Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements

Dan Davidi[a,1], Elad Noor[b,1], Wolfram Liebermeister[c], Arren Bar-Even[d], Avi Flamholz[e], Katja Tummler[f], Uri Barenholz[a], Miki Goldenfeld[a], Tomer Shlomi[g], and Ron Milo[a,2]

[a]Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot 76100, Israel; [b]Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule Zürich, CH-8093 Zürich, Switzerland; [c]Institut für Biochemie, Charité-Universitätsmedizin Berlin, Berlin 10117, Germany; [d]Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany; [e]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; [f]Theoretical Biophysics, Humboldt-Universität zu Berlin, 10115 Berlin, Germany; and [g]Department of Computer Science, Technion, Haifa 32000, Israel

Edited by David Baker, University of Washington, Seattle, WA, and approved February 9, 2016 (received for review July 23, 2015)

Turnover numbers, also known as k_{cat} values, are fundamental properties of enzymes. However, k_{cat} data are scarce and measured in vitro, thus may not faithfully represent the in vivo situation. A basic question that awaits elucidation is: how representative are k_{cat} values for the maximal catalytic rates of enzymes in vivo? Here, we harness omics data to calculate k_{max}^{vivo}, the observed maximal catalytic rate of an enzyme inside cells. Comparison with k_{cat} values from *Escherichia coli*, yields a correlation of $r^2 = 0.62$ in log scale ($p < 10^{-10}$), with a root mean square difference of 0.54 (3.5-fold in linear scale), indicating that in vivo and in vitro maximal rates generally concur. By accounting for the degree of saturation of enzymes and the backward flux dictated by thermodynamics, we further refine the correspondence between k_{max}^{vivo} and k_{cat} values. The approach we present here characterizes the quantitative relationship between enzymatic catalysis in vitro and in vivo and offers a high-throughput method for extracting enzyme kinetic constants from omics data.

kinetic constants | proteomics | turnover number | kcat | flux balance analysis

The rates of enzyme catalysis are fundamental to the understanding of central cellular phenomena, such as metabolite concentrations and fluxes, cell responses to perturbations, and resource investment in enzymes (1–6). Many models of cellular metabolism include k_{cat} values, the maximal turnover rates of enzymes, as key inputs to predict the behavior of metabolic pathways and networks (7–9). However, most values have never been measured experimentally. *Escherichia coli* is the most intensely biochemically characterized organism, but k_{cat} values are available for only about 10% of its $\approx 2,000$ enzyme-reaction pairs (Dataset S1). Indeed, k_{cat} values are missing for several central metabolic enzymes. The scarcity of kinetic data limits the scope of models and necessitates generic parameter assignments that significantly reduce the predictive power of cellular models.

Even if a larger collection of k_{cat} values was made available, their current use poses a major difficulty: k_{cat} values are measured through in vitro enzyme assays, representing the initial rate of the reaction, i.e., full substrates saturation and negligible levels of products. Such assays may underrepresent factors like cellular metabolite concentrations, thermodynamic constraints, posttranslational modifications, chaperones, cellular crowding, and activating and inhibiting molecules, which can substantially alter enzyme kinetics in vivo. These omissions call into question the relevance of k_{cat} measurements in vivo (10–12). Furthermore, an effort to measure a large number of k_{cat} values under in vivo–like conditions presents a daunting challenge, given how many unknown biochemical factors might be involved.

Several studies grapple with missing k_{cat} values by sampling from the distribution of k_{cat} values measured in vitro or by using measurements of the same enzyme from related species (13–16). These approximations systematically ignore any errors resulting from the differences between in vitro and in vivo environments. Approximations of this sort may also introduce significant errors, as k_{cat} values can deviate by orders of magnitude between isozymes in the same organism as well as across organisms (17–19).

Here, we describe an alternative approach that addresses the above challenges by leveraging recent progress in omics studies. The in vivo catalytic rate of an enzyme can be inferred from the flux carried by the enzyme and the enzyme copy number. Proteomics methods now offer quantitative measurements of enzyme levels in several organisms and under a wide range of growth conditions. By dividing flux from computational flux predictions by enzyme abundances from proteomics, we calculate the rate of enzymes in vivo, denoted as k_{app} (20). We perform this analysis over a large set of growth conditions ($n = 31$). By taking the maximum value of k_{app} across many conditions, we obtain an estimate of the maximal turnover rate of an enzyme in vivo, which we define as k_{max}^{vivo}. What is the relationship between k_{cat} and k_{max}^{vivo}? We compare these values and quantitatively analyze the observed differences.

Significance

The k_{cat} values of enzymes are important for the study of metabolic systems. However, the current use of k_{cat} presents major difficulties, as values for most enzymes have not been experimentally measured, and experimentally available values are often measured under nonphysiological conditions, thereby casting doubt on the relevance of k_{cat} under in vivo conditions. We present an approach that utilizes omics data to quantitatively analyze the relationship between in vitro k_{cat} values and the maximal catalytic rate of enzymes in vivo. Our approach offers a high-throughput method to obtain enzyme kinetic constants, which reflect in vivo conditions, and are useful for more accurate and complete cellular metabolic models.

Author contributions: D.D., E.N., W.L., A.B.-E., and R.M. designed research; D.D., E.N., and M.G. performed research; D.D., E.N., and R.M. analyzed data; K.T., U.B., and T.S. contributed new reagents/analytic tools; and D.D., E.N., W.L., A.F., and R.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]D.D. and E.N. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: ron.milo@weizmann.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1514240113/-/DCSupplemental.

