Thermodynamic dissipation constrains metabolic versatility of unicellular growth

Tommaso Cossetto,[1,2] Jonathan Rodenfels,[2,3,4] and Pablo Sartori[1,4]

[1] *Instituto Gulbenkian de Ciência, Oeiras, Portugal*
[2] *The Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany*
[3] *Cluster of Excellence Physics of Life, TU Dresden, Dresden, Germany*
[4] *Correspondence: rodenfels@mpi-cbg.de and psartori@igc.gulbenkian.pt*

From deep hydrothermal vents to artificial laboratory conditions, unicellular organisms grow in almost any environment. This is due to metabolic versatility, the capacity of metabolism to extract energy and matter from vastly different chemical substrates. Since growth occurs far from thermodynamic equilibrium, the second law of thermodynamics, which establishes the unavoidability of energy dissipation, has long been believed to pose key constraints to life. Yet, such constraints remain largely unknown. Here, we integrate published data spanning eight decades of experiments on unicellular chemotrophic growth and compute the corresponding thermodynamic dissipation. Due to its wide span in chemical substrates and microbial species, this dataset samples the versatility of metabolism. We find two empirical thermodynamic rules that constrain metabolic versatility and thus unicellular growth. First, the amount of energy dissipation per unit of biomass grown is largely conserved across metabolic types and domains of life. Second, aerobic respiration exhibits a trade-off between dissipation and growth, which reflects in its high thermodynamic efficiency. Using tools from non-equilibrium thermodynamics, we identify the origin of these rules in the forces that drive and oppose growth. Our results show that dissipation imposes tight constrains to metabolic versatility far from equilibrium. They further suggest that evolutionary pressure on the thermodynamic features of metabolism may have driven the origin and persistence of aerobic respiration. Overall, this work establishes the central role of non-equilibrium thermodynamics in unicellular life.

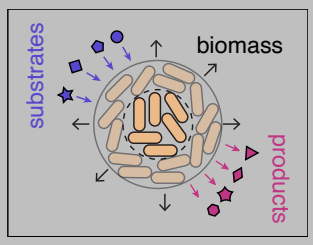
Introduction. A fundamental characteristic of life is its ubiquity across diverse environments. At the core of this ubiquity lies cellular metabolism, an interconnected network of chemical reactions that transforms environmental nutrients into the mass and energy that support growth. Metabolism is highly versatile, which is manifested by life's spread in environments spanning a wide range of nutrient types, temperatures, and other physico-chemical parameters. Because the matter and energy fluxes sustained by metabolism keep cells far from thermodynamic equilibrium, it has long been speculated that the second law of thermodynamics poses essential constraints to life [1–5]. However, far from equilibrium there is no general relation between fluxes and the free energy that these fluxes dissipate. Therefore, despite remarkable efforts [6–13], little is known about the interplay between thermodynamics and metabolism. Are there general quantitative rules that constrain the versatility of metabolism by the free energy it dissipates?

In the last decades, studies of metabolism have mirrored molecular biology's great advances and focused on genome-scale metabolic networks [11, 14–17]. In contrast, pioneering descriptions of metabolism dating as far back as a century ago were macroscopic [18], and summarized cellular growth by the substrates and products exchanged with the environment. This macroscopic approach identifies a *metabolic type* with a "macrochemical equation", e.g. $C_6H_{12}O_6 + O_2 + NH_3 \rightarrow \text{biomass} + CO_2 + H_2O$ for glucose respiration, and is sufficient for a thermodynamic description of steady unicellular growth [19–21] (see Box 1 for a summary of the formalism). Metabolic versatility is then embodied in the large variety of electron donors and acceptors characterizing metabolic types (glucose and oxygen respectively, in the previous example), as well as in the perturbations to growth within each type [22–25]. Applications of this macroscopic approach in biotechnology [26–29] and ecology [30–32] were aimed at predicting biomass yield, yet hinted to a fundamental relationship between thermodynamics and microbial metabolism. However, as metabolic versatility arises from both, evolutionary adaptations as well as plastic response to environmental changes, exploring its thermodynamic constraints necessitates experiments on cellular growth far beyond what a single research group can achieve.

We applied the framework of Box 1 to published measurements from the last eight decades, and created a database spanning 504 instances of unicellular growth with well-characterized non-equilibrium thermodynamics (SI A). This dataset covers more than one hundred metabolic types belonging to all domains of chemotrophic life, vastly expanding previous approaches [28, 30, 32]. Furthermore, it incorporates over 70 species of microbes and a variety of physico-chemical conditions for each type, thus constituting an ideal resource to investigate the interplay between thermodynamics and metabolic versatility.

We tested the thermodynamic accuracy of our dataset by validating three key principles (Fig. Box 1 and SI I). First, the second law of thermodynamics, which establishes the positivity of dissipation, is satisfied in more than 99% of the data that we collected. Second, the heat produced during growth closely matches the enthalpy exchanges, in agreement with Hess' law (Eq. 2), for all data that contained calorimetric heat measurements (98 instances). Third, chemical fluxes predicted by mass

Box 1 — Thermodynamics of cellular growth in a nutshell



Consider a colony of cells growing at a constant rate γ as a dilute solution of chemical constituents $i = 1, 2, \ldots$ Substrates ($i \in$ sb) are imported at fluxes $f_i^+ > 0$ (in units of mol per volume and time) and transformed into biomass. At steady state the composition of biomass as well as its density ρ_C (in units of carbon moles per volume) are constant. Products ($i \in$ pd) are exported at fluxes $-f_i^- < 0$. Singling out the flux of electron donor, $f_{\rm ed}^+$, allows to define the biomass yield, $y \equiv \gamma\rho_C/f_{\rm ed}^+$, in carbon moles per moles.

This non-equilibrium process is constrained by the second law of thermodynamics:

$$\dot{s}_{\rm prod} = -\frac{\dot{q}}{T} - \sum_{i\in{\rm sb}} f_i^+ s_i + \sum_{i\in{\rm pd}} f_i^- s_i + \gamma s \geq 0\,, \quad (1)$$

where $\dot{s}_{\rm prod}$ is the rate at which entropy is produced, $\dot{q}$ the rate at which heat is produced (or absorbed, if positive), T the temperature of the media, s_i the entropy of the i-th constituent, and s the entropy density of biomass.

Energy conservation during growth is expressed by the first law of thermodynamics:

$$\dot{q} = -\sum_{i\in{\rm sb}} f_i^+ h_i + \sum_{i\in{\rm pd}} f_i^- h_i + \gamma h \quad , \quad (2)$$

where h is the biomass enthalpy and h_i the enthalpy of the corresponding chemical species. This equivalence between heat and enthalpy is also known as Hess' law.
Putting together the expressions of the first and second law, and using the definition of chemical potential, $\mu_i \equiv h_i - Ts_i$, we arrive at

$$T\dot{s}_{\rm prod} = -\dot{g}_{\rm diss} = \sum_{i\in{\rm sb}} f_i^+\mu_i - \sum_{i\in{\rm pd}} f_i^-\mu_i - \gamma g \geq 0, \quad (3)$$

which decomposes the total rate of free energy dissipation, $\dot{g}_{\rm diss} \leq 0$, in terms of free energy influx of substrates, $i \in$ sb, free energy outflux of products, $i \in$ pd, and free energy flux due to biomass synthesis, γg. Normalizing by flux of electron donor gives $\sigma \equiv -\dot{g}_{\rm diss}/f_{\rm ed}^+$, to which we refer in the text as dissipation. The free energies imported and exported through exchanges of substrates and products are defined as $\sigma_{\rm sb} \equiv -\sum_{i\in{\rm sb}} f_i^+\mu_i/f_{\rm ed}^+$ and $\sigma_{\rm pd} \equiv -\sum_{i\in{\rm pd}} f_i^-\mu_i/f_{\rm ed}^+$, respectively.

Taken together, the above constitutes a minimal thermodynamic framework that allows us to infer non-equilibrium thermodynamic properties from growth measurements. We validated this framework on our dataset in the figure below and SI I.

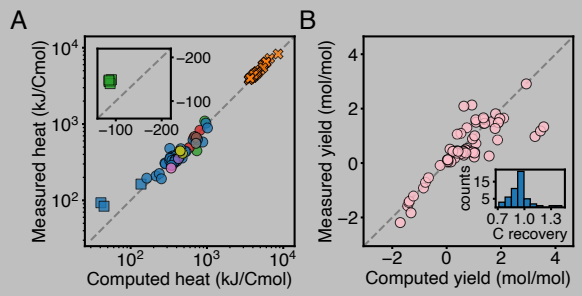


FIG. Box 1 *Thermodynamic validation of the dataset.* **A.** Comparison of (minus) the enthalpy balance with (minus) the heat measured. Hess' law, Eq. (2), is approximately satisfied in 98 entries for which calorimetric heat measurements are available (inset for endothermic data). **B.** Element conservation was used to predict 56 yields that were independently measured. Inset shows 51 measurements of carbon recovery, i.e. global carbon balance of the experiment. Dashed lines are drawn as reference and are not a fit.

conservation closely matched those measured (69 fluxes from 56 instances). Having established the quality of our dataset, we now investigate the interplay between physiological and thermodynamic growth parameters.

Variability in yield and dissipation highlights metabolic versatility. A key physiological parameter of growth is the biomass yield, y, which quantifies the amount of biomass produced per electron donor consumed (Box 1 and SI B 2). We observed wide variability in y within our dataset (Fig. 1A). Yield variability across metabolic types (full list in Fig. SI1) spans three orders of magnitude. In contrast, variability within types spans under a factor of two. This variability reflects changes in experimental conditions (such as temperature, dilution rate, or pH), unicellular species, and experimental uncertainty. The lowest yield, $y \sim 10^{-2}$ Cmol/mol, corresponds to autotrophic methanogenesis, an ancient archeal metabolic type that synthesizes biomass and methane from CO_2 and molecular hydrogen in anoxic environments [33–35]. The highest yields, $y \approx 10$ Cmol/mol, are produced by the respiration

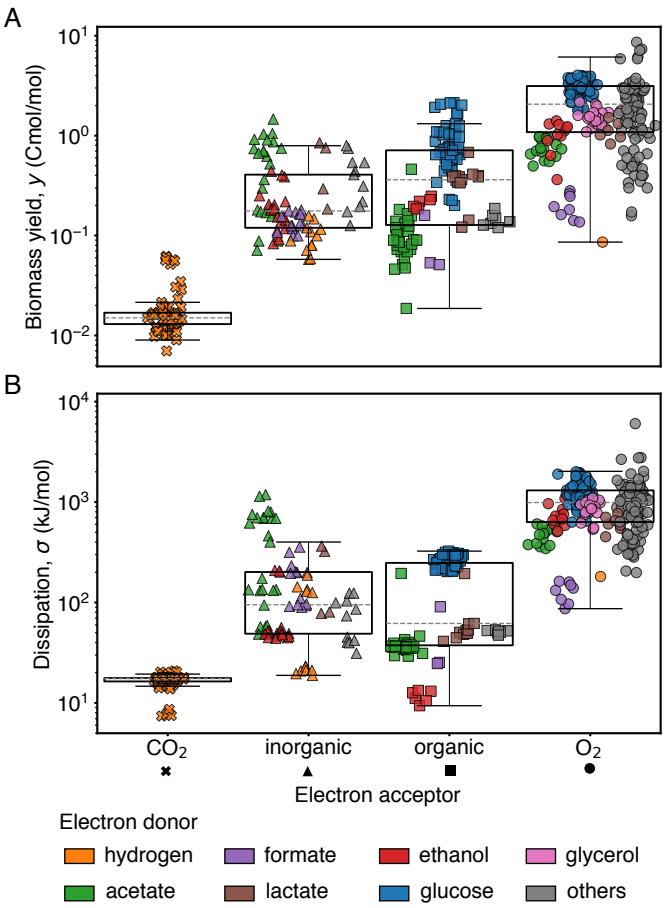


FIG. 1: *Variability of yield and dissipation in unicellular growth.* **A.** The biomass yield y, biomass produced per electron donor consumed, spans across three orders of magnitude. The data includes over one hundred different metabolic types, which we grouped by class of electron acceptor (marker) and electron donor (color for most frequents). This same color-marker scheme is preserved throughout the paper. In the plots, the boxes delimit the first and third quartiles, the dashed lines represent the median and the whiskers are drawn at the farthest data point within 1.5 times the interquartile range from the box. We use the same convention throughout the paper. **B.** The dissipation σ, free energy dissipated per electron donor consumed, displays a variability that spans a range comparable to that of yield.

of disaccharides, such as sucrose and lactose, reaching values close to their carbon limit of 12 Cmol/mol (the number of carbon atoms present in the electron donor).

To account for the thermodynamics of growth, we quantify the free energy dissipation per electron donor consumed, σ, hereafter simply referred to as dissipation (Box 1 and SI B). As in the case of yields, we find that σ varies substantially, also spanning three orders of magnitude. It is similarly bounded by autotrophic methanogenesis, $\sigma \approx 50\,\mathrm{kJ/mol}$, which synthesises biomass from the most oxidised form of carbon, CO_2 [36]; and respiration of sugars, $\sigma \approx 6\,\mathrm{MJ/mol}$, which are highly reduced electron donors. In contrast, no clear distinction in y nor σ was found between archaea, prokarya, and eukarya beyond the distinction due to changes in metabolic type (Fig. SI3). Similarly, temperatures and pH had no systematic effect on y (Fig. SI4). Thus, the data demonstrates that variability in yield and dissipation are dominated by changes across metabolic types.

The dissipative cost of growth is conserved across metabolic types. To investigate the relationship between thermodynamics and physiology of unicellular growth, we plot the dissipation against the yield (Fig. 2A and Fig. SI7). Two salient features become apparent: first, yield and dissipation are strongly correlated over the span of three orders of magnitude; and second, data corresponding to the same metabolic type cluster together in quasi-linear spaces following distinctive trends. These two observations suggest separating the dissipation into a mean contribution specific to a particular metabolic type, $\bar{\sigma}$, and a small deviation due to within type variability, $\delta\sigma$, and analogously for the yield, $\bar{y}$ and δy (Fig. 2A inset and SI C).

Plotting the mean dissipation against the mean yield across types, we find that they are linearly proportional (Fig. 2B). This correlation persists when the axis are expressed per gram or per carbon mole, albeit over a smaller range (Fig SI10). Therefore, we define the coefficient

$$\alpha \equiv \bar{\sigma}/\bar{y} \quad , \tag{4}$$

which quantifies the *dissipative cost of growth* for a certain metabolic type. This cost, which by the second law must be (and is) positive, measures the loss of free energy per biomass grown. The linear relation implies that α is conserved across different metabolic types. The cost has a median value $\alpha \approx 500\,\mathrm{kJ/Cmol}$ or 8 ATP equivalents per carbon atom fixed (Table I SI D), and an interquartile range of 350 kJ/Cmol (Fig. 3 and corresponding section for a discussion on variability). This value of α is consistent with a previously hypothesized constant used to predict biomass yields in the bioengineering literature [28]. However, it is larger than biophysical estimates [37, 38] (SI D), suggesting that bookkeeping of intracellular processes leaves out significant contributions to dissipation. The cost α is conserved irrespective of the species, the domain of life, the growth rate, or other physico-chemical variables (Fig. SI5 and SI4), presenting an overall constraint to metabolic versatility.

Deviations in yield and dissipation within metabolic types. Within a given type, the deviation $\delta\sigma$ relates linearly to δy. This is a direct consequence of stoichiometric constraints, namely the fact that chemical reactions preserve elements (SI B 3). For example, the carbon atoms absorbed from nutrients are either fixed into biomass or expelled in products, which constraints the chemical exchanges. In contrast to the single trend

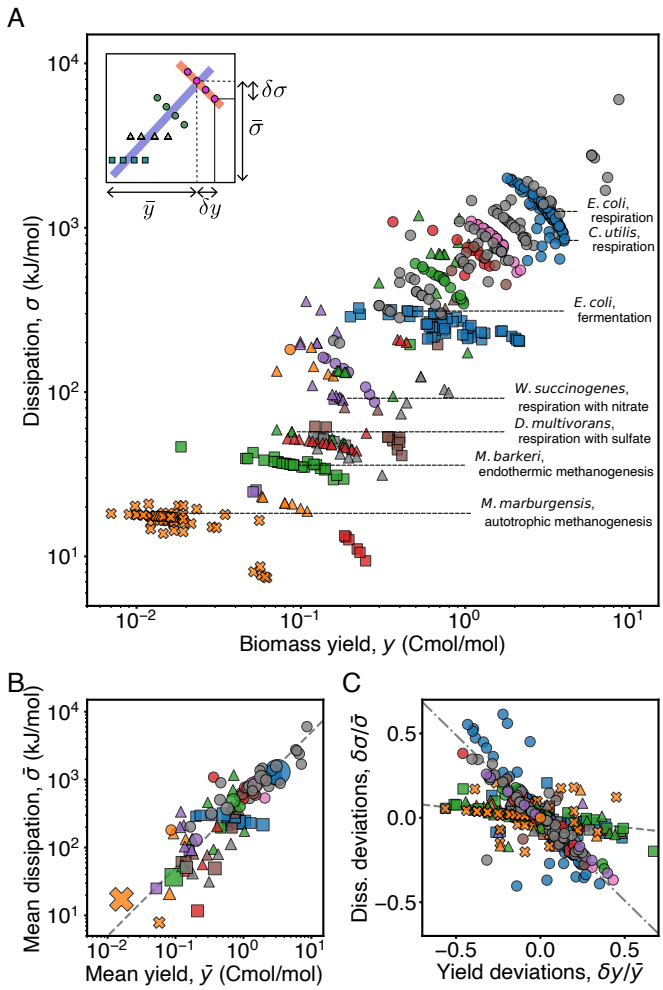


FIG. 2: *Conserved thermodynamic cost of biomass across metabolic types* **A.** Scatter plot of the dissipation, σ, versus the yield, y, for the whole dataset. The variability across types results in a strong correlation of the data over three orders of magnitude (see Fig. SI10 for different units). Data from the same type cluster together, in quasi-linear spaces. Inset depicts the notions of average dissipation/yield and corresponding deviations. **B.** Mean dissipation, $\bar{\sigma}$, plotted against yield, $\bar{y}$, for each type. The marker size represents the amount of data in each type. The dashed line in the background corresponds to, $\bar{\sigma} = \alpha\bar{y}$, fixing $\alpha = 500\,\mathrm{kJ/Cmol}$, which approximates well the data. **C.** Normalized deviations in dissipation, $\delta\sigma/\bar{\sigma}$, are plotted against normalized deviations in yield, $\delta y/\bar{y}$, for each metabolic type. Dashed and dotted lines correspond to linear regressions for anaerobic and aerobic types with slopes -0.11 ± 0.01 and -0.98 ± 0.03, respectively.

observed between $\bar{\sigma}$ and $\bar{y}$ across types, the slope relating $\delta\sigma$ to δy in each type does vary substantially across types. Aerobic types have a steep negative slope while anaerobic types have a shallow negative slope (Fig. 2C and SI9). To quantify the diversity of trends we define a coefficient, β, by the following linear regression within metabolic type,

$$\delta\sigma \approx \beta\delta y \quad . \tag{5}$$

The coefficient β accounts for how deviations in dissipation are adjusted relative to yield within a type. We find that in all instances $\beta < 0$, and thus increases in yield are accompanied by a reduction in dissipation. For anaerobic types we find $\beta \approx -100\,\mathrm{kJ/Cmol}$, and thus the yield can be increased with little effect on the dissipation (Fig. 2C, dashed line). Instead, aerobic types show $\beta \approx -500\,\mathrm{kJ/Cmol}$, and thus there is a significant trade-off between yield and dissipation: relative increases in yield of 50% result in comparable reduction in dissipation (Fig. 2C, dotted line). In summary, we find that while the dissipative cost of growth, quantified by α, is conserved across metabolic types, the deviations within type, quantified by β, distinguish aerobic from anaerobic types.

Cost is conserved despite variability of free energy exchanges. To assess the extent of conservation in the cost, we decomposed it into three terms: free energy absorbed with substrates; released with products; and stored in biomass (Fig. 3A and Box 1). Each of these terms provides a reference for cost variability.

Across metabolic types, the free energy absorbed per biomass produced is high and variable, with a median value of 900 kJ/Cmol and an interquartile range of 2800 kJ/Cmol (Fig. 3A, second box). Similarly, the free energy released with the products is also high and variable, with a median value of 1500 kJ/Cmol and an interquartile range of 2700 kJ/Cmol (Fig. 3A, third box). In contrast, the free energy content of biomass is roughly one order of magnitude lower (Fig. 3A third box and SI H 2). We conclude that the cost α is dominated by the free energy exchanged through substrates and products. Remarkably, while these exchanges are large and variable, α is lower and exhibits a comparatively low variability: median 490 kJ/Cmol and interquartile range 350 kJ/Cmol (Fig. 3A first box, and inset for histograms). Therefore, the cost is conserved over a wide range of free energies imported and exported (Fig. 3B).

The reason for the low variability in α is that, while the free energies absorbed and released are both large variable, they are strongly correlated and compensate each other (Fig. 3C). Such compensation arises from an interplay between element conservation and the free energies of chemicals, which we disentangle in the following sections. In summary, while the free energy flowing through cells varies significantly across metabolic types, the compensation of the influx and outflux of free energy results in a conserved dissipative cost of growth.

Thermodynamic implications of the anaerobic/aerobic transition. Due to the versatility of metabolism, many electron donors can be metabolized both aerobically and anaerobically. To investigate the thermodynamics of this phenomenon, we grouped

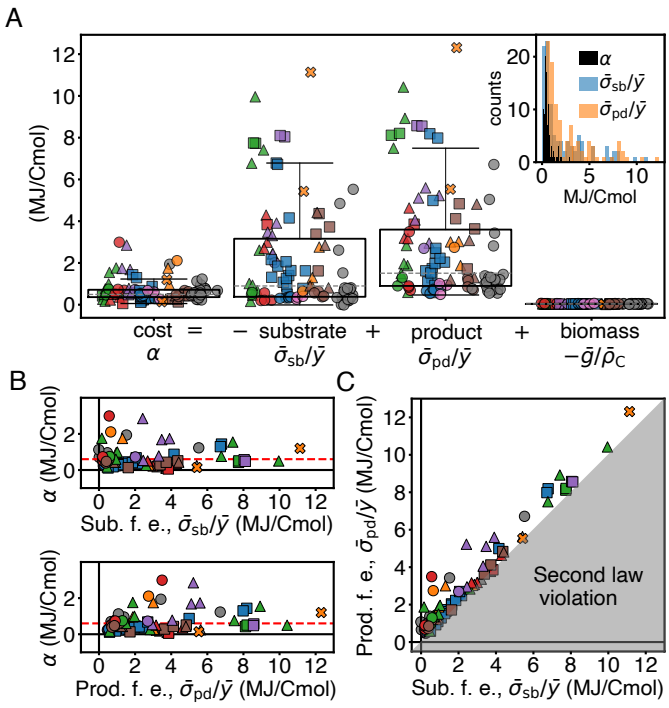


FIG. 3: *Decomposition of the dissipative cost* α. **A.** The dissipative cost, α, is decomposed into the free energy lost due to substrate consumption, $-\bar{\sigma}_{\rm sb}/\bar{y}$, the free energy created due to product formation, $-\bar{\sigma}_{\rm pd}/\bar{y}$, and stored into biomass, $\bar{g}/\bar{\rho}_{\rm C}$, all per unit biomass synthesized. The contribution from free energy storage in biomass is small and its variability is not visible (SI H 2). Substrate and product free energy fluxes are large and variable, yet compensate each other to result in a comparatively small and conserved cost of biomass. **Inset.** Histograms of α, $\bar{\sigma}_{\rm sb}/\bar{y}$, and $\bar{\sigma}_{\rm pd}/\bar{y}$, showing skewness and small variability of α. **B.** The cost as a function of the influx of free energy of substrates (top), and of the outflux of free energy of products (bottom). Compared to the large variability in the free energy influx and outflux, the cost is largely conserved. **C.** Free energy outflux of metabolic products vs free energy influx of substrates. The data covers only a small linear space within the physically realizable region (unshaded area), indicating that the cost of growth is much smaller than influx/outflux of free energy. The shaded area corresponds to violation of the second law (a small offset due to biomass free energy is not visible).

metabolic types based on their electron donor and acceptor, and identified 30 pairs of groups that use the same electron donor both aerobically and anaerobically. For example, one pair consists of a group of metabolic types respiring ethanol with oxygen, and of a group of types respiring ethanol with sulfate. Aerobic groups are exclusively respiratory, using oxygen as electron acceptor, while anaerobic groups include both, fermentation as well as anaerobic respiration, and thus use a variety of electron acceptors other than oxygen. We find that, while in each pair yield and dissipation are generally higher for aerobic groups (Figs. 4A and B), the change in cost across this transition exhibits no clear trend (Fig. 4C). Furthermore, this is true when the transition corresponds to different species (e.g. *C. utilis* respiring ethanol with oxygen or *D. propionicus* respiring ethanol with sulfate), as well as when it corresponds to plastic changes of the same species (e.g. *E. coli* respiring or fermenting glucose). Therefore, irrespective of whether it is evolutionary or plastic, a transition from an anaerobic to an aerobic metabolic type that preserves electron donor increases yield and dissipation proportionally, satisfying the constraint of conserved cost, α.

