Optimization of lag phase shapes the evolution of a bacterial enzyme

Bharat V. Adkar[1], Michael Manhart[1], Sanchari Bhattacharyya[1], Jian Tian[1,2], Michael Musharbash[1], Eugene I. Shakhnovich[1*]

[1]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford St., Cambridge, MA 02138, USA

[2]Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, 12 Zhongguancun South Street, Beijing, 100081, China

[*]Correspondence should be addressed to E.S. (shakhnovich@chemistry.harvard.edu)

Abstract

Mutations provide the variation that drives evolution, yet their effects on fitness remain poorly understood. Here we explore how mutations in the essential enzyme Adenylate Kinase (Adk) of *E. coli* affect multiple phases of population growth. We introduce a biophysical fitness landscape for multiple phases of bacterial growth, which shows how they depend on molecular and cellular properties of Adk. We find that Adk catalytic capacity in the cell (product of activity and abundance) is the major determinant of mutational fitness effects. We show that bacterial lag times are at an optimum for the endogenous enzyme, while exponential growth rates are only weakly affected by variation in Adk. Direct pairwise competitions between strains show how environmental conditions modulate the outcome of a competition where growth rates and lag times show a tradeoff, altogether shedding light on the multidimensional nature of fitness and its importance in the evolutionary optimization of enzymes.

Introduction

Random mutagenesis is often used to assess the distribution of fitness effects in simple experimental models such as propagating viruses and microbes evolving under antibiotic stress[1,2]. However, the enormous size of sequence space severely constrains how much of the fitness landscape can be explored this way, and mechanistic and predictive insights from these experiments are further limited by a lack of knowledge of the molecular effects of mutations. Instead, a more targeted experimental approach relies on the concept of a biophysical fitness landscape, in which fitness effects of mutations are mapped through their effects on molecular traits of the mutated proteins. In this approach, biophysically-rational genetic variation is introduced on the chromosome, and the molecular and phenotypic effects of that variation are analyzed concurrently[3-6]. By mapping fitness effects to variation of molecular properties rather than to sequences of mutated proteins, we can dramatically reduce the dimensionality of the genotype-to-phenotype mapping. The underlying hypothesis is that variation in a small number of properly-selected molecular traits of mutated proteins can explain most of the resulting mutational variation in fitness, and that the relationship between these molecular traits and fitness is smooth and continuous. Several recent studies have supported this approach[5-7].

However, the relationship between sequence variation and fitness is further confounded by the fact that multiple phenotypic traits contribute to fitness, and the relative importance of these traits to the long-term evolutionary fate of a mutation[8] may be highly dependent on environmental and ecological conditions. While a large number of traits (e.g., viability at various life phases, mating success, fecundity, etc.) determine fitness of multicellular organisms, relatively fewer components of fitness, such as the time in lag phase, the exponential growth rate, and the overall yield in stationary phase, determine fitness of unicellular microorganisms like bacteria and yeast. When in competition for limiting resources, all phases of growth contribute towards the outcome, and hence determine fitness[3,9]. The relative importance of these different phases of bacterial growth in sculpting the fitness landscape depends on the conditions of growth and competition[10,11 12].

Overall, the challenge in quantitatively characterizing the fitness landscape is twofold: Understanding fitness in terms of contributions from different phases of growth, and linking each of these components to genotypic properties of cells. In this work, we address both challenges by

introducing biophysically-rational genetic variation in the *adk* locus that encodes the essential *E. coli* enzyme Adenylate Kinase (Adk), and projecting the ensuing variations of phenotypic components onto the biophysical traits of Adk. To that end, we assess a comprehensive set of biophysical properties and fitness effects of Adk mutants. We find that a unique combination of molecular and cellular traits of Adk — a product of intracellular abundance and catalytic activity, which we term catalytic capacity — serves as a reliable predictor of fitness effects covering the full range of genotypic and phenotypic variation. Furthermore, we find that the length of the lag phase is more sensitive to variation in Adk catalytic capacity than is the exponential growth rate, so that the lag phase of the wild-type *E. coli* appears to be optimal with respect to broad variation of Adk catalytic capacity.