SYSTEMS BIOLOGY

Results

Integrating Enzyme Abundances and Fluxes to Calculate the Catalytic Rate of Enzymes in Vivo. Kinetic models of metabolic pathways and networks often describe the maximal rate of an enzyme-catalyzed reaction with the V_{max} parameter:

$$v = \overbrace{E \cdot k_{cat}}^{V_{max}} \cdot \eta, \quad [1]$$

where v is the flux through the reaction in the considered system, E is the overall number of enzyme active sites in the system, and η is a condition-dependent function, ranging between 0 and 1, which describes the decrease in the catalytic rate (relative to the maximum – k_{cat}), due to, for example, incomplete saturation with substrate and backward flux from thermodynamic effects (21).

Here we examine *E. coli* proteomic data measured via mass spectrometry to disentangle k_{cat} from V_{max} (Eq. **1**). We begin by calculating k_{app}, a previously introduced measure of the in vivo catalytic rate of enzymes (20) (Eq. **2**; Fig. 1*A*). For a metabolic reaction catalyzed by a unique enzyme, there is almost always proportionality between v and E, given condition C. As such, we define:

$$k_{app}(C) \equiv \frac{v(C)}{E(C)}. \quad [2]$$

We can thus explicitly see the relationship between V_{max} and k_{app}:

$$v(C) = \overbrace{E(C) \cdot \underbrace{k_{cat} \cdot \eta(C)}_{k_{app}}}^{V_{max}}. \quad [3]$$

Note, that because k_{cat} values are measured in vitro, such projection requires the consideration of effects caused by the shift between in vitro and in vivo, as discussed below.

Proteomic measurements provide abundances of individual polypeptides in cells, but enzymes are often composed of multiple subunits. Therefore, to infer the catalytic rate of enzymes per active site (as traditionally defined), we collected data on subunit and active site stoichiometry for *E. coli* enzymes (22) (Dataset S1). When the enzyme contains a single active site per subunit, E equals the measured abundance of the polypeptide. To calculate k_{app} for multimeric enzymes, we divide the copy number of the polypeptide by the number of chains required to make an active site.

As discussed above, we focus our analysis on reactions catalyzed by unique homomeric enzymes, which constitute about 60% of the enzymatic reactions in iJO1366, the recent genome-scale reconstruction of *E. coli* metabolism (23). Reactions that are catalyzed by multiple distinct enzymes are difficult to analyze in this framework because we do not know how flux is partitioned across isozymes. Similarly, heteromeric enzymes composed of multiple distinct polypeptides complicate the analysis because it is often unclear which subunits contain active sites. To circumvent the challenge posed by heteromeric enzymes, one may consider k_{app} to be the catalytic rate per milligram of enzyme complex rather than the rate per active site. This definition corresponds to the notion of "specific activity" (as opposed to turnover number, k_{cat}). All analyses below give similar results when the maximal specific activity is used instead of k_{cat} (Dataset S1 contains the values of all such cases). Generalization of k_{app} for reactions catalyzed by isozymes is addressed in the *Supporting Information*.

We use polypeptide abundances from recent proteomic studies of *E. coli* in 31 conditions (20, 24, 25), comprising various carbon sources, stress conditions and glucose-limited chemostats (Dataset S1). Given polypeptide abundances, flux measurements or computational flux predictions can be used to calculate k_{app}. To get comprehensive coverage of the metabolic network across conditions, here we use flux balance analysis (FBA) to calculate intracellular fluxes (26, 27). FBA is a mathematical approach for analyzing the rates and stoichiometry of transformations of metabolites through a metabolic network. The reactions are integrated into a stoichiometric matrix to be evaluated through linear programming (28). We chose to use a variant of FBA that minimizes the total flux through the network, and constrains the model by the growth rate and media composition, according to the measurements reported for each condition in the proteomics experiments (see *Materials and Methods* and *Supporting Information* for further details). A parallel analysis using experimental flux measurements (29) is shown in the *Supporting Information*, and is in good agreement with FBA derived fluxes (log correlation of $r^2 = 0.85, p < 10^{-5}$), however, with much smaller coverage than FBA (only 13 enzyme-reaction pairs in four growth conditions). Experimental flux measurements are

