

SOURCE
DATATRANSPARENT
PROCESSOPEN
ACCESSCHECK FOR
UPDATES

Trade-off between flux efficiency and metabolic control shapes enzyme allocation

Kai Sun^{1,4}, Weiyan Zheng^{1,2,4}, Wenchao Fan^{1,3} , Chuyun Ding¹, Dan Huang¹ & Ziwei Dai¹ ✉

Abstract

Efficient use of limited cellular resources is fundamental to metabolism. Although flux optimization is widely recognized as a central objective of metabolic networks, how flux efficiency influences the allocation of the metabolic proteome remains unclear and lacks direct validation. Here, we derive a simple analytical relationship linking the equilibrium constant (K) and the catalytic-abundance quotient (CAQ) of reactions within a pathway that defines the condition for maximal efficiency. By integrating reaction thermodynamics, enzyme kinetics, transcriptomic, and proteomic data, we compared enzyme allocation in glycolysis and the TCA cycle with this K-CAQ relationship across evolutionarily distant species, and showed that deviation from it can be explained by a trade-off between maximizing flux efficiency and concentrating flux control. Moreover, the drive to optimize glycolytic efficiency is strengthened under oncogenic signaling and a limited cellular budget for glycolytic enzymes. These findings establish a principle of Pareto optimality governing enzyme allocation in metabolic pathways and reveal key determinants of efficiency optimality in glycolysis.

Subject Category Metabolism

<https://doi.org/10.1038/s44320-026-00250-5>

Received 29 October 2025; Revised 12 August 2026;

Accepted 19 September 2026

Published online: 04 October 2026

Introduction

From autocatalytic chemical reaction systems in the prebiotic world to complex, compartmentalized metabolic networks of modern cells, the evolution of metabolism has lasted for billions of years. The optimization of fitness during the evolution of metabolism covers multiple facets, including its ability to sustain robust metabolic fluxes, minimize investment of resources, and maintain functions upon environmental perturbations. As one of the earliest biological processes to be mathematically modeled, these metabolic objectives can be formalized to predict metabolic states that maximize fitness (Islam et al, 2025).

Efficient resource utilization is a central goal of metabolism. Among essential resources of metabolism, the cellular proteome is a major investment, as biosynthesis of proteins is among the most energy-consuming cellular processes, contributing to roughly 20% of the basal metabolic rate (Welle and Nair, 1990). Consequently, the proteome is under strong pressure for efficiency optimization. Evidence for this theory spans multiple scales: amino acids with low synthetic costs are preferred in the proteome of unicellular organisms such as bacteria (Akashi and Gojobori, 2002; Heizer et al, 2006) and yeast (Chen and Nielsen, 2022) as well as in cancer cells (Zhang et al, 2018). Dietary amino acid intake matching the proteomic amino acid composition has beneficial health outcomes in multicellular organisms (Piper et al, 2017), probably by saving the cost of amino acid biosynthesis. The goal of minimizing proteome cost is implicitly encoded in the current codon system, which serves as the very fundamental logic of almost all existing organisms to minimize the potential increase in amino acid biosynthetic cost under mutations (Shenhav and Zeevi, 2020). At the whole proteome level, coarse-grained mathematical modeling and quantitative biology experiments have shown that its allocation into major functional groups such as ribosome and metabolism for maximizing proliferation can precisely explain the emergence of growth scaling law, overflow metabolism, and many other phenomena in several bacteria species (Basan et al, 2015; Flamholz et al, 2025; Scott et al, 2010; Zhu et al, 2025).

A key question about such an economic principle is how it governs the operation of individual metabolic pathways. Proteomics data across bacterial species showed that the relative abundance of glycolytic enzymes are conserved across species and related to reaction thermodynamics, implying the existence of such optimality principles (Khana et al, 2025). This can be mathematically framed as maximizing pathway flux under limited enzyme abundance - a constrained optimization problem solvable either analytically or numerically if reaction kinetics and thermodynamics are known. While theoretical works on optimal proteome allocation often adopt a coarse-grained strategy focusing on major functional classes, our understanding of optimal enzyme allocation within individual pathways is still limited. Early theoretical work of metabolic control analysis (MCA) derived general conditions for optimal efficiency (Brown, 1991; Heinrich and Klipp, 1996; Klipp and Heinrich, 1999). For instance, upon optimal pathway efficiency, the flux control coefficients (FCCs), which quantify how the pathway flux responds to perturbations of single enzymes,

¹Department of Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, China. ²Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. ³Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. ⁴These authors contributed equally: Kai Sun, Weiyan Zheng. ✉E-mail: daiwz@sustech.edu.cn

are proportional to the enzyme abundances (Heinrich and Klipp, 1996). However, a major challenge is translating these theoretical principles into testable predictions against modern omics data, because the theory of MCA was developed in the 1970s, long before high-throughput sequencing-based omics technologies were available. A numerical method for solving the optimal enzyme allocation (Noor et al, 2016) and rate-law-dependent analytical solutions have been derived later, but for realistic Michaelis–Menten kinetics, an analytical solution is unavailable as the optimization problem becomes highly nonlinear (Noor and Liebermeister, 2024). Moreover, these theoretical models ignore multi-substrate reactions and isozyme-catalyzed reactions, hence their predictions are difficult to validate robustly due to the noise in experimental measurements and the complexity of in vivo metabolic systems. Therefore, a universal, simple, and experimentally testable quantitative relationship connecting enzyme abundance and reaction characteristics at optimal efficiency is lacking.

In this study, we establish a quantitative principle underlying metabolic enzyme allocation for linear pathways at optimal efficiency. This theory provides a simple, explicit equilibrium constant (K)-catalytic abundance quotient (CAQ) relationship linking reaction thermodynamics, catalytic efficiencies, and enzyme abundances testable with real-world data. Using glycolysis as a model pathway, we compare enzyme allocation profiles and this K-CAQ relationship across bacteria, yeast, and human cells, and investigate deviation of glycolytic reactions from this K-CAQ relationship. We found that the deviation from this K-CAQ relationship can be explained by the trade-off between maximization of flux efficiency and metabolic control, in which cells sacrifice flux efficiency to concentrate flux control into a few key steps. Such theory of flux-control trade-off is validated by enzyme allocation profiles and metabolite concentrations in real metabolic pathways such as glycolysis and the TCA cycle. Finally, we quantified the tendency of efficiency optimization for glycolysis and found that stronger efficiency optimization is associated with enhanced oncogenic signaling in tumors and a limited total amount of glycolysis enzymes across diverse conditions. These findings reveal a simple economic principle by which metabolic pathways balance flux efficiency with flux regulation across phylogenetically distant species.

Results

The allocation of human metabolic enzyme abundances is conserved

To test the hypothesis that metabolic enzyme abundance follows conserved allocation principles, we analyzed quantitative proteomics data for 378 human cancer cell lines in the CCLE project (Nusinow et al, 2020) and transcriptomics data for TCGA tumors and normal tissues. We found that the allocation of glycolytic enzymes was highly conserved across cell lines with different tissues-of-origin (Fig. 1A). The conservation persists when extended to genome-scale metabolic enzymes, with a median pairwise Spearman's rank correlation coefficient of metabolic enzyme abundance exceeding 0.8 at both protein and transcript levels (Fig. 1B–E). A moderate correlation (Spearman correlation around 0.5) was also observed between protein abundances in the

cell lines and transcript levels in the tumors (Appendix Fig. S1), suggesting that the allocation of metabolic enzyme abundance is robust across different omics layers. These results reveal a conserved allocation of metabolic enzymes across diverse human cell and tissue types, indicating optimality principles underlying them.

K-CAQ coupling at optimal efficiency of metabolic pathways

We hypothesize that the conserved allocation of metabolic enzymes is a consequence of efficiency maximization. The efficiency of a metabolic pathway means the ability of it to achieve as high a steady state flux as possible with a limited total amount of metabolic enzymes. To simplify mathematical derivation, we start from the simplest linear metabolic pathway consisting of a chain of subsequent unimolecular reactions. We assume first-order reaction kinetics (Fig. 2A), which can be used to approximate the Michaelis–Menten rate law when the substrate concentration is much lower than the Michaelis constant K_m of the enzyme:

$$v_i = k_i \left([S_{i-1}] - \frac{[S_i]}{K_i} \right)$$

In which k_i is the apparent rate constant for the i th reaction, defined as a product of the enzyme abundance $[E_i]$ and catalytic efficiency a_i for the enzyme catalyzing this reaction. At steady state, when the concentrations of boundary metabolites S_{in} and S_{out} are fixed, the pathway flux J is a function of enzyme abundances $[E_i]$, catalytic efficiencies a_i , and equilibrium constants K_i . We define the flux efficiency of the pathway as the ratio of the flux J to the total enzyme abundance $\sum_{i=1}^n [E_i]$:

$$\eta = \frac{J}{\sum_{i=1}^n [E_i]}$$

We then derived the conditions for the pathway to maximize the flux efficiency η . As the rate of each reaction is linear to its apparent rate constant k , which is proportional to the enzyme abundance $[E]$ (Appendix mathematical derivations, section 2.1), multiplying each enzyme abundance $[E_i]$ by a constant α while keeping $[S_{in}]$, $[S_{out}]$ and K unchanged does not affect the value of η :

$$\eta([S_{in}], [S_{out}], \mathbf{a}, \mathbf{E}, \mathbf{K}) = \eta([S_{in}], [S_{out}], \mathbf{a}, \alpha \mathbf{E}, \mathbf{K})$$

Therefore, without the loss of generality, the problem of maximizing the flux efficiency η is equivalent to the constrained optimization problem below:

$$\begin{cases} \max J, \text{ s.t.} \\ k_i \left([S_{i-1}] - \frac{[S_i]}{K_i} \right) = J \\ k_i = a_i [E_i] \\ \sum_{i=1}^n [E_i] = E_t \end{cases}$$

In which E_t is the total abundance of enzymes available to this pathway. Using Lagrange's multiplier method (Appendix mathematical

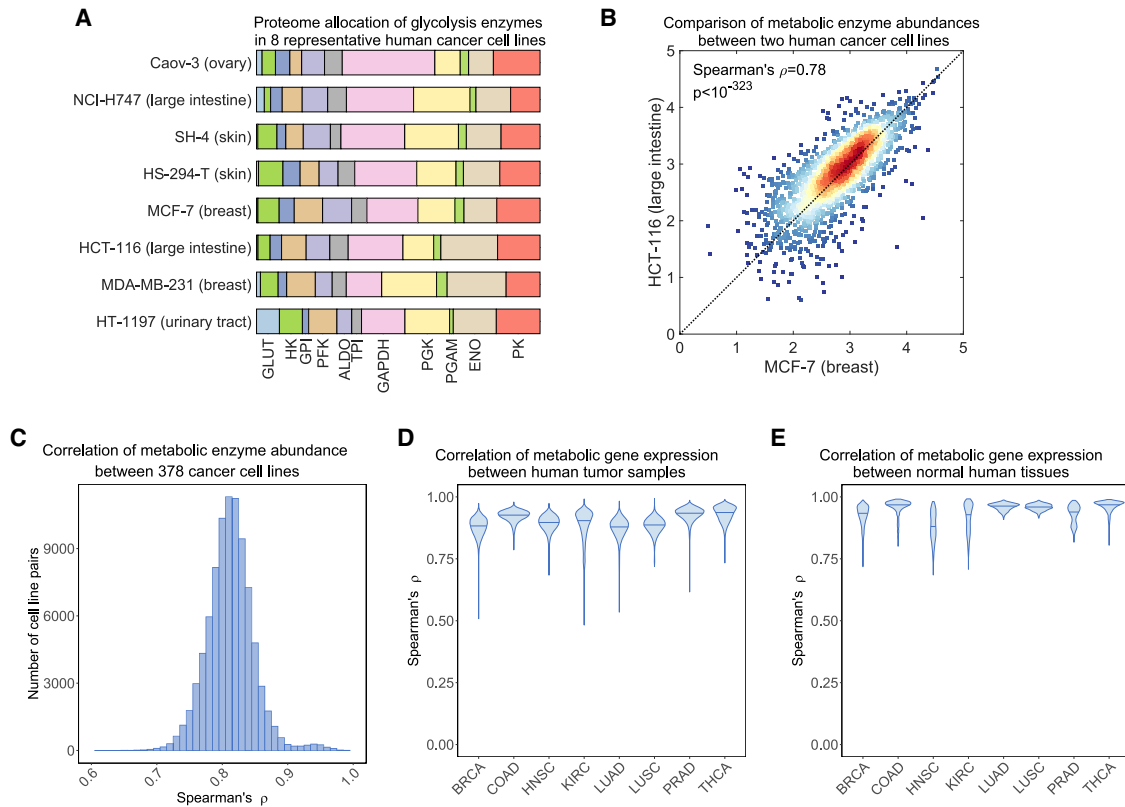


Figure 1. Allocation of human metabolic enzyme abundances is conserved.

(A) Allocation of glycolysis enzyme abundances in 8 human cancer cell lines in the CCLE project. (B) Scatter plot comparing the abundances of metabolic enzymes between two human cancer cell lines, MCF-7 and HCT-116. Colors indicate the density of data points. Sample size: $n = 1830$ enzymes; statistical test: Spearman's rank correlation test. P value $< 10^{-323}$ means that the P value is below the numerical precision limit of MATLAB double-precision calculations. (C) Distribution of Spearman's rank correlation coefficient of metabolic enzyme abundance between 378 different human cancer cell lines in the CCLE project. Sample size: $n = 87,990$ cell line pairs. (D) Distributions of Spearman's rank correlation coefficient of metabolic gene expression between human tumor samples in the TCGA project. Horizontal lines indicate median values. Sample size: $n = 624,403$ (BRCA), 42,196 (COAD), 135,981 (HNSC), 146,611 (KIRC), 146,070 (LUAD), 125,751 (LUSC), 125,751 (PRAD), 131,328 (THCA) sample pairs. (E) Distributions of Spearman's rank correlation coefficient of metabolic gene expression between human normal tissue samples in the TCGA project. Horizontal lines indicate median values. Sample size: $n = 6328$ (BRCA), 820 (COAD), 946 (HNSC), 2556 (KIRC), 1711 (LUAD), 1275 (LUSC), 1326 (PRAD), 1711 (THCA) sample pairs. Source data are available online for this figure.

derivations, section 3.2.1), we can prove that the relationship below holds for every tandem reaction pair at maximal efficiency:

$$K_i = \frac{a_i[E_i]^2}{a_{i+1}[E_{i+1}]^2} \quad (1)$$

This relationship states that the equilibrium constant of the upstream reaction, K_i , must equal the ratio of the catalytic efficiencies multiplied by the squared enzyme abundance ratio, a term we define as the catalytic-abundance quotient (CAQ).