Instead, the thermodynamic signature of such transitions lies in the coefficient β, which determines how dissipation is adjusted relative to yield within metabolic type. Figures 4E-G display scatter plots of the dissipation and yield deviations for three anaerobic/aerobic pairs. As one can see, for the same electron donor aerobic types exhibit a marked trade-off between dissipation and yield, absent in non-aerobic types. This difference in β between anaerobic/aerobic types extends to almost all pairs (Fig. 4D). Therefore, a transition from anaerobic to aerobic metabolism processing the same electron donor allows to trade dissipation for yield.

We identified a few outliers to this rule: metabolic types performing anaerobic respiration with electron acceptors such as nitrate, among few others (Fig. 4D and H). Those outliers harbor a β similar to the value of aerobic types. Overall, we conclude that the thermodynamic implication of a transition from an anaerobic to an aerobic type lies predominantly in the emergence of a dissipation-yield trade-off.

Opposing thermodynamic forces control dissipation and efficiency. So far we have established that the dissipative cost of growth is conserved across types, and that a trade-off between dissipation and yield deviations accompanies aerobic respiration. We now characterize these two metabolic signatures following the formalism of thermodynamics far from equilibrium. First, we note that cellular dissipation is given by a sum of products between chemical fluxes and their corresponding free energies (Eq. 3 in Box 1). However, as already anticipated, these fluxes are not independent due to constraints imposed by the stoichiometry of the metabolic type (i.e. element conservation).

To disentangle these constraints, we decompose dissipation into independent contributions (Fig. 5A). Each contribution is the product of an independent flux f_i (e.g. biomass, electron donor, etc.) coupled to a generalized thermodynamic force r_i. To build intuition, let us show how this framework applies to a simple chemical reaction: ATP hydrolysis. In this case, the dissipation is given by $T\dot{s}_{\rm prod} = f^+_{\rm ADP}\mu_{\rm ADP} + f^+_{\rm Pi}\mu_{\rm Pi} - f^-_{\rm ATP}\mu_{\rm ATP}$ and the stoichiometric constraint is $f^-_{\rm ATP} = f^+_{\rm ADP} = f^+_{\rm Pi}$. Arbitrarily choosing $f^-_{\rm ATP}$ as independent flux, we can write $T\dot{s}_{\rm prod} = f^-_{\rm ATP} r_{\rm ATP}$, with $r_{\rm ATP} = \mu_{\rm ADP} + \mu_{\rm Pi} - \mu_{\rm ATP}$ the free energy change of the reaction which contains the free energies of the dependent chemicals, ADP and Pi. Sim-

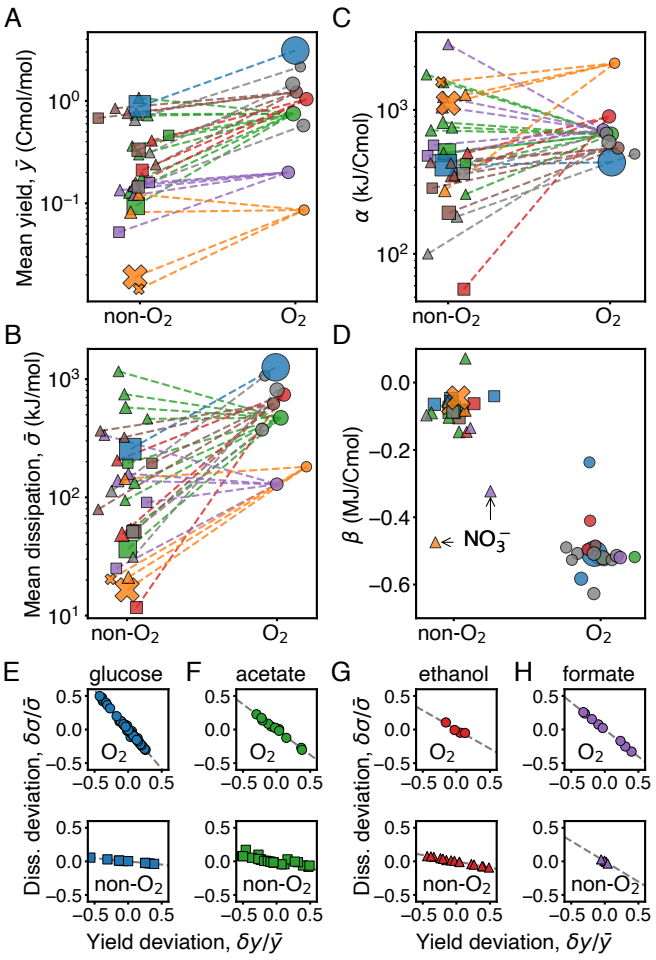


FIG. 4: *Effect of oxygen on the dissipation-yield relation.* **A.** Mean biomass yield, $\bar{y}$, for 30 pairs of groups of metabolic types that use the same electron donor anaerobically (non-oxygen) and aerobically (oxygen). A dashed line is drawn connecting each pair. The marker size represents the amount of data in each group. For most pairs, the aerobic group has a higher yield than the anaerobic counterpart. **B.** The same as A but for dissipation, $\bar{\sigma}$, showing the same trend. **C.** The same as A and B but for the cost, α, which is more conserved within pairs and shows no clear trend. **D.** The same as A, B and C, but for the deviation coefficient, β. For aerobic types β is large and negative, indicating a trade-off between yield and dissipation. Two exceptions correspond to anaerobic respiration of nitrate NO_3^-: reduction to ammonia (purple triangle) shows a value of β intermediate to those of anaerobic and aerobic types, reduction to nitrogen (orange triangle) shows a value of β indistinguishable from that of aerobic types. **E-H.** Examples of deviation of yield and dissipation for three pairs, which agree with the trend in Fig. 2C. Panel H shows an exception, in which the anaerobic electron acceptor nitrate produces a steep slope similar to aerobic types.

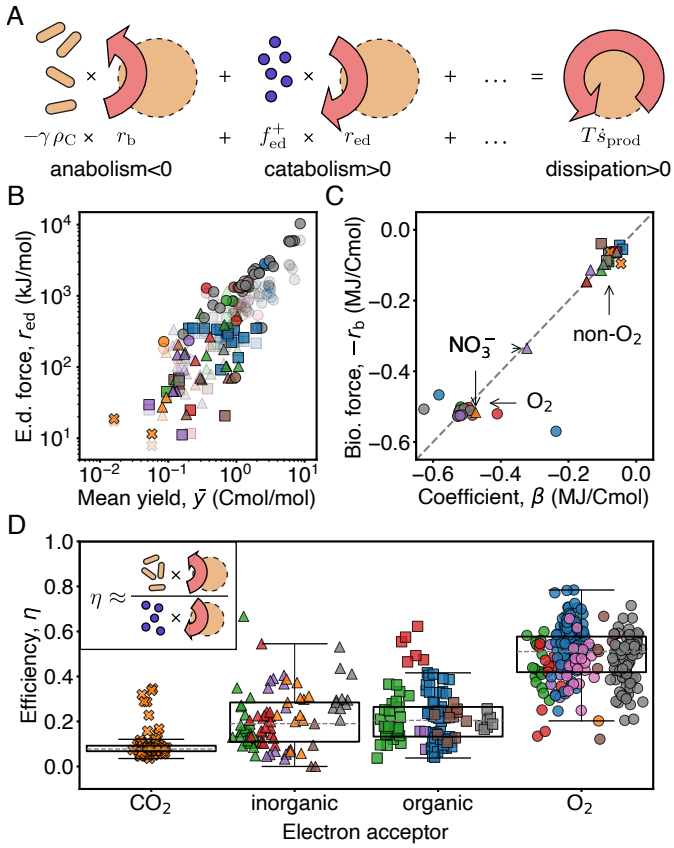


FIG. 5: *Independent thermodynamic forces control the dissipative cost and the efficiency of growth.* **A.** The total free energy dissipation can be decomposed in a sum of independent terms, each corresponding to the product of an independent flux with a generalized thermodynamic force. Schematic shows the dominant ones. **B.** The electron donor force, r_{ed}, for each type is plotted against the biomass yield, $\bar{y}$, showing a strong correlation. In a transparent shade, the dissipation $\bar{\sigma}$ is plotted against the yield, as in Fig. 2B. The proximity between both shades shows that r_{ed} is a good predictor for $\bar{\sigma}$. **C.** The biomass force, r_b, for each type is plotted against the coefficient β. Both terms are essentially equal (dashed line is not a fit). A clear distinction exists between aerobic and anaerobic types, with outlier anaerobes located between them. **D.** The thermodynamic efficiency of growth for all data is shown. Aerobic metabolisms are singled out by their high thermodynamic efficiency. Inset shows a pictorial definition of the efficiency as a ratio of the dissipative terms corresponding to anabolism and catabolism.

ilarly, in a metabolic type the thermodynamic force r_i generalizes the chemical potential difference of a chemical reaction to the fluxes sustaining growth. Each force is itself a linear combination containing the chemical potentials of dependent species, with coefficients determined by the stoichiometry of the metabolic type [39] (SI E 1).

We identify two dominant thermodynamic forces of metabolism: the electron donor force, r_{ed}, coupled to the influx of electron donor molecules, and the biomass force, $-r_b$, that couples to biomass synthesis. Therefore, these

two forces can be heuristically associated with catabolism and anabolism, respectively.

We find that r_{ed} is always positive and approximates well the mean dissipation, i.e. $\bar{\sigma} \approx r_{\mathrm{ed}} > 0$ (Fig. 5B). As a consequence, r_{ed} can be understood as the driving force of growth. The conserved cost α arises from a linear relationship between this force and the biomass yield across metabolic types, which extends and systematizes previous observations for "catabolic reactions" [26, 29]. In contrast, we find that the anabolic force opposes dissipation, $-r_{\mathrm{b}} < 0$, and is approximately equal to the coefficient controlling deviations in yield, i.e. $\beta \approx -r_{\mathrm{b}}$ (Fig. 5C and SI E). As a consequence, the anabolic force can be understood as an opposing force to growth. The equivalence between r_{b} and β allows to trace the trade-off between dissipation and yield in aerobic metabolism to the large and negative free energy content of carbon dioxide (SI E 4). Therefore, dissipation is dominated by two opposing forces, catabolic and anabolic, which control the relationship between dissipation and yield within and across metabolic types.

The decomposition of dissipation into independent terms further allows to formulate a well defined thermodynamic efficiency, η [40]. Intuitively, this efficiency measures the ratio of free energy produced and consumed by independent fluxes. In our case, the independent fluxes are largely dominated by biomass production and electron donor consumption, and so $\eta \approx y r_{\mathrm{b}}/r_{\mathrm{ed}}$, which can be understood as a ratio of anabolic to catabolic dissipation (SI E 3). Figure 5D shows that aerobic types have a much larger efficiency than anaerobic types. This is because, due to the relationships established before, we have that $\eta \approx \beta/\alpha$, and the efficiency measures the ratio between variations across and within metabolic types. Therefore, despite aerobic biomass synthesis being opposed by a greater thermodynamic force, aerobic metabolism is more efficient. We remark that the efficiency η that we defined is different from the common notion of ATP yield, which counts the net number of ATP molecules produced per electron donor catabolized (e.g. ≈ 32 and 2 for glucose respiration and fermentation, respectively).

Discussion. The metabolic versatility characterizing unicellular growth is thermodynamically constrained by two empirical rules. First, the non-equilibrium cost of biomass, α, is largely conserved across adaptive and plastic changes in metabolism; and second, aerobic metabolism results in a trade-off between deviations in dissipation and yield, with coefficient β. These rules are controlled by two separate thermodynamic forces, r_{ed} and r_{b}, associated with catabolism and anabolism. The opposing signs of the catabolic and anabolic terms imply that cellular growth occurs through free energy transduction, akin to the transformation of chemical energy into mechanical work in molecular motors [41]. The decomposition into thermodynamic forces allows to define the efficiency of unicellular growth, dominated by the ratio of β and α, which is consistently higher for aerobic metabolic types.

Deviations from the two thermodynamic rules single out exceptional metabolic types, with evolutionary implications. For example, hydrogenotrophic methanogenesis in archeans (orange crosses in Fig. 2) has a cost two-fold higher than the typical value [42], deviating from the first rule. Likewise, anaerobic respiration of nitrate can have an anabolic opposing force and thermodynamic efficiency comparable to the aerobic one, and thus deviates from the second rule. Interestingly, these metabolic types play a central role in theories of origin of life, eukaryogenesis and the evolution of aerobic metabolisms [22, 35, 43, 44]. The thermodynamic rules and their deviations suggest that thermodynamic features of metabolism may have played a central role in the evolution of unicellular growth.

Our data collection and processing builds on decades of research on the energetics of microbial growth. Early studies aimed at controlling microbial processes in bioreactors by predicting the biomass yield from electron donor properties, such as its degree of reduction [7]. Later meta-analysis suggested a more complex empirical relation between electron donor properties and the dissipative cost of growth [26]; a correlation between the free energy of the "catabolic reaction" and biomass yield [21, 29]; and thermodynamic corrections due to non standard conditions [30]. Remarkably, it was also proposed that a constant value of the dissipative cost of growth results in a good predictor of biomass yields. Our work substantially advances these efforts by crucially distinguishing variability within and across metabolic types, and by applying a general non-equilibrium thermodynamics framework to an extended and validated dataset. For example, r_{ed} generalizes the notion of "catabolic dissipation" to complex metabolic types; r_{b} provides a definition of "anabolic free energy" not related to intracellular processes; and the thermodynamic efficiency η is bounded by the second law for all data. Therefore, our systematic analysis reconciles many previous observations with varying degrees of support into two empirical thermodynamic rules that constrain unicellular growth across the domains of life.

Overall, our work constitutes a step towards connecting non-equilibrium thermodynamics with cellular physiology [24, 46, 47]. The progress we made in this direction has implications in broadly different fields. For example, our dataset can be used as the basis for models of biogeochemical cycles [48], but it can also guide research linking mammalian metabolism, growth, and thermodynamics in healthy and diseased systems [49, 50]. Furthermore, as our dataset is open, it can be augmented to include other metabolic types, such as photosynthesis, thus expanding ecological studies of energetics. Finally, while we made no reference to intracellular processes, our approach can be bridged with thermodynamics of metabolic or signalling networks [11, 51]. Overall, the two empirical rules we established constitute a long sought-after

connection between evolution, metabolism, and thermodynamics.

Acknowledgements

All authors thank G. Falasco, R. Rao, F. Avanzini, V. Bocan, S. Pigolotti, as well as several members of the IGC community for feedback on the manuscript. P.S. would like to thank J. Howard and S. Leibler for insightful discussions. J.R. was supported by core funding form the Max-Planck-Society. This project has received funding from the European Research Council (ERC) under the Horizon 2020 research and innovation programme (Grant agreement No. 949811 — EnBioSys) to J.R.; from laCaixa (LCF/BQ/PI21/11830032) grant and core funding from the Gulbenkian Foundation to P.S.

[1] Alfred J Lotka. Contribution to the energetics of evolution. *Proceedings of the National Academy of Sciences*, 8(6):147–151, 1922.

[2] Erwin Schrödinger. *What is life?: With mind and matter and autobiographical sketches*. Cambridge university press, 2012.

[3] William Martin, John Baross, Deborah Kelley, and Michael J Russell. Hydrothermal vents and the origin of life. *Nature Reviews Microbiology*, 6(11):805–814, 2008.

[4] Nick Lane and William Martin. The energetics of genome complexity. *Nature*, 467(7318):929–934, 2010.

[5] Sergio A Muñoz-Gómez. Energetics and evolution of anaerobic microbial eukaryotes. *Nature Microbiology*, pages 1–7, 2023.

[6] Max Kleiber. Body size and metabolic rate. *Physiological reviews*, 27(4):511–541, 1947.

[7] Joop A Roels. *Energetics and kinetics in biotechnology*. Elsevier biomedical press, 1983.

[8] Urs von Stockar. Biothermodynamics of live cells: a tool for biotechnology and biochemical engineering. 35(4): 415–475, 2010.

[9] Avi Flamholz, Elad Noor, Arren Bar-Even, Wolfram Liebermeister, and Ron Milo. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences*, 110(24): 10039–10044, 2013.

[10] James H Brown, James F Gillooly, Andrew P Allen, Van M Savage, and Geoffrey B West. Toward a metabolic theory of ecology. *Ecology*, 85(7):1771–1789, 2004.

[11] Bastian Niebel, Simeon Leupold, and Matthias Heinemann. An upper limit on gibbs energy dissipation governs cellular metabolism. *Nature Metabolism*, 1(1):125–132, 2019.

[12] Jonathan Rodenfels, Karla M Neugebauer, and Jonathon Howard. Heat oscillations driven by the embryonic cell cycle reveal the energetic costs of signaling. *Developmental cell*, 48(5):646–658, 2019.

[13] Jonathan Rodenfels, Pablo Sartori, Stefan Golfier, Kartikeya Nagendra, Karla M Neugebauer, and Jonathon Howard. Contribution of increasing plasma membrane to the energetic cost of early zebrafish embryogenesis. *Molecular biology of the cell*, 31(7):520–526, 2020.

[14] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature biotechnology*, 28 (3):245–248, 2010.

[15] Daniel A Beard, Shou-dan Liang, and Hong Qian. Energy balance for analysis of complex metabolic networks. *Biophysical journal*, 83(1):79–86, 2002.

[16] Amit Varma and Bernhard O Palsson. Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/technology*, 12(10):994–998, 1994.

[17] Matteo Mori, Chuankai Cheng, Brian R Taylor, Hiroyuki Okano, and Terence Hwa. Functional decomposition of metabolism allows a system-level quantification of fluxes and protein allocation towards specific metabolic functions. *Nature Communications*, 14(1):4161, 2023.

[18] H Tamiya. Material and energy balances of biological synthesis. *Actualities Scientifiques et Industrielles*, 214, 1935.

[19] H. J. Morowitz. *Foundations of bioenergetics*. New York Academic Press, 1978.

[20] J. A. Roels. Application of macroscopic principles to microbial metabolism. *Biotechnology and Bioengineering*, 22(12):2457–2514, 1980.

[21] Edwin H Battley. *Energetics of microbial growth*. Wiley, 1987.

[22] Nicole R Buan. Methanogens: pushing the boundaries of biology. *Emerging Topics in Life Sciences*, 2(4):629–646, 2018.

[23] Gerard Muyzer and Alfons JM Stams. The ecology and biotechnology of sulphate-reducing bacteria. *Nature reviews microbiology*, 6(6):441–454, 2008.

[24] Jacques Monod. The growth of bacterial cultures. *Annual review of microbiology*, 3(1):371–394, 1949.

[25] Matthew Scott and Terence Hwa. Shaping bacterial gene expression by physiological and proteome allocation constraints. *Nature Reviews Microbiology*, 21(5):327–342, 2023.

[26] JJ Heijnen and JP Van Dijken. In search of a thermodynamic description of biomass yields for the chemotrophic growth of microorganisms. *Biotechnology and Bioengineering*, 39(8):833–858, 1992.

[27] Urs Von Stockar, Thomas Maskow, Jingsong Liu, Ian W Marison, and Rodrigo Patino. Thermodynamics of microbial growth and metabolism: an analysis of the current situation. *Journal of Biotechnology*, 121(4):517–533, 2006.

[28] J-S Liu, V Vojinović, Rodrigo Patino, Th Maskow, and Urs von Stockar. A comparison of various gibbs energy dissipation correlations for predicting microbial growth yields. *Thermochimica Acta*, 458(1-2):38–46, 2007.

[29] Eric E Roden and Qusheng Jin. Thermodynamics of microbial growth coupled to metabolism of glucose, ethanol, short-chain organic acids, and hydrogen. *Applied and environmental microbiology*, 77(5):1907–1909, 2011.

[30] Christina M Smeaton and Philippe Van Cappellen. Gibbs energy dynamic yield method (gedym): Predicting microbial growth yields under energy-limiting conditions. *Geochimica et Cosmochimica Acta*, 241:1–16, 2018.

[31] Robbert Kleerebezem and Mark CM Van Loosdrecht. A generalized method for thermodynamic state analysis of environmental systems. *Critical Reviews in Environmental Science and Technology*, 40(1):1–54, 2010.

[32] Salvatore Calabrese, Arjun Chakrawal, Stefano Manzoni, and Philippe Van Cappellen. Energetic scaling in microbial growth. *Proceedings of the National Academy of Sciences*, 118(47):e2107668118, 2021.

[33] Yuichiro Ueno, Keita Yamada, Naohiro Yoshida, Shigenori Maruyama, and Yukio Isozaki. Evidence from fluid inclusions for microbial methanogenesis in the early archaean era. *Nature*, 440(7083):516–519, 2006.

[34] Paul N Evans, Joel A Boyd, Andy O Leu, Ben J Woodcroft, Donovan H Parks, Philip Hugenholtz, and Gene W Tyson. An evolving view of methane metabolism in the archaea. *Nature Reviews Microbiology*, 17(4):219–232, 2019.

[35] William F Martin and Filipa L Sousa. Early microbial evolution: the age of anaerobes. *Cold Spring Harbor perspectives in biology*, 8(2):a018127, 2016.

[36] Bradley E Jackson and Michael J McInerney. Anaerobic microbial metabolism can proceed close to thermodynamic limits. *Nature*, 415(6870):454–456, 2002.

[37] AH Stouthamer. A theoretical study on the amount of atp required for synthesis of microbial cell material. *Antonie van Leeuwenhoek*, 39:545–565, 1973.

[38] Rob Phillips, Jane Kondev, Julie Theriot, and Hernan Garcia. *Physical biology of the cell*. Garland Science, 2012.

[39] Francesco Avanzini, Emanuele Penocchio, Gianmaria Falasco, and Massimiliano Esposito. Nonequilibrium thermodynamics of non-ideal chemical reaction networks. *The Journal of Chemical Physics*, 154(9):094114, 03 2021. ISSN 0021-9606. doi: 10.1063/5.0041225. URL `https://doi.org/10.1063/5.0041225`.

[40] Artur Wachtel, Riccardo Rao, and Massimiliano Esposito. Free-energy transduction in chemical reaction networks: From enzymes to metabolism. *The Journal of Chemical Physics*, 157(2), 2022.

[41] Terrell Hill. *Free energy transduction in biology: the steady-state kinetic and thermodynamic formalism*. Elsevier, 2012.

[42] Natascha A Schill, Jing-Song Liu, and Urs von Stockar. Thermodynamic analysis of growth of methanobacterium thermoautotrophicum. *Biotechnology and bioengineering*, 64(1):74–81, 1999.

[43] William Martin and Miklós Müller. The hydrogen hypothesis for the first eukaryote. *Nature*, 392(6671):37–41, 1998.

[44] Daniel B Mills, Richard A Boyle, Stuart J Daines, Erik A Sperling, Davide Pisani, Philip CJ Donoghue, and Timothy M Lenton. Eukaryogenesis and oxygen in earth history. *Nature Ecology & Evolution*, 6(5):520–532, 2022.

[45] Harold J Morowitz. Energy flow in biology; biological organization as a problem in thermal physics. 1968.

[46] Enrico Fermi. *Thermodynamics*. Courier Corporation, 2012.

[47] Matthew Scott, Carl W Gunderson, Eduard M Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007):1099–1102, 2010.

[48] Emily J Zakem, Martin F Polz, and Michael J Follows. Redox-informed models of global biogeochemical cycles. *Nature communications*, 11(1):5680, 2020.

[49] Benjamin T Jackson and Lydia WS Finley. Metabolic regulation of the hallmarks of stem cell biology. *Cell Stem Cell*, 31(2):161–180, 2024.

[50] Natalya N. Pavlova and Craig B. Thompson. The emerging hallmarks of cancer metabolism. *Cell Metabolism*, 23 (1):27–47, 2016. ISSN 1550-4131. doi: https://doi.org/10.1016/j.cmet.2015.12.006.

[51] Ganhui Lan, Pablo Sartori, Silke Neumann, Victor Sourjik, and Yuhai Tu. The energy–speed–accuracy tradeoff in sensory adaptation. *Nature physics*, 8(5):422–428, 2012.

[52] Riccardo Rao and Massimiliano Esposito. Nonequilibrium thermodynamics of chemical reaction networks: Wisdom from stochastic thermodynamics. *Physical Review X*, 6(4), Dec 2016. ISSN 2160-3308. doi: 10.1103/physrevx.6.041064.

[53] T Bauchop and SR Elsden. The growth of microorganisms in relation to their energy supply. *Microbiology*, 23(3):457–469, 1960.

[54] SJ Pirt. The maintenance energy of bacteria in growing cultures. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 163(991):224–231, 1965.

[55] Ao H Stouthamer and Corry Bettenhaussen. Utilization of energy for growth and maintenance in continuous and batch cultures of microorganisms: A reevaluation of the method for the determination of atp production by measuring molar growth yields. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics*, 301(1):53–70, 1973.