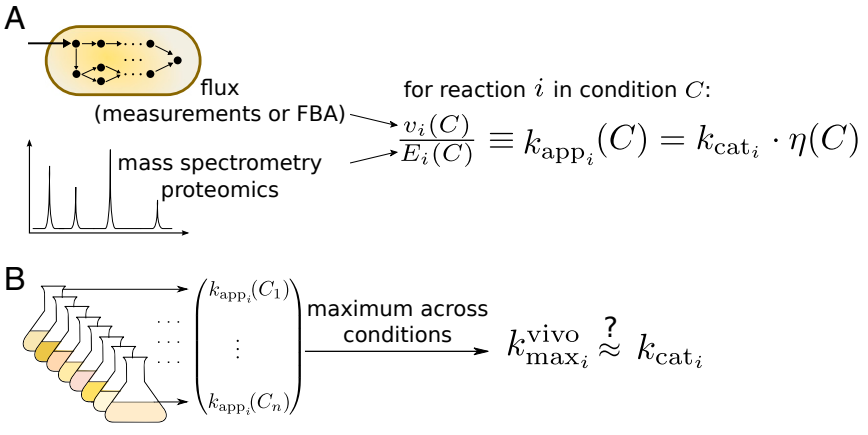

Fig. 1. Using omics data to estimate the catalytic rate of enzymes in vivo. (*A*) Flux (v) and enzymatic active site abundance (E) are integrated to calculate the rate of a single enzyme. Because v and E will change between conditions, the catalytic rate, k_{app}, is defined per condition (C). From Eq. **1**, k_{app} equals the maximal rate (k_{cat}) times a condition-dependent function η. (*B*) For each metabolic reaction in the cell, there exists a condition in which the catalytic rate of enzyme i is maximal. $k_{max_i}^{vivo}$ is the maximal value of k_{app_i} over many growth conditions ($N = 31$ in this study), and represents a lower bound estimate of the maximal catalytic capacity of enzyme i in vivo. The relation between in vivo k_{max}^{vivo} and in vitro k_{cat}, i.e., the maximal value of η, is investigated in this study.

currently limited in their applicability, and could be used in the future as depicted here, when data becomes available.

We calculate k_{app} values for 436 enzymes. To minimize stochastic effects we analyzed only enzymes with at least 10 copies of the polypeptide chain per femto liter (fL) of cytoplasm [roughly the volume of *E. coli* (30); *Materials and Methods*]. Notably, the turnover rate of an enzyme can vary between the forward and backward direction of the reaction and so k_{app} is defined for a specific enzyme-reaction-direction combination.

In Vivo and in Vitro Maximal Rates Show High Similarity. The catalytic capacity of *E. coli* cells varies with their physiology, i.e., with the growth rate and nutrient conditions (20, 31, 32). Given such variation across conditions, how can we infer the maximal catalytic rates of enzymes in vivo? We obtain a proxy for the maximal catalytic rate of enzymes in vivo by calculating k_{app} values in many conditions and taking the maximum value (Fig. 1*B*). Effectively, we look for the condition in which η is maximal (Eq. **3**), and denote the value of k_{app} in this condition as k_{max}^{vivo}.

In vitro k_{cat} measurements are often used to describe the maximal catalytic rate of enzymes in metabolic models. How close are k_{max}^{vivo} values to in vitro k_{cat} measurements? To perform a global analysis, we mined the literature for k_{cat} values of *E. coli* enzymes. We collected a set of 146 reported k_{cat} values from the BRENDA database (33) and other publications, each pertaining to a specific enzyme-reaction pair. Manual curation of the data was required to ensure values were reported per active site and to verify that the literature value was correctly entered into the database (with errors mostly resulting from incorrect unit conversion; list of all curated k_{cat} values with references is provided in Dataset S1).

We calculated k_{max}^{vivo} predictions for 132 of the 146 enzyme-reaction pairs, for which k_{cat} values were available (the other 14 were expressed in too low copy numbers or supported no flux). These 132 enzymatic reactions span a wide range of metabolic pathways, including glycolysis, pentose–phosphate pathway, amino acid metabolism, nucleotide biosynthesis, biotin biosynthesis, and heme biosynthesis. As shown in Fig. 2, we find a correlation of $r^2 = 0.62$ in log scale ($p < 10^{-10}$) between in vivo and in vitro values. Both k_{max}^{vivo} and k_{cat} values span a similar range of roughly six orders of magnitude, with a root mean square difference (RMSD) of 3.5-fold in linear scale. Analysis using the second largest value of k_{app} instead of k_{max}^{vivo} shows similar correlation to k_{cat} ($r^2 = 0.59$, $p = 10^{-10}$), indicating that our k_{max}^{vivo} approximation is quite robust to the specific conditions measured.

In the face of these high global correlations, we also observe large deviations between k_{max}^{vivo} and k_{cat} for specific enzymes. A remarkable example is the enzyme malate dehydrogenase (*mdh*), whose k_{max}^{vivo} is $\approx 10\ s^{-1}$ compared with a k_{cat} of $\approx 1{,}000\ s^{-1}$. Indeed, it has been suggested that *mdh* evolved a high k_{cat} value to compensate for a thermodynamically unfavorable reaction in the direction of the tricarboxylic acid cycle, which causes much backward flux in vivo. This may result in reduction in the k_{max}^{vivo}, which is not accounted for under in vitro conditions (21). Other outliers serve as good candidates for in-depth biochemical analysis. In the *Supporting Information* we also discuss the relation between the number of allosteric regulators and the residual between k_{max}^{vivo} and k_{cat}.

Saturation and Thermodynamics Improve the Correspondence Between in Vitro and in Vivo Rates. To rationalize the difference between k_{max}^{vivo} and k_{cat} values, we consider η, the condition dependent term in the rate law (Eq. **1**), as the product of (*i*) the degree of saturation of the enzyme, (*ii*) an effect of the thermodynamic driving force of the reaction, leading to backward flux, and (*iii*) vivo–vitro effects, which account for regulation and other ambient factors including pH, temperature, crowding agents, cofactors, chaperones, etc. (Fig. 3). Saturation and thermodynamic effects are functions of