Next, we extended our theoretical analysis to zero-order kinetics and reversible Michaelis–Menten kinetics, deriving conditions for optimal pathway efficiency for both. For each of these reaction kinetics, the optimal condition has a generalized form (Fig. S2a; Appendix mathematical derivations, section 3.2.1):

$$K_i\Phi_i = \frac{a_i[E_i]^2}{a_{i+1}[E_{i+1}]^2}$$

In which $\Phi_i = \frac{[S_{i-1}][S_{i+1}]}{K_{i+1}[S_i]^2}$ for zero-order kinetics and $\Phi_i = 1$ for first-order kinetics. For Michaelis–Menten kinetics, the form of Φ_i depends

on the concentrations of metabolites, equilibrium constants, and Michaelis constants related to the two reactions r_i and r_{i+1} . To validate the relationship between $K\Phi$ and CAQ under Michaelis–Menten kinetics, we performed numerical simulations of a 10-reaction pathway (Fig. 2B, “Methods”). For most parameter sets simulated, $\log_{10}(K\Phi)$ and $\log_{10}(\text{CAQ})$ exhibited near-perfect agreement at maximal flux efficiency, with Pearson's correlations close to 1 and RMSE values close to 0 (Appendix Fig. S2b,c). Moreover, despite the complexity in the form of Φ_i , the coupling between K_i and CAQ persists under Michaelis–Menten kinetics, as shown by the strong positive correlation between $\log_{10}(K)$ and $\log_{10}(\text{CAQ})$ across almost all parameter configurations (Fig. 2C) independent of enzyme saturation levels (Fig. 2D; Appendix Fig. S2d,e).

Finally, we extended the analysis to branching pathways (Appendix Fig. S3). Optimization of flux efficiency in these pathways can be solved by minimizing the total enzyme abundance under a given steady state flux configuration, since it is impossible to define a single flux efficiency metric due to the coexistence of multiple branches. We considered both diverging and converging pathways (Appendix Fig. S3b). Although

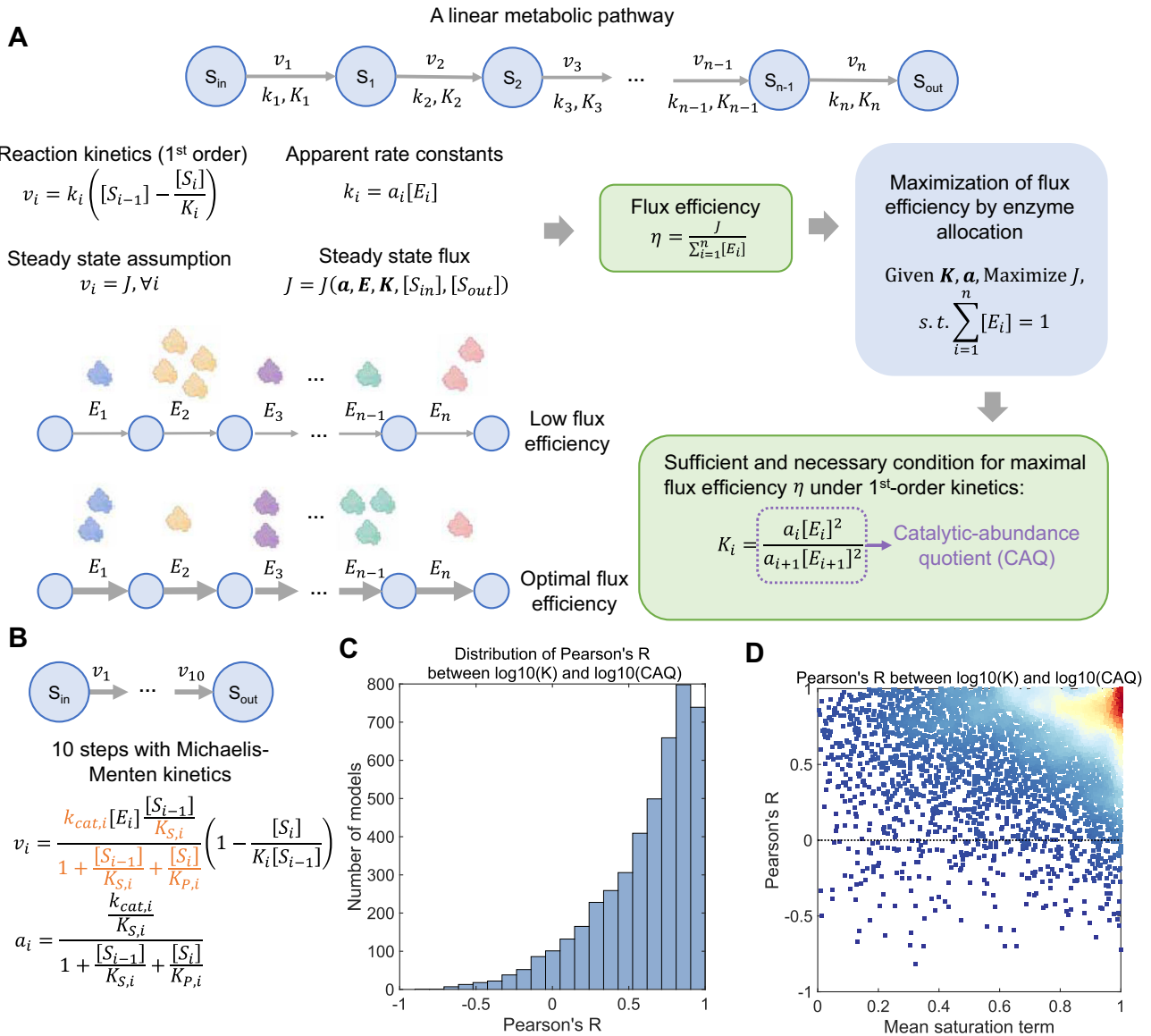


Figure 2. K-CAQ coupling at optimal efficiency of metabolic pathways.

(A) The mathematical model for maximization of flux efficiency in a linear metabolic pathway with first-order kinetics. (B) Scheme of a linear metabolic pathway model with reversible Michaelis-Menten kinetics. (C) Bar plot showing the distribution of Pearson's correlation coefficient between log10(K) and log10(CAQ) at optimal efficiency for the linear metabolic pathway with Michaelis-Menten kinetics. Sample size: $n = 4533$ parameter sets. (D) Scatter plot comparing mean enzyme saturation term and Pearson's correlation coefficient between log10(K) and log10(CAQ) at optimal efficiency for the linear metabolic pathway with Michaelis-Menten kinetics. Colors indicate the density of data points. Sample size: $n = 4533$ parameter sets. Source data are available online for this figure.

strict equality between K and CAQ no longer holds, the optimality solution still obeys inequalities with similar form: for the diverging pathway, the relationship $K_1 < \frac{a_1 [E_1]^2}{\sum_{i=2}^n a_i [E_i]^2}$ holds, while for the converging pathway, the relationship becomes $\sum_{i=1}^{n-1} \frac{a_i [E_i]^2}{K_i} < a_n [E_n]^2$ (Appendix Fig. S3b,c and Appendix mathematical derivations, section 4.3). If we define $CAQ = \frac{a_1 [E_1]^2}{\sum_{i=2}^n a_i [E_i]^2}$ for the diverging pathway, and $CAQ_i = \frac{a_i [E_i]^2}{a_n [E_n]^2}$ for the converging pathway, the inequalities can be written as $K_1 < CAQ$ for the diverging pathway and $\sum_{i=1}^{n-1} \frac{CAQ_i}{K_i} < 1$ for converging pathway. Therefore, optimal enzyme allocation profiles in these pathways are still constrained by the relationship between K and CAQ.

To summarize, our theoretical analysis reveals a relationship of K-CAQ coupling that links enzyme abundance, efficiency, and reaction thermodynamics at optimal efficiency of metabolic pathways. The mathematical simplicity of this relationship allows direct validation of it with omics data across diverse biological contexts.

Validation of the K-CAQ relationship using omics data

Validation of the K-CAQ relationship at maximal flux efficiency requires the equilibrium constant K of metabolic reactions, and catalytic efficiency and abundance of metabolic enzymes. We estimated these parameters using a multi-step workflow (Fig. 3A,

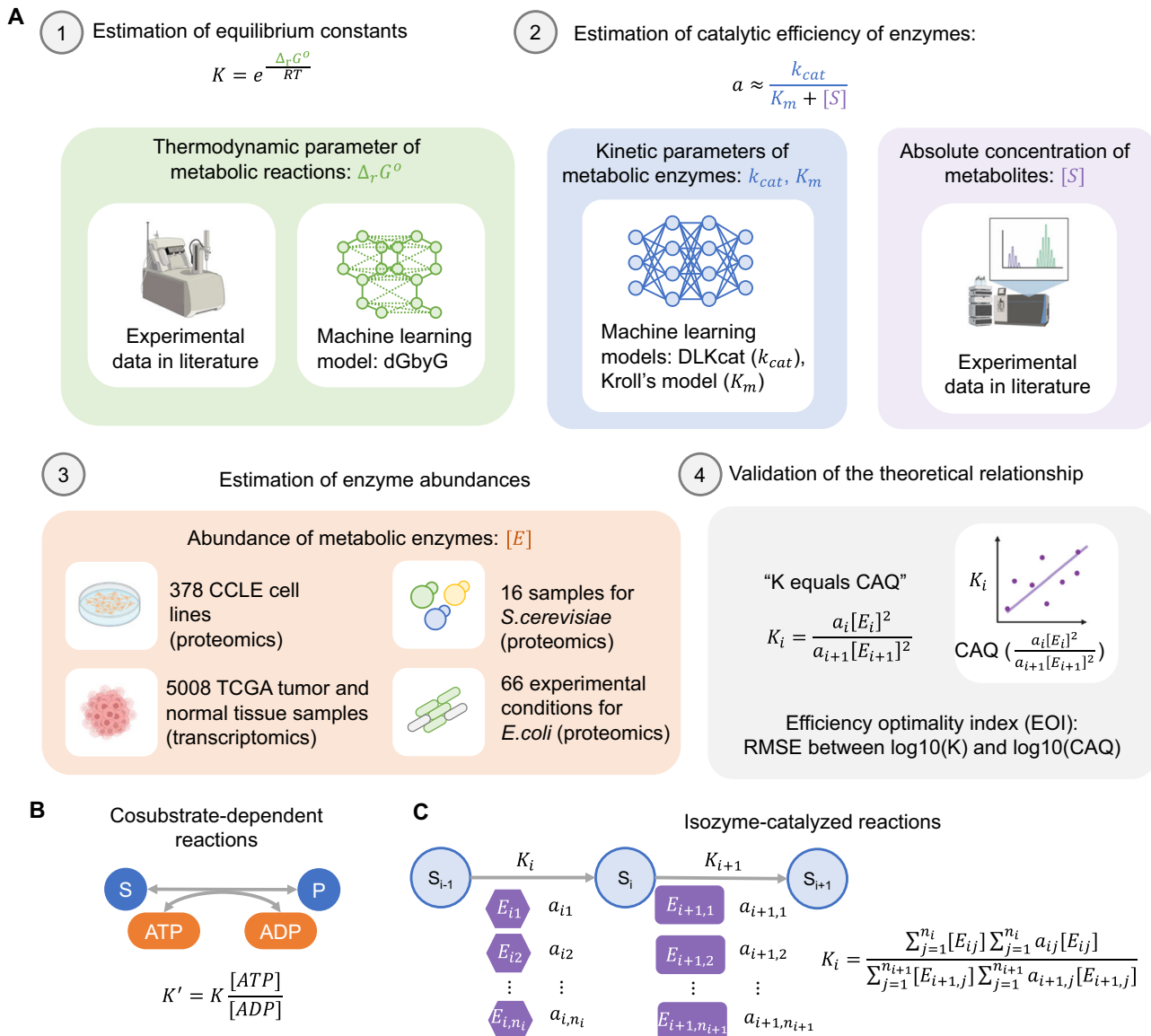


Figure 3. Validation of the K-CAQ relationship at maximal flux efficiency.

(A) Workflow for validation of the K-CAQ relationship with thermodynamics, enzyme kinetics, and omics datasets. (B) Adjustment of equilibrium constants for reactions depending on cofactors such as ATP/ADP. (C) K-CAQ relationship at maximal flux efficiency for reactions catalyzed by isozymes.

"Methods"). We computed the equilibrium constants from their standard Gibbs free energy changes, estimated catalytic efficiency of the enzymes from their k_{cat} and K_m values and substrate concentrations, and obtained enzyme abundances from proteomics datasets for 378 CCLE human cell lines (Nusinow et al, 2020), *S. cerevisiae* (Ho et al, 2018), and *E. coli* (Mori et al, 2021). Moreover, acknowledging that the protein abundances and transcript levels are substantially correlated at the cross-gene level (Buccitelli and Selbach, 2020; Schwanhaussner et al, 2011) (Appendix Fig. S1), we also used transcriptomics datasets for human tumor samples and normal tissue samples in the TCGA project to validate our theory. These datasets were combined to compute the CAQ for each

tandem reaction pair, which was finally compared to the equilibrium constants of upstream reactions.