Results

Biophysical properties of Adk mutants

We chose a set of 21 missense mutations at 6 different positions of *adk* designed to sample a broad range of molecular and cellular traits of the protein (Table S1 and Fig. 1). We selected residues such that their accessible surface area was less than 10% and they were at least 6 Å away from the catalytically-active sites of Adk, so that mutations at these residues were likely to destabilize the protein[13]. For most mutants, we chose amino acid mutations that appeared only at low frequency in an alignment of 895 homologous sequences of Adk. As intended, the purified mutant proteins were destabilized over a wide range (~17 °C in terms of T_m, and ~7.5 kcal/mol in terms of folding ΔG) (Table S1, Figs. 1B, S1, S2). In only one case (L209I) did we change the *E. coli* sequence to the consensus amino acid at that position, and we found it in fact stabilized the protein by ~1 kcal/mol (Table S1). Although most of the Adk mutants were less stable than the wild-type, they nevertheless existed predominantly as monomers in solution (Fig. S3). However, several mutations in one position — V106H, V106N, and V106W — did have significant fractions of proteins present in higher oligomeric forms, in addition to the predominant monomeric species (Fig. S3). These proteins bound 4,4'-Dianilino-1,1'-binaphthyl-5,5'-Disulfonate (Bis-ANS) dye to a higher degree compared to the rest of the mutants (Fig. S4), indicating the presence of possible molten globule states in solution[14]. The proteostat dye that reports on protein aggregation[4,15] also bound these mutants more strongly compared to others

(Fig. S4), clearly indicating a higher fraction of aggregated species. The catalytic efficiency (k_{cat}/K_M) of the mutant Adk proteins was distributed broadly with most mutants showing a lower activity than *E. coli* WT (Table S1, Figs. 1C, S5).

Intracellular abundance of Adk follows prediction from Boltzmann distribution

We then incorporated each of the 21 *adk* mutations one-by-one into the *E. coli* chromosome using a genome-editing approach based on homologous recombination[3,4]. We measured the *total* intracellular abundance of WT and mutant Adk proteins using a quantitative western blot (Table S2). The sigmoidal dependence of total intracellular Adk abundance on folding stability (ΔG) (Fig. 1D) is well-described by the Boltzmann distribution for two-state unfolding proteins:

$$P_F = \frac{1}{1+\exp(\beta \Delta G)} \quad (1)$$

where P_F is the fraction of folded molecules in the ensemble of intracellular Adk and $\beta = 1/k_B T$, with the Boltzmann constant k_B and the growth temperature T. The total measured abundance of a protein is its amount in the cytoplasm at steady-state, achieved by a balance between production and degradation. Since Adk is expressed from a constitutive promoter in the cells, it is generally safe to assume that the rates of production of all mutants are similar. Under this assumption, the sigmoidal dependence of abundance on stability clearly indicates that the unfolded protein is degraded in the active medium of the cytoplasm.

Mutations in Adk cause more variation in lag times than exponential growth rates

Mutations in Adk affect both intracellular abundance (via folding stability) and catalytic activity of the protein. Flux dynamics theory predicts, and experiments have confirmed, that the key enzymatic parameter determining the flux through an enzymatic reaction chain is the quantity which we call "catalytic capacity," defined as the product of intracellular abundance and enzymatic efficiency k_{cat}/K_M [5,6,16]. To that end, we determined how two key components of bacterial growth — the exponential growth rate and the lag time (Fig. 2A) — depend on the total catalytic capacity of Adk in *E. coli* cells (Fig. 2B,C; also see *Methods* and Fig. S6-S8 for estimation of growth parameters). We find that the variance in lag times across all strains is