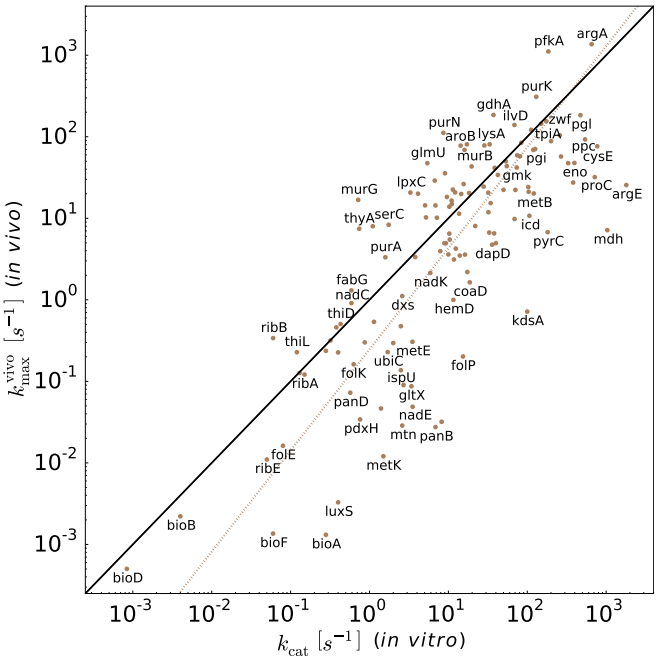

Fig. 2. In vivo and in vitro maximal catalytic rates. Log–log plot of k_{max}^{vivo} values verses in vitro k_{cat} values for all measured enzymes in *E. coli* ($N = 132$). Each data point represents an enzyme-reaction-direction combination and is labeled by the name of the enzyme (if not overlapping with other labels; data labels appear just above the data points for points that are above the $y = x$ line, and just below the data points for points that are below the line). The $y = x$ line is shown in black. Values show a correlation of $r^2 = 0.62$ ($p < 10^{-10}$) and span a similar range of 6 orders of magnitude. Dashed brown line represents the best fit by total least squares in $\log_{10}$ scale with slope = 1.23 ± 0.07, intercept = −0.6 ± 0.1, and an RMSD of 0.54.

metabolite concentrations and will range between 0 and 1, representing how much of the in vitro capacity (i.e., k_{cat}) is used (34–36). Regulatory and environmental effects, however, may result in higher rates in vivo compared with in vitro (e.g., due to absence of chaperones in the test tube) or vice versa (e.g., suboptimal pH in the cytosol). Fig. 3 shows how one can apply this rationale to interpret the residual between k_{cat} and k_{max}^{vivo}.

Although the vivo–vitro effects may represent key characteristics of enzyme catalysis inside the cell, current knowledge is not enough to fully account for them. We can, however, calculate the decrease in rate caused by nonsaturating substrate levels and backward flux, thereby assessing these differences attributed to the cellular environment. For this we use metabolite concentrations measured in *E. coli* (37, 38) (Dataset S1). For each enzyme, we sought metabolite concentrations measured in the same conditions as the proteomics data ultimately used to calculate k_{max}^{vivo}.

To estimate the degree of saturation, we collected K_M values of enzymes from the BRENDA (33) and EcoCyc (39) databases (Dataset S1), and used a relation derived from the irreversible Michaelis–Menten rate law (40, 41) for multiple substrates (42):

$$\text{saturation term} = \frac{\prod_j \left(\frac{s_j}{K_j}\right)^{m_j}}{1 + \prod_j \left(\frac{s_j}{K_j}\right)^{m_j}}, \qquad [4]$$

where s_j is the concentration of substrate j, and K_j and m_j are its K_M and stoichiometric coefficient, respectively.

None of *E. coli* enzymes have fully measured (*i*) k_{cat} values, (*ii*) K_M values for both substrates and products, and (*iii*) substrate and product concentrations. Hence, we calculate the saturation