[56] Michael Lynch and Georgi K Marinov. The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences*, 112(51):15690–15695, 2015.

[57] Gita Mahmoudabadi, Rob Phillips, Michael Lynch, and Ron Milo. Defining the energetic costs of cellular structures. *bioRxiv*, page 666040, 2019.

[58] TM McCollom and JP Amend. A thermodynamic assessment of energy requirements for biomass synthesis by chemolithoautotrophic micro-organisms in oxic and anoxic environments. *Geobiology*, 3(2):135–144, 2005.

[59] Jan P Amend, Douglas E LaRowe, Thomas M McCollom, and Everett L Shock. The energetics of organic synthesis inside and outside the cell. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368 (1622):20120255, 2013.

[60] U Von Stockar and J-S Liu. Does microbial life always feed on negative entropy? thermodynamic analysis of microbial growth. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1412(3):191–211, 1999.

[61] Robbert Kleerebezem and Mark C. M. Van Loosdrecht. A generalized method for thermodynamic state analysis of environmental systems. *Critical Reviews in Environmental Science and Technology*, 40(1):1–54, 2010.

[62] Edwin H Battley. Calculation of thermodynamic properties of protein in escherichia coli k-12 grown on succinic acid, energy changes accompanying protein anabolism, and energetic role of atp in protein synthesis. *Biotechnology and bioengineering*, 40(2):280–288, 1992.

[63] Sybren Ruurds De Groot and Peter Mazur. *Nonequilibrium thermodynamics*. Courier Corporation, 2013.

[64] Edwin H Battley. The development of direct and indirect methods for the study of the thermodynamics of microbial growth. *Thermochimica Acta*, 309(1-2):17–37, 1998.

[65] Walter P Hempfling and Stanley E Mainzer. Effects of varying the carbon source limiting growth on yield and maintenance characteristics of escherichia coli in continuous culture. *Journal of bacteriology*, 123(3):1076–1087, 1975.

[66] Klaus B Andersen and Kaspar von Meyenburg. Are growth rates of escherichia coli in batch cultures limited by respiration? *Journal of bacteriology*, 144(1):114–123, 1980.

[67] Werner Badziong and Rudolf K Thauer. Growth yields and growth rates of desulfovibrio vulgaris (marburg) growing on hydrogen plus sulfate and hydrogen plus thiosulfate as the sole energy sources. *Archives of microbiology*, 117:209–214, 1978.

[68] Werner Badziong, Rudolf K Thauer, and J Gregory Zeikus. Isolation and characterization of desulfovibrio growing on hydrogen plus sulfate as the sole energy source. *Archives of Microbiology*, 116:41–49, 1978.

[69] Astrid Brandis and Rudolf K Thauer. Growth of desulfovibrio species on hydrogen and sulphate as sole energy source. *Microbiology*, 126(1):249–252, 1981.

[70] HB Chua and JP Robinson. Formate-limited growth of methanobacterium formicium in steady-state cultures. *Archives of microbiology*, 135:158–160, 1983.

[71] M Clarens and R Molleta. Kinetic studies of acetate fermentation by methanosarcina sp. msta-1. *Applied microbiology and biotechnology*, 33:239–244, 1990.

[72] Pia M Crabbendam, OM Neijssel, and DW Tempest. Metabolic and energetic aspects of the growth of clostridium butyricum on glucose in chemostat culture. *Archives of microbiology*, 142:375–382, 1985.

[73] L Dijkhuizen, M Wiersma, and W Harder. Energy production and growth of pseudomonas oxalaticus ox1 on oxalate and formate. *Archives of Microbiology*, 115:229–236, 1977.

[74] Abraham Esteve-Núñez, Mary Rothermich, Manju Sharma, and Derek Lovley. Growth of geobacter sulfurreducens under nutrient-limiting conditions in continuous culture. *Environmental microbiology*, 7(5):641–648, 2005.

[75] I Goldberg, JS Rock, A Ben-Bassat, and RI Mateles. Bacterial yields on methanol, methylamine, formaldehyde, and formate. *Biotechnology and bioengineering*, 18(12):1657–1668, 1976.

[76] Eovaldo Hernandez and Marvin J Johnson. Energy supply and cell yield in aerobically grown microorganisms. *Journal of Bacteriology*, 94(4):996–1001, 1967.

[77] M Heyndrickx, P De Vos, and J De Ley. Fermentation characteristics of clostridium pasteurianum lmg 3285 grown on glucose and mannitol. *Journal of Applied Microbiology*, 70(1):52–58, 1991.

[78] Beat A Huser, Karl Wuhrmann, and Alexander JB Zehnder. Methanothrix soehngenii gen. nov. sp. nov., a new acetotrophic non-hydrogen-oxidizing methane bacterium. *Archives of Microbiology*, 132:1–9, 1982.

[79] Kjeld Ingvorsen, Alexander JB Zehnder, and Bo B Jørgensen. Kinetics of sulfate and acetate uptake by desulfobacter postgatei. *Applied and environmental microbiology*, 47(2):403–408, 1984.

[80] Hendrikus J Laanbroek, Harm J Geerligs, Lolke Sijtsma, and Hans Veldkamp. Competition for sulfate and ethanol among desulfobacter, desulfobulbus, and desulfovibrio species isolated from intertidal sediments. *Applied and Environmental Microbiology*, 47(2):329–334, 1984.

[81] Pei-Ying Lin, Liang-Ming Whang, Yi-Ru Wu, Wei-Jie Ren, Chia-Jung Hsiao, Shiue-Lin Li, and Jo-Shu Chang. Biological hydrogen production of the genus clostridium: metabolic study and mathematical model simulation. *International Journal of Hydrogen Energy*, 32(12):1728–1735, 2007.

[82] RW Lovitt, DB Kell, and JG Morris. The physiology of clostridium sporogenes ncib 8053 growing in defined media. *Journal of applied bacteriology*, 62(1):81–92, 1987.

[83] William Roy Mayberry, George John Prochazka, and William Jackson Payne. Growth yields of bacteria on selected organic compounds. *Applied Microbiology*, 15 (6):1332–1338, 1967.

[84] Charles R Myers and Kenneth H Nealson. Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science*, 240(4857):1319–1321, 1988.

[85] GB Patel. Characterization and nutritional properties of methanothrix concilii sp. nov., a mesophilic, aceticlastic methanogen. *Canadian Journal of Microbiology*, 30(11): 1383–1396, 1984.

[86] V Peters, PH Janssen, and R Conrad. Efficiency of hydrogen utilization during unitrophic and mixotrophic growth of acetobacterium woodii on hydrogen and lactate in the chemostat. *FEMS Microbiology Ecology*, 26 (4):317–324, 1998.

[87] Norbert Pfennig and Hanno Biebl. Desulfuromonas acetoxidans gen. nov. and sp. nov., a new anaerobic, sulfur-reducing, acetate-oxidizing bacterium. *Archives of Microbiology*, 110:3–12, 1976.

[88] Joseph A Robinson and James M Tiedje. Competition between sulfate-reducing and methanogenic bacteria for h 2 under resting and growing conditions. *Archives of Microbiology*, 137:26–32, 1984.

[89] JM Macy and S Lawson. Cell yield (ym) of thauera selenatis grown anaerobically with acetate plus selenate or nitrate. *Archives of microbiology*, 160:295–298, 1993.

[90] Eric E Roden and Derek R Lovley. Dissimilatory fe (iii) reduction by the marine microorganism desulfuromonas acetoxidans. *Applied and Environmental Microbiology*, 59(3):734–742, 1993.

[91] Michiel Rutgers, Hanneke ML van der Gulden, and Karelvan Dam. Thermodynamic efficiency of bacterial growth calculated from growth yield of pseudomonas oxalaticus ox1 in the chemostat. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 973(2):302–307, 1989.

[92] Robert A Sanford, Qingzhong Wu, Youlboong Sung, Sara H Thomas, Benjamin K Amos, Emily K Prince, and Frank E Löffler. Hexavalent uranium supports growth of anaeromyxobacter dehalogenans and geobacter spp. with lower than predicted biomass yields. *Environmental microbiology*, 9(11):2885–2893, 2007.

[93] Henrik Sass, Jörg Overmann, Heike Rütters, Hans-Dietrich Babenzien, and Heribert Cypionka. Desulfosporomusa polytropa gen. nov., sp. nov., a novel sulfate-reducing bacterium from sediments of an oligotrophic lake. *Archives of microbiology*, 182:204–211, 2004.

[94] Neil L Schauer and James G Ferry. Metabolism of formate in methanobacterium formicicum. *Journal of Bacteriology*, 142(3):800–807, 1980.

[95] Michael R Smith and Robert A Mah. Growth and methanogenesis by methanosarcina strain 227 on acetate and methanol. *Applied and Environmental Microbiology*, 36(6):870–879, 1978.

[96] AJB Zehnder and K Wuhrmann. Physiology of a methanobacterium strain az. *Archives of Microbiology*, 111:199–205, 1977.

[97] PJ Weimer and JG Zeikus. One carbon metabolism in methanogenic bacteria: cellular characterization and growth of methanosarcina barkeri. *Archives of Microbiology*, 119:49–57, 1978.

[98] Marion Stieb and Bernhard Schink. Anaerobic degradation of isobutyrate by methanogenic enrichment cultures and by a desulfococcus multivorans strain. *Archives of microbiology*, 151:126–132, 1989.

[99] Tobin O Strohm, Ben Griffin, Walter G Zumft, and Bernhard Schink. Growth yields in bacterial denitrification and nitrate ammonification. *Applied and environmental microbiology*, 73(5):1420–1424, 2007.

[100] Regine Szewzyk and Norbert Pfennig. Competition for ethanol between sulfate-reducing and fermenting bacteria. *Archives of microbiology*, 153:470–477, 1990.

[101] Yinjie J Tang, Adam L Meadows, and Jay D Keasling. A kinetic model describing shewanella oneidensis mr-1 growth, substrate consumption, and product secretion. *Biotechnology and bioengineering*, 96(1):125–133, 2007.

[102] IA Vasiliadou, S Siozios, IT Papadas, K Bourtzis, S Pavlou, and DV Vayenas. Kinetics of pure cultures

of hydrogen-oxidizing denitrifying bacteria and modeling of the interactions among them in mixed cultures. *Biotechnology and bioengineering*, 95(3):513–525, 2006.

[103] Christina Wallrabenstein, Elisabeth Hauschild, and Bernhard Schink. Syntrophobacter pfennigii sp. nov., new syntrophically propionate-oxidizing anaerobe growing in pure culture with propionate and sulfate. *Archives of Microbiology*, 164:346–352, 1995.

[104] Friedrich Widdel and Norbert Pfennig. A new anaerobic, sporing, acetate-oxidizing, sulfate-reducing bacterium, desulfotomaculum (emend.) acetoxidans. *Archives of Microbiology*, 112:119–122, 1977.

[105] Friedrich Widdel and Norbert Pfennig. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids: I. isolation of new sulfate-reducing bacteria enriched with acetate from saline environments. description of desulfobacter postgatei gen. nov., sp. nov. *Archives of microbiology*, 129:395–400, 1981.

[106] Friedrich Widdel and Norbert Pfennig. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids ii. incomplete oxidation of propionate by desulfobulbus propionicus gen. nov., sp. nov. *Archives of Microbiology*, 131:360–365, 1982.

[107] Shang-Tian Yang and MR Okos. Kinetic study and mathematical modeling of methanogenesis of acetate using pure cultures of methanogens. *Biotechnology and bioengineering*, 30(5):661–667, 1987.

[108] Sukhwan Yoon, Robert A Sanford, and Frank E Löffler. Shewanella spp. use acetate as an electron donor for denitrification but not ferric iron or fumarate reduction. *Applied and environmental microbiology*, 79(8):2818–2822, 2013.

[109] Linda MI De Poorter, Wim J Geerts, and Jan T Keltjens. Coupling of methanothermobacter thermautotrophicus methane formation and growth in fed-batch and continuous cultures under different h2 gassing regimens. *Applied and environmental microbiology*, 73(3): 740–749, 2007.

[110] Wytske de Vries, HGD Niekus, Marian Boellaard, and AH Stouthamer. Growth yields and energy generation by campylobacter sputorum subspecies bubulus during growth in continuous culture with different hydrogen acceptors. *Archives of Microbiology*, 124:221–227, 1980.

[111] Philippe Duboc. Transient growth of saccharomyces cerevisiae. Technical report, EPFL, 1997.

[112] R Brettel, I Lamprecht, and B Schaarschmidt. Microcalorimetric investigations of the metabolism of yeasts vii. flow-calorimetry of aerobic batch cultures. *Radiation and Environmental Biophysics*, 18:301–309, 1980.

[113] Bernard Birou, Ian W Marison, and Urs Von Stockar. Calorimetric investigation of aerobic fermentations. *Biotechnology and bioengineering*, 30(5):650–660, 1987.

[114] Laurent Dejean, Bertrand Beauvoit, Bernard Guérin, and Michel Rigoulet. Growth of the yeast saccharomyces cerevisiae on a non-fermentable substrate: control of energetic yield by the amount of mitochondria. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1457(1-2):45–56, 2000.

[115] Edwin H Battley. Enthalpy changes accompanying the growth of saccharomyces cerevisiae (hansen). *Physiologia Plantarum*, 13(4):628–640, 1960.

[116] Z Dermoun and JP Belaich. Microcalorimetric study of escherichia coli aerobic growth: kinetics and experimental enthalpy associated with growth on succinic acid. *Journal of Bacteriology*, 140(2):377–380, 1979.

[117] Z Dermoun and JP Belaich. Microcalorimetric study of cellulose degradation by cellulomonas uda atcc 21399. *Biotechnology and bioengineering*, 27(7):1005–1011, 1985.

[118] A Belaich and JP Belaich. Microcalorimetric study of the anaerobic growth of escherichia coli: growth thermograms in a synthetic medium. *Journal of bacteriology*, 125(1):14–18, 1976.

[119] ANNE Belaich and JEAN-PIERRE Belaich. Microcalorimetric study of the anaerobic growth of escherichia coli: measurements of the affinity of whole cells for various energy substrates. *Journal of bacteriology*, 125 (1):19–24, 1976.

[120] Yasufumi Ishikawa, Yukio Nonoyama, and Makoto Shoda. Microcalorimetric study of aerobic growth of escherichia coli in batch culture. *Biotechnology and Bioengineering*, 23(12):2825–2836, 1981.

[121] Yasufumi Ishikawa and Makoto Shoda. Calorimetric analysis of escherichia coli in continuous culture. *Biotechnology and Bioengineering*, 25(7):1817–1827, 1983.

[122] Ian Marison and Urs Von Stockar. A calorimetric investigation of the aerobic cultivation of kluyveromyces fragilis on various substrates. *Enzyme and microbial technology*, 9(1):33–43, 1987.

[123] Syunzi Yamagata. *Über die elementare Zusammensetzung des Schimmelpilzkörpers*. 1934.

[124] Rita Whelton and Michael Doudoroff. Assimilation of glucose and related compounds by growing cultures of pseudomonas saccharophila. *Journal of Bacteriology*, 49 (2):177–186, 1945.

[125] H Samejima and J Myers. On the heterotrophic growth of chlorella pyrenoidosa. *Microbiology*, 18(1):107–117, 1958.

[126] Sam R Hoover and Franklin E Allison. The growth metabolism of rhizobium, with evidence on the interrelations between respiration and synthesis. *Journal of Biological Chemistry*, 134(1):181–192, 1940.

[127] J-S Liu, IW Marison, and U Von Stockar. Microbial growth by a net heat up-take: a calorimetric and thermodynamic study on acetotrophic methanogenesis by methanosarcina barkeri. *Biotechnology and bioengineering*, 75(2):170–180, 2001.

[128] Bernard Birou and Urs Von Stockar. Application of bench-scale calorimetry to chemostat cultures. *Enzyme and microbial technology*, 11(1):12–16, 1989.

[129] Rodrigo Patino, Marcel Janssen, and Urs von Stockar. A study of the growth for the microalga chlorella vulgaris by photo-bio-calorimetry and other on-line and off-line techniques. *Biotechnology and bioengineering*, 96(4):757–767, 2007.

[130] J Zhao, T Baba, H Mori, and K Shimizu. Global metabolic response of escherichia coli to gnd or zwf gene-knockout, based on 13 c-labeling experiments and the measurement of enzyme activities. *Applied microbiology and biotechnology*, 64:91–98, 2004.

[131] Stanley E Mainzer and Walter P Hempfling. Effects of growth temperature on yield and maintenance during glucose-limited continuous culture of escherichia coli. *Journal of bacteriology*, 126(1):251–256, 1976.

[132] Cornelis Verduyn. Energetic aspects of metabolic fluxes in yeasts. 1992.

[133] Cornelis Verduyn, Erik Postma, W Alexander Schef-

fers, and Johannes P van Dijken. Physiology of saccharomyces cerevisiae in anaerobic glucose-limited chemostat culturesx. *Microbiology*, 136(3):395–403, 1990.

[134] Sébastien Bernacchi, Simon Rittmann, Arne H Seifert, Alexander Krajete, and Christoph Herwig. Experimental methods for screening parameters influencing the growth to product yield (y (x/ch4)) of a biological methane production (bmp) process performed with methanothermobacter marburgensis. *AIMS Bioengineering*, 1(2): 72–87, 2014.

[135] Jing-Song Liu, Natascha Schill, Walter M van Gulik, Damien Voisard, Ian W Marison, and Urs von Stockar. The coupling between catabolism and anabolism of methanobacterium thermoautotrophicum in h2-and iron-limited continuous cultures. *Enzyme and microbial technology*, 25(10):784–794, 1999.

[136] Robert J Silbey, Robert A Alberty, George A Papadantonakis, and Moungi G Bawendi. *Physical chemistry*. John Wiley & Sons, 2022.

[137] Robert A Alberty. *Thermodynamics of biochemical reactions*. John Wiley & Sons, 2005.

[138] Moritz E Beber, Mattia G Gollub, Dana Mozaffari, Kevin M Shebek, Avi I Flamholz, Ron Milo, and Elad Noor. equilibrator 3.0: a database solution for thermodynamic constant estimation. *Nucleic acids research*, 50 (D1):D603–D609, 2022.

[139] William R Cannon and Lionel M Raff. The formulation of chemical potentials and free energy changes in biochemical reactions. *Physical Chemistry Chemical Physics*, 23(27):14783–14795, 2021.

[140] Antonio Sabatini, Alberto Vacca, and Stefano Iotti. Balanced biochemical reactions: a new approach to unify chemical and biochemical thermodynamics. *Plos one*, 7 (1):e29529, 2012.

[141] Elad Noor, Hulda S. Haraldsdóttir, Ron Milo, and Ronan M. T. Fleming. Consistent estimation of gibbs energy using component contributions. *PLOS Computational Biology*, 9(7):1–11, 07 2013. doi: 10.1371/journal.pcbi.1003098.

[142] Edwin H. Battley and Handling Editor Daniel E. Dykhuizen. A theoretical study of the thermodynamics of microbial growth using saccharomyces cerevisiae and a different free energy equation. *The Quarterly Review of Biology*, 88(2):69–96, 2013. ISSN 00335770, 15397718.

[143] Robert W Sterner and James J Elser. Ecological stoichiometry. In *Ecological stoichiometry*. Princeton university press, 2017.

[144] T Gurakan, IW Marison, U Von Stockar, L Gustafsson, and E Gnaiger. Proposals for a standardized sample handling procedure for the determination of elemental composition and enthalpy of combustion of biological material. *Thermochimica acta*, 172:251–266, 1990.

[145] Marko Popovic. Thermodynamic properties of microorganisms: determination and analysis of enthalpy, entropy, and gibbs free energy of biomass, cells and colonies of 32 microorganism species. *Heliyon*, 5(6), 2019.

[146] Ron Grosz and Gregory Stephanopoulos. Statistical mechanical estimation of the free energy of formation of e. coli biomass for use with macroscopic bioreactor balances. *Biotechnology and bioengineering*, 25(9):2149–2163, 1983.

[147] Urs von Stockar, Lena Gustafsson, Christer Larsson, and Ian Marison. Thermodynamic considerations in constructing energy balances. *Biochimica et Biophysica Acta*, 1183:221–240, 1993.

[148] WM Thornton. Xv. the relation of oxygen to the heat of combustion of organic compounds. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 33(194):196–203, 1917.

[149] Jean-Louis Cordier, Bertram M Butsch, Bernard Birou, and Uros von Stockar. The relationship between elemental composition and heat of combustion of microbial biomass. *Applied Microbiology and Biotechnology*, 25: 305–312, 1987.

[150] Snehal A Patel and LE Erickson. Estimation of heats of combustion of biomass from elemental analysis using available electron concepts. *Biotechnology and Bioengineering*, 23(9):2051–2067, 1981.

[151] E Gnaiger and G Bitterlich. Proximate biochemical composition and caloric content calculated from elemental chn analysis: a stoichiometric concept. *Oecologia*, 62: 289–298, 1984.

[152] Robert Hołyst and Andrzej Poniewierski. *Thermodynamics for chemists, physicists and engineers*, volume 344. Springer, 2012.

[153] Edwin H. Battley, Robert L. Putnam, and Juliana Boerio-Goates. Heat capacity measurements from 10 to 300 k and derived thermodynamic functions of lyophilized cells of saccharomyces cerevisiae including the absolute entropy and the entropy of formation at 298.15 k. *Thermochimica Acta*, 298(1):37–46, 1997. ISSN 0040-6031.

[154] Edwin H. Battley. An empirical method for estimating the entropy of formation and the absolute entropy of dried microbial biomass for use in studies on the thermodynamics of microbial growth. *Thermochimica Acta*, 326:7–15, 1999.

[155] Marko Popovic, Gavin B.G. Stenning, Axel Göttlein, and Mirjana Minceva. Elemental composition, heat capacity from 2 to 300 k and derived thermodynamic functions of 5 microorganism species. *Journal of Biotechnology*, 331:99–107, 2021. ISSN 0168-1656.

[156] Peter Linstrom and William Mallard. The nist chemistry webbook: A chemical data resource on the internet. (46), retrieved September 27, 2023.

Supplementary Information

Contents

A. Summary of supplementary files 15

B. Non-equilibrium thermodynamics of unicellular growth 16
1. Mass conservation and steady balanced growth 16
2. Definition of yield 17
3. Element conservation and macrochemical equation 18
4. Thermodynamics 18

C. Defining the cost across and the coefficient within metabolic types 20

D. Alternative definitions for the thermodynamic cost of growth 21

E. The origin of the cost across types, α, and of the within type coefficient, β 24
1. Analytical relation between yield and dissipation 24
2. Variations of yield and dissipation across and within metabolic type 24
3. Efficiency 25
4. Origin of the differences in deviation coefficient β for aerobic and anaerobic types 25

F. A thermodynamic database of microbial growth 26
1. Data curation and validation from primary literature sources 26
2. Units and typical values of variables used 27
3. Comments on data curation and parsing from primary literature 27

G. Thermodynamic properties of substrates and products 33

H. Thermodynamic properties of biomass 34
1. Biomass composition 34
2. Enthalpy, entropy and free energy of biomass 34

I. Dataset validation 36

J. Supplementary figures 37

Appendix A: Summary of supplementary files

The results of this work are supported by multiple data-files appended to it, which we describe below.

- `measured_yields.csv`: This file contains the experimentally measured yields that we parsed from the primary sources. Each row corresponds to an individual entry, i.e. an experiment where one or more yields were measured. We identify the entries by a number, denoted "entry", which is consistently used across all the data-files. Each column corresponds to a chemical species, denoted by its stoichiometric formula. The formula is used consistently throughout the data-files and in the file `thermo_chem.csv` we also report the name of the chemical species. We report the values of the yields as moles per moles of electron donor (carbon moles in the case of biomass). When the primary source reported the yields in a different form, e.g. in different units or as individual fluxes, we applied the transformation needed. In this data-file, the sign is negative for the chemical species consumed and positive for the ones produced.

- `yields.csv`: This file contains all the yields of biomass and chemical species exchanged as metabolic substrates and products. The rows (entries) contain the yields of the corresponding experiment, whose values were either measured or computed as explained below. Each column corresponds to a chemical species identified by its stoichiometric formula. For each entry, we enforced element conservation, which provides a set of equations for the yields, SI B 3 and I. To solve these equations, we used the measured yield of biomass as input. When the system of equations needed multiple yields to be determined, we chose these extra yields among the measured yields that we parsed. The remaining yields were computed solving the equations for the balance of elements. In some occasions, we could parse more measured yields than those needed to determine the system of equations. We decided which yields to use to determine each system on the basis of the measurements methods reported in the original reference, trying to select the more reliable ones. The yields chosen are reported as "extra yields used". The extra yields not used to determine the system were used to validate the dataset, SI I and Fig. Box 1B. In all cases, the yield of chemical species is computed relative to the electron donor, and reported in units of moles per moles of electron donor. In this data-file, the sign is negative for the chemical species consumed and positive for the ones produced. For example, the yield of electron donor trivially takes the value -1, whereas the yield of biomass is always a positive number. Finally, the file contains the biomass stoichiometry of each entry, used to compute the element balance.