We next address two complexities of real-world metabolism that are not present in our initial theoretical analysis: the involvement of cofactors and isozymes. For reactions depending on cofactors (ATP/ADP, NADP⁺/NADPH, NAD⁺/NADH, and so on), we combined experimentally measured intracellular concentrations of these cofactors in the literature to compute adjusted equilibrium constants K' which were used for validation (Fig. 3B, "Methods"). For reactions catalyzed by isozymes, whose relative abundances are largely conserved across tissue types (Appendix Fig. S4), we treated their relative abundances as constant. Under this assumption, we

derived a generalized form of the K-CAQ relationship, in which the CAQ term involves the catalytic efficiencies and abundances of all isozymes (Fig. 3C, Appendix mathematical derivations, section 3.2.2):

$$K_i = \frac{\sum_{j=1}^{n_i} [E_{ij}] \sum_{j=1}^{n_i} a_{ij} [E_{ij}]}{\sum_{j=1}^{n_{i+1}} [E_{i+1,j}] \sum_{j=1}^{n_{i+1}} a_{i+1,j} [E_{i+1,j}]}$$

In which a_{ij} and $[E_{ij}]$ are the catalytic efficiency and concentration of the j th isozyme catalyzing the i th reaction. This formulation preserves the core principle that K equals the CAQ, which is a quotient linking catalytic efficiencies and abundances of enzymes participating in two connected reactions.

Finally, we quantified the agreement between theory and data by calculating root mean squared error (RMSE) between $\log_{10}(K)$ and $\log_{10}(\text{CAQ})$ across all tandem reaction pairs in a pathway. We term this value the Efficiency Optimality Index (EOI). Although correlation-based metrics, such as Pearson's R, can also quantify the similarity between K and CAQ, those correlation-based metrics cannot distinguish different proportional relationships, while RMSE directly assesses deviation from equality. The theory predicts that at optimal efficiency, K equals CAQ for every pair, resulting in a perfect alignment between these two quantities (EOI = 0). A smaller EOI indicates a closer agreement with the theoretical state of maximal flux efficiency.

Comparison between enzyme allocation profiles and the K-CAQ relationship for glycolysis

Glycolysis is an ancient pathway conserved across the kingdoms of life, producing energy from carbohydrates. It is of particular interest to systems biologists because it is conserved across species, plays a central role in carbon metabolism, and is subject to complex regulations (Nemkov et al, 2024) that give rise to rich dynamic patterns such as oscillation (Chandra et al, 2011) and multistability (Mulukutla et al, 2014). In addition, the well-characterized steps and enzymes in glycolysis (Fig. 4A), along with its near-linear structure and the availability of quantitative experimental data across biological contexts, make it ideal for mathematical modeling. Here, we use omics and enzyme kinetics data from three phylogenetically distant species, *H. sapiens*, *S. cerevisiae*, and *E. coli*, to assess whether the allocation of glycolytic enzymes is consistent with the K-CAQ relationship at maximal flux efficiency.

We first retrieved proteomics data for 378 human cell lines from the CCLE database and transcriptomics data for 5008 human tumor and normal tissue samples from the TCGA project covering eight different tissues-of-origin, including breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), and thyroid carcinoma (THCA). We computed the CAQ of all tandem reaction pairs in glycolysis, compared them to the upstream equilibrium constants (K_i) (Fig. 4B–D), and computed the EOI of glycolysis by comparing these two for each sample. The results reveal a moderate deviation between K and CAQ across human cell and tissue types (median EOI around 2.5 for both TCGA samples and CCLE cell lines, Fig. 4E). A similar pattern was observed in the unicellular organisms *S. cerevisiae* (Fig. 4F) and *E. coli* (Fig. 4G) across diverse experimental conditions (Fig. 4H). Pearson's

correlation coefficients between $\log_{10}(K)$ and $\log_{10}(\text{CAQ})$ also suggest moderate positive correlation between these two (Pearson's R ranging from 0 to 0.5 for most human, yeast, and bacteria datasets, Appendix Fig. S5). Higher RMSE values were observed if substrate concentrations were not used in estimating the catalytic efficiencies (Appendix Fig. S6, "Methods"), highlighting the importance of quantitative metabolite concentrations in validation of the theory. These results together suggest that K and CAQ are moderately coupled with each other for glycolytic enzymes across the three species tested, suggesting that maximization of flux efficiency can serve as one of the objectives underlying the organization of glycolysis.

Next, we investigated potential mechanisms enabling the optimization of flux efficiency. There are two strategies to establish the coupling between K and CAQ: (1) by tuning enzyme abundance $[E]$ under a given set of catalytic efficiencies a , or (2) by evolving enzyme kinetic parameters such as k_{cat} and K_m (thus the corresponding a) so that the thermodynamics-kinetics coupling remains approximately satisfied over a broad range of enzyme abundance profiles. Since the evolution of DNA and protein sequences can change both the expression level (strategy 1) and kinetic parameters (strategy 2) of an enzyme, optimization of flux efficiency can involve both strategies. To characterize their contributions, we performed permutation tests in which the catalytic efficiencies and abundances of glycolytic enzymes were individually or jointly permuted. If the cell relies on strategy 1 only, random permutation of enzyme abundance alone should disrupt the K-CAQ coupling as strongly as joint permutation. Otherwise, if optimization relies primarily on kinetic parameter evolution, permuting catalytic efficiencies alone should have a comparable effect. We found that compared to the scenario where catalytic efficiencies and enzyme abundances were jointly permuted, disrupting enzyme abundances or catalytic efficiencies alone yielded lower EOI values in all human (Appendix Fig. S7a–c), bacteria (Appendix Fig. S7d), and yeast (Appendix Fig. S7e) datasets. Therefore, the evolution of metabolic pathways likely utilizes both strategies: enzyme kinetic parameters are pre-optimized to provide robustness (strategy 2), and cells further fine-tune enzyme expression (strategy 1) to increase flux efficiency.

A trade-off between flux efficiency and control efficiency explains the deviation between K and CAQ

As the cross-species omics data showed moderate deviation from the theoretical $K = \text{CAQ}$ relationship, we next asked whether such deviation was dominated by a few key reaction pairs. The comparison between K and CAQ for glycolytic enzymes across the three species showed that the largest deviations repeatedly occur at the reaction pairs GLUT-HK, HK-GPI, and PFK-ALDO in the *H. sapiens* and *S. cerevisiae* datasets (Fig. 4B–D,F) and GLCptspp-GPI and PFK-ALDO in the *E. coli* dataset (Fig. 4G). Upstream reactions of these reaction pairs, such as glucose transport and phosphorylation, are known as key control points of glycolysis. At these steps, we repeatedly observed $\text{CAQ} < K$, suggesting that at key reactions controlling the glycolytic flux, abundances of enzymes were lower than the theoretically predicted values at maximal flux efficiency.

From the perspective of MCA, reducing the abundance of an enzyme can increase its flux control coefficient (FCC), which

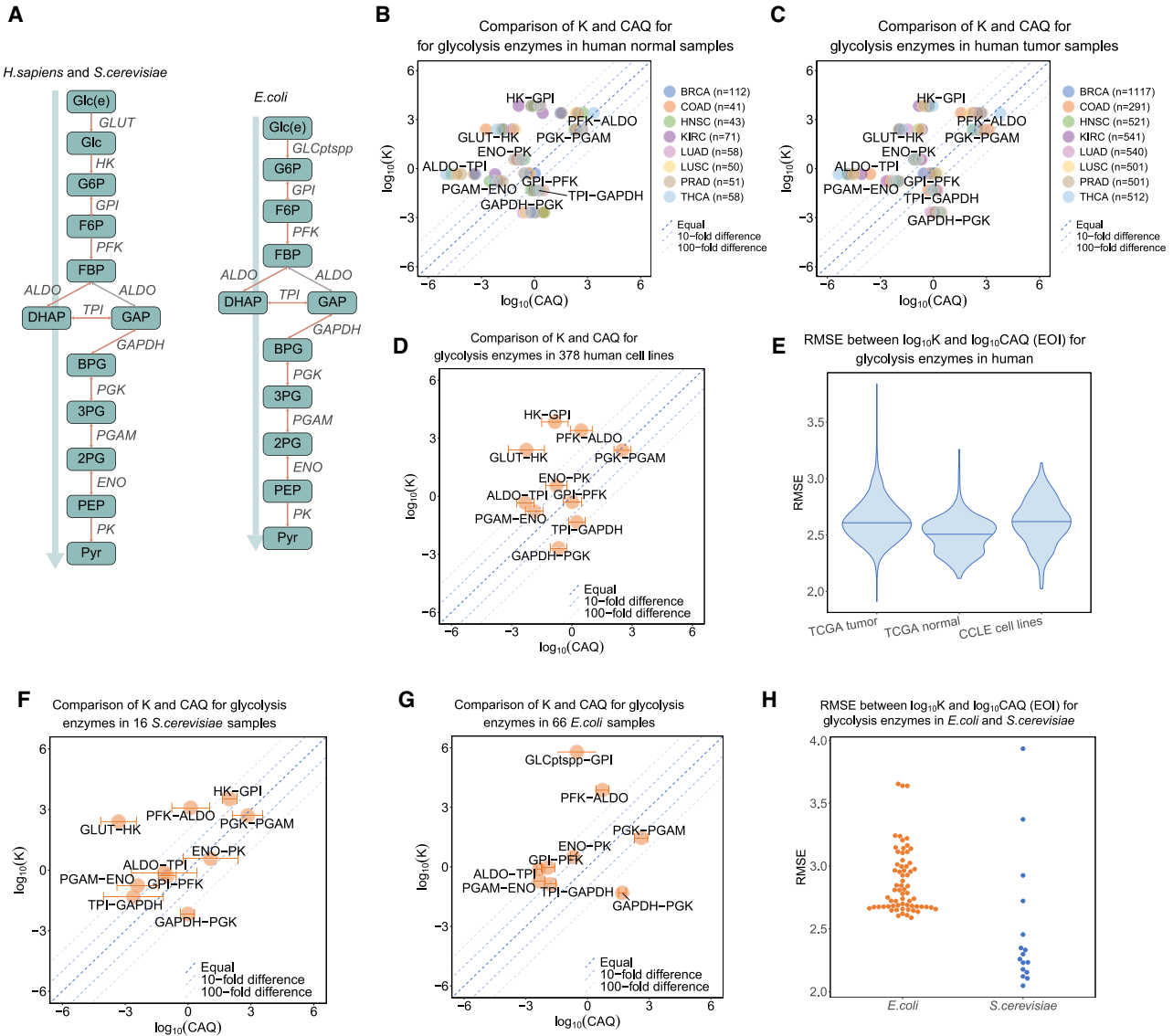


Figure 4. Comparison between enzyme allocation profiles and the K-CAQ relationship for glycolysis.

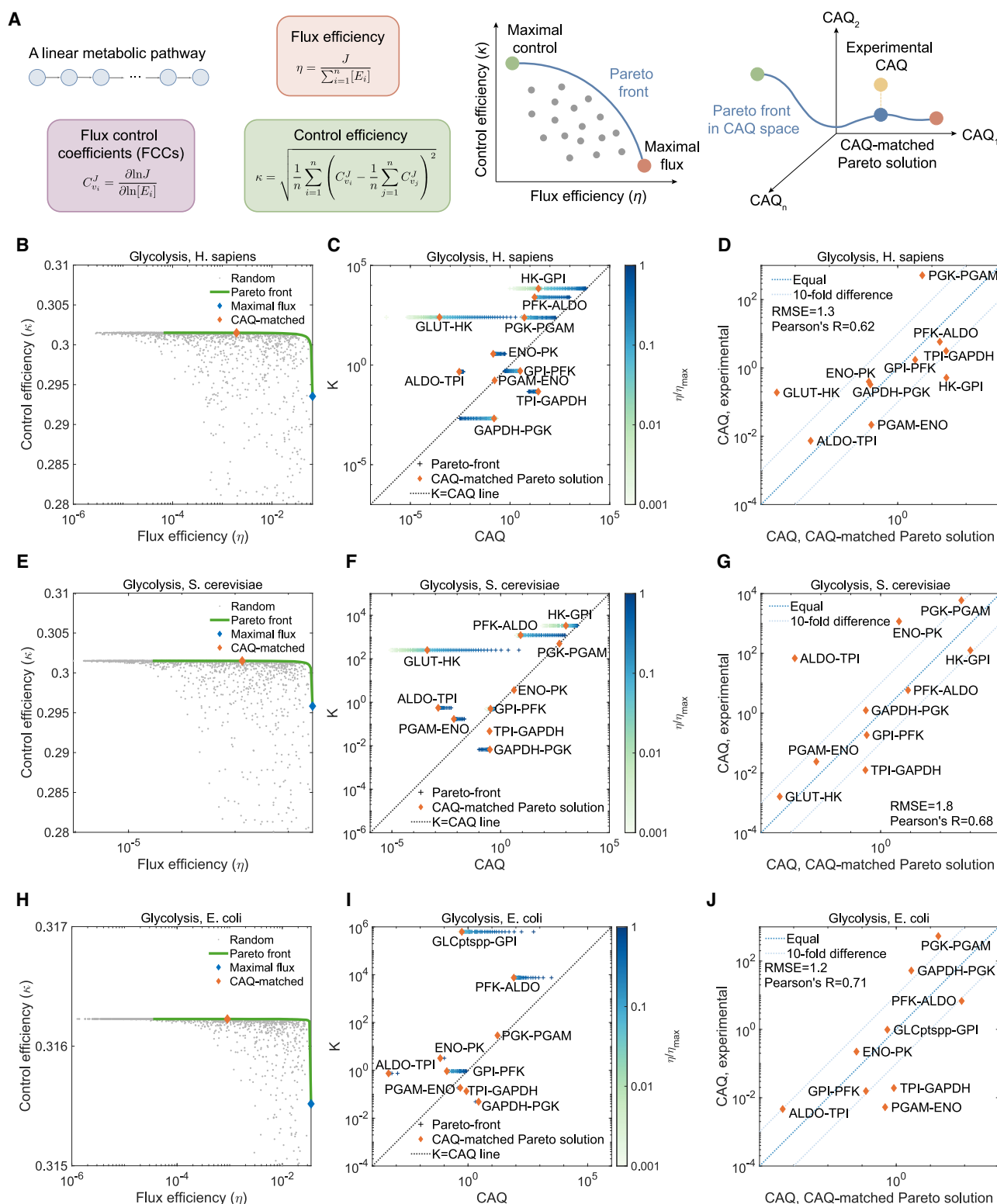
(A) Diagram of the glycolysis pathway in *H. sapiens*, *S. cerevisiae*, and *E. coli*. (B) Scatter plot comparing K and average CAQ ($a_i[E_i]^2/(a_{i+1}[E_{i+1}]^2)$) of tandem reaction pairs in glycolysis for normal human tissue samples in TCGA. Sample size: $n = 112$ (BRCA), 41 (COAD), 43 (HNSC), 71 (KIRC), 58 (LUAD), 50 (LUSC), 51 (PRAD), 58 (THCA) samples. (C) Same as in (B) but for TCGA human tumor samples. Sample size: $n = 1117$ (BRCA), 291 (COAD), 521 (HNSC), 541 (KIRC), 540 (LUAD), 501 (LUSC), 501 (PRAD), 512 (THCA) samples. (D) Same as in (B) but for 378 CCL human cell lines. Dots indicate the mean of $\log_{10}(CAQ)$, and horizontal bars indicate the standard deviation of $\log_{10}(CAQ)$ across different cell lines. (E) Distribution of RMSE between $\log_{10}(K)$ and $\log_{10}(CAQ)$, i.e., EOI of glycolysis, for the human cell lines, tumors, and normal tissue samples. Horizontal lines indicate median values. Sample size: $n = 4531$ (TCGA tumor), 491 (TCGA normal), 378 (CCL cell lines). (F) Same as in (B) but for the budding yeast *Saccharomyces cerevisiae*. Dots indicate the mean of $\log_{10}(CAQ)$, and horizontal bars indicate the standard deviation of $\log_{10}(CAQ)$ across different samples. Sample size: $n = 16$ experimental conditions. (G) Same as in (B), but for the bacteria *Escherichia coli*. Dots indicate the mean of $\log_{10}(CAQ)$, and horizontal bars indicate the standard deviation of $\log_{10}(CAQ)$ across different samples. Sample size: $n = 66$ experimental conditions. (H) Distribution of the EOI of glycolysis in the *E. coli* and *S. cerevisiae* samples. Sample size: $n = 16$ (*S. cerevisiae*), 66 (*E. coli*). Source data are available online for this figure.