significantly greater than the variance in exponential growth times (reciprocal growth rates) (Brown-Forsythe test, $p = 9\times10^{-4}$), and the mean change in lag time (relative to wild-type) from each mutation is significantly greater than the mean change in growth time (Welch's t-test, $p = 3\times10^{-7}$) (see *Methods* for details). This suggests that the mutations in Adk affect the lag phase more significantly than the exponential growth phase. One mechanism for producing longer apparent lag times is when a greater proportion of cells that come out of stationary phase are simply nonviable, as described in a recent study[17]. However, this appears not to be the major cause in our case, as lag times are fairly consistent across replicates (error bars in Fig. 2C), and do not negatively correlate with the number of viable cells (Fig. S9). We also find that the variation in total catalytic capacity of Adk correlates better with the variation in lag times (Spearman rank correlation $\rho = -0.44$, $p = 0.057$) than with the variation in growth rates (Spearman rank correlation $\rho = -0.08$, $p = 0.737$) (Fig. S10). The variation in lag times is also better explained by the variation in catalytic capacity than with any of the Adk properties separately (stability, abundance, or activity) (Fig. S10). Surprisingly, growth rate appears to tolerate a rather large drop in catalytic capacity of Adk, while lag time does not.

WT E. coli is positioned at the cusp of the biophysical fitness landscape for lag time

Since almost all the mutants have lower catalytic capacity than *E. coli* WT, they only provide sampling in the lower range of catalytic capacity. To determine the dependence of growth rate and lag time on catalytic capacity above WT levels, we over-expressed WT Adk from a pBAD plasmid (see Supplementary Methods). We observed no significant change in either growth rate or lag time at higher than endogenous catalytic capacity (Fig. 2B,C, Table S3). This means that while the growth rate appears to be insensitive to large changes in Adk catalytic capacity both below and above the wild-type level, WT catalytic capacity appears to be situated at the threshold of optimizing lag time. Next, we quantitatively compared the position of WT on these two fitness landscapes. To that end, we used a simple Michaelis-Menten-like function to fit the data in Fig 2B and C (see Eq. 3 and 4 and *Methods*). The fitting parameter 'c' which characterizes the onset of curvature on the landscape (analogous to K_M in Michaelis-Menten equation for enzymatic rate) reports proximity of WT to the cusp on the landscape (see *Methods*). It was 0.005 for growth rate, and 0.12 for lag time as compared to normalized catalytic capacity of 1 for WT. This clearly shows that WT is situated close to the cusp in terms of lag

time and well inside the plateau in terms of growth rate. This also suggests that selection for lag time, rather than growth rate, was the predominant determinant of WT Adk catalytic capacity.

Shorter lag imparts advantage at low carrying capacity: A computational model

This data highlights the pleiotropic effects of mutations on different phases of bacterial population growth, which raises the question of how pleiotropy shapes the evolutionary fate of a mutation. We explore this issue by considering the outcome of binary competition between strains[18]. We first simulated binary competitions over a wide range of growth rates and lag times in media conditions that allow for either 5-fold (low carrying capacity) or 500-fold (high carrying capacity) increase over the initial population (Fig. 3A) (See *Methods*). We found that there is a significant tradeoff between lag times and growth rates in determining the winners of binary competitions, with lag playing a more important role at low carrying capacity (Fig. 3A), implying that beneficial lag provides a greater fitness advantage under strongly nutrient-limiting conditions.

Shorter lag imparts advantage at low carrying capacity: Experimental evidence

To realize varying nutrient conditions in binary competition experiments, we explored the growth of *E. coli* over a range of glucose concentrations, mimicking the variation of carrying capacity in simulations, and found that only the carrying capacities are proportional to glucose concentration with minimal effects on lag time and growth rate (Fig. 4). This suggests that observing the outcome of the competition at different time snapshots in a nutrient-rich medium is equivalent to running the competition at different glucose concentrations (carrying capacities). To evaluate the predictions from simulations, we carried out two sets of binary competition experiments based on the overall distribution of growth rates and lag times (Fig. 3B). First, we selected strains exhibiting a tradeoff between growth rate (μ) and lag time (λ) ($\mu_1 > \mu_2$ and $\lambda_{,1} > \lambda_{,2}$) (inset of Fig. 5B). Second, we tested competition between strains that differ in their lag times but have nearly indistinguishable growth rates ($\mu_1 \approx \mu_2$ and $\lambda_1 > \lambda_2$) (inset of Fig. 5C). In all cases a strain with shorter lag time is expected to dominate at lower carrying capacity conditions (corresponding to the competition outcome at early time points), however this advantage would be lost at later time points if its growth rate is lower than that of the competing strain (Fig. 5A). In the second scenario, the advantage due to short lag is expected to persist even