- `thermo_growth_dat.csv`: This file contains the auxiliary information for each entry, including microbial species, temperature, pH, growth rate, a list of chemical products and substrates of metabolism, heat measured and carbon balance (when available). This information was retrieved from the original sources of the data, with possibly the exception of the products. In some cases indeed, we completed the list of metabolic products beyond what explicitly stated in the papers, on the basis of well known metabolic pathways (e.g. if glucose was metabolised respiring oxygen, we included carbon dioxide as product even if not mentioned).

- `biomass_composition.csv`: This file contains data on the stoichiometric composition of biomass. Each row refers to an organism identified by the name of the species and optionally the strain or additional qualifiers. We reported the hydrogen, oxygen and nitrogen content of biomass normalized by carbon content. We discarded other elements when (very rarely) reported. The phylum and domain of the organism, as well as the reference where the stoichiometric data was taken from, were also reported.

- `thermo_chem.csv`: This file contains the thermodynamic properties of all chemical species in our dataset. Each row corresponds to a chemical species, identified by its stoichiometric formula, consistently with the other files, and its name. The free energies of formation are reported in standard conditions, i.e. 1 M concentration at 298 Kelvin. We further reported the enthalpies of formation, used to validate the data that provided calorimetric measurements and to correct the use of standard free energies at non-standard temperature, according to the Gibbs-Helmholtz equation (G2).

- `thermo_growth_computed.csv` This file contains the main thermodynamic quantities computed (excluding biomass yield) as described in the text and SI, including dissipation (free energy dissipated per electron donor consumed), cost (free energy dissipated per biomass produced), anabolic and catabolic force, thermodynamic efficiency, free energy and enthalpy of formation of biomass. In this file, we associated a name to each metabolic type, providing a coarser grouping than the macrochemical equation. For example, "anaerobic fermentation of glucose" does not distinguish the different fermentation products. We remark that a metabolic type is defined by the set of substrates and products (i.e. macrochemical equation), but with no distinction between (de)protonated forms of chemical species (e.g. we considered equivalent acetic acid and acetate, ignoring charges and protons) nor between carbon dioxide and bicarbonate.

Appendix B: Non-equilibrium thermodynamics of unicellular growth

1. Mass conservation and steady balanced growth

We describe a cellular colony as a dilute solution of P biomass constituents, $i = 1, \ldots, P$, at abundances N_i. The constituents span any cellular component, including metabolites, nucleic acids and proteins, and are characterized by their mass m_i. The mass abundances are thus defined as $M_i = m_i N_i$ and the total biomass as $M = \sum_i M_i$. This biomass occupies a volume V at constant pressure Π (not to be confused with the volume of the vessel where the colony grows, V^{v}, the two related through the volume fraction $\phi \equiv V/V^{\mathrm{v}}$), which allows to define the dry biomass density $\rho = M/V$ (i.e. weight of dry mass per volume of biomass).

To sustain steady growth, cells import substrates from the surrounding media into the colony at fluxes $F_i^+ > 0$ and export products at fluxes $-F_i^- < 0$. Furthermore, inside the cell matter is transformed through reaction fluxes K_j spanning chemical reactions $j = 1, \ldots, R$. This network of chemical reactions is characterized by the stoichiometric matrix a_{ij}. Within this framework, mass conservation for each constituent is captured by

$$\dot{N}_i = F_i^+ - F_i^- + \sum_j a_{ij} K_j \quad . \tag{B1}$$

Notice that for each chemical i, either $F_i^+ = 0$ if i is a product ($i \in$ pd), $F_i^- = 0$ if it is a substrate ($i \in$ sb), or $F_i^+ = F_i^- = 0$ for all the other biomass constituents.

The growth rate of the colony is the amount of biomass produced per unit biomass, and so for a closed vessel we have $\gamma \equiv \dot{M}/M$. Similarly, the volume of the colony, V, expands as $\dot{V} = \gamma V$. Taking this into consideration, we can write an intensive form of Eq. (B1) by defining the molar concentrations of chemicals inside the colony as $c_i = N_i/V$ (we similarly define densities as $\rho_i = m_i c_i$, with $\rho = \sum_i \rho_i$). Noting that $\dot{c}_i = \dot{N}_i/V - \gamma c_i$, we have

$$\dot{c}_i = f_i^+ - f_i^- - \gamma c_i + \sum_j a_{ij} \nu_j \quad , \tag{B2}$$

where we have defined the molar import/export flux densities as $f_i^{+/-} = F_i^{+/-}/V$, and the molar reaction flux densities in chemical reactions as $\nu_j = K_j/V$. In this work, we only consider the case of steady balanced growth, in which the growth rate γ is time independent and equal for all constituents, i.e. $\dot{N}_i = \gamma N_i$. This corresponds to exponential growth conditions in which the composition of the cells is not changing. In that case, we have that $\dot{c}_i = 0$, and Eq. (B2) reduces to

$$\gamma c_i = f_i^+ - f_i^- + \sum_j a_{ij} \nu_j \quad . \tag{B3}$$

Since the fluxes $F_i^{+/-}$ and K_j scale as $\sim N_i$, the quantities $f_i^{+/-}$ and ν_j are intensive and time-independent. We note that this description of microbial growth is analogous to that which forms the basis of Flux Balance Analysis [16].

In a *batch culture*, steady balanced growth is achieved during the exponential stage of cell growth in a vessel, i.e. when $\phi \ll 1$, prior to the cellular colony reaching carrying capacity. Steady balanced growth can also be achieved in a *chemostat*. To see this, note that in a chemostat an amount of volume is constantly added and removed per unit time, G, with the dilution rate defined as $d = G/V^{\mathrm{v}}$. The volume added contains nutrients, whereas the volume removed contains all constituents present in the vessel, including biomass. The balance of biomass is, therefore, $\dot{M} = \gamma M - M d$, which reaches a steady state for $d = \gamma$. Equation (B1) is also modified to

$$\dot{N}_i = F_i^+ - F_i^- + \sum_j a_{ij} K_j - N_i d \quad , \tag{B4}$$

and a steady state is reached when the abundances of cellular components are constant within the vessel, i.e. $\dot{N}_i = 0$. Applying this condition in Eq. (B4), identifying $d = \gamma$, and dividing by the volume of the colony V results in Eq. (B3). Therefore, the equation of steady balanced growth applies equally to chemostats at steady state and batch cultures during exponential growth. Note that in writing Eq. (B4) we took into consideration that the exchange fluxes $F_i^{+/-}$ refer to the cellular colony, and not to the vessel.

We emphasize that the concentration c_i represents the amount of constituent i per volume of the cellular colony. Alternatively, one can define concentrations per dry-mass weight of cellular colony, the difference between two being a rescaling by the density ρ. Since our results are presented as ratios of per colony volume quantities, this choice is arbitrary. We further discuss this point and other unit choices in section F 2.

2. Definition of yield

The biomass yield is an important physiological quantity that characterizes growth and plays a central role in our work. Broadly speaking, the growth yield or biomass yield, y, quantifies the amount of biomass grown per unit of nutrient consumed. We now review the precise definition of growth yield, and how it is inferred for two different experimental scenarios: chemostated cultures and batch cultures.

We first start with the case of a *chemostated culture*. First, we write the kinetics of biomass and nutrients in the chemostat. Mass conservation in a chemostat is expressed through Eq. (B4). Defining the exchange mass fluxes $\mathcal{M}_i^{+/-} = m_i F_i^{+/-}$, as well as the net exchange mass fluxes $\mathcal{M}^{+/-} = \sum_i \mathcal{M}_i^{+/-}$, leads to

$$\dot{M} = \mathcal{M}^+ - \mathcal{M}^- - dM \quad , \tag{B5}$$

where we can identify the growth rate as $\gamma = (\mathcal{M}^+ - \mathcal{M}^-)/M$. For the case of any chemical i consumed or produced by the cell i, and in particular for the electron donor $i = \mathrm{ed}$, we have that the total abundance in the vessel is composed of the N_{ed} molecules inside the colony plus the N'_{ed} diluted in the surrounding media. Mass conservation for the abundances in the media is therefore given by

$$\dot{M}'_{\mathrm{ed}} = \rho^{\mathrm{in}}_{\mathrm{ed}} G - \rho'_{\mathrm{ed}} G - \mathcal{M}^+_{\mathrm{ed}} \quad , \tag{B6}$$

where $\rho^{\mathrm{in}}_{\mathrm{ed}}$ is the density of electron donor in the media that is flowed into the chemostat, and we have defined the mass of electron donor outside the colony $M'_{\mathrm{ed}} = m_{\mathrm{ed}} N'_{\mathrm{ed}}$ as well as the density outside the colony $\rho'_{\mathrm{ed}} = M'_{\mathrm{ed}}/V^{\mathrm{v}}$, now relative to the volume of the vessel. A steady state for a chemostat is achieved for $d = \gamma$ and a concentration of electron donor in the vessel $\rho'_{\mathrm{ed}} = \rho^{\mathrm{in}}_{\mathrm{ed}} - \mathcal{M}^+_{\mathrm{ed}}/G$.

In this chemostat setting the yield is defined as

$$Y \equiv \frac{\gamma M}{\mathcal{M}^+_{\mathrm{ed}}} \quad , \tag{B7}$$

which is the expression commonly used in the primary sources of our study. Note that at steady state we have $\mathcal{M}^+_{\mathrm{ed}} = G(\rho^{\mathrm{in}}_{\mathrm{ed}} - \rho'_{\mathrm{ed}})$, which allows to determine the denominator directly from measuring the concentration in outflux from the chemostat. Often, nutrients are limiting in the media and so most nutrients are depleted, $\rho'_{\mathrm{ed}} \ll \rho^{\mathrm{in}}_{\mathrm{ed}}$. In this case, we can approximate $\mathcal{M}^+_{\mathrm{ed}} \approx G\rho^{\mathrm{in}}_{\mathrm{ed}}$. Note also that the stability of the steady state requires further assumptions on the dependence of the growth rate on the electron donor abundance, e.g. Monod law $\gamma = \gamma_{\max} M'_{\mathrm{ed}}/(K + M'_{\mathrm{ed}})$, with K an effective constant and $\gamma_{\max}$ the maximal growth rate [24]. The definition of yield allows to write the chemostat equations in the more familiar form

$$\dot{M} = \gamma M - dM \tag{B8}$$

$$\dot{M}'_{\mathrm{ed}} = \rho^{\mathrm{in}}_{\mathrm{ed}} G - \rho'_{\mathrm{ed}} G - \gamma M/Y \quad . \tag{B9}$$

The expression of yield in Eq. (B7) is written using extensive quantities, but it is straightforward to divide by the cellular volume and obtain

$$Y = \frac{\gamma \rho}{m_{\mathrm{ed}} f^+_{\mathrm{ed}}} \quad . \tag{B10}$$

Furthermore, while the yield is a dimensionless quantity, it was calculated as a ratio of mass fluxes. Instead, we can define an expression of the yield as a ratio of carbon molar flux for biomass, and molar flux for electron donor. To achieve this, we simply multiply the above yield by the conversion factor $m_{\mathrm{ed}}/m_{\mathrm{b}}$, with m_{b} the molecular weight of biomass, see also F 2. The resulting form for the yield is then

$$y = \frac{\gamma \rho_{\mathrm{C}}}{f^+_{\mathrm{ed}}} \quad , \tag{B11}$$

with $\rho_{\mathrm{C}} = \rho/m_{\mathrm{b}}$ the dry biomass density in carbon moles. This is the expression that was used in the main text.

We now discuss the case of *batch culture*. Conservation for the biomass of the colony and the mass of electron donor outside the colony is simply described by

$$\dot{M} = \mathcal{M}^+ - \mathcal{M}^- \tag{B12}$$

$$\dot{M}'_{\mathrm{ed}} = -\mathcal{M}^+_{\mathrm{ed}} \quad , \tag{B13}$$

where as before the growth rate is given by $\gamma = (\mathcal{M}^+ - \mathcal{M}^-)/M$. At early stages of growth the electron donor is in great abundance, and the growth rate can be considered independent of $M'_{\rm ed}$ (for Monod's law we have $\gamma \approx \gamma_{\rm max}$). In this case, we can use the definition of yield in Eq. (B7) to obtain the time-dependent solutions $M(t) = M(0)\exp(\gamma t)$ and $M'_{\rm ed} = M'_{\rm ed}(0) - (M(t) - M(0))/Y$. From this, we obtain that during the exponential phase the yield can also be written as

$$Y = -\frac{M(t) - M(0)}{M'_{\rm ed} - M'_{\rm ed}(0)} \quad , \tag{B14}$$

which is the expression used for the batch culture experiments that we analyzed (transforming to y is, as before, straightforward). This expression is strictly correct for the case in which the experiment was arrested during exponential growth (note that the lag-phase, in which nutrient is not depleted and biomass does not grow, does not affect this result). However, Eq. (B14) can also be used as an approximation for the yield when the experiment is run until depletion of substrate. The accuracy of this approximation then depends on the duration of the transition from exponential growth to saturation.

We note that, besides the biomass yield, the yield of any chemical can be computed. We always take the electron donor flux as reference, and from the flux $f_i^{+/-}$ one can define the yield $y_i \equiv f_i^{+/-}/f_{\rm ed}^+$, analogous to the biomass yield. Even heat fluxes are often reported as heat yields, i.e. using $y_{\dot{q}} \equiv \dot{q}/f_{\rm ed}^+$, where $\dot{q}$ is defined in Box 1 of the main text and below in Eq. (B17). Furthermore, with this definition the dissipation, σ, is nothing but the dissipation yield, i.e. $\sigma = y_{\dot{g}_{\rm diss}} \equiv -\dot{g}_{\rm diss}/f_{\rm ed}^+$, where $\dot{g}_{\rm diss}$ is defined in Box 1 (Eq. (3)). Importantly, this allows to treat in a consistent manner both, chemostat and batch experiments. For this reason, in our work we used yields, instead of fluxes, to characterize the cellular growth. See however Fig. SI5 for a description of chemostated data using fluxes.

3. Element conservation and macrochemical equation

Each chemical species is characterized by its element composition. The matrix of entries e_{ik} contains this information, as it expresses the number of elements of type k (C, H, O, N, ...) present in chemical species i (e.g. glucose, acetate, ammonia, etc). Since chemical reactions preserve elements, we have that $\sum_i e_{ik} a_{ij} = 0$. Using this in Eq. (B3), we obtain a balance between the element fluxes towards the colony, and those away from the colony (including dilution):

$$\sum_{\rm i\in sb} e_{ik} f_i^+ = \sum_{\rm i\in pd} e_{ik} f_i^- + \gamma c_k \quad , \tag{B15}$$

where the first/second sum runs over substrates/products, and $c_k = \sum_i e_{ik} c_i$ refers to the concentration of elements in the colony. Importantly, Eq. (B15) constitutes a set of constraints (as many as elements) on the biomass exchange fluxes, which can be used to determine them [20].

More specifically, each entry in our dataset involves a relatively small number of chemicals exchanged, C. This number is typically of the same order and larger than the set of elements spanned by all substrates and products combined, E. For example, microbial growth through glucose respiration can be schematically summarized as the following *macrochemical equation*:

$$\mathrm{C_6H_{12}O_6 + O_2 + NH_3 \rightarrow biomass + CO_2 + H_2O} \quad , \tag{B16}$$

where the above "equation" is not stoichiometrically balanced. In this example, the number of chemicals exchanged is $C = 5$ (3 substrates: glucose, oxygen and ammonia; and 2 products: carbon dioxide and water), and the number of elements is $E = 4$ (C, H, O and N). Eq. (B15) then constitutes a set of four equations, one per element, with six unknowns: the five exchange fluxes and the biomass growth. Provided the measurement of two fluxes, e.g. the biomass growth rate and glucose consumption rate, the rest can be determined by solving the linear system Eq. (B15).

Most of the primary sources that we used to construct the dataset contain very sparse information on the exchange fluxes, and so element conservation was used to complete unknown fluxes in this manner. When redundant information was available, this was used to verify the accuracy of element conservation, see section I.

4. Thermodynamics

Having discussed mass and element conservation, we now construct a non-equilibrium thermodynamic description under steady balanced growth conditions, i.e. conditions under which Eq. (B3) applies. Assuming knowledge of

matter fluxes and thermodynamic properties of biomass and chemicals, the steady state formulation of the second law takes the form:

$$\dot{s}_{\text{prod}} = -\frac{\dot{q}}{T} + \dot{s}_{\text{ex}} + \gamma s \geq 0 \quad , \tag{B17}$$

where $\dot{s}_{\text{prod}}$ is the rate at which entropy is produced, $\dot{q}$ the rate at which heat is produced (or absorbed, if positive), T the temperature set by the media, $\dot{s}_{\text{ex}}$ the exchange entropy rate, and γs the amount of entropy produced in the form of biomass per unit time, with s the biomass entropy density (see [52] for a comprehensive treatment of thermodynamics of chemical reaction networks). All fluxes are expressed per unit volume of biomass, unless otherwise specified. The exchange entropy rate is given by

$$\dot{s}_{\text{ex}} = \sum_{i \in \text{pd}} f_i^- s_i - \sum_{i \in \text{sb}} f_i^+ s_i \quad , \tag{B18}$$

and it expresses a balance between the fluxes of entropy due to incoming and outgoing chemicals, with s_i the molar entropy of the corresponding chemical species.

So far Eq. (B17) is incomplete, as it requires an expression for the heat. In other words, it can not be directly used to compute entropy production. To obtain an expression for the heat we make use of Hess' law, which connects heat with enthalpy exchanges. At steady state, this gives

$$\dot{q} = \dot{h}_{\text{ex}} + \gamma h \quad , \tag{B19}$$

where γh is the enthalpy flux due to biomass synthesis and $\dot{h}_{\text{ex}}$ is the exchange enthalpy, given by

$$\dot{h}_{\text{ex}} = \sum_{i \in \text{pd}} f_i^- h_i - \sum_{i \in \text{sb}} f_i^+ h_i \quad , \tag{B20}$$

with h_i the enthalpy of the chemical species. We note that Eq. (B19) is a direct expression of energy conservation, and thus a reformulation of the first law of thermodynamics.

Putting together the expressions of the first and second law, and using the definition of chemical potential, $\mu_i = h_i - T s_i$, we arrive at

$$T\dot{s}_{\text{prod}} = \underbrace{\sum_{i \in \text{sb}} f_i^+ \mu_i}_{\dot{g}_{\text{sb}}} - \underbrace{\sum_{i \in \text{pd}} f_i^- \mu_i}_{\dot{g}_{\text{pd}}} - \gamma g \geq 0 \quad , \tag{B21}$$

which decomposes the total rate of free energy dissipation, $\dot{g}_{\text{diss}} = -T\dot{s}_{\text{prod}}$, in terms of free energy influx of substrates, $\dot{g}_{\text{sb}}$, free energy outflux of products, $\dot{g}_{\text{pd}}$, and free energy flux due to biomass synthesis, γg, with $g = h - Ts$ the free energy of biomass. In a substantial portion of our dataset growth rates are not reported, and instead only yields are provided. We therefore introduce the dissipation per electron donor, $\sigma = T\dot{s}_{\text{prod}}/f_{\text{ed}}^+$, which in the main text is simply referred to as dissipation. The second law can then be written in terms of yields as

$$\sigma = \underbrace{\sum_{i \in \text{sb}} y_i \mu_i}_{\sigma_{\text{sb}}} - \underbrace{\sum_{i \in \text{pd}} y_i \mu_i}_{\sigma_{\text{pd}}} - yg/\rho_{\text{C}} \quad , \tag{B22}$$

where $y_i = f_i^{+/-}/f_{\text{ed}}^+$ are the yields of all exchanged chemicals and $y_{\text{ed}} = 1$.

Remarkably, Eq. (B19) provides a prediction for growth heat in terms of enthalpy of chemicals and exchange fluxes, including biomass. This prediction can be tested through calorimetry experiments, as is shown in Fig. Box 1A. We also note that $\dot{s}_{\text{prod}}$ and $\dot{q}$ are determined solely by the fluxes of exchanged chemicals, their enthalpies/entropies, the enthalpy/entropy of biomass and the growth rate. Therefore, under steady state conditions, a complete thermodynamic description is possible without any knowledge of the internal reaction fluxes ν_j. We conclude that the thermodynamic properties of cellular growth are determined by three sets of data: the fluxes of matter, which we discussed in B 1 and B 3; the thermodynamic properties of the corresponding chemicals, i.e. their free energy of formation, μ_i, and enthalpy of formation, h_i, which we will discuss in H 2; and the thermodynamic properties of biomass, which we will discuss in H.

Appendix C: Defining the cost across and the coefficient within metabolic types

In the main text we grouped the data in terms of metabolic types, $t = 1, \ldots, T$, with each type t containing several instances of data, $d_t = 1, 2, \ldots, D_t$. The metabolic type, also referred to in the literature as macrochemical equation, denotes the set of substrates and products of cellular growth. For a particular type, different data refer to different species, the same species at different temperatures or dilution rates, or experimental variation. In Fig. SI1 we provide a comprehensive list of all metabolic types that our database contains with legends used in figures SI7, SI8 and SI9. In the main text, for simplicity, only some types were highlighted and the index notation was dropped.

With this grouping, the biomass yield of a single data point can be denoted as $y(t, d)$, which refers to the dth instance of data in the type t. Similarly, the dissipation of a data point can be denoted as $\sigma(t, d_t)$. We computed the mean dissipation of a type as $\bar{\sigma}(t) = \sum_{d_t=1}^{D_t} \sigma(t, d_t)/D_t$, and similarly for the mean yield, $\bar{y}(t)$. The yield of a particular data point can therefore be written as $y(t, d_t) = \bar{y}(t) + \delta y(d_t)$. Using these definitions, we identified the across-type cost as $\alpha(t) = \bar{\sigma}(t)/\bar{y}(t)$. Within a type, element conservation establishes a linear relation between $\delta\sigma(d_t)$ and $\delta y(d_t)$, or quasi-linear in case of multiple metabolic products. Therefore, we defined the within type coefficient $\beta(t)$ by performing a linear regression of $\delta\sigma(d_t)$ against $\delta y(d_t)$. Our two key findings are simply that $\alpha(t)$ is highly conserved across types, while $\beta(t)$ takes two different values depending on whether the electron acceptor is oxygen or not.

Appendix D: Alternative definitions for the thermodynamic cost of growth

In this work, we have defined the cost of cellular growth using a clear non-equilibrium thermodynamic model, see section B. In particular, the cost α is the amount of free energy dissipation per carbon mole of biomass synthesized. This is in the spirit of classical works such as Von Stockar et al. [27] and reference within. We remark that our method does not introduce putative catabolic and anabolic reactions to compose the macrochemical equation, contrary to what is usually done in the aforementioned body of literature. Our use of the terms catabolic and anabolic follows *after* the computation of the dissipation. Indeed, our definition is based on whether a fundamental force drives or opposes dissipation, and does not refer to intracellular processes. We will now compare our definition and value of cost to alternatives proposed in the literature. To this end, it is useful to recast the value we obtained, $\alpha \approx 500\,\mathrm{kJ/Cmol}$, in different units (see Table I).

Energy units / per biom. grown	kJ	$k_\mathrm{B}T(= RT/N_\mathrm{A})$	ATPmol$(= N_\mathrm{A}$ATP)	ATP (#)
carbon mole (/Cmol)	$500\,\mathrm{kJ/Cmol}$	$1.25\cdot 10^{26}\,k_\mathrm{B}T/\mathrm{Cmol}$	$8\,\mathrm{ATPmol/Cmol}$	$5\cdot 10^{24}\,\mathrm{ATP/Cmol}$
gram (/g)	$20\,\mathrm{kJ/g}$	$5\cdot 10^{24}\,k_\mathrm{B}T/\mathrm{g}$	$0.3\,\mathrm{ATPmol/g}$	$1.8\cdot 10^{23}\mathrm{ATP/g}$
carbon atom (/C)	$8\cdot 10^{-22}\,\mathrm{kJ/C}$	$210\,k_\mathrm{B}T/\mathrm{C}$	$1.3\cdot 10^{-23}\mathrm{ATPmol/C}$	$8\,\mathrm{ATP/C}$
cell (/cell)	$6\cdot 10^{-12}\,\mathrm{kJ/cell}$	$1.5\cdot 10^{12}\,k_\mathrm{B}T/\mathrm{cell}$	$1\cdot 10^{-13}\,\mathrm{ATPmol/cell}$	$6\cdot 10^{10}\,\mathrm{ATP/cell}$

TABLE I: **Unit conversion of the dissipative cost of growth (α).** Unit conversion of $\alpha \approx 500\,\mathrm{kJ/Cmol}$ to various energy scales: kilo Joules (kJ), thermal energy ($k_\mathrm{B}T$), moles of ATP hydrolyzed (ATPmol), and individual ATP hydrolysis (ATP). Energy scales were normalized either per carbon mole of biomass (Cmol), per gram of biomass (g), per carbon atom of biomass (C), and per cell. The following conversion factors were used: $k_\mathrm{B}T \approx 4\cdot 10^{-21}\,\mathrm{J}$ (and thus $RT \approx 2.45\,\mathrm{kJ/mol}$), hydrolysis of one ATP molecule $\approx 25\,k_\mathrm{B}T$, hydrolysis of one mole of ATP $\approx 60\,\mathrm{kJ}$, average molar biomass weight discussed before $m_\mathrm{b} \approx 24.5\,\mathrm{g/Cmol}$, and Avogadro's constant $N_\mathrm{A} \approx 6\cdot 10^{23}$. Per cell normalization was achieved by assuming an average cell size of $1\,\mu\mathrm{m}^3$ with a dry weight of $0.3\,\mathrm{pg}$, similar to an *E. coli* cell (BNID:103904).

