects of CO_2 and light on the N2-fixing cyanobacterium *Trichodesmium* IMS101: A mechanistic view. *Plant Physiol* 154:346–356.
28. Whitehead L, Long BM, Price GD, Badger MR (2014) Comparing the in vivo function of α-carboxysomes and β-carboxysomes in two model cyanobacteria. *Plant Physiol* 165:398–411.
29. Mackenzie TDB, Burns RA, Campbell DA (2004) Carbon status constrains light acclimation in the cyanobacterium *Synechococcus elongatus*. *Plant Physiol* 136:3301–3312.
30. Heureux AMC, et al. (2017) The role of Rubisco kinetics and pyrenoid morphology in shaping the CCM of haptophyte microalgae. *J Exp Bot* 68:3959–3969.
31. Brading P, Warner ME, Smith DJ, Suggett DJ (2013) Contrasting modes of inorganic carbon acquisition amongst *Symbiodinium* (Dinophyceae) phylotypes. *New Phytol* 200:432–442.
32. Young JN, Goldman JAL, Kranz SA, Tortell PD, Morel FMM (2015) Slow carboxylation of Rubisco constrains the rate of carbon fixation during Antarctic phytoplankton blooms. *New Phytol* 205:172–181.
33. Wu Y, Campbell DA, Irwin AJ, Suggett DJ, Finkel ZV (2014) Ocean acidification enhances the growth rate of larger diatoms. *Limnol Oceanogr* 59:1027–1034.
34. Davidi D, et al. (2016) Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements. *Proc Natl Acad Sci USA* 113:3401–3406.
35. Beer C, et al. (2010) Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science* 329:834–838.
36. Welp LR, et al. (2011) Interannual variability in the oxygen isotopes of atmospheric CO_2 driven by El Niño. *Nature* 477:579–582.
37. Ma Z, Cooper C, Kim H-J, Janick-Buckner D (2009) A study of rubisco through western blotting and tissue printing techniques. *CBE Life Sci Educ* 8:140–146.
38. Bender M, Orchardo J, Dickson M-L, Barber R, Lindley S (1999) In vitro O_2 fluxes compared with 14C production and other rate terms during the JGOFS Equatorial Pacific experiment. *Deep Sea Res Part I Oceanogr Res Pap* 46:637–654.
39. Duarte CM, Cebrián J (1996) The fate of marine autotrophic production. *Limnol Oceanogr* 41:1758–1766.
40. Halsey KH, Milligan AJ, Behrenfeld MJ (2010) Physiological optimization underlies growth rate-independent chlorophyll-specific gross and net primary production. *Photosynth Res* 103:125–137.
41. Juranek LW, Quay PD (2013) Using triple isotopes of dissolved oxygen to evaluate global marine productivity. *Annu Rev Mar Sci* 5:503–524.
42. Flamholz A, et al. (2018) Revisiting tradeoffs in Rubisco kinetic parameters. *bioRxiv* 10.1101/470021. Preprint, posted November 15, 2018.
43. Mori M, Schink S, Erickson DW, Gerland U, Hwa T (2017) Quantifying the benefit of a proteome reserve in fluctuating environments. *Nat Commun* 8:1225.
44. Raven JA, et al. (2017) The possible evolution and future of CO2-concentrating mechanisms. *J Exp Bot* 68:3701–3716.
45. Liang S, Li X, Wang J, eds (2012) Leaf area index. *Advanced Remote Sensing* (Academic, Boston), pp 347–381.
46. Song X-P, et al. (2018) Global land change from 1982 to 2016. *Nature* 560:639–643, and erratum (2018) 563:E26.
47. Monfreda C, Ramankutty N, Foley JA (2008) Farming the planet, 2: Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochem Cycles* 22:GB1022.
48. Still CJ, Berry JA, Collatz GJ, DeFries RS (2003) Global distribution of C3 and C4 vegetation: Carbon cycle implications. *Global Biogeochem Cycles* 17:6-1–6-14.

SYSTEMS BIOLOGY