quantifies the sensitivity of steady state pathway flux to local changes in activity of this reaction. Therefore, the reduced abundance of enzymes such as GLUT, HK, and PFK may further enhance their control over the pathway flux, allowing efficient regulation through a limited number of key reaction steps. We therefore define a quantity κ named control efficiency (Fig. 5A), which quantifies the extent by which the control of pathway flux is concentrated within a few reaction steps or uniformly distributed over all reactions in the pathway. Control efficiency is defined as

the standard deviation of FCCs:

$$\kappa = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(c_{v_i}^j - \frac{1}{n} \sum_{j=1}^n c_{v_j}^j \right)^2}$$

According to our previous derivations, maximal flux efficiency requires a specific enzyme allocation profile satisfying $K = CAQ$. As



decreasing enzyme abundance can increase FCC of a reaction (see Appendix mathematical derivations, section 2.4.3 for a mathematical proof under the first-order rate law), concentrating flux control into a few key reactions requires reducing the abundance of enzymes catalyzing these reactions below the flux efficiency optimum to increase their FCCs. Because such redistribution of

enzyme abundance deviates from the $K = CAQ$ condition that maximizes flux efficiency, a trade-off naturally arises between flux efficiency η and control efficiency κ .

Next, we sought to determine how the trade-off between flux efficiency η and control efficiency κ shapes enzyme allocation profiles in linear metabolic pathways. The condition for maximal

Figure 5. A trade-off between flux efficiency and control efficiency explains the deviation between K and CAQ.

(A) Definition of flux efficiency and control efficiency, and the Pareto front balancing the needs to maximize these two objectives. (B) Scatter plot showing flux efficiency and control efficiency in randomly sampled enzyme allocation profiles (gray dots), Pareto front balancing these two (green curve), and the CAQ-matched Pareto solution (orange diamond) for human glycolysis. Sample size: $n = 2000$ (random), 101 (Pareto). (C) Relationship between K and CAQ on the Pareto front for glycolysis in *H. sapiens*. Orange diamonds indicate the CAQ-matched Pareto solution. Sample size: $n = 101$ Pareto solutions. (D) Scatter plot comparing CAQ values between the CAQ-matched Pareto solution and experimental values computed from proteomics data for glycolysis in *H. sapiens*. (E) Same as in (B), but for the budding yeast *S. cerevisiae*. Sample size: $n = 2000$ (random), 101 (Pareto). (F) Same as in (C), but for the budding yeast *S. cerevisiae*. Sample size: $n = 101$ Pareto solutions. (G) Same as in (D) but for the budding yeast *S. cerevisiae*. (H) Same as in (B), but for the bacteria *E. coli*. Sample size: $n = 2000$ (random), 101 (Pareto). (I) Same as in (C), but for the bacteria *E. coli*. Sample size: $n = 101$ Pareto solutions. (J) Same as in (D), but for the bacteria *E. coli*. Source data are available online for this figure.

flux efficiency requires that the FCCs are proportional to enzyme abundances (Appendix mathematical derivations):

$$C_{v_i}^J \propto [E_i]$$

In which $C_{v_i}^J = \frac{[E_i]}{J} \frac{\partial J}{\partial [E_i]}$ is the FCC for the i th reaction in the pathway. It is worth noting that the derivation of this relationship only requires the reaction rate to be linear to enzyme abundance, which is satisfied by the first-order, zero-order, and Michaelis–Menten rate laws considered here. Simulation with a linear metabolic pathway under Michaelis–Menten kinetics confirms the proportionality between $C_{v_i}^J$ and $[E_i]$ when the pathway achieves maximal flux efficiency (Appendix Fig. S8). According to the summation theorem in MCA, which states $\sum_{i=1}^n C_{v_i}^J = 1$, the relationship below holds at maximal flux efficiency:

$$\frac{[E_i]}{J} \frac{\partial J}{\partial [E_i]} = \frac{[E_i]}{\sum_{i=1}^n [E_i]}$$

Which means that for any reaction in the pathway,

$$\frac{\partial J}{\partial [E_i]} = \frac{J}{\sum_{i=1}^n [E_i]} = \eta$$

Therefore, at maximal flux efficiency, a small reduction δ in the abundance of any enzyme E_i will reduce the pathway flux J by approximately $\delta\eta$, yet not all deviations from the flux efficiency optimum contribute equally to the improvement of control efficiency. For a given reduction in flux efficiency, the increase in control efficiency κ is maximized when enzyme abundance is reduced at reactions that already possess relatively large FCCs at maximal flux efficiency (e.g., GLUT, HK, and PFK, Appendix Fig. S9), thereby increasing their control over pathway flux. The framework of flux-control trade-off therefore provides a plausible explanation for why physiological glycolysis deviates from the theoretical proportionality between FCC and enzyme abundance at maximal flux efficiency, with enzymes catalyzing key control points such as GLUT and HK expressed at relatively low abundance.

To illustrate such a trade-off between flux efficiency and control efficiency, we applied an ϵ -constraint method to compute the Pareto front balancing the objectives of maximizing flux efficiency and maximizing control efficiency in human glycolysis (Fig. 5B,C, “Methods”). We then identified a representative solution on the Pareto front, named the CAQ-matched Pareto solution (Fig. 5A), by selecting the Pareto solution with the smallest Euclidean distance to the physiological CAQ profile computed from the proteomics data. The simulated Pareto front for the flux-control trade-off (Fig. 5C) clearly aligned with the observed deviation between K and CAQ in human glycolysis (Fig. 4B–D): reaction pairs GLUT–HK, HK–GPI, and PFK–ALDO

consistently showed $CAQ < K$ in the Pareto front, while reaction pairs TPI–GAPDH and GAPDH–PGK exhibited $CAQ > K$. The directional agreement in the deviation between K and CAQ was above 80% for almost all samples and dropped to around 50% after randomly permuting the CAQ profiles (Appendix Fig. S10, “Methods”), indicating that the Pareto front captures properties of K–CAQ deviation when the pathway moves away from the maximal flux efficiency state. The CAQ-matched Pareto solution is highly consistent with the experimental CAQ (RMSE = 1.3, Pearson’s $R = 0.62$, Fig. 5D) and was used as a representative physiological solution along the Pareto front. We then extended the analysis to glycolysis in *S. cerevisiae* and *E. coli*, and found that the Pareto front correctly predicted directions of deviation between K and CAQ across these species (Appendix Fig. S10), and that the CAQ-matched Pareto solution consistently showed close agreement with the experimental CAQ (RMSE = 1.8 in *S. cerevisiae* and 1.2 in *E. coli*, Pearson’s $R = 0.68$ in *S. cerevisiae* and 0.71 in *E. coli*, Fig. 5E–J). Finally, the CAQ-matched Pareto solution for glycolysis in *H. sapiens* also aligned with the experimental observation that GLUT, HK, and PFK are key control points for glycolysis (Appendix Fig. S11) (Tanner et al, 2018). These results together support the theoretical framework of flux-control trade-off.

The theory of flux-control trade-off correctly predicts metabolite concentrations

Using a mathematical model of a linear metabolic pathway representative of glycolysis, we computed steady-state metabolite concentrations under two scenarios: enzyme allocation profiles maximizing flux efficiency and CAQ-matched Pareto solutions (“Methods”). We then compared the predicted metabolite concentrations with experimental values collected from the literature (Fig. 6). The results showed that while the flux-efficiency-maximizing enzyme allocation profiles substantially overestimated metabolite concentrations and exhibited poor agreement with experimental measurements (RMSE ≈ 4 , Pearson’s $R \approx 0.2$, Fig. 6A–C), enzyme abundances from the CAQ-matched Pareto solution showed good agreement with experimental metabolite concentrations in *H. sapiens* (RMSE = 1.4, Pearson’s $R = 0.67$, Fig. 6D), *S. cerevisiae* (RMSE = 0.5, Pearson’s $R = 0.90$, Fig. 6E), and *E. coli* (RMSE = 1.0, Pearson’s $R = 0.69$, Fig. 6F). The systematic overestimation produced by the flux-efficiency-maximizing solution suggests that real glycolytic pathways operate away from the maximal-flux optimum. Incorporating the flux-control trade-off through CAQ-matched Pareto solutions substantially improved the agreement with experimentally measured metabolite concentrations, therefore further supporting the theory of flux-control trade-off.

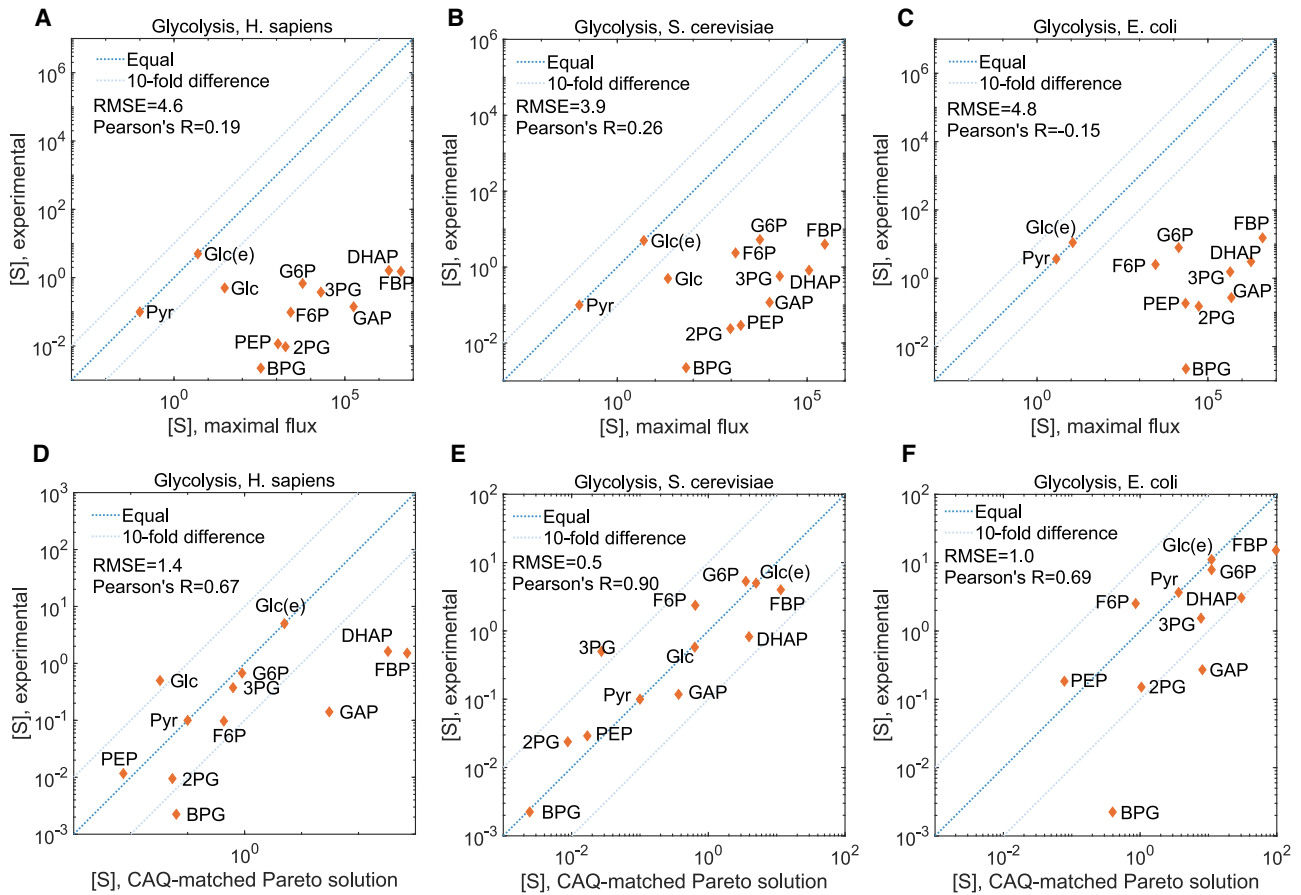


Figure 6. The theory of flux-control trade-off correctly predicts metabolite concentrations.