at high carrying capacity conditions because the growth rates of the competing strains do not differ. We estimated the relative proportions of the two strains by a qPCR-based mismatch amplification mutation assay (MAMA) approach[19] (see *Methods* and Fig. S11). As expected in the first scenario, L083F and V106H dominated at earlier time points when competed against A093I and L209I, respectively, due to their shorter lag times ($\lambda_{L083F} < \lambda_{A093I}$ and $\lambda_{V106H} < \lambda_{L209I}$) (Fig. 5B). Eventually their fraction dropped below 0.5 at later time points (equivalent to high carrying capacity) where the growth rates determine the competition output ($\mu_{L083F} < \mu_{A093I}$ and $\mu_{V106H} < \mu_{L209I}$) (Fig. 5B). Similarly, for the second scenario, despite having similar growth rates ($\mu_{WT} \approx \mu_{Y182V} \approx \mu_{L209A}$), the fraction of WT was always maintained above 0.5 as it spends a shorter time in the lag phase compared to Y182V and L209A (Fig. 5C). The early advantage to WT due to its shorter lag phase determined the competition fitness throughout the whole growth cycle.

Discussion

A complete mapping of mutational fitness effects would ideally require sampling a practically infinite number of mutations, an impossible proposition. Instead, we can project fitness onto a fairly small number of molecular properties of proteins. Within this paradigm, the identity of a particular mutation does not matter as much as its effect on essential biochemical and biophysical properties of the proteins in question. Our data, as well as previous studies[5-7,20], validate this approach by showing that we can collapse several molecular phenotypes into a single effective parameter – the product of protein abundance and activity k_{cat} / K_M (catalytic capacity) – which quantitatively determines the biophysical fitness landscape to a great extent (Fig. 2B,C). That is, Fig. 2 indicates that the fitness effects of mutations can largely be predicted from their biophysical effects over a broad range of catalytic capacity, validating the utility of a biophysical fitness landscape to map variation in the *adk* locus to the phenotype. The 21 engineered mutations, along with the Adk overexpression data, allow us to outline the biophysical fitness landscape comprehensively, covering a wide range of variation of the physical parameters of Adenylate Kinase.

These results illustrate how the evolutionary endpoint of molecular traits may depend fundamentally on the multidimensional nature of fitness, with the relative importance of different

components of fitness depending on the environment and lifestyle of the organism. It has been argued that endogenous molecular traits are established as a result of mutation-selection balance[21], with the final outcome depending on the relative strengths of selection and genetic drift as determined by the population structure[22,23]. Here we encounter a more complex situation where mutations in the essential enzyme Adk change multiple traits with different effects on fitness. Apparently the mutation-selection balance resulted in disparate outcomes for the two traits, placing lag time at the cusp while keeping the exponential growth rate farther within the plateau region of its respective biophysical fitness landscape. Such an outcome can reflect different strengths of selection and drift as applied to different phenotypic traits. It is therefore possible that ecological conditions of *E. coli* put stronger emphasis on survival in low carrying-capacity or fluctuating environments, leading to the balance of selection and drift that keeps lag phase just on the cusp in Fig. 2, i.e., optimal with respect to point mutations that decrease catalytic capacity. Our studies of binary competitions (Figs. 3 and 5) highlight this scenario by showing how the environmental parameter of carrying capacity can determine winners and losers in evolutionary dynamics. Although the lag time of a population can depend not only on the environment but also on the population's specific history (e.g., how long it was previously in stationary phase), the fundamental role of Adk in metabolism suggests that its effects on lag time are likely to be common across conditions and histories. The deep connection between ecological history of species and optimization of biophysical traits of their proteins is a subject for interesting future studies.