These unit transformations (see Table I) will help us relate the magnitude of our thermodynamic cost of growth to alternative ones introduced in the literature.

1. **The cost as dissipation in the metabolic network.** Whereas early attempts at modeling metabolism using Flux Balance Analysis are thermodynamically inconsistent [16], the solution to enforce the preservation of the second law was soon noted [15]. Using this approach, Niebel et al. [11] studied the free energy dissipated by the whole metabolic network of aerobically fermenting *S. cerevisae* and *E. coli*. The authors arrived at maximal values of $\approx 9\,\mathrm{kJ/g}$ for *S.cerevisae* and $\approx 12\,\mathrm{kJ/g}$ for *E.coli*. In our database, for analogous conditions and microbial species, we find values ranging from the same, up to three or four folds higher. Using Flux Balance Analysis on a model of *E. coli* growing on glucose minimal media, Mori et al. [17] computed the flux of ATP production in aerobic and anaerobic conditions. These fluxes correspond to ≈ 7 and $3\,\mathrm{kJ}$ per gram of biomass, respectively. In our database, for the same species in analogous conditions and with similar biomass yield, we find values of dissipation per biomass ≈ 16 and $7\,\mathrm{kJ/g}$, respectively.

2. **The cost of growth as ATP equivalents.** The main cellular energy currency is adenosine-triphosphate (ATP). Under physiological conditions, ATP hydrolysis to adenosine diphosphate (ADP) and inorganic phosphate results in $\approx 50-60\,\mathrm{kJ/mol}$ of free energy (BNID: 101964). Many metabolic reactions and pathways that transform energy and mass for cellular growth have been biochemically characterized in the last decades. This knowledge allows for experimental and theoretical estimation of the cost of growth in terms of ATP molecules being hydrolized. However, some biochemical reactions do not use ATP as a main energy source, e.g. protein translation uses guanine-triphosphate (GTP) instead of ATP. Similarly, cellular oxidation and reducing power required for growth is provided in the form of adenine-dinucleotides such as NAD(H), NADP(H), and FAD(H_2). For simplicity, ATP costs are usually estimated as ATP equivalents by converting other currencies such as GTP and adenine dinucleotides to ATP equivalents.

   There are two complementary approaches to estimate the ATP growth requirements – one theoretical and one experimental. The theoretical approach uses biochemical knowledge of metabolic pathways to calculate the amount of ATP equivalents required to build and polymerize precursors to form cellular macromolecules compatible with the known composition of a cell. These theoretical estimates can be compared to experimental methods that commonly use ATP yields (Y_ATP) as a measure of the cost of growth. Y_ATP is defined as the amount of biomass produced per amount of ATP hydrolyzed and its calculation depends on how much ATP is produced per substrate consumed for a given metabolic type. The notion of Y_ATP was first introduced by Bauchop and Elsden [53], and refined by Stouthamer [37] to consider the cost of maintenance [54].

- **Experimental.** Soon after its definition, an experimental value of $Y_{\mathrm{ATP}} \approx 10.5\,\mathrm{g/ATPmol}$ has been generally accepted [55]. Further experiments and calculations by Stouthamer and Bettenhaussen [55] extended the range of values from ≈ 2 to $12\,\mathrm{g/ATPmol}$. A comparison with $1/\alpha \approx 3\,\mathrm{g/ATPmol}$ shows that our measure of cost tends to be compatible or higher by a few folds.
- **Theoretical.** Stouthamer [37] provided a theoretical estimate of $Y_{\mathrm{ATP}} \approx 28.8\,\mathrm{g/ATPmol}$, or $\approx 51\,\mathrm{kJ/Cmol}$, starting from glucose and inorganic salts and excluding maintenance. The value changes to $\approx 10.5\,\mathrm{g/ATPmol}$ or $\approx 140\,\mathrm{kJ/Cmol}$ when acetate is the substrate. More recent estimates in the same spirit, e.g. [38] p.199, have shown the cost of *E. coli* growth under aerobic conditions to be $\sim 10^{10}$ ATP/cell or $\sim 80\,\mathrm{kJ/Cmol}$, excluding maintenance. Although a direct comparison suggests that our result is an order of magnitude larger, we stress that our definition of cost encompasses the overall dissipation of the growth process, maintenance included. Lynch and Marinov [56] use a conversion factor to transform volume into ATP yield. This was done so to partition the cost of a gene, a concept introduced in [4]. Such conversion factors are common in the case of oxygenic metabolisms, often referred to as P:O ratio. The cost obtained in this manner in [56] for *E. coli*, $\approx 3 \cdot 10^{10}$ ATP/cell is similar, smaller by a factor of two, to what we obtained. These estimates take into account explicitly the maintenance cost, and implicitly the opportunity cost of cellular metabolism [57].

3. **The cost as free energy required by biomass formation.** In an early attempt, Morowitz [45] estimated the entropy and enthalpy changes of transforming simple substrate molecules into biomass. This gives the value of the free energy required by microbial biosynthesis. Considering a minimal media, Morowitz writes the coarse-grained chemical reactions to form cellular building blocks, such as amino acids and nucleotides, from the substrates available in the media. This allows to estimate the net enthalpy and entropy change of synthesizing the precursors required for 1 g of biomass, on the basis of a representative biomass composition. By further using polymer statistics and known hydrolysis enthalpies of peptides and sugar-phosphate bonds, he estimates the entropy and enthalpy change of those precursor polymerizing into macromolecules. Such estimate, for the anaerobic synthesis of *E. coli* biomass from glucose, results in a free energy change $\approx -1\,\mathrm{kJ/g}$, denoting an exergonig process. More recently, in the spirit of Morowitz and using similar approaches of precursor reaction thermodynamics, McCollom and Amend [58], Amend et al. [59] estimated the energy requirement of aerobic and anaerobic chemolithoautotrophic synthesis of monomeric cellular components. They found the minimal free energy required to amount to ≈ 18 and $1\,\mathrm{kJ/g}$ of biomass, for aerobic and anaerobic conditions respectively. McCollom and Amend [58] suggest that the inclusion of all anabolic processes would increase the energy requirement and reduce the difference between the aerobic and anaerobic results. Their values of free energy required by anabolism are therefore compatible with our results.

4. **The cost from the macrochemical equation approach.** Several works have used the macrochemical equation approach to study thermodynamic aspects of microbial growth. Although these works and our both rely on the notion of macrochemical equation, the theoretical foundation for this latter is fundamentally different. In most cases, the thermodynamics cost of growth have been implicitly defined as a combination of a catabolic and an anabolic reaction [21, 26, 28–32, 60]. There, the catabolic and anabolic reaction are postulated on the basis of biochemical intuition. The net free energy change of biomass growth is then computed balancing one or two [30, 61] parameters, that account for the ratios of both the anabolic and catabolic reaction. The results of these approaches, which often used a "standard" biomass stoichiometry, suggest a cost of $\approx 300 - 600\,\mathrm{kJ/Cmol}$, in agreement with our estimates, see Fig. 3B in [30] and Table VG in [26]. We remark that we went to the primary references of these works, and many of these references constitute a subset of our database.

5. **The cost as free energy of formation of biomass.** Another definition of the cost of growth is the free energy of formation of biomass, defined as the energy change of a putative chemical reaction of forming biomass from its constituting elements. Estimating the free energy of formation of biomass is experimentally very challenging, involving measurements of the heat released by biomass combustion to calculate the enthalpy of formation, and low-temperature differential scanning calorimetry to estimate its entropy content. In our work, we incorporated the free energy of biomass formation as part of the cost, and found that its contribution is very small (typically $\leq 10\%$). Indeed, Battley [62] and McCollom and Amend [58] estimated that $\leq 10\%$ of the total free energy is conserved in biomass or cellular protein as a proxy for total biomass. For more detail on the free energy of formation of biomass see section H.

Taken together, the above-mentioned approaches all rely on particular assumptions and mostly underestimate the dissipative cost of growth by a factor of $2 - 10$. Our approach is based on a formal notion of cost in thermodynamics, i.e. the amount of free energy that is dissipated (or total entropy produced). It does not require any conversion factors based on oxygen respiration or substrate-phosphorylation, nor does it rely on details of the metabolic pathway

or arbitrary formulations of anabolic reactions, all details that are often not known. Instead, our method does depend on metabolism through the known substrates and products (see Fig. SI7), their chemical properties (see H 2), and the biomass yield and composition. The content of biomass, both its enthalpic and entropic contributions, is included and species-specificity is considered by its stoichiometric composition, see H. Our approach naturally encompassed different unicellular sizes and doubling times, as the final results are given per unit biomass synthesized. Converting to size and time is straightforward. While some of these features of our model are shared with some of the aforementioned approaches, none of them combine them all.

Appendix E: The origin of the cost across types, α, and of the within type coefficient, β

1. Analytical relation between yield and dissipation

The total dissipation in a non-equilibrium system can be reduced to a sum of products between independent fluxes and corresponding generalized thermodynamic forces. For the case of chemical systems, this can be done by identifying conserved quantities [39, 52], which in our case correspond to elements. We now develop this decomposition.

Following B 3, we label with i_i the $C-E$ independent yields and i_d the remaining dependent ones. Since the choice is arbitrary, we chose to treat as independent yields the measured ones, which almost always include biomass. Once the element matrix is restricted to the dependent species only (entries $e_{i_\mathrm{d},k}$), it can be inverted (entries $(e^{-1})_{i_\mathrm{d},k}$). Using element conservation, Eq. (B15), to express dependent yields as a function of independent ones, we arrive at

$$y_{i_\mathrm{d}} = E_{i_\mathrm{d},\mathrm{ed}} - yE_{i_\mathrm{d}} + \sum_{i_\mathrm{i}\in\mathrm{sb}} y_{i_\mathrm{i}} E_{i_\mathrm{d},i_\mathrm{i}} - \sum_{i_\mathrm{i}\in\mathrm{pd}} y_{i_\mathrm{i}} E_{i_\mathrm{d},i_\mathrm{i}} \quad , \tag{E1}$$

where $E_{i_\mathrm{d},i_\mathrm{i}} = \sum_k (e^{-1})_{i_\mathrm{d},k} e_{i_\mathrm{i},k}$, $E_{i_\mathrm{d}} = \sum_k (e^{-1})_{i_\mathrm{d},k} c_k/\rho_c$, and we assumed $i_\mathrm{d} \in \mathrm{pd}$ (in the case $i_\mathrm{d} \in \mathrm{sb}$ signs are opposite). Using Eq. (E1) in Eq. (B22) we obtain an expression for the dissipation as a function of the independent yields,

$$\sigma = r_\mathrm{ed} - yr_\mathrm{b} + \Delta \quad , \text{with} \quad \Delta = \sum_{i_\mathrm{i}\in\mathrm{sb}} y_{i_\mathrm{i}} r_{i_\mathrm{i}} - \sum_{i_\mathrm{i}\in\mathrm{pd}} y_{i_\mathrm{i}} r_{i_\mathrm{i}} \quad . \tag{E2}$$

The coefficients $r_{i_\mathrm{i}} = \mu_{i_\mathrm{i}} - \sum_{i_\mathrm{d}} E_{i_\mathrm{d},i_\mathrm{i}} \mu_{i_\mathrm{d}}$, with dimensions of a chemical potential, relate the independent yields y_{i_i} to the dissipation (there are analogous expressions to those of r_{i_i} for the biomass coefficient, r_b, and the electron donor coefficient, r_ed). From a thermodynamic point of view, Eq. (E2) is an expression of the dissipation as a product of generalized and independent fluxes (the yields) and forces (the coefficients) [40, 63]. We note that for the cases in which there is a single independent yield we have $\Delta = 0$, and so

$$\sigma = r_\mathrm{ed} - yr_\mathrm{b} \quad . \tag{E3}$$

Next, we will relate these expressions of the dissipation to the dissipative cost α and to the coefficient β used in the main text.

2. Variations of yield and dissipation across and within metabolic type

To gain further insight into how the decomposition above relates to variation within and across metabolic types, it is important to remark the different sources of data variation. Across types, variations are due to different metabolic pathways being activated or deactivated in presence of different nutrient sources. In contrast, variations of yield for a given type arise from diverse sources: changes in chemical concentrations (that affect the chemical potentials, although weakly, see SI G), changes in temperature (that also affect the chemical potentials weakly through Eq. (G2), with a few percent effect), changes in microbial species using the same metabolic type (which affect the biomass composition and through it the stoichiometry of the equation and the free energy of biomass, also with minor consequences, see Fig. SI11), or experimental variability. Therefore, the sources of variations within and across types are very different, and have different effects on the parameters and variables in Eq. E2.

Taking the above into account, we now decompose within and across variability for the simple case of a single independent yield. In this case, the mean dissipation of a type t is given by

$$\bar{\sigma}(t) \approx r_\mathrm{ed}(t) - \bar{y}(t) r_\mathrm{b}(t) \quad , \tag{E4}$$

where the approximation arises from $r(t,i) \approx r(t)$, and thus we are ignoring changes in temperature and biomass composition within type, which as we said are small. Using Eq. E4 we then have that variations in dissipation within type are given by

$$\delta\sigma(i_t) \approx -r_\mathrm{b}(t)\delta y(i_t) \quad . \tag{E5}$$

This second expression explains the origin of the linear relationships for yield-dissipation variations within metabolic type observed in Figs. 2C and Fig. 4, see also Fig. SI9. Overall, the linear relationship within type arises directly from element conservation for the case of a single independent yield, due to the small variability on driving forces. In the

case of multiple independent yields, deviations from this linearity are expected and observed, and these deviations reflect the role of the additional yields, which is often also small.

Taking into account the approximations above, it is now easy to identify the origins of the deviation coefficient β and the across cost α. For the coefficient, we simply have that $\beta \approx -r_{\rm b}$, which we emphasize again only depends on stoichiometry and chemical potentials. We note that, in all metabolic types of our database, the force $r_{\rm b}$ is positive. Therefore, the yield of biomass is a thermodynamically unfavorable *anabolic* flux that needs to be driven by the remaining *catabolic* fluxes. The across cost instead stems from Eq. E4. Since in a large part of the data the second term on the right hand side is sub-dominant, the force $r_{\rm ed}$ approximates the driving *catabolic* force of growth. We then have that the conservation of the cost across metabolic types arises from the relationship $r_{\rm ed}(t) \approx \alpha(t)\bar{y}(t)$, which is shown in Fig. 5.

3. Efficiency

Having performed the decomposition in Eq. E2, it is possible to define a thermodynamic efficiency for the overall growth process. We follow the approach of [40] to define the efficiency as the ratio of the output dissipation over the input dissipation. The output dissipation is simply the dissipation in which the generalized thermodynamic forces oppose the corresponding fluxes, and the input dissipation is the opposite. For substrates, input dissipation is the sum of terms with positive force, $r^+_{i_{\rm i}} \equiv r_{i_{\rm i}} > 0$ for $i_{\rm i} \in$ sb, and output dissipation with negative force, $r^-_{i_{\rm i}} \equiv r_{i_{\rm i}} < 0$ for $i_{\rm i} \in$ sb. For products and biomass, the signs are opposite because of the minus sign in Eq. (E2).

In our case, we empirically found that the electron donor force is always positive, $r_{\rm ed} > 0$, and the biomass flux opposes its corresponding force, i.e. $r_{\rm b} > 0$ (except in three cases). Therefore, we define the efficiency as

$$\eta = -\frac{-y r_{\rm b} + \sigma_{\rm out}}{r_{\rm ed} + \sigma_{\rm in}} \quad , \tag{E6}$$

with $\sigma_{\rm in} = \sum_{i\in{\rm sb}} r^+_{i_i} y_{i_i} - \sum_{i\in{\rm pd}} r^-_{i_i} y_{i_i}$ and $\sigma_{\rm out} = \sum_{i\in{\rm sb}} r^-_{i_i} y_{i_i} - \sum_{i\in{\rm pd}} r^+_{i_i} y_{i_i}$. According to this definition, the thermodynamic efficiency is always positive and smaller than one, and it recapitulates the intuition of a coupling between anabolism (in the numerator) and catabolism (in the denominator). Notice that, for the three exceptional cases with negative $r_{\rm b}$, the corresponding term appears in the denominator of Eq. (E6).

4. Origin of the differences in deviation coefficient β for aerobic and anaerobic types

We now discuss the typical values of $r_{\rm b}$, which approximates the coefficient β, for different metabolic types. For most types, we have that $r_{\rm b} \sim 100$ kJ/Cmol. As pointed out in the main text, aerobic respiration has an exceptionally higher value, ≈ 500 kJ/Cmol. Such a thermodynamic force opposing biomass synthesis implies that a greater biomass yield is achieved with a sharp decrease in dissipation. However, the reduction of oxygen to water is not the only metabolic process associated with a higher value of the force $r_{\rm b}$. The metabolic reduction of nitrate, NO_3^-, or nitrite, NO_2^-, to ammonium, NH_4^+, can give $r_{\rm b} \approx 300$ kJ/Cmol. Moreover, when the reduction of nitrate to nitrogen, N_2, is coupled to the autotrophic reduction of CO_2, we find that $r_{\rm b} \approx 500$ kJ/Cmol, as in aerobic respiration.

To understand the origin of the greater force opposing aerobic biosynthesis, we can look at the definition given in SI E 1,

$$r_{\rm b} = g/\rho_{\rm C} - \sum_{i_{\rm d}} E_{i_{\rm d}} \mu_{i_{\rm d}} \quad . \tag{E7}$$

The value of $r_{\rm b}$ is the result of a summation. We find the main contribution to come from the chemical potential of carbon dioxide (i.e. bicarbonate), at ≈ -400 kJ/mol (≈ -600 kJ/mol respectively). This role of carbon dioxide extends to the anaerobic respiration cases where $r_{\rm b}$ has the exceptionally low values cited above.

However, while necessary, the presence of carbon dioxide is not sufficient to determine the value of $r_{\rm b}$. Indeed, carbon dioxide is produced in many anaerobic processes and yet the corresponding value of $r_{\rm b}$ is modest. The reason is that each chemical potential μ_{i_d} in Eq. (E7) is multiplied by the stoichiometric pre-factor E_{i_d}. In respiration with oxygen or nitrate/nitrite, $E_{CO_2} = 1$. In most cases of anaerobic fermentation, $E_{CO_2} \ll 1$, suppressing the contribution of the chemical potential. For example, the anaerobic fermentation of glucose to acetate produces carbon dioxide. The pre-factor suppressing the chemical potential of this latter reaches $E_{CO_2} \sim 0.01$ mol/Cmol. Finally, in most cases of anaerobic respiration $E_{CO_2} = 1$, but the effect of carbon dioxide is counterbalanced by the other terms in the sum. For example, when sulfate, SO_4^{2-}, is reduced to sulfide, H_2S, the contribution of carbon dioxide's free energy is fully counterbalanced by the term due to sulfate itself, with $\mu_{SO_4^{2-}} E_{SO_4^{2-}} \sim 400$ kJ/Cmol.

Appendix F: A thermodynamic database of microbial growth

1. Data curation and validation from primary literature sources

Our thermodynamic analysis of growth is based on data compiled from diverse sources. The database contained in the file `thermo_growth_dat.csv` consists of unicellular growth parameters across a wide range of species, environmental conditions, and metabolic types. All data used in this work was collected from published literature, and with minimal exceptions, we used primary sources (see references below). These primary sources were found by extensive literature search, using also as basis previously published meta-analyses such as [30, 64]. We note that such previous meta-analysis were only used as a source of references, and in all cases the data was curated directly from the original publications. For each experiment, we curated the following information and parameters when available:

- **Organism.** Species name.
- **Growth condition.** Whether the experiments were performed under conditions of continuous growth (i.e. chemostat), fed-batch culture, or batch cultures, see section B 1.
- **Environmental condition.** Chemical constituents involved in growth: carbon source, electron donor, electron acceptor, nitrogen source, and diverse metabolic products (these were indexed $i = 1, 2, \ldots$ in B 1). Also the growth temperature T and pH.
- **Macrochemical equation**. Macrochemical equations characterizing unicellular growth were curated when reported by the literature source, see section B 3. Missing macrochemical equations were constructed using curated information on trophism, metabolic substrates, products, and biomass stoichiometry. Whenever necessary we wrote the corresponding ionization states of chemicals, adding hydrogen ions to conserve charge.
- **Growth rate.** For batch experiments, maximal growth rates during exponential growth, γ; and for continuous culture experiments, dilution rates were curated, d (with $d = \gamma$, see B 2).
- **Yields and concentrations**. The reported yields of chemical species, y_i, including biomass, y. In addition, for chemostat experiments, concentrations of biomass or substrate chemicals (together with the dilution rate this allows to compute the chemical fluxes, $f_i^{+/-}$)
- **Stoichiometry**. Stoichiometric element composition of biomass, x_k, see section B 3. When stoichiometric information was available in the primary source, it was parsed. When it was not available, the stoichiometry of the particular unicellular species was taken from a stoichiometry dataset constructed from diverse sources, see H 1. When the species was not available in the dataset, stoichiometry was assumed to be similar to the closest phylogenetic relative (by genus or domain) with available information on its stoichiometry, and matched accordingly.
- **Element recovery**. Whenever element recovery experiments were performed, we annotated the fraction of recovered elements (typically carbon).
- **Heat**. Calorimetric measurements of heat production, $\dot{q}$, often in the form of a yield $y_{\dot{q}} \equiv \dot{q}/f_{\mathrm{ed}}^{+}$

As we attempted to find thermodynamic patterns persistent across varied types of experiments, we excluded only a very small fraction of the data that we encountered. We only excluded data when the growth medium was such that we were not able to write down a sensible macrochemical equation for the growth process. We also excluded data when, upon balancing fluxes, there was a substantial violation of the macrochemical equation, with fluxes of particular chemicals taking the opposite sign to that reported in the literature, i.e. products becoming substrates. This for example occurred with data from [65], in the case of anaerobic fermentation.

Overall, our database contains 504 data points. Approximately half, 296 data points correspond to continuous culture experiments. The remaining instances correspond to batch cultures with 13 points corresponding to fed-batch culture. 98 data points contained calorimetric heat dissipation measurements, 56 instances reported one or more redundant yields and 51 data on carbon recovery. On occasion, data curation involved digitizing plots, adjusting mistyped units, or the inference of underlying macrochemical equations of unicellular growth. In section F 3, we provide details and comments on the data parsing of each primary literature source. Data curation was extensive and done with great care in order to standardize and convert the different reported units such, that they are directly comparable and usable for our minimal non-equilibrium thermodynamic framework. The units and typical values of variables used in this work are described in the section below, section F 2.

2. Units and typical values of variables used

Multiple choices of units for the thermodynamics of unicellular growth are used in the literature, depending on the particular scientific community. We now make some remarks on the choices of units and provide typical values of the different quantities that we used in this work. Unless explicitly stated otherwise, densities are provided per unit volume of the chemostat, and not per unit volume of the cellular colony. Whereas the latter was used to develop the formalism, both are equivalent at the steady state (results in the main text are given as ratios of densities, so this choice becomes irrelevant).

- **Chemical fluxes**. In chemostat experiments the exchange fluxes are measured from concentrations in the fermentation vessel, i.e. the environment, times the dilution rate. A typical value is $F^{+}_{\rm ed} \sim 10^{-1}\,{\rm g/L\,h}$ for the glucose exchange flux.

- **Biomass stoichiometry**. Biomass is often characterized by a stoichiometric formula, ${\rm CH}_{x_{\rm H}}{\rm O}_{x_{\rm O}}{\rm N}_{x_{\rm N}}\ldots$. Here $x_k = c_k/c_{\rm C}$ measures how many atoms of the corresponding element are present per carbon atom, with c_k the elementary concentration defined below Eq. (B15). Typical values for stoichiometry used in the database are $x_{\rm H} = 1.7$, $x_{\rm O} = 0.5$, $x_{\rm N} = 0.2$.

- **"Molar" biomass weight**. It is common to measure biomass in units of carbon moles (Cmol). Provided a stoichiometric formula and the atomic masses, m_k, the mass per Cmol of biomass is given by $m_{\rm b} = (m_{\rm C} + m_{\rm H}x_{\rm H} + m_{\rm O}x_{\rm O} + m_{\rm N}x_{\rm N}\ldots)$ with a typical value of $m_{\rm b} = 24.5\,{\rm g/Cmol}$.