(A) Scatter plots comparing metabolite concentrations predicted by the maximal-flux solution to experimental measurements for glycolysis in *H. sapiens*. (B) Same as in (A) but for the budding yeast *S. cerevisiae*. (C) Same as in (A), but for the bacteria *E. coli*. (D) Scatter plots comparing metabolite concentrations predicted by the CAQ-matched Pareto solution to experimental measurements for glycolysis in *H. sapiens*. (E) Same as in (D) but for the budding yeast *S. cerevisiae*. (F) Same as in (D), but for the bacteria *E. coli*. Source data are available online for this figure.

Flux-control trade-off in *E. coli* TCA cycle

We next asked whether the flux-control trade-off theory can explain enzyme allocation in metabolic pathways beyond glycolysis. To test its generality, we analyzed the TCA cycle in *E. coli* as an additional system (Fig. 7A). We illustrated the trade-off between flux efficiency η and control efficiency κ by constructing the Pareto front (Fig. 7B). We then compared κ and CAQ for reaction pairs in the TCA cycle and found that κ and CAQ correlated poorly with each other (Fig. 7C). The largest deviations were observed for the reaction pairs AKGDH-SUCOAS and ICDHyr-AKGDH (Fig. 7C).

Notably, the reaction pair AKGDH-SUCOAS exhibited $\text{CAQ} < \kappa$. This reflects the same reaction-specific pattern observed in glycolysis: deviations from $\kappa = \text{CAQ}$ were concentrated at known control points rather than being randomly distributed across the pathway. Because AKGDH is a known control point of the TCA cycle (Green and Storey, 2020; Hansen and Gibson, 2022), this observation is consistent with the prediction of the flux-control trade-off theory that enzymes catalyzing key control steps may be expressed below the flux-efficiency optimum to concentrate flux control into a small number of reactions. This interpretation is further supported by the Pareto front and the CAQ-matched Pareto

solution (Fig. 7D), which reproduced the large deviations at AKGDH-SUCOAS and ICDHyr-AKGDH observed in the experimental data. CAQ profile of the CAQ-matched Pareto solution showed good agreement with the experimental CAQ (RMSE = 1.3, Pearson's $R = 0.85$, Fig. 7E). Finally, the CAQ-matched Pareto solution also improved the accuracy of metabolite concentration prediction (RMSE = 0.8, Pearson's $R = 0.78$, Fig. 7F) compared to the solution maximizing flux efficiency (RMSE = 1.0, Pearson's $R = 0.75$, Fig. 7G). Taken together, these results indicate that the flux-control trade-off may also shape enzyme allocation in metabolic pathways outside glycolysis.

Validation of the Pareto front independent of proteomics-based calibration

As the identification of CAQ-matched Pareto solution uses proteomics data to infer the physiological position of enzyme allocation along the Pareto front, we further tested the predictive power of the Pareto front independently of this calibration procedure. First, we showed that experimental CAQ profiles were consistently closer to the Pareto front than randomly permuted CAQ profiles, with permutation P values ranging from 0.01 to 0.07

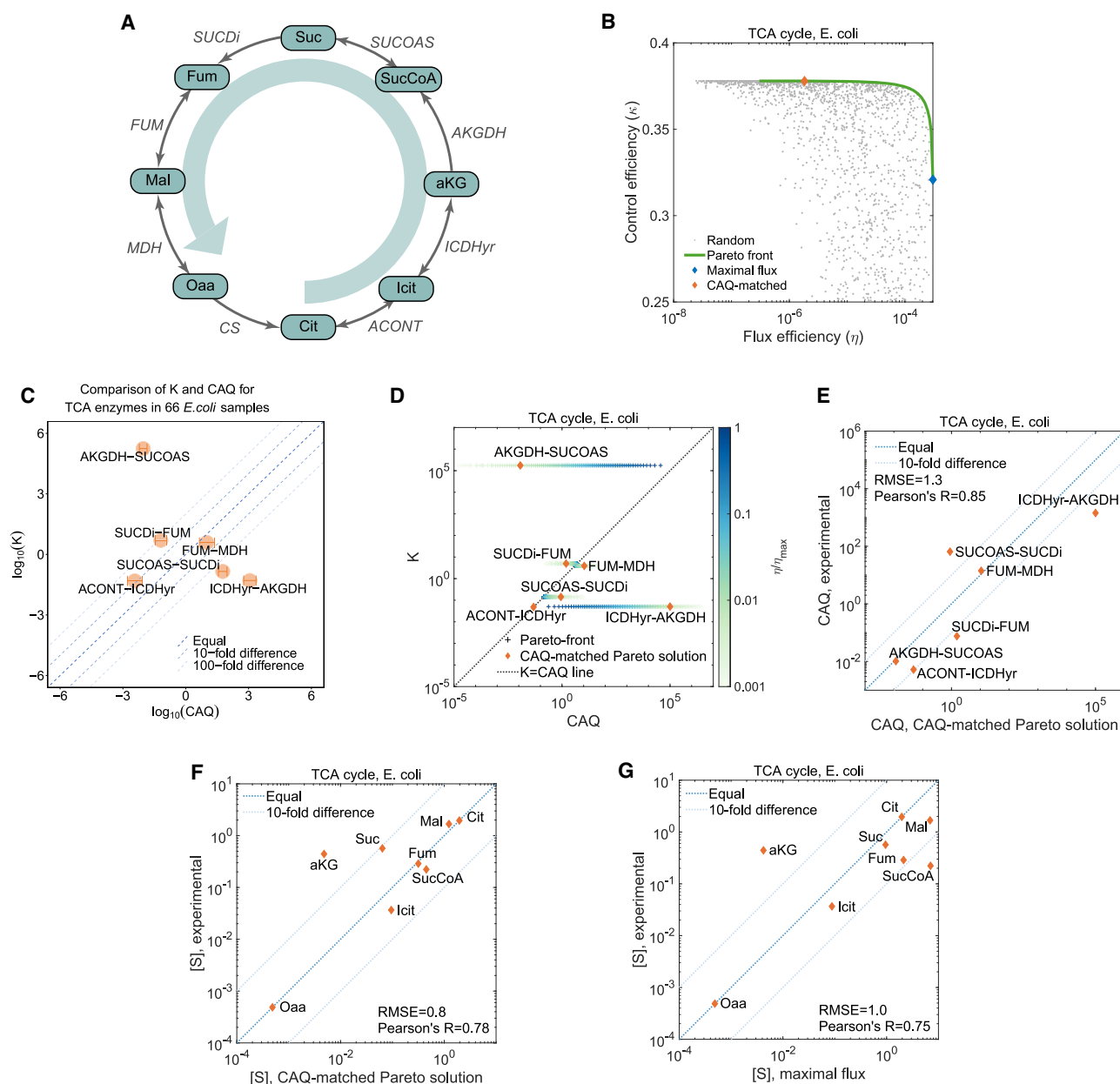


Figure 7. Flux-control trade-off in *E. coli* TCA cycle.

(A) Scheme of the TCA cycle in *E. coli*. (B) Scatter plot showing flux efficiency and control efficiency in randomly sampled enzyme allocation profiles (gray dots), Pareto front balancing these two (green curve), and the CAQ-matched Pareto solution (orange diamond). Sample size: $n = 2000$ (random), 101 (Pareto). (C) Scatter plot comparing K and CAQ for reaction pairs in *E. coli* TCA cycle. CAQ values were estimated from proteomics data for *E. coli*. Dots indicate the mean of $\log_{10}(CAQ)$, and horizontal bars indicate the standard deviation of $\log_{10}(CAQ)$ across different samples. Sample size: $n = 66$ experimental conditions. (D) Relationship between K and CAQ on the Pareto front balancing flux efficiency and control efficiency. Sample size: $n = 101$ Pareto solutions. (E) Scatter plot comparing CAQ values between the CAQ-matched Pareto solution and experimental values computed from proteomics data. (F) Scatter plot comparing metabolite concentrations predicted by the CAQ-matched Pareto solution to experimental measurements. (G) Scatter plot comparing metabolite concentrations predicted by the maximal-flux solution to experimental measurements. Source data are available online for this figure.

(0.04 for glycolysis in *H. sapiens*, 0.06 for glycolysis in *S. cerevisiae*, 0.01 for glycolysis in *E. coli*, and 0.07 for the TCA cycle in *E. coli*, Appendix Fig. S12, “Methods”). Second, to test whether the Pareto front can predict experimental CAQ profiles without using proteomics data to pre-select a physiological Pareto solution, we used experimental metabolite concentrations instead of proteomics to select a metabolite-matched Pareto solution, defined as the

Pareto solution with predicted metabolite concentrations closest to the experimental measurements in log10 space (“Methods”). The metabolite-matched Pareto solution showed moderate-to-strong agreement with the experimental CAQ profiles for glycolysis across all three species (Pearson’s R ranging from 0.58 to 0.71 and RMSE ranging from 1.2 to 1.8, Appendix Fig. S13). The correlation was weaker for the *E. coli* TCA cycle (Appendix Fig. S13), largely due to

the discrepancy for the AKGDH-SUCOAS reaction pair, whose CAQ varies substantially along the Pareto front (Fig. 7D). Nevertheless, 21 of 35 reaction pairs differed by less than 10-fold between the predicted and experimental CAQ values, supporting the general agreement between the Pareto front and experimental CAQ profiles. Importantly, because the metabolomics and proteomics datasets were obtained from independent studies and were not sample-matched, this analysis provides cross-dataset evidence that these two independent omics modalities converge on a shared underlying region of the Pareto front. Taken together, these results show that the agreement between experimental CAQ profiles and the Pareto front is not an artifact of the CAQ-matching procedure, and that the predictive power of the Pareto front can be tested without using proteomics data to select a physiological Pareto solution.

Determinants of glycolytic efficiency optimization

The theory of flux-control trade-off predicts that, in real cells, metabolic pathways operate within the Pareto front, balancing the objectives of flux efficiency maximization and control efficiency maximization. The preference between the two objectives could vary across cell types and conditions, as the EOI values varied across cell types (Fig. 4E) and experimental conditions (Fig. 4H). Furthermore, using the mathematical model of glycolysis we developed previously (“Methods”), we computed several candidate metrics for the EOI alongside the Pareto front for the flux-control trade-off (Appendix Fig. S14) and confirmed that the currently used EOI metric, RMSE between $\log_{10}(K)$ and $\log_{10}(\text{CAQ})$, increases monotonically and continuously alongside the Pareto front as flux efficiency η decreases. Therefore, the EOI value quantitatively captures the position of a solution along the Pareto front between flux-efficiency maximization and concentrated flux control.

Here, we sought to identify potential determinants of the EOI of glycolysis. We correlated gene expression levels to EOI of glycolysis in each type of TCGA tumors, and performed gene set enrichment analysis (GSEA) to identify KEGG pathways associated with lower EOI (i.e., better agreement with the $K = \text{CAQ}$ relationship) in each cancer type (Fig. 8A). We found that lower glycolytic EOI was consistently associated with the activation of multiple oncogenic signaling pathways across all cancer types (Fig. 8B,C; Appendix Fig. S15), suggesting that aggressive tumors have a stronger drive to optimize glycolytic efficiency. Another universal trend observed in the TCGA tumors is the association between higher EOI of glycolysis and enhanced expression of ribosome (Appendix Fig. S16a,b) and oxidative phosphorylation (OXPHOS) genes (Appendix Fig. S16c,d). This trend indicates a potential trade-off between glycolytic efficiency and other energy-consuming processes and alternative energy-producing routes.

Another factor that may affect the optimization of glycolytic efficiency is the total cellular investment in glycolytic enzymes. We computed the fraction of glycolytic enzymes in the total cellular proteome and correlated it with the EOI of glycolysis. A positive correlation between these two was observed in most of the samples (16 positive correlations out of 19 groups of samples, Fig. 8D; Appendix Fig. S17), suggesting that a tighter constraint on total available glycolytic enzyme abundance is a key driver for maximization of the pathway efficiency. This mirrors a principle

of rational economics in which optimal efficiency was necessitated under scarcity of resources.

Taken together, these results highlight the heterogeneous nature of glycolytic efficiency optimization and reveal potential determinants of it. The link to oncogenic signaling suggests that pathological processes could change the economic decision of cells, while the inverse relationship with proteomic investment provides a clue that resource constraints may shape metabolic efficiency.

Discussion

In this study, we derived a quantitative K-CAQ coupling relationship linking thermodynamics and kinetics underlying optimal flux efficiency of metabolic pathways and compared it with experimentally measured enzyme allocation profiles across species and cell types. The moderate agreement between this relationship and omics data across species suggests that optimization of flux efficiency serves as one objective of metabolic pathways, while the deviation from this relationship in key control points of glycolysis and the TCA cycle indicates the existence of alternative objectives. The simplicity of this K-CAQ relationship also opens a route for rationally engineering synthetic pathways towards maximal efficiency, addressing a key challenge where engineered metabolic pathways often suffer from extremely low efficiency due to the lack of a predictive theoretical framework.

Regarding the K-CAQ coupling relationship, several points need to be clarified. First, it should not be interpreted as a biochemical necessity imposed directly by thermodynamics or enzyme biology. Rather, it emerges as the optimal solution with the objective of maximizing flux efficiency under enzyme resource constraints, instead of a mandatory physical law that all pathways must satisfy under all conditions. Alignment of *in vivo* metabolic pathways to this relationship is feasible because cells can adjust CAQ values through evolution of enzyme kinetic parameters and regulation of enzyme abundances to match the K values.