Much of our current understanding of microbial cultures and fitness comes from experiments done in the laboratory, where strains are typically grown under a large supply of nutrients. The situation might be very different in the wild, however, where bacteria and other microbes have to survive under harsh conditions of nutrient starvation, extreme temperature, and other environmental stresses[24-26]. In these circumstances, organisms are likely to spend only a minute fraction of their life-cycle in the exponential growth phase, while undergoing many cycles of lag-growth-saturation as new resources become available and old ones are exhausted. It is therefore intuitive to expect that there has been strong selection in favor of organisms that can not only divide rapidly during exponential growth, but that can also wake up quickly from their lag phase and respond to newly available resources. Our study demonstrates how this selection may shape individual molecular traits.

Our study highlights the relationship between various components of fitness and the molecular properties of modern enzymes — the endpoint of evolutionary selection. An interesting question which is beyond the scope of current work is how modern variants emerged in evolutionary dynamics. To that end mapping temporary reconstructed ancestral species onto biophysical fitness landscape of Adk (and other enzymes) appears a promising direction of future research.

Methods

Selection of mutations: Mutations at relatively-buried positions generally result in decreased stability and lower fitness[13,27]. Hence we selected the sites for mutagenesis with side-chain accessibility of less than 10%. In addition, the selected sites were also away from the active-site residues, or active-site contacting residues, and a minimum of 6 Å away from the inhibitor Ap5A binding sites (pdb 1ake). We define the active-site residues as those whose accessible surface area changes by at least 5 $Å^2$ in the presence of the inhibitor Ap5A. A similar criterion was used to define the residues contacting the active site. Altogether 4 residues from the LID domain, 3 from the NMP domain, and 28 from the Core domain satisfy these criteria. Of the 28 sites from the Core domain, we randomly chose 6 to mutate. We chose the identities of the mutations to span various sizes of the side chains and a range of conservation. We derived the conservation from the multiple sequence alignment of 895 sequences for Adk collated from ExPASy database (as of Nov 2012).

Generation of mutant strains: We generated the strains with WT and mutant *adk* with chloramphenicol- and kanamycin-resistance genes on either end of the *adk* gene using the genome editing approach as described previously[3]. Since the *adk* gene is flanked by two repeat regions (REPt44 and REPt45) on the wild-type chromosome, we extended the homology required for recombination up to the middle of the adjacent genes.

Growth curve measurements and media conditions: WT and mutant strains were grown overnight at 30 °C from single colonies in a supplemented M9 medium (0.2 % glucose, 1 mM $MgSO_4$, 0.1 % casamino acids, and 0.5 μg/ml Thiamine). OD600 was measured for all the strains and then the cultures were normalized to whichever had the lowest OD. The normalized cultures were diluted 1:100 in fresh supplemented M9 media and the growth curves were

monitored in triplicates using Bioscreen C at 37 °C. We derived the growth parameters by fitting ln(OD) versus time with the four-parameter Gompertz function (see below). The error in replicates was found to be between 2-3% on an average, and it did not improve significantly upon increase in number of replicates.

Fitting growth data and estimation of growth parameters: In our study, we define lag time (λ) as the time required to achieve the maximum growth rate (μ) (Fig. 2A). We used two different methods to infer these parameters: A) direct analysis of growth curve derivatives and B) fits to the Gompertz function (Fig. S6).

In method A, we took the growth rate as the maximum value of $\frac{\ln\left(OD(t)/OD(t-\Delta t)\right)}{\Delta t}$ where Δt is 15 minutes. The lag time was then the earliest time at which this maximum growth rate was achieved.

For method B we used the following four-parameter Gompertz function to fit ln(OD) vs. time plots (considering only points with OD600 >= 0.02):

$$\ln(OD) = \ln(OD_0) + K\exp\left[-\exp\left(-\frac{t-\lambda}{b}\right)\right] \quad (1)$$

where the carrying capacity is K, the maximum growth rate is $\mu = K/\left(b\cdot\exp(1)\right)$, and the lag time λ is the time taken to achieve the maximum growth rate.

The μ and λ estimated from the above methods are strongly correlated (Pearson's r=0.80, p=1.4e-5 for μ, and r=0.71, p=3.0e-4 for λ) (Fig. S7). However, the uncertainty in the fitted parameters appears to be less than the uncertainty in the parameters obtained from the derivatives, which are limited by the low time-resolution of the experimental data (acquired at an interval of 15 min).