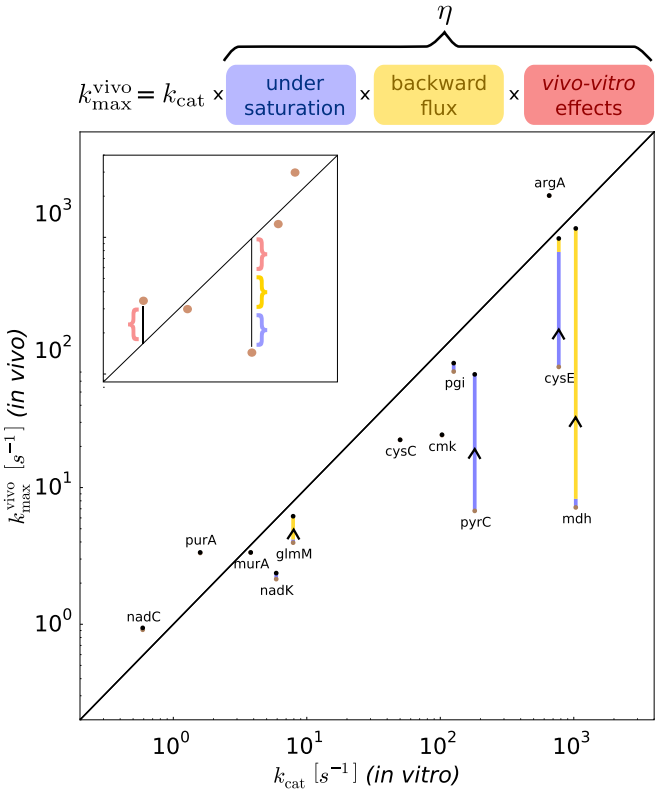

Fig. 3. The rate of an enzyme can be described as k_{cat} times a factor (η) decomposed into saturation effects, backward flux (from thermodynamic effects), and vivo–vitro effects (e.g., regulation, pH, crowding, cofactors). In log scale, the residual between k_{max}^{vivo} and k_{cat} on the *y* axis is the sum of saturation, thermodynamic, and ambient effects, as shown in the *Inset*. The first two effects can only reduce the catalytic rate of the enzyme relative to its maximum, k_{cat}. Ambient-specific effects can result in slower or faster catalytic rates. Correction of k_{max}^{vivo} by saturation and thermodynamic effects shows improved correlation, for the relatively small dataset for which such information is available. Brown dots represent k_{max}^{vivo} values achieved via taking the maximum across all conditions, as shown previously in Fig. 2; black points represent k_{max}^{vivo} values divided by the saturation and thermodynamic terms. The correlation between k_{max}^{vivo} and k_{cat} is $r^2 = 0.55$ and $r^2 = 0.92$ before and after the correction, respectively ($N = 13$). Stacked bars represent the relative contribution of undersaturation (blue) and backward-flux (yellow) effects to the residual between k_{max}^{vivo} and k_{cat}.

term based on substrate saturation only, without accounting for the effect of products. Overall, sufficient data were available for only 13 of 132 enzyme-reaction pairs and so we proceeded to analyze only this subset.

To estimate the backward-to-forward flux ratio (thermodynamic term), we use the component contribution method (43–45) to calculate the standard Gibbs free energy of a reaction, $\Delta_r G'^{\circ}$, and correct for cellular metabolite concentrations using:

$$\Delta_r G' = \Delta_r G'^{\circ} + RT \ln \prod_i c_i^{m_i}. \quad [5]$$

This gives us the thermodynamic term that accounts for the extent of backward flux:

$$\text{thermodynamic term} = \left(1 - e^{\Delta_r G'/RT}\right), \quad [6]$$

where R is the ideal gas constant, T is the absolute temperature, c_i is the concentration of reactant i, and m_i is its stoichiometric coefficient (21, 36). For unmeasured products we use a default characteristic concentration of 100 μM (37) (*Materials and Methods*).

In Fig. 3 we show the effect of undersaturation and backward flux on the residual between k_{max}^{vivo} and k_{cat} values. We find that enzymes with $k_{cat} \gg k_{max}^{vivo}$ (i.e., *mdh, pyrC,* and *cysE*) can be explained by undersaturation and poor thermodynamic driving force. After the correction of k_{max}^{vivo} values, the change in k_{max}^{vivo} values yields an improved correlation to k_{cat}, for this limited set of 13 enzymes, with $r^2 = 0.92$, $p < 10^{-4}$ (compared with $r^2 = 0.55$ before the correction).

Discussion

Currently, there is no systematic method for measuring in vivo kinetic constants. Using in vitro enzyme assays to cover a representative set of metabolic enzymes would be a gargantuan effort and still, in some cases, may not reflect in vivo kinetics very well. In this study we propose a widely applicable method to approximate the maximal catalytic rate of enzymes in vivo using proteomic data and flux estimates. By taking the maximum apparent catalytic rate k_{app} across a large set of conditions, we obtain a proxy for the maximal in vivo rate, k_{max}^{vivo}, and validate its accuracy using in vitro k_{cat} measurements.

The correlation between k_{max}^{vivo} and k_{cat} shown in Fig. 2 supports two independent claims. First, in vivo and in vitro maximal rates are similar, that is, in vitro enzyme assays usually accurately reflect the maximal rates of enzyme in vivo. Second, our approach can serve as a tool to predict k_{cat} values. The RMSD between k_{max}^{vivo} and k_{cat} is 3.5-fold, compared with a standard deviation of 15-fold in the k_{cat} dataset. Thus, k_{max}^{vivo} values are about fourfold better estimators for missing catalytic constants than using the average k_{cat} value as a representative maximal rate [$\approx 10\ s^{-1}$ (19)].

This analysis is shown for *E. coli*, however the presented method can be extended straightforwardly to other organisms and metabolic pathways. Although k_{cat} data are very limited, proteomics data are increasingly available and offer a high-throughput alternative for estimating the maximal catalytic rates of enzymes. Our approach provides much higher coverage of the metabolic network than k_{cat} and, for some enzymes, may be a more useful predictor of the rate limit than k_{cat}, as it is based on in vivo measurements. To achieve good coverage in other organisms, a parallel essential effort is a more comprehensive mapping between enzymes and their associated metabolic reactions. Although in *E. coli* most of the proteome has been characterized (46), limited gene annotations in other organisms may lead to biased predictions of k_{app}, as one must know which enzymes support flux through a given metabolic reaction to infer the rate of the enzymes.