- **Biomass**. The biomass density ρ is the amount of dry biomass per volume and is measured in g/L, whereas the carbon density $\rho_{\rm C}$ is measured in units of Cmol/L. The former units are sometimes indicated by "g.d.w." (gram dry weight). The biomass concentration in the vessel can be measured sampling a given volume of the media and weighing the corresponding amount of dry biomass. A typical biomass density for a continuous culture of *E. coli* is $\sim 0.1\,{\rm g/L}$ or $\sim 10^{-2}\,{\rm Cmol/L}$, with a vessel volume of $\sim 0.1\,{\rm L}$.

- **Growth rate**. The growth rate γ takes values ranging from $\sim 1\,{\rm h}^{-1}$ in fast growing aerobic respiration experiments, to $\sim 10^{-2}\,{\rm h}^{-1}$ for slowly growing cells using inorganic electron acceptors.

- **Yield**. Yields are defined as the ratio of fluxes and their units are dimensionless. Since the fluxes change units in different ways, it does make a difference whether fluxes were measured in grams, moles, or else. To keep the same notation y without ambiguity, the units of the fluxes are often specified as g/g, mol/mol, Cmol/mol, etc.

  For example, the biomass yield is $y = \gamma\rho_{\rm C}/f^{+}_{\rm ed}$ in Cmol/mol and $Y = \gamma M/\mathcal{M}^{+}_{\rm ed}$ in g/g. The two definitions are related by the change of units $m_{\rm b}/m_{\rm ed}$, see also section B 2. For the case of *E. coli* growing on glucose, we have $y \sim 1\,{\rm Cmol/mol} \sim 0.1\,{\rm g/g}$. In Fig. SI10 we show how changes in the units of yield have little effect in our results.

- **Free energy, enthalpy, entropy**. Thermodynamic potentials are measured in standard units per mol (or per Cmol, in the case of biomass) using free energies, enthalpies, and entropies of formation. For example, the biomass enthalpy of formation is $h \sim -10^2\,{\rm kJ/Cmol}$, as discussed in detail in section H, and the free energy of formation of glucose is $\mu_{\rm gl} \sim -10^2\,{\rm kJ/mol}$.

- **Heat**. The heat density flux, $\dot{q}$, has units of energy per unit volume and time, i.e. power density. In experiments, it is reported the rate of heat produced per biomass grown, with typical values of $\dot{q}/\gamma\rho_{\rm C} \sim 10^2\,{\rm kJ/Cmol} \sim 10\,{\rm kJ/g}$ for *E. coli* growing aerobically on glucose, where the grams refer to dry weight. Because the yield of *E. coli* on glucose is $y \sim 1\,{\rm Cmol/mol}$, we find that $\dot{q}/f^{+}_{\rm ed} \sim 10^2\,{\rm kJ/mol}$, which is the normalization used in the text to refer to heat as a part of the dissipation, σ.

- **Cost**. The dissipative cost of biomass, $\alpha = \bar{\sigma}/\bar{y}$, is measured in units of energy per amount of biomass synthesized, measured in either grams or moles of carbon. A typical value for the cost is $\alpha \sim 500\,{\rm kJ/Cmol}$, which can also be expressed as $\alpha \sim 210\,k_{\rm B}T/{\rm C}$ (per atom of carbon in the biomass). See section D for different unit choices.

3. Comments on data curation and parsing from primary literature

We now provide some comments on the parsing procedure. Due to the diversity of experimental methods and conventions used in the literature, a significant effort was dedicated to standarizing the data. The comments below

are not meant to be an exhaustive description of the parsing done in each paper, but to provide the necessary cues to address the peculiarities from each paper. For instance, when multiple yields reported could result in ambiguity, we comment which ones were taken; we specify the constituents for which yields were reported, when these were not biomass; we specify which references contained heat measurements, and how these were obtained; we specify some unit conversions; etc.

- Andersen and von Meyenburg [66]: We obtained yields by dividing reported substrate consumption and product formation yield rate by the growth rate.
- Badziong and Thauer [67]: We parsed the yields at the reported pH optima for growth.
- Badziong et al. [68]: We curated the yields obtained by C-14 acetate labeling (Table 4). Biomass yields were taken from Table 3 reporting the incorporation of CO_2 and acetate carbon into biomass.
- Brandis and Thauer [69]: Two experiments were not performed in minimal media and required low concentrations of yeast extract. However, C-14 labelling revealed those sources are not used for energy or as carbon sources and that acetate is fully incorporated into biomass. Biomass yield were reported as intervals. We computed the mean biomass yields (e.g. $9 - 10\,\mathrm{g/mol}$ as $9.5\,\mathrm{g/mol}$).
- Chua and Robinson [70]: This work contains both batch and chemostat experiments at various pH and dilution rates. We calculated the yields using Pirt's equation and data provided in Table 1 using a dilution rate of $D = 0.08\,\mathrm{h}^{-1}$. Minimal media contained additional cysteine.
- Clarens and Molleta [71]: The paper reported product yields (methane on acetate) and growth yield in response to variations of acetate concentration, temperature, and pH. Data points are averages of triplicate measurements which were parsed. From Tables 2&3 and Figures 4&5, we curated the corresponding growth rates.
- Crabbendam et al. [72]: This study uses chemostat cultures to investigate growth in response to variation in glucose concentration, pH, and dilution rate. Product yields are reported as rates. We obtained yields by dividing the reported yield rates by the dilution rates. The experiments in Table 4 contained mannitol in addition to glucose. We assumed glucose as the main electron donor and attribute the total concentration of mannitol and glucose to glucose for simplicity.
- Dijkhuizen et al. [73]: We parsed the yields reported in Table 1. The authors declare a good agreement between predictions based on element balance and measurements. The authors measured and reported biomass stoichiometry and found it to be independent of the dilution rate.
- Esteve-Núñez et al. [74]: We parsed yields and the steady-state concentration of biomass, electron donor, electron acceptor, fumarate, mannitol, and succinate.
- Goldberg et al. [75]: We parsed the yields reported in Table 1.
- Hernandez and Johnson [76]: We curated the yield of biomass and oxygen reported in Tables 1, 3, and 5 as well as cell concentration. We converted the reported generation time, t_{g}, to growth rate through $\gamma = \ln 2/t_{\mathrm{g}}$.
- Heyndrickx et al. [77]: Product yields and carbon recovery data from experiments varying dilution rate, pH, and substrate concentration were parsed from Tables 2, 3, and 5. We excluded the yields containing mixed carbon sources and data that was insufficient to solve the stoichiometry of the marco-chemical equation. The media for the batch cultures used in this work was not minimal and therefore we did not parse the data.
- Huser et al. [78]: We parsed the reported yields of methane produced per acetate consumed.
- Ingvorsen et al. [79]: The parsed yield is the average of the reported batch cultures.
- Laanbroek et al. [80]: The basal media contains yeast extract. The biomass is estimated by carbon measurement. We parsed the biomass yields from Table1. We converted units of grams of carbon per moles of electron donor to carbon moles dividing by carbon molecular weight. We also parsed the yields of acetate produce per ethanol consumed.
- Lin et al. [81]: The media contained peptone and thioglycolate. All product yields were measured and we curated them accordingly.
- Lovitt et al. [82]: The minimal media contains essential amino acids and fatty acids. We parsed all chemostat yield data reported in Table 3 as well as the reported carbon recovery.

- Mayberry et al. [83]: We parsed Table 2, which reports averages from repeated experimental runs for concentrations and yields of biomass and oxygen.
- Myers and Nealson [84]: The defined media contains amino acids. We parsed the yield reported in the text and shown in Figure 3A. The yield was obtained by measurements of cell count and converted by the authors to grams of cells assuming a conversion factor of $2.8 \cdot 10^{-13}$ g/cell.
- Patel [85]: The media contains cysteine. We parsed the biomass and methane yields for two acetate concentrations reported in the text.
- Peters et al. [86]: We parsed data from Table 1 of growth on either lactate or H_2 at different dilution rates and temperatures.
- Pfennig and Biebl [87]: The authors measured biomass dry weight by cellular nitrogen measurement. We parsed biomass and reduced electron acceptor yields from Tables 2 & 3 as averages from individual reported replicates.
- Robinson and Tiedje [88]: The media contains cysteine. The yield of proteins on electron donor reported is obtained by fitting an integrated Monod equation for substrate utilization. We used a protein content of biomass of 55%, in analogy with [89] to obtain biomass yields. For sulfidogens and methanogens, we parsed the initial concentration of H_2 taken from Figures 4 and 5 respectively.
- Roden and Lovley [90]: We only parsed yields of electron acceptor reduced per electron donor consumed of cultures growing in the presence of MnO_2, in which yeast extract was not present.
- Rutgers et al. [91]: We parsed yields and dilution rates reported in Table 1 which have been calculated using Pirt equation from experimental data. The biomass is expressed in carbon moles and the stoichiometry has been reported in the paper.
- Sanford et al. [92]: Media contains additional cysteine. Measurements of 16S rRNA gene copy numbers served as a proxy for the number of cells. Transformation to dry weight was done by the authors assuming $3.8 \cdot 10^{-14}$ gr RNA/cell. We parsed data reported in Table 1 for growth of *A. dehalogenans* and the two yields for the *Geobacter* species reported in the text (in cell numbers).
- Sass et al. [93]: We parsed the data reported in Table 2.
- Schauer and Ferry [94]: The minimal growth media contained small amounts of yeast extract and cysteine. We parsed the biomass yield for methane and formate.
- Smith and Mah [95]: The minimal growth media contained small amounts of yeast extract and cysteine. We parsed the yields reported in Table 1. Based on the data provided, we assumed acetate as the sole energy and carbon source.
- Zehnder and Wuhrmann [96]: The media contained additional cysteine. The authors conclude cysteine is an essential sulphur source contributing 0.04 mg per mg of biomass. We assumed the contribution of cysteine to the overall thermodynamic and element balance to be negligible and parsed data reported in Table 3 for "enrichment media" (not containing yeast extract).
- Weimer and Zeikus [97]: We parsed yields of biomass over methane produced. CO_2 and H_2 were reported to be the sole carbon and energy sources.
- Stieb and Schink [98]: We parsed data reported in Table 2. We derived the yield from reported measurements of product and biomass concentrations.
- Strohm et al. [99]: We parsed data from Tables 1, 2 and 3.
- Szewzyk and Pfennig [100]: We parsed data from Table 1 (batch cultures), and Table 2, (chemostat experiments). The authors report a carbon recovery between 85 to 108 percent for performed experiments.
- Tang et al. [101]: We parsed the biomass yields from Table 1 (for chemostat experiments). The authors used biomass stoichiometry data from previous literature.
- Vasiliadou et al. [102]: The media contains tap water (synthetic wastewater). *Acinetobacter* grows in two phases, with nitrite accumulation and consumption. We parsed the yields from Table 5, which were obtained by fitting of the data with a kinetic model by the authors. We converted grams of nitrate nitrogen to moles using the molecular weight of 76 grams per mole. We converted grams of nitrite nitrogen to moles using the molecular weight of 60 grams per mole.

- Wallrabenstein et al. [103]: We parsed data reported in Table 1 containing growth yields, biomass, and product concentrations as well as the reported carbon recovery.
- Widdel and Pfennig [104]: The biomass is measured directly as dry weight. We parsed the data reported in Table 1 reporting differences in concentrations of biomass, electron donor used and electron acceptor reduced at different stages of growth compared to controls. We computed the yield from a least squares errors linear fit of the reported concentrations.
- Widdel and Pfennig [105]: We parsed the data in Table 2 reporting differences in concentrations of biomass, electron donor used and electron acceptor reduced at different stages of growth compared to controls. We computed the yield from a least squares errors linear fit of the reported concentrations.
- Widdel and Pfennig [106]: We parsed the data in Table 2 reporting differences in concentrations of biomass, electron donor used and electron acceptor reduced at different stages of growth compared to controls and computed the yields from the total concentrations of electron donor used. Table 3 contained concentrations of electron donor used and electron acceptor reduced. We parsed these data taking the ratios as yields.
- Yang and Okos [107]: The media contained small amounts of yeast extract. We parsed biomass yield from Figure 3 and text, which was computed from the biomass formed at the end of the exponential phase with respect to acetate consumption.
- Yoon et al. [108]: The dry weight of biomass was obtained by correlating optical density and gene copy number. We parsed data reported in Table 1.
- De Poorter et al. [109]: Media contained additional cysteine. The continuous cultures were operated at steady-state, whereas the fed-batch experiments had lag-exponential-linear phases of growth. Table 3 reports the maximum yields or the yield in the linear growth phase. In the linear phase, which we parsed, the fluxes have the same time dependency, which make them consistent with our framework, see section B 2.
- de Vries et al. [110]: The media contained a small amount of trypsin digested broth. The dry weight of biomass was measured directly. We parsed yields from Table 1.
- Von Stockar and Liu [60]: This paper reviews previously published results by the authors. We parsed two data points reported in Table 5. The original reference is Duboc [111]. This work contained calorimetric heat measurements.
- Brettel et al. [112]: This work reported the overall yield of biphasic aerobic growth on glucose, with production and subsequent utilization of fermentation products such as ethanol. We parsed biomass yields, reported in grams. We converted them to carbon moles using the average value of $\approx 25\,\mathrm{g/Cmol}$ based on a standard biomass stoichiometry. We also parsed heat measurements done with a flow micro-calorimeter.
- Birou et al. [113]: We parsed the yields from Table IV and one data point from Table VII. When product formation was reported, we computed missing yields using the reported carbon recovery (Table VII). In three instances, we further used the reported measurements of heat dissipation yields (energy recovery, ER). We excluded these data points in our validation of Hess' law. Yields using the energy recovery were computed according to the formula: $f_{\mathrm{p}}^{\mathrm{b}} = \frac{(1-\mathrm{ER})h_{\mathrm{ed}}^{\mathrm{c}}}{h_{\mathrm{p}}^{\mathrm{c}} f_{\mathrm{b}}^{\mathrm{ed}}}$, where ER is the value reported in the original paper, $f_{\mathrm{p}}^{\mathrm{b}}$ is the product yield on biomass and $h_{\mathrm{p}}^{\mathrm{c}}$ and $h_{\mathrm{ed}}^{\mathrm{c}}$ are the enthalpies of combustion of product and electron donor, respectively. Heat measurements were performed with a modified bench-scale calorimeter.
- Dejean et al. [114]: We parsed the growth yields and heat measurements reported in the main text and in Fig. 2 of the reference. The yeast has two exponential growth phases, for which data are reported separately. Therefore, we constructed two macrochemical equations for the two phases. We converted the measured biomass from grams to carbon moles using the molecular weight 24.8 g/Cmol and ash fraction reported in the paper. We obtained the products yields from the enthalpy yields using the combustion enthalpies reported, together with the aid of Fig. 3 of the reference. From the calorimetric-respirometric ratio reported in the reference, we could compute the oxygen yield. We compared this latter to the oxygen yield obtained enforcing element conservation.
- Battley [115]: We parsed the yields, biomass composition, macrochemical equation, and heat dissipation measurements corrected for side processes other than growth.
- Dermoun and Belaich [116]: We parsed the reported yields, macro chemical equation, and heat dissipation measurement (performed on a modified differential scanning calorimeter).

- Dermoun and Belaich [117]: We converted the biomass yield, reported in grams, to carbon moles using the reported biomass stoichiometries. We excluded data for growth on acetate due to mismatching enthalpy recovery and undetected product formation reported by the authors. Heat measurements performed by using a modified differential scanning calorimeter were also parsed.

- Belaich and Belaich [118]: We used only data for growth on glucose as other experiments provide no information on product formation. Part of the reported data can be originally found in an accompanying paper [119]. Calorimetry experiments were performed using a modified differential scanning calorimeter. The initial concentrations of substrates are provided as an upper bound. We converted the biomass yield to carbon moles using 22.8 g/Cmol.

- Ishikawa et al. [120]: The yields of biomass and heat reported correspond to ratios of integrated quantities. Of note, the authors ignore two growth phases, first acetate formation and then acetate utilization. The heat measurements were performed with a twin-type heat conduction micro-calorimeter. We remark that using the reported measured heat dissipation in this work to validate the predicted enthalpy dissipation based on the reported yield showed a large deviation. This work and other works by the authors [121] have been criticized in Battley [21] for being a scenario of non-balanced growth. Marison and Von Stockar [122] argue that the observed enthalpy balance mismatched are due to undetected metabolic products.

- Ishikawa and Shoda [121]: We digitized Fig. 2 and 4 and parsed Table I of the paper to extract the data. We matched the heat measurements with the biomass/nutrient concentrations according to the dilution rate. We established a threshold of $0.003\,\mathrm{h}^{-1}$ discrepancy in dilution rate to reject a match. At a high dilution rate, the production of acetate is reported but no measurements are provided. Taken together, we only considered two out of 13 reported data points. We converted grams of biomass to carbon moles using Table I of the paper.

- Tamiya [18]: Data was parsed from Battley [64], as there was no access to the primary source. Of note, Tamiya et al. are the first to elaborate on the concept of degree of reduction and anabolic growth equations. The element analysis of biomass was performed by Yamagata [123].

- Whelton and Doudoroff [124]: Data was parsed indirectly from Battley [64].

- Samejima and Myers [125]: Data was parsed indirectly from Battley [64].

- Hoover and Allison [126]: Data was parsed indirectly from Battley [64].

- Liu et al. [127]: We parsed yields and heat dissipation measurements performed with a bench-scale calorimeter. Acetate (electron donor) concentration was maintained constant to a value reported as $5 - 10\,\mathrm{g/L}$, pulsed in at a very low rate (we used 5 g/L).

- Birou and Von Stockar [128]: We digitized Fig. 1 to obtain concentrations of biomass, heat dissipated (measured with a bench-scale heat flux calorimeter), and residual glucose. We converted the biomass from grams to carbon moles using the molecular weight of *K. marxianus* biomass stoichiometry. We excluded data for the highest dilution rate data due to detection of uncharacterized products.

- Marison and Von Stockar [122]: We parsed the biomass and product yields from Table IV, and heat and concentration of chemicasls from Table II. Heat measurements were performed with a bench-scale heat flux calorimeter. We excluded the first two data points because the calorimeter operated close to the sensitivity limit. We used the biomass composition reported in the paper.

- Patino et al. [129]: We parsed reported yields, heat dissipation, and macrochemical equations with biomass composition. Heat measurements were performed with a modified commercial reaction calorimeter. We excluded data where biomass growth measurements were out of the sensitivity range of the measuring device.

- Zhao et al. [130]: We parsed yields and fluxes of biomass, CO_2, and acetate from continuous cultures. The organisms used in the experiments include genetic variants and mutants of the same species.

- Hempfling and Mainzer [65]: We used 24 g/Cmol to convert the biomass measurements. We took aerobic data from Tables 2 and 3 (converting grams of oxygen to moles). We excluded fermentative data from Table 1 because we could not write a reliable macrochemical equation. Either element conservation was violated, or the second law was violated, or we did not have enough yields measured to determine the system.

- Mainzer and Hempfling [131]: We treated the data analogously to that in Hempfling and Mainzer [65]. We parsed anaerobic data from Table 1 and aerobic data from Table 3.

- Verduyn [132]: We parsed the data on *C. utilis* growth and biomass composition reported in Table 8 p.143. The composition of biomass for *S. cerevisiae* is taken from Eq. 3 in Verduyn et al. [133]. In experiments with alanine and glutamate, we included ammonia as a product to allow for nitrogen balance.
- Bernacchi et al. [134]: Growth media contained additional cysteine, which was described as not being a carbon source. We parsed the balance of carbon, nitrogen, and degree of reduction from Table 1 and used a macrochemical equation based on Schill et al. [42].
- Liu et al. [135]: We parsed reported data by digitizing Figs. 3, 4A, 4B, 6 and 7. We computed yields from fluxes and took the biomass composition and macrochemical equations from Schill et al. [42].
- Schill et al. [42]: We parsed reported data by digitizing Fig. 1. The heat measurements were performed with a bench-scale heat flux calorimeter.

Appendix G: Thermodynamic properties of substrates and products

According to the framework in B 4, a complete thermodynamic description for a growing cellular colony requires detailed knowledge of the free energies of formation for all substrates and products of metabolism (as well as of biomass, see H). The free energy per molecule for an aqueous solution of a chemical species i is given by its chemical potential

$$\mu_i = \mu_i^{\mathrm{o}} + RT \ln\left(\gamma_i \frac{c_i}{c^{\mathrm{o}}}\right) \quad , \tag{G1}$$

where μ_i^{o} is the standard chemical potential measured in some standard state, c_i is the concentration, c^{o} the concentration at the standard state, and γ_i is the activity coefficient that measures deviations from the ideal approximation ($\gamma_i = 1$ in the ideal case).

The experiments that we study were normally performed under constant pH conditions. Therefore, many chemical species existed in one (or multiple) ionization states. This has two implications for the precise computation of the chemical potential. First, the activity coefficient γ_i deviates from unity due to the presence of Coulomb interactions, and is given by the Debye-Hückel equation for ionic solutions [136]. Second, the existence of multiple ionization states for chemicals requires that either (i) the Alberty's method is used for correcting the chemical potential [136–138], or (ii) ionized equations are written for each ionization state subspecies keeping track of charge balance and including the pH value in the free energy of the charged hydrogens [139, 140]. Concerning the Debye-Hückel correction to the free energy, in practice its role is minor. Concerning the role of multiple ions, it is often the case that one ionization state dominates over the rest, so that keeping track of the dominant ion provides a good approximation.

For completeness, in our work we used two different approaches to deal with the role of ionization. On the first approach (contained in the file `thermo_chem.csv`), which is presented in the main text, we neglected the Debye-Hückel correction and considered only the charge-balanced macrochemical equation for the dominant ion. We therefore did not transform the chemical potentials using Alberty's method, and instead included the free energy of hydrogen ions that compensate the charge imbalance, $\mu_{\mathrm{H}^+} = -RT\mathrm{pH}\ln 10$. On the second approach, we used the Alberty's method and therefore included both types of corrections. The uncorrected values of the free energy, (i), were taken from the references indicated in `thermo_chem.csv` file. The corrected values of the free energy, (ii), were taken from the eQuilibrator database [141], setting the corresponding value of pH, a typical ionic strength of 0.25 M and pMg value of 3. As we lacked the value of the concentrations of chemicals in most instances, we used $c_i = 1\,\mathrm{mM}$ as commonly done in the literature [142]. As one can see in Fig. SI12, the differences between both approaches are only quantitative. Furthermore, an increase/decrease in concentration of three orders of magnitude changes the chemical potential by $\approx \pm 17$kJ/mol, which is often a negligible contribution compared to the value of the chemical potential itself.

Key to the validation of our dataset, we compared the heat dissipation measurements available with the enthalpy balance according to Eq. (B19), see I. To compute the balance, we collected the enthalpies of formation in aqueous solution of substrates and products of metabolism, as well as of biomass (see H). The enthalpies used are reported in the file `thermo_chem.csv` together with the free energies and the reference to the sources of the data. In the choice of the sources, we prioritized using the same source for as many entries as possible, so to guarantee self-consistency within the database. The enthalpies were also used to correct the use of standard free energies at non-standard temperature, according to the Gibbs-Helmholtz equation

$$\mu_i^0(T) = \mu_i^0(T^0)\frac{T}{T^0} + h_i\left(1 - \frac{T}{T^0}\right) \quad , \tag{G2}$$

where $T^0 = 298$ K and T is the actual temperature of the experiment. The effect of such correction is typically small.

Appendix H: Thermodynamic properties of biomass

1. Biomass composition

The biomass composition is characterized by the concentration of each of its constituents, c_i, with $i = 1, \ldots, M$, see section B 3. However, as we have seen in section F 2, for our purposes it is more convenient to describe biomass by its elementary composition normalized to carbon content, $x_k = c_k/c_{\mathrm{C}} = \sum_i e_{ik}c_i / \sum_i e_{i\mathrm{C}}c_i$, where e_{ik} is the elementary composition matrix introduced before. The elementary composition of biomass, also referred to as biomass stoichiometry [143], has been extensively measured through elementary analysis [144]. From these experiments, the elementary composition of different species is available. Because elementary composition of biomass is necessary for our analysis, the data-file `yields.csv` contains a specific biomass composition for each entry. We now describe how these biomass compositions were determined.

For data in which the primary sources described in section F 3 contained specific biomass composition, the corresponding stoichiometry was used, and inserted in our main data-file. However, many of the primary sources we used did not contain biomass stoichiometry data. We therefore compiled a comprehensive dataset of biomass stoichiometry that consists of: (*i*) all the stoichiometric data in the primary sources described in F 3, and (*ii*) the data in the meta-analysis study [145] that referred to species in our study. Our biomass stoichiometry dataset is in the file `biomass_composition.csv`, and includes dozens of different species. For some of the species there are multiple entries, from measurements performed by different groups or under different conditions. We filled the missing biomass compositions in our main data-file using the following approach: for species in our study that appear in `biomass_composition.csv`, we took the corresponding value (or the average, when multiple values appear in the biomass dataset); and for species in our study that did not appear in `biomass_composition.csv`, we took the average value of the genus or domain in `biomass_composition.csv`. While this approach leads to inaccuracies, the variations in biomass stoichiometry observed across species did not significantly influence the results (Fig. SI11).