The growth rate (μ) and lag time (λ) appear to be statistically independent of each other across the Adk mutant strains (Spearman $\rho = 0.31$, $p = 0.15$, Fig. 3B). Hence it is conceivable that selection can act separately on these two traits, which is further illustrated by the different fitness landscapes observed when projected onto the axis of catalytic capacity (Fig. 2B,C).

Statistical tests of mutational variation in growth and lag phases: We compare the relative effects of mutations on growth and lag phases in two ways. First, we calculate the variances in exponential growth time (reciprocal $1/\mu$ of growth rate, proportional to the maximum cell division time) and lag time; we use growth time rather than growth rate since it must have the same units (i.e., minutes) as lag time for comparison. These variances tell us how much each strain's growth or lag time differs from the average across all strains. We then use the Brown-Forsythe test (since the growth and lag times are not normally distributed) to determine whether these variances are significantly different. Second, we calculate the mean absolute deviation of each mutant's growth and lag times relative to the wild-type values. This tells us the average change in growth or lag time caused by a mutation.

The strength of selection on the growth or lag phase should be proportional to the difference in growth or lag time between the two competing strains (Manhart, Adkar, Shakhnovich, *unpublished results*). Therefore the variances in traits are proportional to the average selection coefficient between all pairs of strains for that trait, while the mean absolute deviation relative to wild-type is proportional to the average selection coefficient between each mutant and the wild-type. The statistical tests above, which determine whether the variances and mean absolute deviation are significantly different between growth time and lag time, also indicate which trait is under stronger average selection between the strains.

Quantification of the location of WT on the fitness landscapes: As in previous works from our lab [5,6] as well as earlier work [16,20] we used the following Michaelis-Menten-like elasticity curve functions to fit the landscape of growth rate vs catalytic capacity (Fig. 2B):

$$\text{Relative Growth Rate} = \frac{a \cdot \text{Catalytic Capacity}}{c + \text{Catalytic Capacity}} \tag{2}$$

where, a is the saturation parameter, and c is the catalytic capacity at $a/2$. For a similar landscape with lag time (Fig. 2C), the reciprocal of Eq. 3 was used in the following form:

$$\text{Relative lag time} = \frac{b \cdot \left(c + \text{Catalytic capacity}\right)}{\text{Catalytic Capacity}} \tag{3}$$

where, b is the asymptote parameter, and c is the catalytic capacity at $2b$. In both the equations 3 and 4, c is a characteristic value of catalytic capacity at which the landscape transitions from

the plateau to the curved part. Since catalytic capacity is normalized by WT c serves as a measure of how close to the cusp the WT on the respective landscapes.