The k_{max}^{vivo} represents a lower bound estimate of the maximal catalytic rate. However, about 10 enzymes showed k_{max}^{vivo} values fivefold or more higher than their respective k_{cat}. It is possible that technical limitations led biochemists to underestimate k_{cat} for specific enzymes. In the case of *purN* and *murG*, the two enzymes with the highest k_{max}^{vivo}/k_{cat} ratio, we found that the original k_{cat} measurements used cofactor analogs to allow better stability of the enzyme or because the actual cofactor was not available, respectively (47, 48). This can explain the lower k_{cat} values. In several other cases, tagged enzymes were used for kinetic measurements (49, 50), which may have resulted in enzyme instability or interference with catalysis (51). We also found a few cases where enzymes were refolded from aggregated protein (52), suggesting that the measurement buffer might contain misfolded enzyme and, therefore, underestimate k_{cat}. For other enzymes, we suspect that an unknown allosteric regulator is responsible for an artificially low k_{cat} measurement. For example, several of the enzymes showing $k_{max}^{vivo} > k_{cat}$ are dehydrogenases (e.g., *gdhA*), which are known to have complex kinetics and many allosteric regulators (AMP, ADP, ATP, amino acids, etc.) (53). As such, we suggest that enzymes with $k_{max}^{vivo} \gg k_{cat}$ are promising candidates for the identification of unknown allosteric activators,

which may prove useful for better understanding the regulation of metabolic networks (54–56).

The k_{max}^{vivo} value is obtained directly from flux prediction and enzyme abundance data and requires no parameter fitting. Large-scale absolute proteome quantification is advancing in great strides and becoming ubiquitous. It may soon be possible to consider hundreds of conditions in calculating the maximum k_{app} value. Genomewide estimation of maximal rates could prove invaluable for metabolic engineering applications, for example, to select the fastest enzyme from a collection of orthologs (57, 58), or compare between alternative synthetic pathways (59–61). In addition, tuning the expression levels of enzymes in a recombinantly expressed pathway will necessarily require information on their specific activities. As demonstrated here, the availability of absolute proteome quantification provides the missing information required to accelerate the collection of enzyme kinetic data.

Materials and Methods

Computational Flux Predictions and Proteomics Data as Inputs for k_{app} Estimates. Flux estimations were performed using parsimonious FBA (pFBA) (62, 63) using the COBRApy package (64). The pFBA analysis minimizes the total flux through the network and enforces reaction directionality constraints that exist in the model, and was shown to be consistent with measured fluxes (63, 65). We constrained the model by the growth rate and media composition, as was measured in the proteomics experiment. For each reaction, we analyzed the variation in flux via Flux Variability Analysis [FVA (66)]. For 99% of reactions the flux variability was negligible, representing less than 1% of the net flux through the reaction (*Supporting Information*). The flux through the ATP maintenance reaction was also shown to have negligible effect on the results (*Supporting Information*). For integration of proteomics data we used gene annotations from the iJO1366 model to map polypeptide chains to metabolic reactions. In Valgepea et al. (20) and Peebo et al. (25) data were reported in copies per fL cytoplasm, given condition specific cell size measurements. In Schmidt et al.'s dataset (24), values were given in copies per cell, hence we converted to copies per fL using a characteristic value of 180 fg/μm^3 protein concentration in *E. coli* (67). Though this value may vary on the order of 10% across conditions, there is currently no comprehensive information on this variation across conditions, and in any case this will be a minor effect given the variation in expression levels. Out of 436 homomeric enzymes expressed at a level of at least 10 polypeptide chains per fL, 257 support flux in at least one condition, resulting in $k_{app} > 0$. Notably, the 179 apparently idle enzymes might be an artifact of minimizing the total flux, and in reality may carry flux. By using a flux prediction method that suggests many zero values, we restrict ourselves to values that are relatively reliable. Enzymes that are expressed and do not support flux are not included in the analysis. All numerical data are available at Dataset S1. Codes are open-source and available at https://github.com/milo-lab/in-vivo-catalytic-rates (implemented in Python and tested mainly on Linux).

Data Collection and Unit Conversions for Comparison Between k_{cat} and k_{max}^{vivo} Values. We collected a comprehensive set of k_{cat} values from BRENDA and primary literature. Where multiple values were reported for the same k_{cat} we took the maximal reported value. The units of k_{cat} are 1/s, hence flux values were converted from millimole per gram cell dry weight per hour (mmol/gCDW/h) to millimole per gram cell dry weight per second (mmol/gCDW/s). Abundance was converted from molecules/fL to mmol/gCDW, assuming dry weight fraction of 30% [BNID:109838 (68, 69), and a cell density of 1.1 g/mL (BNID:103875 (69, 70)]. Thus, dividing flux by abundance results in rate in units of 1/s. For specific activities, we used the molecular weight of enzymes to infer the rate in units of μmol product per mg enzyme per minute (see *Supporting Information* for further details).

Calculating the Backward Flux of Reactions. To calculate $\Delta_r G'^{\circ}$ we used the component contribution method, which integrates measured thermodynamic data to calculate the formation energies of metabolites (43, 44). By including measured metabolite concentrations we approximate the in vivo Gibbs free energy of a reaction, which is a more accurate way to assess reversibility and the thermodynamic driving force than examining the standard Gibbs energy (43).

ACKNOWLEDGMENTS. We thank Matthias Heinemann for supplying data and help with analysis. We thank Dan Tawfik, Leeat Keren, Niv Antonovsky, Daniel Segré, Christopher Marx, Lianet Noda, Martin Lercher, Nathan Lewis, Kasper Valgepea, and Shlomi Reuveni for fruitful discussions. This work was funded by the European Research Council (Project SYMPAC 260392); Dana and Yossie Hollander; Helmsley Charitable Foundation; Israel Ministry of Science; and The Larson Charitable Foundation. R.M. is the Charles and Louise Gartner Professional Chair and an European Molecular Biology Organization Young Investigator Program member.