Second, the coupling between K and CAQ at maximal flux efficiency reflects a strategy of kinetics-thermodynamics matching. For each reaction, its rate can be decoupled into the product of the forward flux and a thermodynamic term $1 - \frac{[P]}{K[S]}$ (Appendix Fig. S2a, Appendix mathematical derivations, section 2.1). Two reasons can lead to a low reaction rate: (1) the reaction rate is limited by kinetics, in which the forward flux is low because of insufficient substrate, low expression or activity of the enzyme; (2) the reaction rate is limited by thermodynamics, in which it operates in a near-equilibrium state and carries high backward flux ($1 - \frac{[P]}{K[S]} \approx 0$). Reactions with large K are generally less likely to be limited by thermodynamics, hence allocating more enzymes to these reactions would help overcome the limitations by kinetics and boost the reaction rate. This intuition explains why reactions with larger equilibrium constants tend to receive greater enzyme investment under maximal flux efficiency. Nevertheless, as CAQ depends on enzyme abundance and kinetic parameters of two adjacent reactions, equilibrium constants alone do not completely determine the amount of enzyme resource allocated into a reaction step. As an example, we compared the theoretically predicted enzyme abundances following the K-CAQ relationship and equilibrium constants of reactions in glycolysis and the TCA cycle and showed that reactions with greater K do not necessarily have higher enzyme

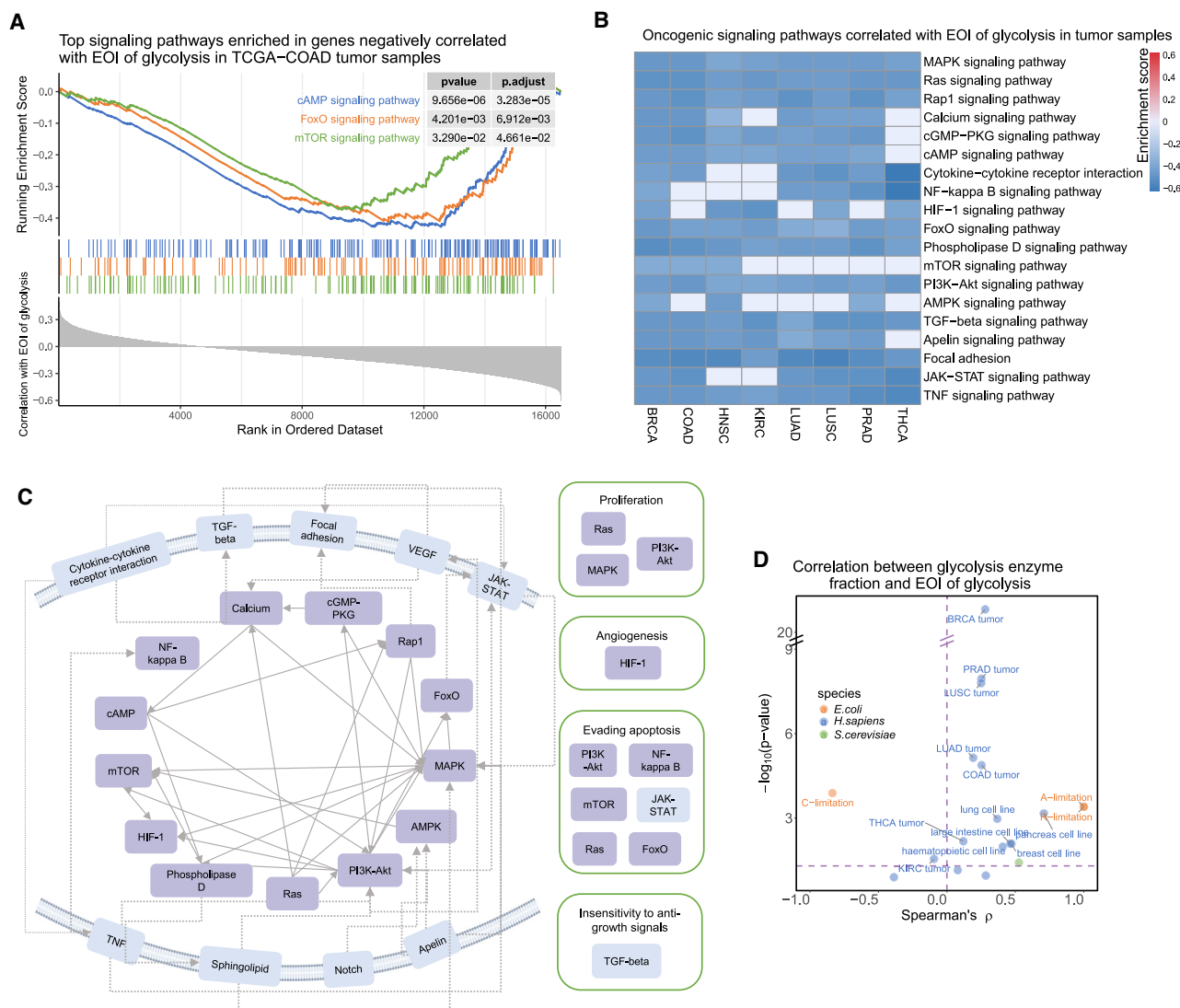


Figure 8. Determinants of glycolytic efficiency optimization.

(A) Top pathways enriched in genes negatively correlated with EOI of glycolysis in TCGA-COAD. Statistical test: permutation test. (B) Enrichment scores for oncogenic signaling pathways negatively associated with EOI of glycolysis. (C) Regulatory links between oncogenic signaling pathways associated with EOI of glycolysis. Blue blocks indicate membrane-associated pathways and purple blocks indicate membrane-unassociated pathways. (D) Spearman's rank correlation between the fraction of glycolytic enzyme in the total cellular proteome and EOI of glycolysis. Sample size: $n = 19$ groups of samples. Source data are available online for this figure.

abundance, although a positive correlation between these two exists (Appendix Fig. S18).

Acknowledging maximization of flux efficiency as one important metabolic objective, previous studies have investigated conditions for optimal efficiency from the perspectives of enzyme kinetics, reaction thermodynamics, and flux control, showing that the state of maximal flux efficiency is associated with evenly distributed thermodynamic driving force along the pathway (Noor et al, 2014; Park et al, 2016) and proportionality between enzyme abundances and FCCs (Heinrich and Klipp, 1996). While previous studies progressively established the conceptual connection between thermodynamics, enzyme cost, and metabolic control, our work extends this line of research by deriving an explicit, closed-form optimality condition for enzyme allocation. First, unlike previous theoretical studies that rely on complex numerical

solutions (e.g., the ECM method (Noor et al, 2016)), our work establishes a simple relationship readily applicable to realistic metabolic pathways with biological complexities such as cofactors and isozymes. Numerical simulations confirm its robustness across different rate laws, in which the core coupling between K and CAQ holds under the most complex reversible Michaelis-Menten kinetics.

Second, our theory about the K - CAQ relationship addresses a distinct level of proteome optimization. While previous coarse-grained models focus on global allocation among major functional sectors such as ribosome and metabolism (Basan et al, 2015; Scott et al, 2010; Zhu et al, 2025), our theory operates at the local pathway level, determining how enzymes are optimally distributed among reaction steps to maximize the efficiency of each pathway. Hence, the distinction parallels that between

macro- and micro-economics, together illustrating how economic principles operate at multiple scales to achieve optimal cellular resource allocation.

Moving beyond consideration of flux maximization as the only objective, our theoretical framework of flux-control trade-off provides an explanation of the deviation from the K-CAQ relationship: metabolic pathways tend to downregulate enzymes catalyzing reactions with larger FCCs at maximal flux efficiency, yielding a more concentrated flux control through a few key control points. Such concentrated flux control might enhance the responsiveness of the cell to changes in activity and abundance of enzymes catalyzing these key control steps and reduce the complexity of the upstream regulatory network. A similar idea underlies the theory of network controllability, which seeks to identify a minimal set of driver nodes capable of controlling the dynamics of a complex network (Liu et al, 2011). The phenomenon that a few key regulators control a large network is observed in many biological systems – examples include operons that involve a cluster of genes controlled by a single promoter, master regulators controlling the developmental landscape of cell fates (Davis and Rebay, 2017), and command neurons controlling animal behaviors (Flood et al, 2013). The broad existence of these controllers suggests that concentrating control within a reduced set of regulators may represent a common design principle across biological systems, yet concentrating control might not be free of cost. Our results raise the possibility that metabolic pathways adopt a similar strategy by concentrating flux control into a small number of reactions at the expense of reduced flux efficiency.

Another important output of our study is the EOI metric that quantifies the position of a pathway within the Pareto front between maximal flux efficiency and concentrated flux control. This enables systematic investigation of efficiency optimality across biological conditions. Applying this metric, we found that two key patterns emerge: first, EOI was negatively correlated with oncogenic signaling in human tumors, suggesting that malignant cells not only upregulate glycolysis, but also optimize its efficiency; second, EOI was positively correlated with total proteomic investment in glycolysis, revealing a “scarcity-driven efficiency” where tighter resource constraints necessitate stronger efficiency optimization. This pattern is reminiscent of human rational decision-making under budget limitations, indicating an intrinsic similarity between fundamental principles of natural and societal processes.

Finally, our theoretical framework of flux-control trade-off does not exclude other competing metabolic objectives. Other goals, such as robustness to environmental fluctuations (Basan et al, 2020; Shen et al, 2024), minimization of metabolite load, and facilitation of metabolic switch, also have profound effects, resulting in a pre-optimized thermodynamic pattern of pathways (Fan et al, 2025), precise tuning of metabolite concentrations (Tepper et al, 2013), and scaling between metabolites and enzymes (Dourado et al, 2021). A complete understanding of how cells balance these potential objectives requires both the construction of a higher-dimensional Pareto front and integration of multi-omics data to map how different cellular states navigate this Pareto landscape. Looking forward, such analysis should extend to genome-scale metabolic networks to provide a quantitative description of how cells balance multiple competing objectives under varying physiological conditions.

Methods

Reagents and tools table

Reagent/ resource	Reference or source	Identifier or catalog number
Experimental models		
Recombinant DNA		
Antibodies		
Oligonucleotides and other sequence-based reagents		
Chemicals, enzymes, and other reagents		
Software		
MATLAB R2024b	MathWorks (https://www.mathworks.com/)	
R 4.3.2	R Project (https://www.r-project.org/)	
Other		

Linear pathway model with Michaelis-Menten kinetics

A linear pathway model with 10 reaction steps was used to simulate the optimal enzyme allocation maximizing the flux efficiency using MATLABR2024b. The dynamics of the intermediate metabolite concentrations in the pathway can be modeled using the ordinary differential equations (ODEs) below:

$$\frac{d[S_i]}{dt} = v_i - v_{i+1}$$

In which $i = 1, 2, \dots, 9$. For the first reaction in the pathway, its substrate is the input metabolite S_{in} with a fixed concentration $[S_{in}]$. For the last reaction in the pathway, its product is the output metabolite S_{out} with a fixed concentration $[S_{out}]$. Otherwise, the substrate and product of the i th reaction is S_{i-1} and S_i , respectively.

Reaction rate of the i th reaction in the pathway is calculated from the reversible Michaelis-Menten rate law (Appendix mathematical derivations) (Cornish-Bowden, 2012):

$$v_i = \frac{k_{cat,i}[E_i] \frac{[S_{i-1}]}{K_{S,i}}}{1 + \frac{[S_{i-1}]}{K_{S,i}} + \frac{[S_i]}{K_{P,i}}} \left(1 - \frac{[S_i]}{K_i[S_{i-1}]} \right)$$

In which $[S_{i-1}]$ is the concentration of the substrate, $[S_i]$ is the concentration of the product, $k_{cat,i}$ is the turnover number of the enzyme, $K_{S,i}$ is the Michaelis constant for the substrate, $K_{P,i}$ is the Michaelis constant of the product, K_i is the equilibrium constant, and $[E_i]$ is the concentration of the enzyme. The reaction thermodynamic and kinetic parameters and boundary metabolite concentrations were randomly sampled from realistic datasets for human metabolites and metabolic enzymes. For the turnover number k_{cat} and Michaelis constant K_m , k_{cat} values predicted by DLKcat (Li et al, 2023; Li et al, 2022) and K_m values predicted by the machine learning model developed by Kroll et al (Kroll et al, 2021) were used. For the equilibrium constants K , standard Gibbs free energy change values ($\Delta_r G^\circ$) for human metabolic reactions predicted by the dGbyG algorithm were used. For concentrations of the input and output metabolites $[S_{in}]$ and $[S_{out}]$, experimentally measured absolute concentrations of metabolites in a human cell

line were used. We randomly sampled 10,000 combinations of enzyme parameters and metabolite concentrations from these datasets. Only parameter sets producing positive pathway flux were kept for the following analyses.