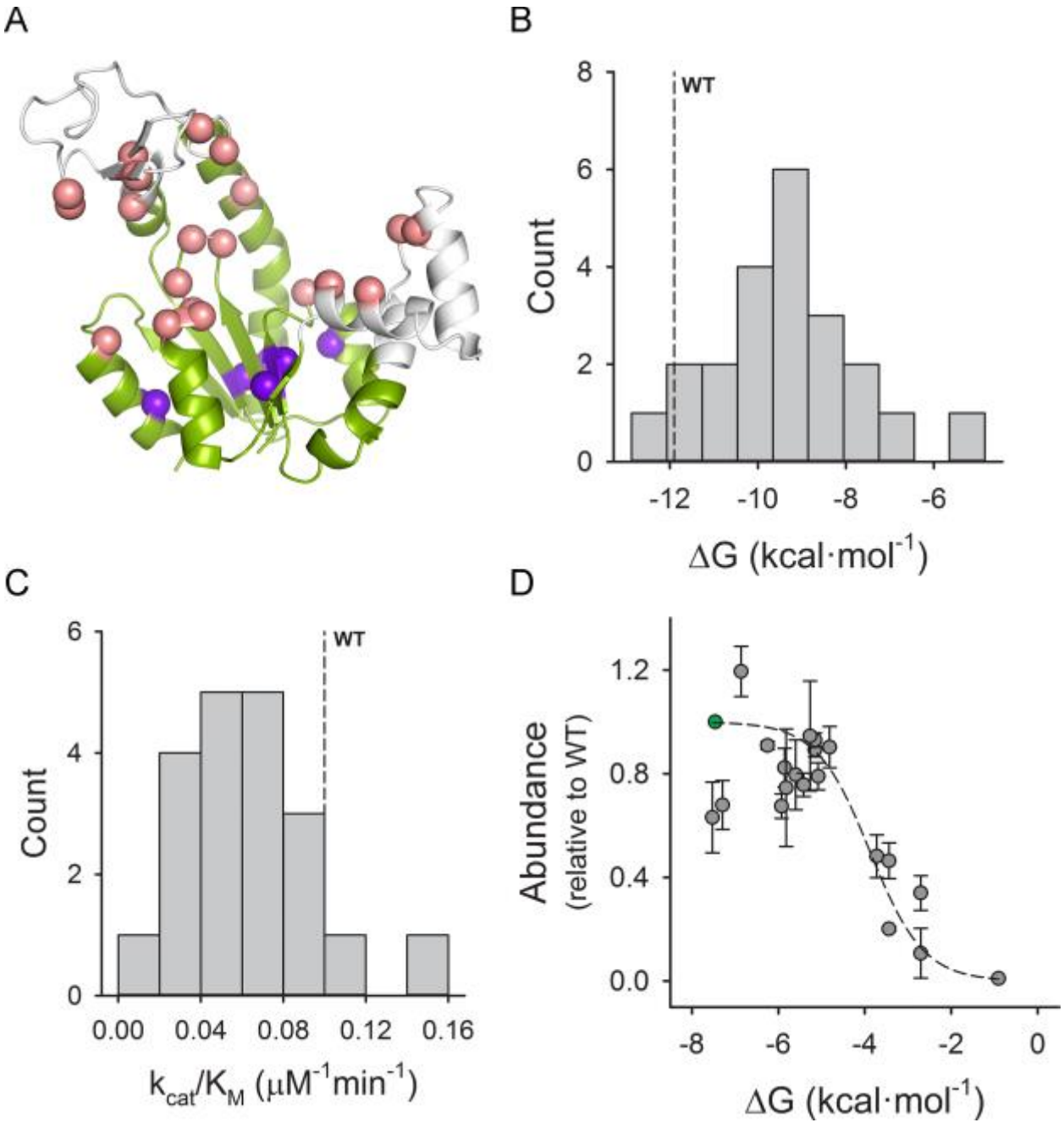
Simulation of binary competition: We simulated the competition of two strains by using the Gompertz function (Eq. 2) to model the growth of individual strains. The initial population (OD0) for both strains was equal, and growth ceases when $\Sigma\left(OD_t/OD_0\right)_i = K$, where K is the carrying capacity. We considered two different values of carrying capacities (5 and 500). We set μ_1 and λ_1 to values derived experimentally for WT Adk strain (Table S2), while the growth rates and lag times for the second competing strain were varied randomly across the intervals 0.005 to 0.030 min^{-1} (for growth rate) and 50 to 300 min (for lag time).

Binary growth competition and quantification: The overnight cultures for individual strains were grown for 16 hours at 30 °C. These cultures were mixed in 1:1 proportion, diluted to an OD of 0.01 in fresh supplemented M9 media, and then regrown at 37 °C. The samples were drawn at different time points, and the OD was adjusted to 2.0, either by concentration or dilution. 5ul of OD 2.0 culture was eventually diluted in 45ul of lysis solution (QuickExtract DNA extraction solution (Epicentre)) to reach OD 0.2. Genomic DNA extracted from 50ul of OD 0.2 culture was diluted 5000 times and used as template. The individual strains in the competition were differentially amplified using allele-specific primers and quantified by a qPCR-based mismatch amplification mutation assay method [19] using QuantiTect SYBR Green PCR kit (Qiagen). A 150 bp long non-mutagenic amplicon of *adk* gene was amplified as a reference to quantify total genomic DNA. The fraction of the competing strains was determined using the following equation:

$$fraction = 2^{\left(c_{t,ref} - c_{t,1}\right)_{competition} - \left(c_{t,ref} - c_{t,1}\right)_{pure}} \quad (4)$$

where C_t represents threshold cycle of qPCR, ref and 1 are the PCR reactions for amplifying the reference and the first allele in competition, while "competition" and "pure" represent the condition of culture.