1. Wolfenden R, Snider MJ (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* 34(12):938–945.
2. Dill KA, Ghosh K, Schmit JD (2011) Physical limits of cells and proteomes. *Proc Natl Acad Sci USA* 108(44):17876–17882.
3. Poolman MG, Assmus HE, Fell DA (2004) Applications of metabolic modelling to plant metabolism. *J Exp Bot* 55(400):1177–1186.
4. Imam S, Schäuble S, Brooks AN, Baliga NS, Price ND (2015) Data-driven integration of genome-scale regulatory and metabolic network models. *Front Microbiol* 6:409.
5. Kim J, Reed JL (2012) RELATCH: Relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Genome Biol* 13(9):R78.
6. Goelzer A, et al. (2015) Quantitative prediction of genome-wide resource allocation in bacteria. *Metab Eng* 32:232–243.
7. Klipp E, et al. (2009) *Systems Biology: A Textbook* (John Wiley & Sons).
8. Labhsetwar P, Cole JA, Roberts E, Price ND, Luthey-Schulten ZA (2013) Heterogeneity in protein expression induces metabolic variability in a modeled Escherichia coli population. *Proc Natl Acad Sci USA* 110(34):14006–14011.
9. Karr JR, et al. (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* 150(2):389–401.
10. van Eunen K, et al. (2010) Measuring enzyme activities under standardized in vivo-like conditions for systems biology. *FEBS J* 277(3):749–760.
11. Goel A, Santos F, Vos WM, Teusink B, Molenaar D (2012) Standardized assay medium to measure Lactococcus lactis enzyme activities while mimicking intracellular conditions. *Appl Environ Microbiol* 78(1):134–143.
12. García-Contreras R, Vos P, Westerhoff HV, Boogerd FC (2012) Why in vivo may not equal in vitro - New effectors revealed by measurement of enzymatic activities under the same in vivo-like assay conditions. *FEBS J* 279(22):4145–4159.
13. Liebermeister W, Klipp E (2005) Biochemical networks with uncertain parameters. *IEE Proc Systems Biol* 152(3):97–107.
14. Lubitz T, Schulz M, Klipp E, Liebermeister W (2010) Parameter balancing in kinetic models of cell metabolism. *J Phys Chem B* 114(49):16298–16303.
15. Shlomi T, Benyamini T, Gottlieb E, Sharan R, Ruppin E (2011) Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect. *PLOS Comput Biol* 7(3):e1002018.
16. Singh VK, Ghosh I (2006) Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in Mycobacterium tuberculosis, and its application to assessment of drug targets. *Theor Biol Med Model* 3:27.
17. Smallbone K, et al. (2013) A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes. *FEBS Lett* 587(17):2832–2841.
18. Gautam A, Rishi P, Tewari R (2011) UDP-N-acetylglucosamine enolpyruvyl transferase as a potential target for antibacterial chemotherapy: Recent developments. *Appl Microbiol Biotechnol* 92(2):211–225.
19. Bar-Even A, et al. (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50(21):4402–4410.
20. Valgepea K, Adamberg K, Seiman A, Vilu R (2013) Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. *Mol Biosyst* 9(9):2344–2358.
21. Noor E, et al. (2014) Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLOS Comput Biol* 10(2):e1003483.
22. Furnham N, et al. (2014) The Catalytic Site Atlas 2.0: Cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res* 42(Database issue):D485–D489.
23. Orth JD, et al. (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. *Mol Syst Biol* 7:535.
24. Schmidt A, et al. (2016) The quantitative and condition-dependent Escherichia coli proteome. *Nat Biotechnol* 34(1):104–110.
25. Peebo K, et al. (2015) Proteome reallocation in Escherichia coli with increasing specific growth rate. *Mol Biosyst* 11(4):1184–1193.
26. Edwards JS, Ibarra RU, Palsson BO (2001) In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 19(2):125–130.
27. Fell DA, Small JR (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem J* 238(3):781–786.
28. Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? *Nat Biotechnol* 28(3):245–248.
29. Ishii N, et al. (2007) Multiple high-throughput analyses monitor the response of E. coli to perturbations. *Science* 316(5824):593–597.
30. Kubitschek HE, Friske JA (1986) Determination of bacterial cell volume with the Coulter Counter. *J Bacteriol* 168(3):1466–1467.
31. Chubukov V, et al. (2013) Transcriptional regulation is insufficient to explain substrate-induced flux changes in Bacillus subtilis. *Mol Syst Biol* 9:709.
32. Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3:318–356.
33. Schomburg I, et al. (2004) BRENDA, the enzyme database: Updates and major new developments. *Nucleic Acids Res* 32(Database issue):D431–D433.
34. Hofmeyr JHS (1995) Metabolic regulation: A control analytic perspective. *J Bioenerg Biomembr* 27(5):479–490.