For each parameter set and a candidate enzyme abundance profile E , the steady state concentrations of intermediate metabolites $[S_i]$ was computed by solving the steady state equations:

$$\frac{d[S_i]}{dt} = v_i - v_{i+1} = 0$$

As the reaction rates v_i and v_{i+1} can be written as functions of the kinetic parameters k_{cat} , K_S , K_P , the equilibrium constants K , intermediate metabolite concentrations S , enzyme abundances E , and the boundary metabolite concentrations $[S_{in}]$ and $[S_{out}]$, the steady state equations can be written as the form below:

$$v_i(k_{cat}, K_S, K_P, K, E, S, [S_{in}], [S_{out}]) = v_{i+1}(k_{cat}, K_S, K_P, K, E, S, [S_{in}], [S_{out}])$$

Because directly solving the nonlinear steady state equations using the MATLAB function `fsolve[]` has numerical issues, we converted the problem into a least-squares optimization problem:

$$\min_S ||v_i(k_{cat}, K_S, K_P, K, E, S, [S_{in}], [S_{out}]) - v_{i+1}(k_{cat}, K_S, K_P, K, E, S, [S_{in}], [S_{out}])||_2$$

This problem was solved using the MATLAB function `lsqnonlin[]` to address numerical difficulties of directly solving the nonlinear equations. Steady state pathway flux J was then computed from the steady-state metabolite concentrations based on the rate law. As the whole procedure computes pathway flux J from k_{cat} , K_S , K_P , K , E , $[S_{in}]$, $[S_{out}]$, the pathway flux J can also be written as a function of these parameters and concentrations:

$$J = J(k_{cat}, K_S, K_P, K, E, [S_{in}], [S_{out}])$$

Optimal enzyme allocation profile maximizing flux efficiency under a given parameter set was computed using the MATLAB constrained optimization solver `fmincon[]` to maximize the pathway flux with the upper bound of total enzyme abundance set to 1:

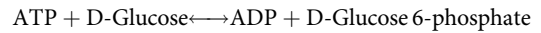
$$\max_E J(k_{cat}, K_S, K_P, K, E, [S_{in}], [S_{out}])$$

$$\text{s.t.} \begin{cases} 0 \leq [E_i] \leq 1 \\ \sum_{i=1}^n [E_i] \leq 1 \end{cases}$$

Estimation of equilibrium constant

Equilibrium constants (K) of each reaction were calculated from standard Gibbs energy changes ($\Delta_r G^\circ$) based on the equation $\Delta_r G^\circ = -RT \ln K$, with $R = 8.314472 \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$, $T = 298.15 \text{ K}$ (*H. sapiens* and *S. cerevisiae*), or $T = 310.15 \text{ K}$ (*E. coli*). $\Delta_r G^\circ$ of glycolysis reactions for eukaryotes were obtained from experimental data in literature (Mathews, 2013). For *E. coli*, $\Delta_r G^\circ$ values were predicted using the machine learning model dGbyG (Fan et al, 2025), under physiological conditions compiled from literature (Hurwitz and Rosano, 1967; Noor et al, 2013) (Dataset EV1). Subsequently, the adjusted equilibrium constant (K') was calculated from the original equilibrium constant K and intracellular concentrations of cofactors by multiplying the ratio of the product of concentrations of cofactors serving as products to

the product of concentrations of cofactors serving as substrates by the original equilibrium constant K . For example, the phosphorylation of glucose catalyzed by hexokinase has the chemical equation below:



Here we treated ATP and ADP as cofactors and used their intracellular concentrations to compute the adjusted equilibrium constant:

$$K' = K \frac{[\text{ATP}]}{[\text{ADP}]}$$

The adjusted equilibrium constants in the three organisms are available in Dataset EV2. Substrate concentrations were obtained from published datasets (Bashir et al, 2015; Chowdhury, 2020; Park et al, 2016; Thorfinnsdottir et al, 2023) (Dataset EV3).

Estimation of catalytic efficiency

We estimated the catalytic efficiency of metabolic enzymes using reversible Michaelis–Menten kinetics, considering substrate saturation. In this case, the catalytic efficiency of an enzyme can be computed from its kinetic parameters k_{cat} and K_m , together with the concentration of substrate:

$$a \approx \frac{k_{cat}}{K_m + [S]}$$

For the enzyme kinetic parameters k_{cat} and K_m , since experimental values from the BRENDA enzyme database (Chang et al, 2021) are missing for the wildtype form of several enzymes in glycolysis and individual measurements from different sources vary a lot, we used values predicted based on amino acid sequences of the catalytic subunit of each enzyme and molecular structures of its substrates by two deep learning algorithms, including k_{cat} values predicted by DLKcat (Li et al, 2023; Li et al, 2022) and K_m values predicted by the machine learning model developed by Kroll et al (Kroll et al, 2021). Amino acid sequences of the enzymes obtained from UniProt and SMILES strings encoding structures of the metabolites retrieved from MetaNetX (Moretti et al, 2021) were used in the prediction (Consortium, 2025). The predicted values of enzyme kinetic parameters were available in Dataset EV4. For the TCA cycle in *E. coli*, these kinetic parameters, as well as thermodynamic parameters and metabolite concentrations, are available in Dataset EV5. Experimentally measured absolute metabolite concentrations (Bashir et al, 2015; Chowdhury, 2020; Park et al, 2016; Thorfinnsdottir et al, 2023) in human cells, the bacteria *E. coli*, and the unicellular eukaryote *S. cerevisiae* were used together with the enzymatic parameters to estimate the catalytic efficiency a .

A specific case that needs to be addressed is the GLCptspp reaction catalyzed by the phosphotransferase system (PTS) in *E. coli*. The PTS consists of three components: enzyme I (EI), histidine phosphocarrier protein (HPr), and membrane-bound sugar-specific permeases (enzymes II, EII) (Kottrba et al, 2001). In the iML1515 model, the gene-reaction rules for GLCptspp include the following genes: *ptsI* (EI), *ptsH* (HPr), *crr* (EII), *ptsG* (EII), *malX* (EII), *manX* (EII), *manY* (EII), and *manZ* (EII). We used the amino acid sequence of the subunit *ptsG* responsible for specific glucose

binding and catalysis (EC:2.7.1.199) in the prediction of the kinetic parameters.

Omics data processing

Information about reactions in glycolysis, including their substrates, products, and gene-protein-reaction relationships, was extracted from the following genome-scale metabolic models (GEMs): Recon3D (*H. sapiens*) (Brunk et al, 2018), iML1515 (*E. coli*) (Monk et al, 2017), and iMM904 (*S. cerevisiae*) (Mo et al, 2009). This study utilized three published proteome datasets. The proteomics data for 378 human cancer cell lines were obtained from the Cancer Cell Line Encyclopedia (<https://sites.broadinstitute.org/ccle/datasets>) (Nusinow et al, 2020). The mass spectrometry intensities of the proteins were normalized into molar fractions using the molecular mass of the enzymes. The proteomics datasets for *E. coli* and *S. cerevisiae* were retrieved from two published datasets summarizing absolute protein copy numbers in the two species across independent proteomics studies (Ho et al, 2018; Mori et al, 2021).

Transcriptomics datasets in human tumors and normal tissue samples were obtained from The Cancer Genome Atlas (TCGA) database (<https://www.cancer.gov/tcga>). Transcripts Per Million (TPM) data for BRCA, HNSC, KIRC, LUAD, LUSC, PRAD, and THCA were downloaded directly. Original paired-end FASTQ files for COAD were downloaded and mapped to the human genome (GRCh38) using STAR (version 2.7.9a) (Dobin et al, 2013). Read counts were then calculated using featureCounts (version 1.5.3) (Liao et al, 2014), and subsequently normalized into TPM using R (version 4.3.2). Zero expression levels were replaced with a small positive number of 10^{-6} to avoid zero or infinity values of the resulting CAQ. The fraction of glycolytic enzymes in the entire cellular proteome was computed by dividing the total protein copy number of glycolytic enzymes by the total copy number of all proteins (proteomics datasets) or approximated by computing the total TPM of all glycolytic genes (transcriptomics data).

Mathematical models for glycolysis and the TCA cycle

Linear metabolic pathway models with first-order kinetics were constructed to represent glycolysis and the TCA cycle. Each linear metabolic pathway starts from an input metabolite S_{in} that is converted to an output metabolite S_{out} through a sequence of reaction steps. For the glycolysis models, extracellular glucose (Glc(e)) was treated as the input metabolite, and pyruvate (Pyr) was treated as the output metabolite. For the TCA cycle model, citrate (Cit) was treated as the input metabolite, and oxaloacetate (Oaa) was treated as the output metabolite. The rate for the i th reaction was computed as follows:

$$v_i = a_i[E_i] \left([S_{i-1}] - \frac{[S_i]}{K_i} \right)$$

In which a_i is the catalytic efficiency estimated for the i th enzyme (see “Estimation of catalytic efficiency” for details), $[E_i]$ is the concentration of enzyme, $[S_{i-1}]$ is the concentration of the substrate S_{i-1} , $[S_i]$ is the concentration of the product S_i , K_i is the equilibrium constant adjusted by cofactor concentrations (see “Estimation of equilibrium constants” for details).

We first computed steady state concentrations of intermediate metabolites by solving the steady state equations below:

$$a_i[E_i] \left([S_{i-1}] - \frac{[S_i]}{K_i} \right) = a_{i+1}[E_{i+1}] \left([S_i] - \frac{[S_{i+1}]}{K_{i+1}} \right), i = 1, 2, \dots, n-1$$

These are linear equations for the intermediate metabolite concentrations S and were solved using the MATLAB matrix operator “\”. After computing the steady state metabolite concentrations, the pathway flux J can be computed as follows:

$$J = a_1[E_1] \left([S_{in}] - \frac{[S_1]}{K_1} \right)$$

Finally, FCCs $C_{v_i}^J$ can be computed as follows (Dai and Locasale, 2018) (Appendix mathematical derivations, section 2.4.3):

$$C_{v_i}^J = \frac{1}{a_i[E_i] \left(\prod_{l=1}^{i-1} K_l \right) \left(\sum_{j=1}^n \frac{1}{a_j[E_j] \prod_{m=1}^{j-1} K_m} \right)}$$

Construction of the Pareto front

The Pareto front for the trade-off between flux efficiency and control efficiency was constructed using ε -constraint algorithm. Briefly, we first computed the maximal flux efficiency from the enzyme allocation profile satisfying the K-CAQ relationship. We assumed that the total abundance of enzymes is equal to 1, as the flux efficiency is invariant upon scaling enzyme abundances. Furthermore, because the flux efficiency η is defined as the ratio of flux J to total enzyme abundance, we have $\eta = J$ when the total enzyme abundance equals 1, therefore can simplify the following analyses. The solution of enzyme abundance at maximal flux efficiency satisfies the equations below:

$$\begin{cases} K_i = \frac{a_i[E_i]^2}{a_{i+1}[E_{i+1}]^2} \\ \sum_{i=1}^n [E_i] = 1 \end{cases}$$

According to these equations, enzyme abundances maximizing flux efficiency can be computed from the catalytic efficiencies and equilibrium constants:

$$[E_i] = \frac{1}{\sqrt{a_i \prod_{l=1}^{i-1} K_l} \left(\sum_{j=1}^n \frac{1}{\sqrt{a_j \prod_{l=1}^{j-1} K_l}} \right)}$$

We computed the maximal flux $J_{\max}$ from the enzyme abundances above. A series of constrained optimization problems were then constructed to compute solutions within the Pareto front:

$$\begin{aligned} \max_E \kappa &= \sqrt{\frac{1}{n} \sum_{i=1}^n \left(C_{v_i}^J - \frac{1}{n} \sum_{j=1}^n C_{v_j}^J \right)^2} \\ \text{s.t. } &\begin{cases} \sum_{i=1}^n [E_i] = 1 \\ J \geq \alpha J_{\max} \end{cases} \end{aligned}$$

In which $C_{v_i}^J$ is the flux control coefficient for the i -th reaction, $\alpha \in (0, 1)$ is a value quantifying the minimal allowed flux J relative

to the maximal flux $J_{\max}$. A larger value of α indicates a stronger tendency to maximize flux J , while a smaller value of α indicates a stronger tendency to maximize the control efficiency κ . 100 different values of α ranging from 10^{-3} to $10^{-0.03}$ were used to solve the optimization problem and trace the Pareto front:

$$\alpha_i = 10^{-0.03i}, i = 1, 2, \dots, 100$$

The resulting constrained optimization problems were solved using the MATLAB function `fmincon()`, generating 100 Pareto solutions corresponding to those different values of α . These Pareto solutions were then compared to real CAQ values estimated from proteomics data to define the CAQ-matched Pareto solution.

Quantification of directional agreement in K-CAQ deviation

For the i -th reaction pair in a species-pathway combination (glycolysis in *H. sapiens*, glycolysis in *S. cerevisiae*, glycolysis in *E. coli*, or TCA cycle in *E. coli*), because the deviation between K and CAQ is unidirectional in the Pareto front, we defined a directional label for the deviation between K and CAQ along the Pareto front:

$$s_{\text{Pareto},j} = \begin{cases} -1, & K_j < \text{CAQ}_j \\ 1, & K_j > \text{CAQ}_j \end{cases}$$

Next, for each experimental CAQ profile computed from proteomics data, $\text{CAQ}_{\text{prot},i} = (\text{CAQ}_{i1}, \text{CAQ}_{i2}, \dots, \text{CAQ}_{i,n-1})$, we also defined directional labels for the deviation between K and CAQ in this CAQ profile:

$$s_{\text{prot},ij} = \begin{cases} -1, & K_j < \text{CAQ}_{ij} \\ 1, & K_j > \text{CAQ}_{ij} \end{cases}$$

The directional agreement A_i in K-CAQ deviation between an experimental CAQ profile $\text{CAQ}_{\text{prot},i}$ and the Pareto front was then defined as the percentage of directional labels consistent between the experimental CAQ profile and the Pareto front:

$$A_i = \frac{\#(s_{\text{Pareto},j} = s_{\text{prot},ij})}{n-1}$$

In which $n-1$ is the number of tandem reaction pairs in the pathway. For the randomized controls, each CAQ profile was permuted once by randomly shuffling the labels of reaction pairs. Directional labels and directional agreement for the deviation between K and CAQ were then computed for the shuffled CAQ profiles using the same procedure described above.

Computation of CAQ-matched and metabolite-matched Pareto solutions

For each species-pathway combination, we first computed the average CAQ across all available proteomics datasets. This average CAQ profile served as the reference for identifying the CAQ-matched Pareto solution.

Let CAQ_{prot} denote the average CAQ profile computed from proteomics, and CAQ_i denote the CAQ vector computed from the i th Pareto solution (i.e., the Pareto solution corresponding to α_i in the section “Construction of Pareto front”), we computed the

Euclidean distance between CAQ_{prot} and CAQ_i in the log10-space:

$$d_i = \|\log_{10}(\text{CAQ}_{\text{prot}}) - \log_{10}(\text{CAQ}_i)\|_2$$

The Pareto solution with the smallest d_i among all 100 Pareto solutions was selected as the CAQ-matched Pareto solution.

Similarly, to select the metabolite-matched Pareto solution, we computed steady-state metabolite concentrations for each Pareto solution. Let S_{exp} denote the experimentally measured metabolite concentration vector, and S_i denote the metabolite concentration vector computed from the i th Pareto solution, we computed the Euclidean distance between S_{exp} and S_i in log10-space, and selected the Pareto solution with smallest Euclidean distance as the metabolite-matched Pareto solution.