2. Enthalpy, entropy and free energy of biomass

The thermodynamic description developed in section B 4 requires knowledge of the enthalpy, h, entropy, s, and free energy, g, of biomass. Because direct measurement of these quantities is challenging, they are often replaced by detailed theoretical estimates [45, 58, 146]. Despite this, we now provide an overview of existing measurements for enthalpy and entropy of biomass, and how these can be combined with the elementary composition of biomass to provide empirical estimates for the thermodynamic properties of biomass.

The enthalpy of biomass is obtained by direct measurement of the heat released during the combustion of dry biomass [147]. This "heat of combustion", q_{c}, is also called "enthalpy of combustion", as it refers to the enthalpy difference of the following (unbalanced) combustion reaction:

$$\mathrm{O_2} + \text{biomass} \rightarrow \mathrm{CO_2} + \mathrm{H_2O} + \mathrm{N_2} \quad . \tag{H1}$$

Typical values of the heat released are $q_{\mathrm{c}} \approx 500\,\mathrm{kJ/Cmol}$ for dry biomass, see Table III in [147] (as explained in section F 2, for practical purposes biomass properties are better reported on a per Cmol basis.). The heat released during the combustion of organic material has been shown to depend on the stoichiometry of the organic sample, $\{x_k\}$, and to be well approximated by the following empirical formula:

$$q_{\mathrm{c}} \approx p\, d(\{x_k\}) \quad . \tag{H2}$$

Here d measures the number of electrons transferred from the organic material to oxygen, and $p = 115\,\mathrm{kJ/Cmol}$ is an effective proportionality constant [148]. For biomass, we have that $d = 4x_{\mathrm{C}} + x_{\mathrm{H}} - 2x_{\mathrm{O}}$, which is therefore proportional to the so called degree of reduction, $d/4$, see [149] and references therein. The validity of this expression, when applied to biomass, has been verified in [149–151]. Once the combustion reaction has been determined, the enthalpy of biomass can be obtained by solving the corresponding enthalpy balance:

$$h = a_{\mathrm{N_2}}h_{\mathrm{N_2}} + a_{\mathrm{H_2O}}h_{\mathrm{H_2O}} + a_{\mathrm{CO_2}}h_{\mathrm{CO_2}} - a_{\mathrm{O_2}}h_{\mathrm{O_2}} - q_{\mathrm{c}} \quad , \tag{H3}$$

with a_i the appropriate stoichiometric coefficients determined by the biomass stoichiometry. Together, Eqs. (H2) and (H3) relate the composition of biomass to its enthalpy. We note that Eq. (H3) requires knowledge of the enthalpy of the chemical species participating in the reaction ($\mathrm{O_2}, \mathrm{CO_2}, \mathrm{H_2O}$, and $\mathrm{N_2}$). Here, we use the enthalpy of formation for these chemical species, which is the enthalpy change of a putative formation reaction of each molecule. The reactants of these formation reactions, chosen to be the elements in their more stable form at standard conditions, provide a

reference (zero) value for the enthalpies [152]. As a result, the enthalpy of biomass will be an enthalpy of biomass formation, with a typical value of $h^{\mathrm{f}} \approx -90$ kJ/Cmol. We also note that we have used measurements of combustion for dry biomass. Clearly, under physiological conditions, biomass is hydrated. However, the enthalpy from hydration is small, measured in [147] to be on the order of $q_{\mathrm{c}} \approx 1\,\mathrm{kJ/Cmol}$, and can therefore be neglected.

For the entropy of biomass, direct measurements rely on low-temperature calorimetry of pelleted dry biomass. This determines the heat capacity, which integrated returns the entropy, with a typical value $s \approx 30\,\mathrm{J/g\,K} \approx 1\,\mathrm{J/Cmol\,K}$ [153]. As hydration entropy has been estimated to have a minor contribution to the overall entropy [154], the value measured for dry biomass, as well as similar ones obtained more recently in the same way [155], can be used as approximate measurement of biomass entropy. One key point is that these entropy values have been shown to correlate well with the entropy of the elementary constituents of biomass [142, 154]. As a result, the following empirical formula that relates biomass entropy to its elementary composition has been proposed:

$$s \approx 0.2 \sum_k c_k s_k \quad , \tag{H4}$$

where s_k is the entropy per atom of the corresponding elementary form (e.g., gaseous H_2 for hydrogen, graphite for carbon, etc.). In particular, we used $s_{\mathrm{C}} = 5.8$, $s_{\mathrm{H}} = 130.7/2$, $s_{\mathrm{O}} = 205.1/2$, $s_{\mathrm{N}} = 191.6/2$, all units of J/mol K, taken from [156]. Because we have used enthalpies of formation, we also will use entropy of formation. We therefore relate the entropy in Eq. (H4) to that of a putative formation reaction. The resulting entropy, which we used in our work, is

$$s^{\mathrm{f}} \approx -0.8 \sum_k c_k s_k \quad . \tag{H5}$$

Note that, for consistency, entropies and enthalpies of all chemical species involved in the growth process will also have to be the ones of the formation reaction. To simplify notation, in the main text we omit the superscript. Finally, the free energy of biomass is obtained as $g = h - Ts$, with a typical value of ≈ -40 kJ/Cmol.

Appendix I: Dataset validation

Our work includes a large variety of data that we cross-referenced, including biomass stoichiometric data, growth yield data, and thermodynamic data of chemicals. Furthermore, each of these data types was obtained from a hetergeneous corpus of literature. Due to this diversity of sources and cross-referencing, validation of our dataset is crucial. Three key validation tests were performed: energy conservation, element conservation, and validity of the second law.

- *Energy conservation.* We validated energy conservation using Hess' law, which establishes that the heat due to growth equals the balance of enthalpies between substrates and products of metabolism, see Box 1 in main text and SI B 4. We therefore computed the enthalpy balance for a subset of the data (total of 98 data points) that contained calorimetric measures, and compared it to the reported heat. This comparison was done using per biomass quantities, i.e. yields as described in section B 2. The results of the comparison is shown in Fig. Box 1A. As one can see, over several orders of magnitude the heat yield predicted through enthalpy balance agrees well with calorimetric measures (see also Fig. SI6 for details on legends). Furthermore, this agreement extends to reported measures of endothermic growth, see inset. We note that in few instances, calorimetry (referred to as "energy recovery") was used to determine specific yields. These instances, for which there are comments in section F 3, were removed from our validation plot.

- *Element conservation.* In this work we tested element conservation in two different ways. First, we noted the instances in which the number of yields measured was higher than the number of yields needed to resolve all yields using element conservation (SI B 3). We then used element conservation to predict those yields that had already been measured. The results, which correspond to a total of 69 yields from 56 experiments and 13 different sources, appear in Fig. Box 1B. As one can see, there is a very good agreement, with most deviations occurring for very small yields (see also Fig. SI6 for details on legends). Second, we parsed the results of element (carbon) recovery experiments done in some of the growth data that we used (a total of 51 data points). The data appear as a histogram in the inset of Fig. Box 1B, and show that carbon recovery typically resulted in a very good validation of element conservation. As in the case of heat, in few instances the primary sources used element recovery experiments to determine a particular yield. These instances were removed from our validation plot.

- Finally, while no direct measurements of free energy dissipation have so far been reported in the literature, we tested whether the dissipation σ was positive for the collected data. Indeed, we found that 504 out of 509 data points complied with the second law. In 5 cases, the original reference did not report full information about products and substrates of the experiment. The tentative macrochemical equation that we postulated resulted in a negative dissipation and we did not include these data points in our study.

Overall, three different and independent tests were performed to validate the data collected in a total of 302 data points out of the 504 that we used in our study. Given the large variety of data included in this test-set, we find the results of the tests remarkably positive. This validation success supports the validity of our approach to the whole dataset.

Appendix J: Supplementary figures

FIG. SI1: *Legend of all metabolic types in our dataset.* This is a comprehensive list of all metabolic types in our dataset. Each symbol/color denotes the macrochemical equation next to it, written as `[substrates]-->[products]`, with the letter `b` denoting biomass. This legend is later used in Figs. SI7, SI8 and SI9

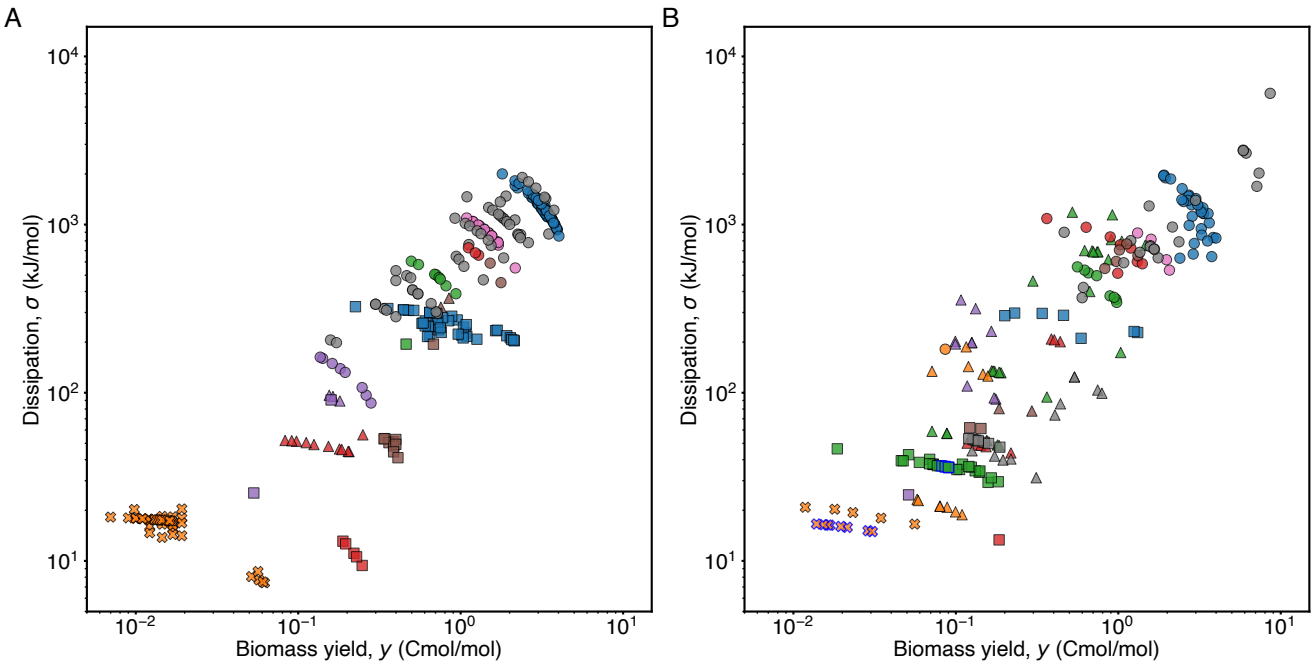


FIG. SI2: *Chemostated and batch cultures exhibit the same behavior.* **A.** Subset of Fig 2A for chemostated cultures. **B.** Subset of Fig 2A for batch cultures. Blue edges: fed-batch. While our theoretical framework is suited to chemostated cultures and batch cultures arrested in exponential phase, we have also used it on batch cultures allowed to saturate, where it constitutes a rougher approximation. As one can see, both sets of data exhibit similar behavior, showing that our conclusions are robust to difference in growth conditions.

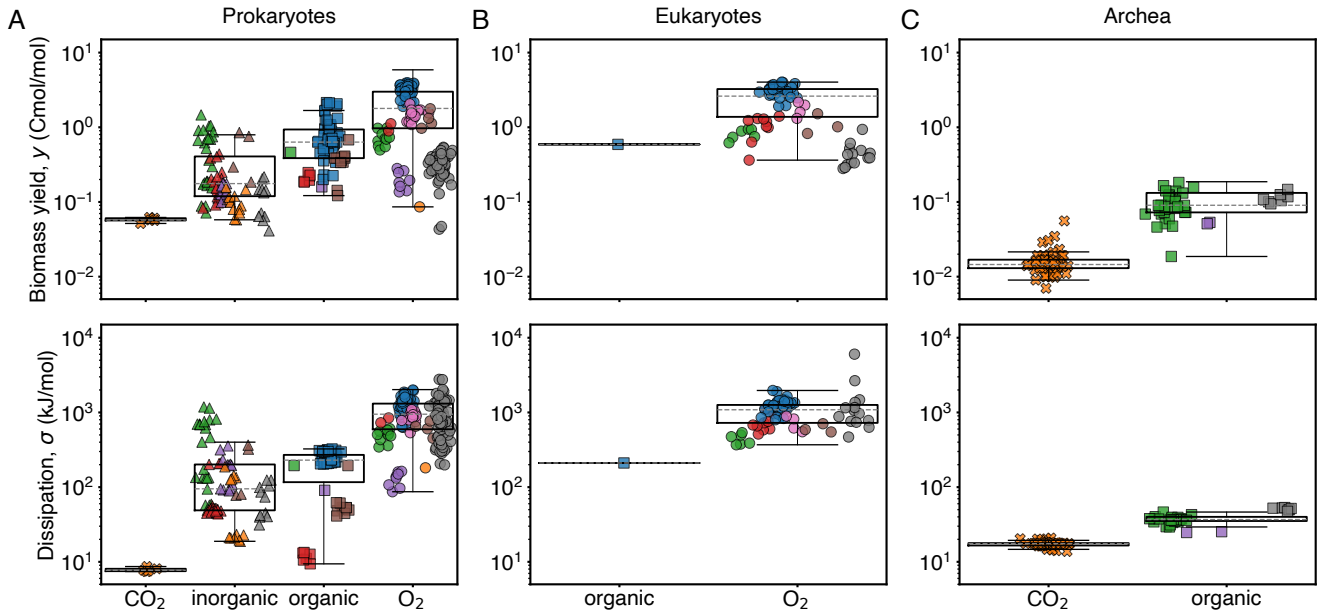


FIG. SI3: *Yield and dissipation separated by domains of life.* **A-C.** Biomass yield variability (upper row) as in Fig 1A, and dissipation variability (lower row) as in Fig 1B, for prokaryotes (A), eukaryotes (B), and archea (C). While eukaryotes have typically higher yield and dissipation than prokaryotes, the values are comparable to those of prokaryotes with similar metabolic types (e.g. aerobic respiratory types). While archae have generally lower yield and dissipation than prokaryotes, there are many prokaryotic data with comparably low yields. Moreover, autotrophic methanogenesis type is found exclusively in archea, and is the reponsible for the lowest yields and dissipations in the domain. Overall, we can not ascribe the observed variability to differences in domains of life.

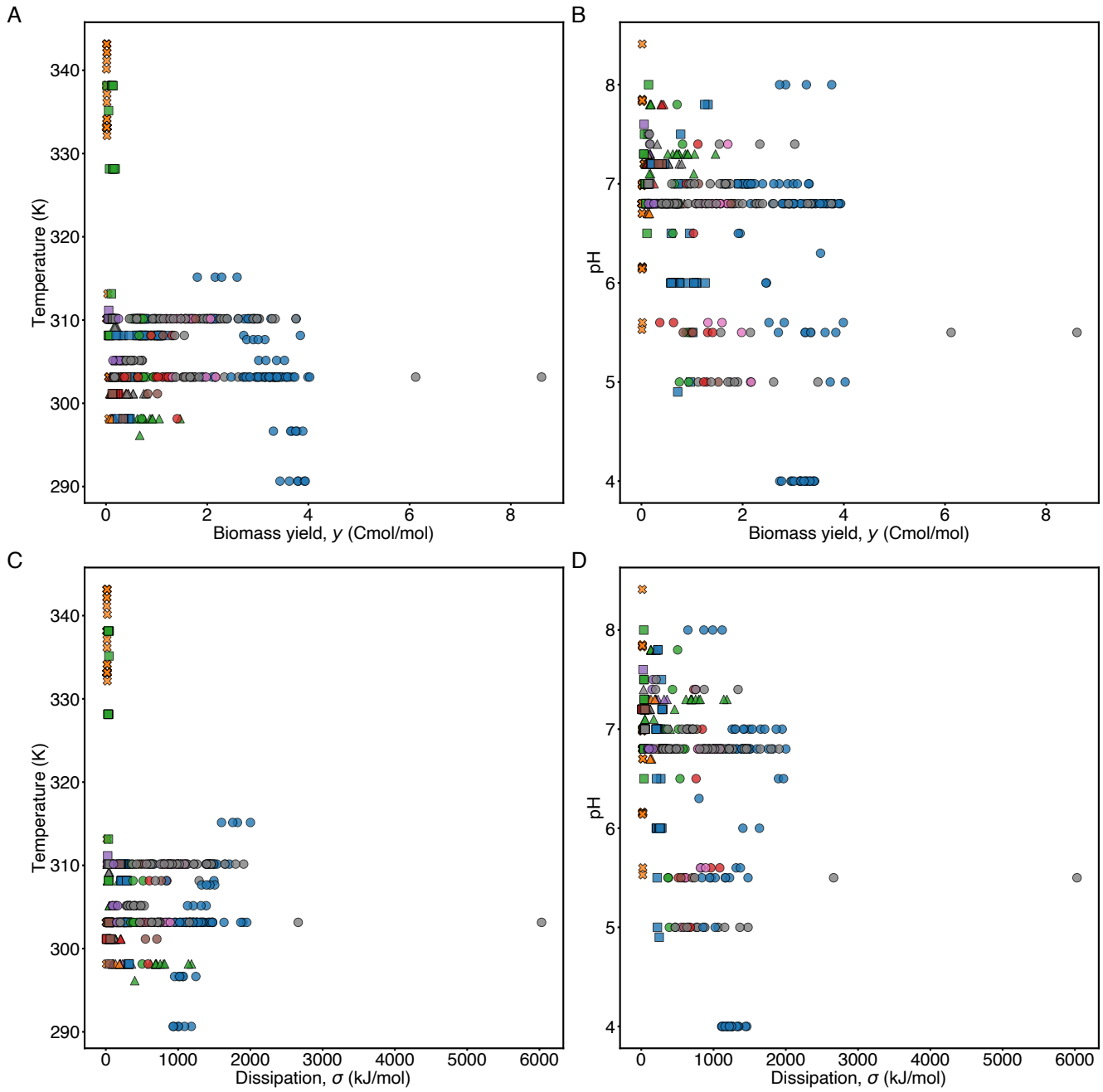


FIG. SI4: *Dissipation and yield as functions of temperature and pH.* **A.** Scatter plot of yield against temperature at which the corresponding experiment was performed. **B.** Scatter plot of yield vs pH of the media. **C.** Scatter plot of dissipation against temperature. **D.** Scatter plot of dissipation against pH. The extent of temperatures explored is large, $\approx 50^{\circ}$C, as also is the range of pH conditions, ≈ 4. Despite this, no consistent relationship between yield and temperature or pH is observed. The same is true for dissipation, despite the fact that pH and temperature corrections were included in computing the dissipation, see SI G.

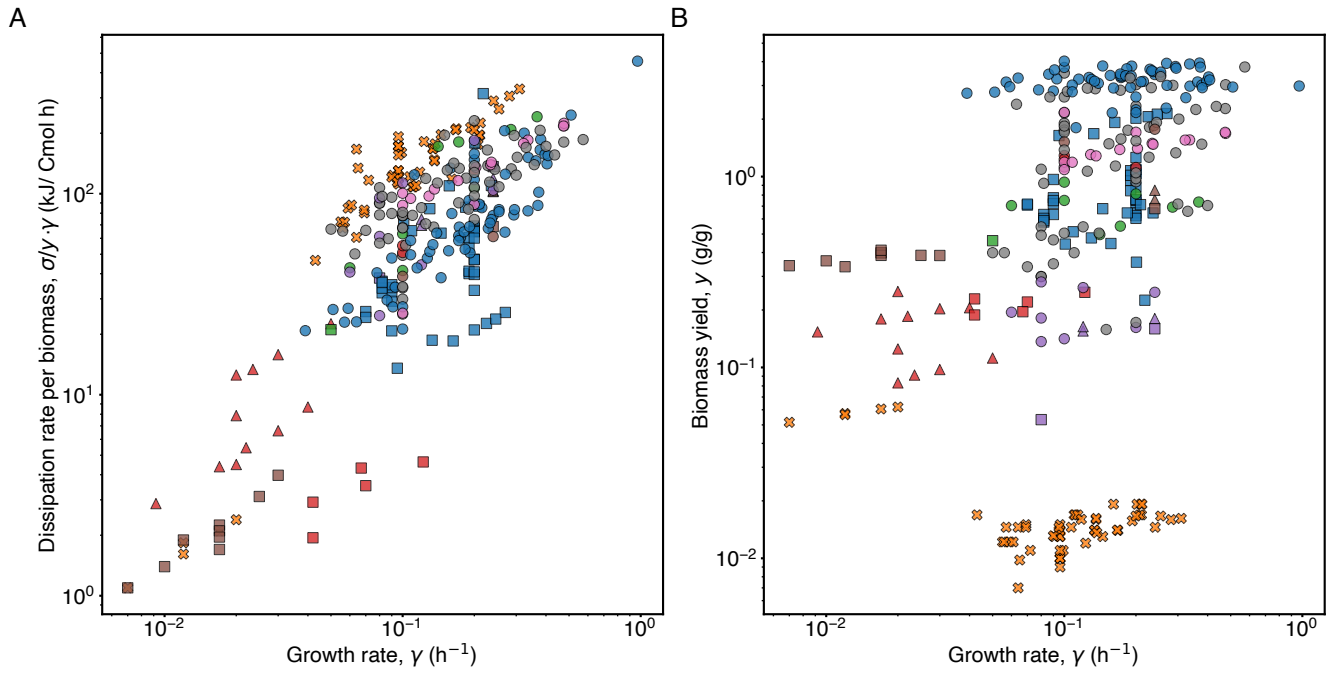


FIG. SI5: *Characterization of dissipation per unit time for chemostated data.* **A.** Dissipation per unit biomass and time, $(\sigma/y)\gamma$, plotted against growth rate. We find a roughly linear scaling for about two decades of growth rate. This is in agreement with the notion of cost conservation, which is robust to changes in growth rate, and compatible with observations in [11]. **B.** Biomass yield plotted against growth rate. We do not find any clear relation (i.e. no yield-growth rate trade-off), not within metabolic type nor across metabolic types.

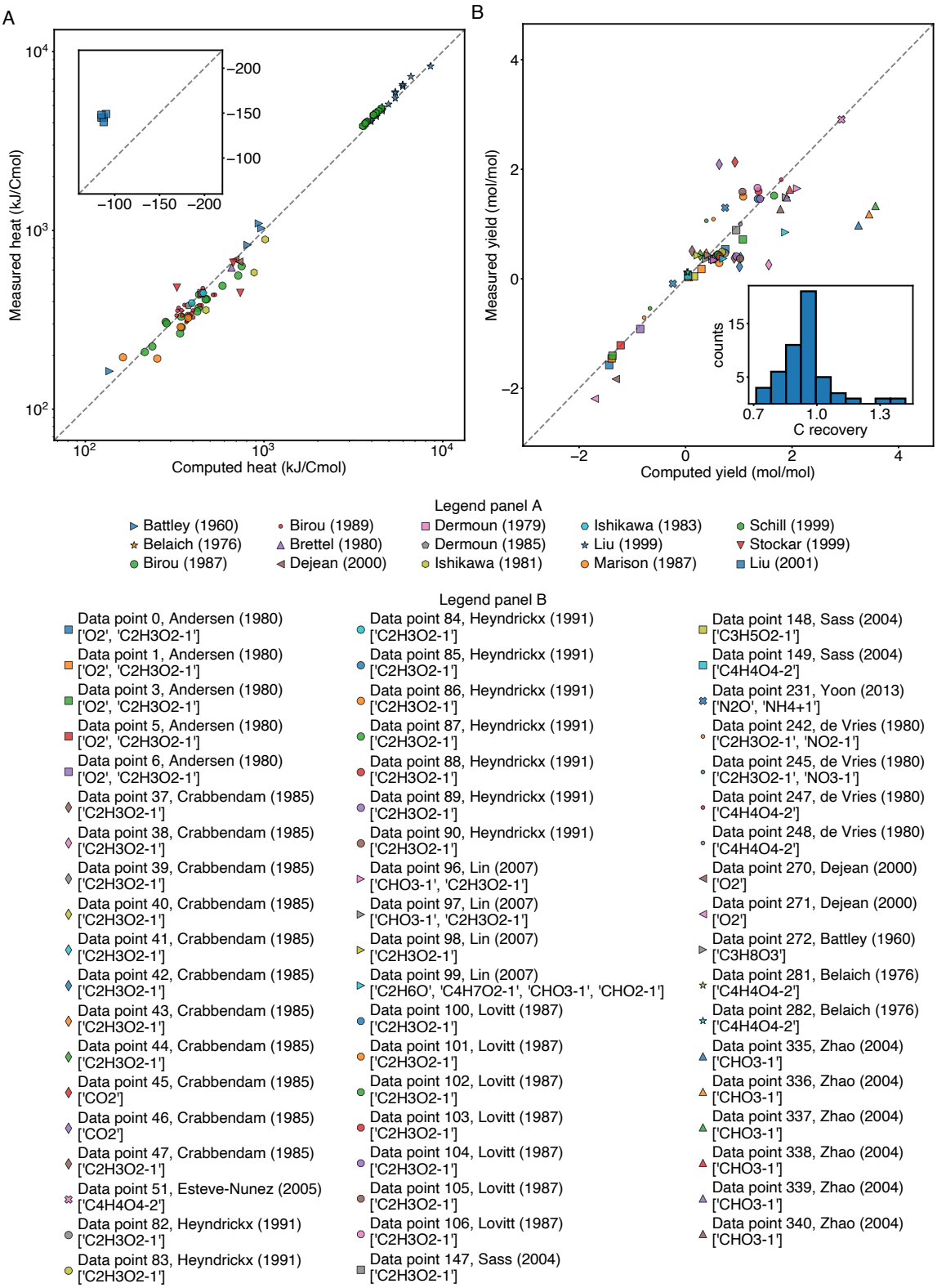


FIG. SI6: *Validation of Hess' law and element conservation.* **A.** (Minus) the heat per Cmol of biomass predicted using enthalpy balance is plotted against direct calorimetric measurements. The legend, below, includes the bibliographic reference (see main text for labelling by types). The inset corresponds to endothermic measurements. The deviations from the dashed line, which is not a fit, are small. This constitutes a validation of our approach. **B.** Measured yields of substrates and products, reported in the file `measured_yields.csv`, are plotted against the corresponding yields predicted using element conservation, reported in the file `yields.csv`. The sign distinguishes substrates, positive in this plot, from products, negative. The yields are reported in moles per moles of electron donor. The legend below indicates the "entry" number of the experiment where the fluxes were measured, the bibliographic reference of the experiment, and the stoichiometric formula of the yields measured, all of the above consistently with the notation used in the data-files. Values of carbon recovery, parsed from the primary sources used in this work, are plotted as an inset. As in the case of heat, predictions match measurements closely, which validates our usage of element conservation.