35. Rohwer JM, Hofmeyr JHS (2010) Kinetic and thermodynamic aspects of enzyme control and regulation. *J Phys Chem B* 114(49):16280–16289.
36. Noor E, Flamholz A, Liebermeister W, Bar-Even A, Milo R (2013) A note on the kinetics of enzyme action: A decomposition that highlights thermodynamic effects. *FEBS Lett* 587(17):2772–2777.
37. Bennett BD, et al. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol* 5(8):593–599.
38. Zimmermann M, Sauer U, Zamboni N (2014) Quantification and mass isotopomer profiling of α-keto acids in central carbon metabolism. *Anal Chem* 86(6):3232–3237.
39. Keseler IM, et al. (2013) EcoCyc: Fusing model organism databases with systems biology. *Nucleic Acids Res* 41(Database issue):D605–D612.
40. Michaelis L, Menten ML (1913) Die kinetik der invertinwirkung [The kinetics of invertase action]. *Biochem Z* 49:352. German.
41. Briggs GE (1925) A further note on the kinetics of enzyme action. *Biochem J* 19(6): 1037–1038.
42. Liebermeister W, Uhlendorf J, Klipp E (2010) Modular rate laws for enzymatic reactions: Thermodynamics, elasticities and implementation. *Bioinformatics* 26(12): 1528–1534.
43. Alberty RA (2006) Biochemical thermodynamics: Applications of Mathematica. *Methods Biochem Anal* 48:1–458.
44. Goldberg RN, Tewari YB, Bhat TN (2004) Thermodynamics of enzyme-catalyzed reactions–A database for quantitative biochemistry. *Bioinformatics* 20(16):2874–2877.
45. Noor E, Haraldsdóttir HS, Milo R, Fleming RMT (2013) Consistent estimation of Gibbs energy using component contributions. *PLOS Comput Biol* 9(7):e1003098.
46. Liebermeister W, et al. (2014) Visual account of protein investment in cellular functions. *Proc Natl Acad Sci USA* 111(23):8488–8493.
47. Inglese J, Johnson DL, Shiau A, Smith JM, Benkovic SJ (1990) Subcloning, characterization, and affinity labeling of Escherichia coli glycinamide ribonucleotide transformylase. *Biochemistry* 29(6):1436–1443.
48. Crouvoisier M, Auger G, Blanot D, Mengin-Lecreulx D (2007) Role of the amino acid invariants in the active site of MurG as evaluated by site-directed mutagenesis. *Biochimie* 89(12):1498–1508.
49. Drewke C, et al. (1996) 4-O-Phosphoryl-l-threonine, a substrate of the pdxC (serC) gene product involved in vitamin B 6 biosynthesis. *FEBS Lett* 390(2):179–182.
50. Gráczer E, et al. (2009) Symmetrical refolding of protein domains and subunits: Example of the dimeric two-domain 3-isopropylmalate dehydrogenases. *Biochemistry* 48(5):1123–1134.
51. Waugh DS (2011) An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr Purif* 80(2):283–293.
52. Schauer S, Lüer C, Moser J (2003) Large scale production of biologically active Escherichia coli glutamyl-tRNA reductase from inclusion bodies. *Protein Expr Purif* 31(2): 271–275.
53. Smith TJ, Stanley CA (2008) Untangling the glutamate dehydrogenase allosteric nightmare. *Trends Biochem Sci* 33(11):557–564.
54. Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. *Nat Chem Biol* 4(8):474–482.
55. Chubukov V, Gerosa L, Kochanowski K, Sauer U (2014) Coordination of microbial metabolism. *Nat Rev Microbiol* 12(5):327–340.
56. Kochanowski K, Sauer U, Noor E (2015) Posttranslational regulation of microbial metabolism. *Curr Opin Microbiol* 27:10–17.
57. Notebaart RA, van Enckevort FHJ, Francke C, Siezen RJ, Teusink B (2006) Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics* 7:296.
58. Eckdahl TT, et al. (2015) Programmed evolution for optimization of orthogonal metabolic output in bacteria. *PLoS One* 10(2):e0118322.
59. Bar-Even A, Noor E, NE, Milo R (2010) Design and analysis of synthetic carbon fixation pathways. *Proc Natl Acad Sci USA* 107(19):8889–8894.
60. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R (2013) Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc Natl Acad Sci USA* 110(24): 10039–10044.
61. Siegel JB, et al. (2015) Computational protein design enables a novel one-carbon assimilation pathway. *Proc Natl Acad Sci USA* 112(12):3704–3709.
62. Holzhütter HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur J Biochem* 271(14):2905–2922.
63. Lewis NE, et al. (2010) Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* 6:390.
64. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR (2013) COBRApy: Constraints-based reconstruction and analysis for Python. *BMC Syst Biol* 7:74.
65. Machado D, Herrgård M (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLOS Comput Biol* 10(4):e1003580.
66. Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 5(4):264–276.
67. Booth I (1988) *Escherichia coli* and *Salmonella typhimurium*: Cellular and molecular biology, Vol. 1 (of 2). *Trends Biochem Sci* 13(12):493–494.
68. Sauer U, et al. (1999) Metabolic flux ratio analysis of genetic and environmental modulations of Escherichia coli central carbon metabolism. *J Bacteriol* 181(21): 6679–6688.
69. Milo R, Jorgensen P, Moran U, Weber G, Springer M (2010) BioNumbers–The database of key numbers in molecular and cell biology. *Nucleic Acids Res* 38(Database issue): D750–D753.
70. Baldwin WW, Myer R, Powell N, Anderson E, Koch AL (1995) Buoyant density of Escherichia coli is determined solely by the osmolarity of the culture medium. *Arch Microbiol* 164(2):155–157.
71. Zhao G, Winkler ME (1995) Kinetic limitation and cellular amount of pyridoxine (pyridoxamine) 5′-phosphate oxidase of Escherichia coli K-12. *J Bacteriol* 177(4): 883–891.
72. Boeker EA (1981) Fundamentals of enzyme kinetics (Cornish-Bowden, Athel). *J Chem Educ* 58(3):A114.