Permutation analysis of experimental CAQ profiles

For each species-pathway combination, the cross-sample average CAQ profile was permuted 1000 times by randomly shuffling the labels of reaction pairs, generating 1000 permuted CAQ profiles. Let $\text{CAQ}_{\text{perm},i}$ denote the i th permuted CAQ profile, its distance to the Pareto front was computed by the smallest distance between it and all 100 Pareto solutions in log10-space:

$$d_{\text{perm},i} = \min_j \|\log_{10}(\text{CAQ}_{\text{perm},i}) - \log_{10}(\text{CAQ}_j)\|_2$$

Let d_{prot} denote the Euclidean distance between the unpermuted experimental CAQ profile and the Pareto front:

$$d_{\text{prot}} = \min_j \|\log_{10}(\text{CAQ}_{\text{prot}}) - \log_{10}(\text{CAQ}_j)\|_2$$

The permutation P value was defined as the fraction of $d_{\text{perm},i}$ that fell below d_{prot} in all 1000 values of $d_{\text{perm},i}$.

Gene set enrichment analysis (GSEA)

All transcriptome datasets were divided into tumor and normal groups based on sample information. Within each group, the mean TPM value for each gene was calculated for samples sharing the same EOI value. Spearman's correlation coefficient between each gene's expression values and EOI was then computed, representing the degree to which the gene's expression is coupled with glycolytic efficiency. Genes were ranked based on this coefficient. KEGG pathway enrichment analysis was performed in R (version 4.3.2) using the clusterProfiler package (version 4.10.1) (Wu et al, 2021) with an adjusted P value (P_{adj}) cutoff of <0.05 .

Data availability

Thermodynamics, kinetics, and omics data used in this study and scripts for validation of the theory are available at the GitHub repository: <https://github.com/bioinf-kud/Optimality-in-glycolysis-enzyme-allocation>. Scripts for simulating enzyme allocation, flux efficiency, and flux-control trade-offs in linear metabolic pathways under different rate laws are available at the GitHub repository: https://github.com/ziweidai/Linear_metabolic_pathway_optimization_and_control.

The source data of this paper are collected in the following database record: [biostudies:S-SCDT-10_1038-S44320-026-00250-5](https://biostudies.org/studies/S-SCDT-10_1038-S44320-026-00250-5).

Expanded view data, supplementary information, appendices are available for this paper at <https://doi.org/10.1038/s44320-026-00250-5>.

Peer review information

A peer review file is available at <https://doi.org/10.1038/s44320-026-00250-5>

References

- Akashi H, Gojobori T (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci USA* 99:3695–3700
- Basan M, Honda T, Christodoulou D, Horl M, Chang YF, Leoncini E, Mukherjee A, Okano H, Taylor BR, Silverman JM et al (2020) A universal trade-off between growth and lag in fluctuating environments. *Nature* 584:470–474
- Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, Hwa T (2015) Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* 528:99–104
- Bashir A, Gropler R, Ackerman J (2015) Absolute quantification of human liver phosphorus-containing metabolites in vivo using an inhomogeneous spoiling magnetic field gradient. *PLoS ONE* 10:e0143239
- Brown GC (1991) Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. *J Theor Biol* 153:195–203
- Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, Gatto F, Nilsson A, Preciat Gonzalez GA, Aurich MK et al (2018) Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* 36:272–281
- Buccitelli C, Selbach M (2020) mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* 21:630–644
- Chandra FA, Buzi G, Doyle JC (2011) Glycolytic oscillations and limits on robust efficiency. *Science* 333:187–192
- Chang A, Jeske L, Ulbrich S, Hofmann J, Koblit J, Schomburg I, Neumann-Schaal M, Jahn D, Schomburg D (2021) BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res* 49:D498–D508
- Chen Y, Nielsen J (2022) Yeast has evolved to minimize protein resource cost for synthesizing amino acids. *Proc Natl Acad Sci USA* 119:e2114622119
- Chowdhury HH (2020) Differences in cytosolic glucose dynamics in astrocytes and adipocytes measured by FRET-based nanosensors. *Biophys Chem* 261:106377
- Consortium TU (2025) UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res* 53:D609–D617
- Cornish-Bowden A (2012) Fundamentals of enzyme kinetics. Wiley-Blackwell, Weinheim, Germany
- Dai Z, Locasale JW (2018) Thermodynamic constraints on the regulation of metabolic fluxes. *J Biol Chem* 293:19725–19739
- Davis TL, Rebay I (2017) Master regulators in development: views from the *Drosophila* retinal determination and mammalian pluripotency gene networks. *Dev Biol* 421:93–107
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21
- Dourado H, Mori M, Hwa T, Lercher MJ (2021) On the optimality of the enzyme-substrate relationship in bacteria. *PLoS Biol* 19:e3001416
- Fan W, Hao Y, Hou X, Ding C, Huang D, Zheng W, Dai Z (2025) Unraveling principles of thermodynamics for genome-scale metabolic networks using graph neural networks. *Cell Syst* 16:101393
- Flamholz AI, Goyal A, Fischer WW, Newman DK, Phillips R (2025) The proteome is a terminal electron acceptor. *Proc Natl Acad Sci USA* 122:e2404048121
- Flood TF, Iguchi S, Gorczyca M, White B, Ito K, Yoshihara M (2013) A single pair of interneurons commands the *Drosophila* feeding motor program. *Nature* 499:83–87
- Green SR, Storey KB (2020) Regulation of the alpha-ketoglutarate dehydrogenase complex during hibernation in a small mammal, the Richardson's ground squirrel (*Urocyon richardsonii*). *Biochim Biophys Acta Proteins Proteom* 1868:140448
- Hansen GE, Gibson GE (2022) The alpha-ketoglutarate dehydrogenase complex as a hub of plasticity in neurodegeneration and regeneration. *Int J Mol Sci* 23:12403
- Heinrich R, Klipp E (1996) Control analysis of unbranched enzymatic chains in states of maximal activity. *J Theor Biol* 182:243–252
- Heizer EM Jr, Raiford DW, Raymer ML, Doom TE, Miller RV, Krane DE (2006) Amino acid cost and codon-usage biases in 6 prokaryotic genomes: a whole-genome analysis. *Mol Biol Evol* 23:1670–1680
- Ho B, Baryshnikova A, Brown GW (2018) Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome. *Cell Syst* 6:192–205.e193
- Hurwitz C, Rosano CL (1967) The intracellular concentration of bound and unbound magnesium ions in *Escherichia coli*. *J Biol Chem* 242:3719–3722
- Islam MM, Shao W, Batavia M, Ford RM, Pálsson BO, Nielsen J, Maranas CD, Lee SY, Papin JA (2025) Reframing the role of the objective function in its proper context for metabolic network modeling. *Cell Syst* 16:101298
- Khana DB, Jen A, Shishkova E, Fisher K, Thusoo E, Williams J, Henkel A, Stevenson DM, Coon JJ, Amador-Noguez D (2025) Thermodynamics shapes the in vivo enzyme burden of glycolytic pathways. *mBio* 16:e0183725
- Klipp E, Heinrich R (1999) Competition for enzymes in metabolic pathways: implications for optimal distributions of enzyme concentrations and for the distribution of flux control. *Biosystems* 54:1–14
- Kotrba P, Inui M, Yukawa H (2001) Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *J Biosci Bioeng* 92:502–517
- Kroll A, Engqvist MKM, Heckmann D, Lercher MJ (2021) Deep learning allows genome-scale prediction of Michaelis constants from structural features. *PLoS Biol* 19:e3001402
- Li F, Chen Y, Anton M, Nielsen J (2023) GotEnzymes: an extensive database of enzyme parameter predictions. *Nucleic Acids Res* 51:D583–D586
- Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MKM, Kerkhoven EJ, Nielsen J (2022) Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nat Catal* 5:662–672
- Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930
- Liu YY, Slotine JJ, Barabasi AL (2011) Controllability of complex networks. *Nature* 473:167–173
- Mathews CK (2013) Biochemistry. Pearson, Toronto
- Mo ML, Pálsson BØ, Herrgård MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* 3:37
- Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W, Zhang Z, Mori H et al (2017) iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol* 35:904–908
- Moretti S, Tran VDT, Mehl F, Ibberson M, Pagni M (2021) MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Res* 49:D570–D574
- Mori M, Zhang Z, Banaei-Esfahani A, Lalanne JB, Okano H, Collins BC, Schmidt A, Schubert OT, Lee DS, Li GW et al (2021) From coarse to fine: the absolute

Escherichia coli proteome under diverse growth conditions. *Mol Syst Biol* 17:e9536

Mulukutla BC, Yongky A, Daoutidis P, Hu W-S (2014) Bistability in glycolysis pathway as a physiological switch in energy metabolism. *PLoS ONE* 9:e98756

Nemkov T, Stephenson D, Earley EJ, Keele GR, Hay A, Key A, Hairman ZB, Erickson C, Dzieciatkowska M, Reisz JA et al (2024) Biological and genetic determinants of glycolysis: Phosphofructokinase isoforms boost energy status of stored red blood cells and transfusion outcomes. *Cell Metab* 36:1979-1997.e1913

Noor E, Bar-Even A, Flamholz A, Reznik E, Liebermeister W, Milo R (2014) Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput Biol* 10:e1003483

Noor E, Flamholz A, Bar-Even A, Davidi D, Milo R, Liebermeister W (2016) The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. *PLoS Comput Biol* 12:e1005167

Noor E, Haraldsdóttir HS, Milo R, Fleming RMT (2013) Consistent estimation of Gibbs energy using component contributions. *PLoS Comput Biol* 9:e1003098

Noor E, Liebermeister W (2024) Optimal enzyme profiles in unbranched metabolic pathways. *Interface Focus* 14:20230029

Nusinow DP, Szpyt J, Ghandi M, Rose CM, McDonald ER 3rd, Kalocsay M, Jane-Valbuena J, Gelfand E, Schweppe DK, Jedrychowski M et al (2020) Quantitative proteomics of the cancer cell line encyclopedia. *Cell* 180:387-402.e316

Park JO, Rubin SA, Xu Y-F, Amador-Noguez D, Fan J, Shlomi T, Rabinowitz JD (2016) Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat Chem Biol* 12:482-489

Piper MDW, Soultoukis GA, Blanc E, Mesaros A, Herbert SL, Juricic P, He X, Atanassov I, Salmonowicz H, Yang M et al (2017) Matching dietary amino acid balance to the in silico-translated exome optimizes growth and reproduction without cost to lifespan. *Cell Metab* 25:1206

Schwanhaussner B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. *Nature* 473:337-342

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* 330:1099-1102

Shen Y, Dinh HV, Cruz ER, Chen Z, Bartman CR, Xiao T, Call CM, Ryseck RP, Pratas J, Weilandt D et al (2024) Mitochondrial ATP generation is more proteome efficient than glycolysis. *Nat Chem Biol* 20:1123-1132

Shenhav L, Zeevi D (2020) Resource conservation manifests in the genetic code. *Science* 370:683-687

Tanner LB, Goglia AG, Wei MH, Sehgal T, Parsons LR, Park JO, White E, Toettcher JE, Rabinowitz JD (2018) Four key steps control glycolytic flux in mammalian cells. *Cell Syst* 7:49-62.e48

Tepper N, Noor E, Amador-Noguez D, Haraldsdóttir HS, Milo R, Rabinowitz J, Liebermeister W, Shlomi T (2013) Steady-state metabolite concentrations reflect a balance between maximizing enzyme efficiency and minimizing total metabolite load. *PLoS ONE* 8:e75370

Thorfinnsdóttir LB, Garcia-Calvo L, Bo GH, Bruheim P, Rost LM (2023) Optimized fast filtration-based sampling and extraction enables precise and absolute quantification of the *Escherichia coli* central carbon metabolome. *Metabolites* 13:150

Welle S, Nair KS (1990) Relationship of resting metabolic rate to body composition and protein turnover. *Am J Physiol* 258:E990-E998

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L et al (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* 2:100141

Zhang H, Wang Y, Li J, Chen H, He X, Zhang H, Liang H, Lu J (2018) Biosynthetic energy cost for amino acids decreases in cancer evolution. *Nat Commun* 9:4124

Zhu M, Mori M, Hwa T, Dai X (2025) Distantly related bacteria share a rigid proteome allocation strategy with flexible enzyme kinetics. *Proc Natl Acad Sci USA* 122:e2427091122

Acknowledgements

The authors thank all members of the Dai lab for helpful discussions. We thank Prof. Yu Liu at Beijing Normal University for valuable advice and discussions during the revision process. We thank the Guangdong Major Project of Basic Research (2026B0303000001 to ZD), National Natural Science Foundation of China (12371489 to ZD), National Key Research and Development Program of China (2021YFA0911300 to ZD), Guangdong Program (2021QN02Y856 to ZD), and the Shenzhen Science and Technology Program (KQTD2018041114343237 to ZD) for their generous support.

Author contributions

Kai Sun: Conceptualization; Resources; Data curation; Software; Formal analysis; Validation; Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing. **Weiyang Zheng:** Conceptualization; Resources; Data curation; Software; Formal analysis; Validation; Investigation; Methodology; Writing—original draft. **Wenchao Fan:** Resources; Formal analysis. **Chuyun Ding:** Formal analysis. **Dan Huang:** Data curation. **Ziwei Dai:** Conceptualization; Resources; Formal analysis; Supervision; Funding acquisition; Investigation; Visualization; Methodology; Writing—original draft; Project administration; Writing—review and editing.

Source data underlying figure panels in this paper may have individual authorship assigned. Where available, figure panel/source data authorship is listed in the following database record: [biostudies:S-SCDT-10_1038-S44320-026-00250-5](https://biostudies.org/studies/S-SCDT-10_1038-S44320-026-00250-5).

Disclosure and competing interests statement

The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. Creative Commons Public Domain Dedication waiver <http://creativecommons.org/public-domain/zero/1.0/> applies to the data associated with this article, unless otherwise stated in a credit line to the data, but does not extend to the graphical or creative elements of illustrations, charts, or figures. This waiver removes legal barriers to the re-use and mining of research data. According to standard scholarly practice, it is recommended to provide appropriate citation and attribution whenever technically possible.

© The Author(s) 2026