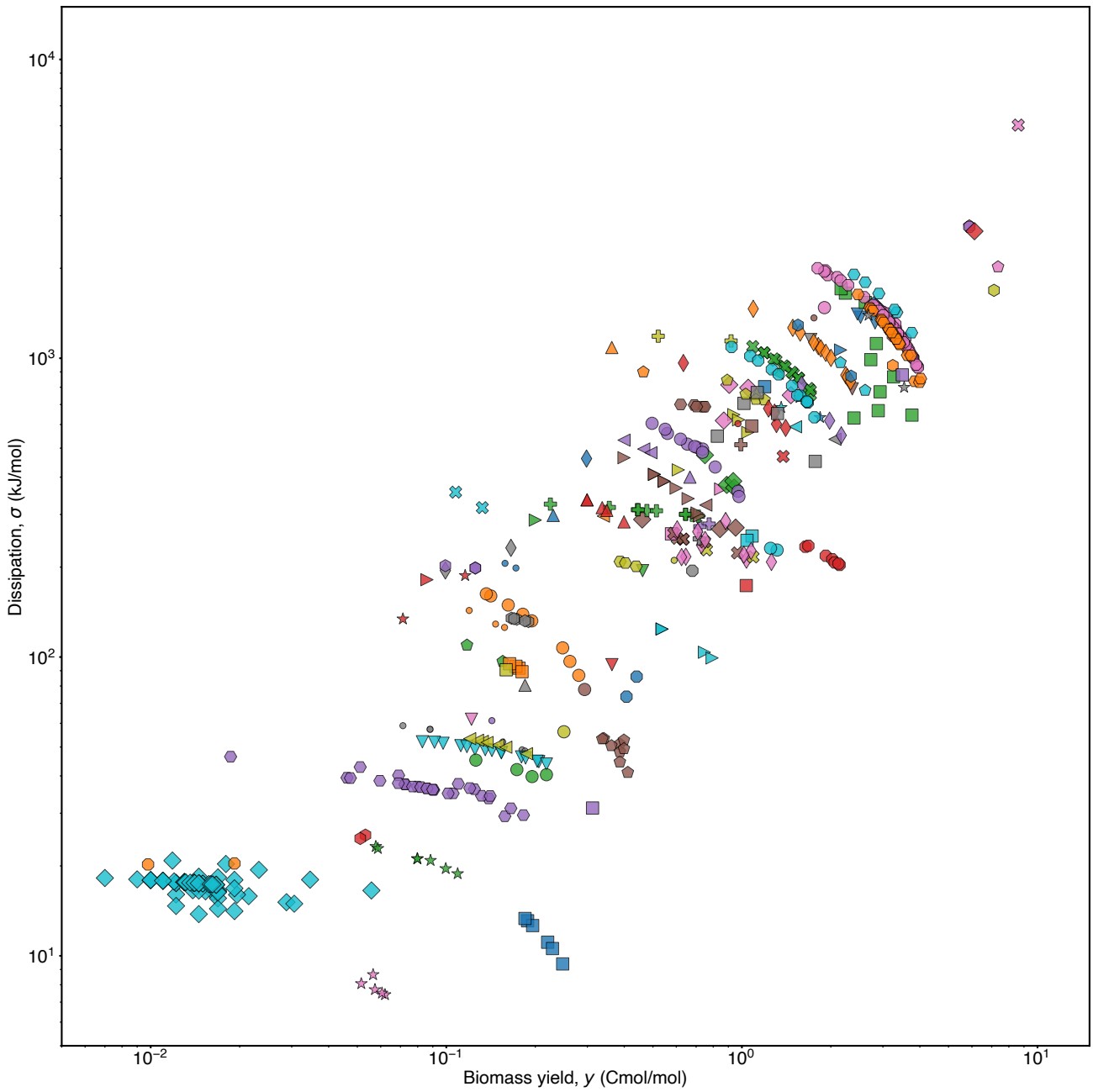


FIG. SI7: *Yield-dissipation relation for our complete dataset.* Analogue of Fig 2A using the comprehensive labeling from Fig. SI1, which distinguishes all different metabolic types included in our dataset.

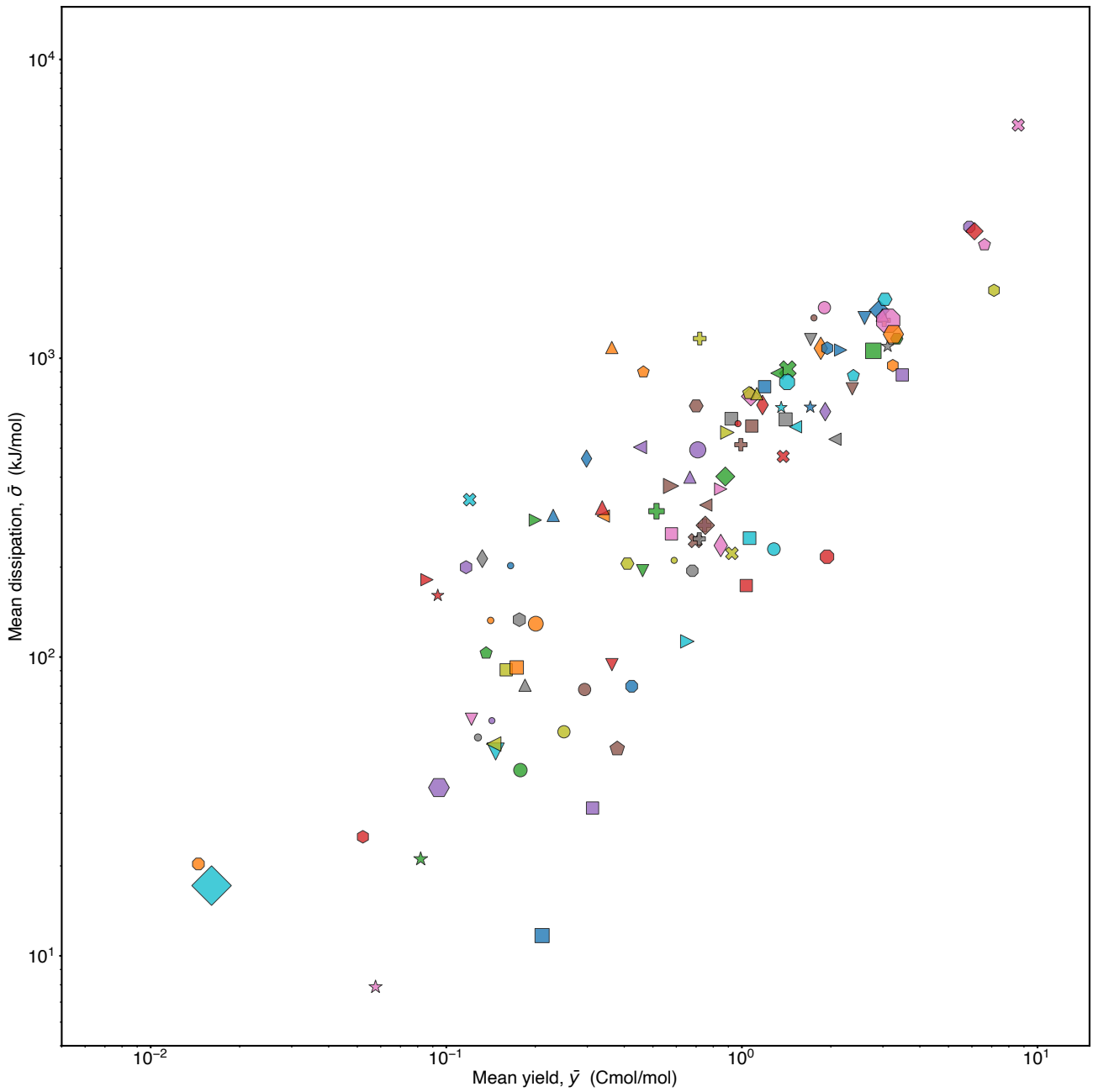


FIG. SI8: *Yield-dissipation relation across metabolic types.* Analogue of Fig 2B using the comprehensive labeling from Fig. SI1, which distinguishes all different metabolic types included in our dataset. The size of the marker is proportional to the number of data points.

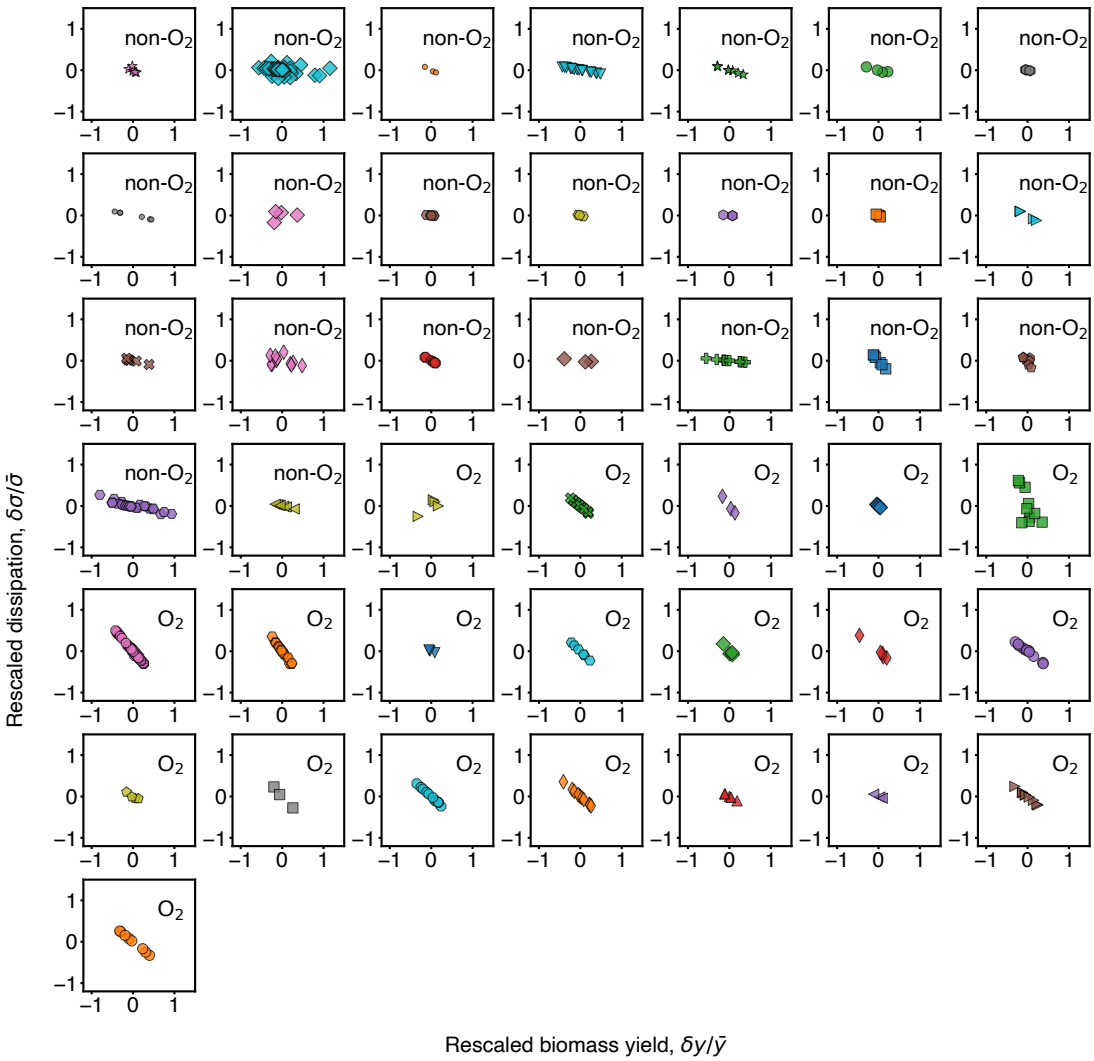


FIG. SI9: *Examples of the relationship between deviations in dissipation and yield of different metabolic types.* Analogue of Fig. 2C, dedicating an individual panel to each metabolic type with three or more data points. On the x-axis, the normalized deviation in yield, and on the y-axis, the normalized deviation in dissipation. The legend is that in Fig. SI1.

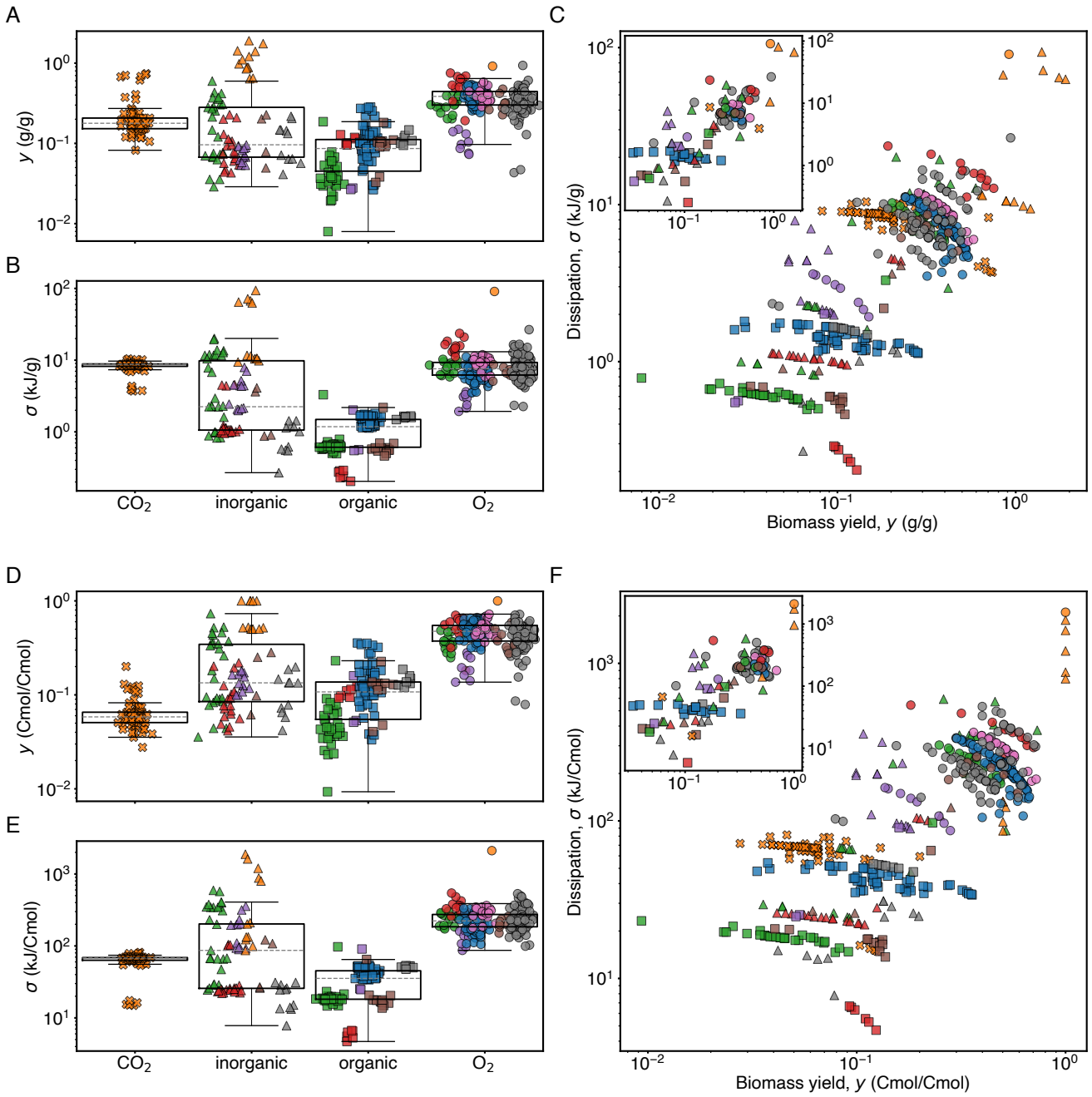


FIG. SI10: *Dissipation, yield and their relationship for two alternative choices of units.* **A** and **B.** Analogue of Fig. 1 using units of gram per gram for the biomass yield and kJ per gram for the dissipation. Note that the range of the data is somewhat reduced, but still spans over two orders of magnitude for yield, and three for dissipation. **C.** Analogue of Fig. 2A repeated for the units of panels A and B, which clearly preserves the correlation between dissipation and yield. Note that the data using H_2 as electron donor has substantially moved within the plot, due to the low weight of this substrate. **C inset.** Analogue of Fig. 2B, mean dissipation $\bar{\sigma}$ plotted against mean yield $\bar{y}$, for each type using the units of panel C. **D**, **E** and **F.** The same as panels A-C but for units of Cmol per Cmol, with similar conclusions. Note that the data using H_2 as electron donor is now expressed per carbon mole of carbon source (CO_2) instead.

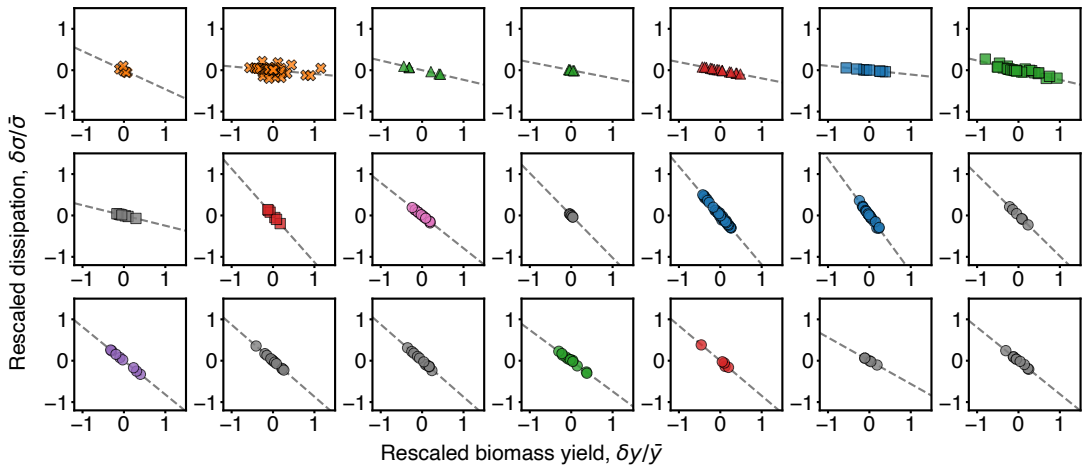


FIG. SI11: *Effect of temperature and biomass composition within metabolic type.* The panels show data for macrochemical equations with one independent yield and more than four data-points per equation. The dashed lines represent the linear relation $y(x) = -\frac{r_{\rm b}\bar{y}}{\bar{\sigma}}x$, relating the rescaled yield to the rescaled dissipation, see Eq. E5. In each panel, deviations of the data points from the line are due to: changes in temperature affecting the chemical potentials, changes in pH affecting the chemical potential of hydrogen ions, changes in biomass stoichiometry affecting r (through both the free energy of biomass and the element balance of the macrochemical equation). These changes affect both $r_{\rm ed}$ and $r_{\rm b}$, and, as one can see, the effect is minor. On the contrary, variations along the line are due to the influence of these parameters on the physiology of the organisms (i.e. on the biomass yield).

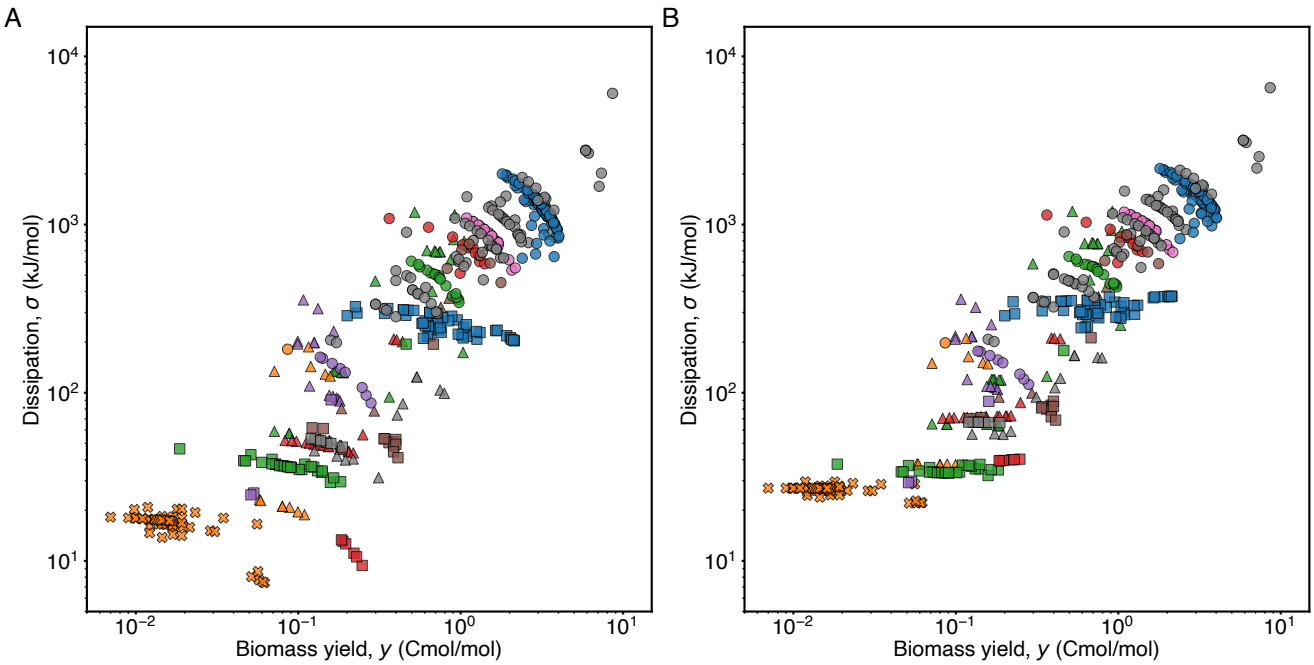


FIG. SI12: *Dissipation against yield using two approaches to correct the chemical potentials for pH changes.* **A.** Dissipation against yield for our whole dataset, as in Fig. 2A. In this case, the chemical potentials of each chemical was taken from `thermo_chem.csv`, with no correction for the pH of the media. Instead, we balanced charges by using hydrogen ions in the macrochemical equation. The pH correction to the dissipation came by giving a chemical potential to the proton ions related to the pH, as explained in SI G. This was the approach followed in the main text, i.e. this panel is identical to Fig. 2A. **B** Analogous plot to that in A, with a different approach to compute the chemical potentials. In this case, we used the free energies of formation from eQuilibrator [141], which include the pH correction performed using the approach introduced by R.A. Alberty [137]. The hydrogen ions were therefore devoid of chemical potential in the dissipation balance. The free energies are taken at a concentration of 1 mM, pMg = 3, ionic strength = 0.25 M and pH of the experiment. When the pH of the experiment was not available in the original reference, we set the value to 7. For bicarbonate, we used the free energy of CO2(total). The data for MnO_2, Mn^{+2}, S (s) and $C_{16}H_{34}$ were not available in eQuilibrator and therefore we took them from `thermo_chem.csv`. We transformed the free energy of hexadecane according to Alberty's method [137]. As one can see, both approaches yield similar results.