335 **Figures:**



336

337 Fig 1: **Biophysical and intracellular properties.** (A) Crystal structure of Adenylate Kinase
338 from *E. coli* (PDB ID 4ake[28]). The core domain is colored in green, while the LID and NMP
339 domains are shown in white. The C_α atoms of active-site residues are shown in pink, and the blue
340 spheres represent the C_α atoms of the 6 buried positions which were mutated in this study. (B)
341 Histogram showing the distribution of folding free energies for all mutant proteins, as

determined by isothermal urea denaturation at 25 °C. The stability of WT is marked by a dashed line. (C) Histogram of the catalytic activity parameter k_{cat}/K_M for all mutants. The dashed line indicates the WT value. (D) Total intracellular abundance of mutant Adk proteins as a function of ΔG at 37 °C. The abundances are normalized by the WT value. Each data point represents the mean and error bars are standard deviation over two experiments. The dashed line represents the fit to the Boltzmann distribution function described in Eq. 1. See related Figs. S1-S5 and Table S1.

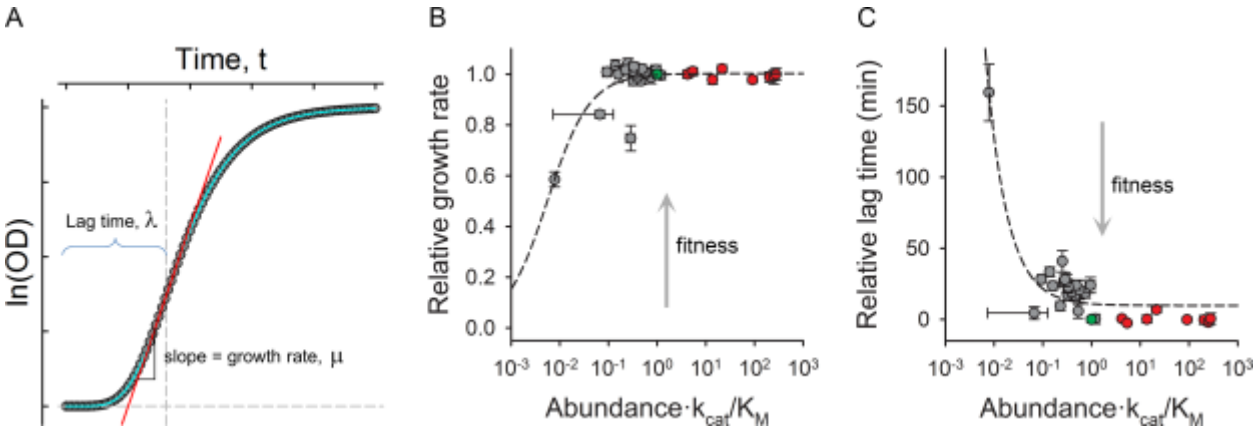


Fig 2: **Traits of population growth.** (A) Schematic of estimation of lag time and growth rate. The representative data points (solid gray circles) were plotted as ln(OD) vs time and was fitted to a four parameter Gompertz function (Eq. 2) (cyan line). The red line is a tangent at the inflection point of the function. The slope of the tangent is considered as the growth rate (μ) and the time required to reach the maximum growth rate or the inflection point is taken as the lag time (λ) (vertical dashed line). (B) Growth rate and (C) lag time as functions of catalytic capacity which is defined as abundance $\times k_{cat}/K_M$. The mutant data is shown in gray circles, whereas red circles represent the BW27783 strain with varying degrees of overexpression of WT Adk from a pBAD plasmid. Data for WT is shown in green. The data points represent mean and error bars represent standard deviation of parameters derived from growth curves of 3 colonies (biological replicates) in triplicates (9 curves). See related Figs. S6-S10 and Tables S2-S3. The dashed black line in (B) and (C) represents a fit to a Michaelis-Menten-like function (see *Methods* for details). The solid gray arrow indicates the direction of increasing fitness.

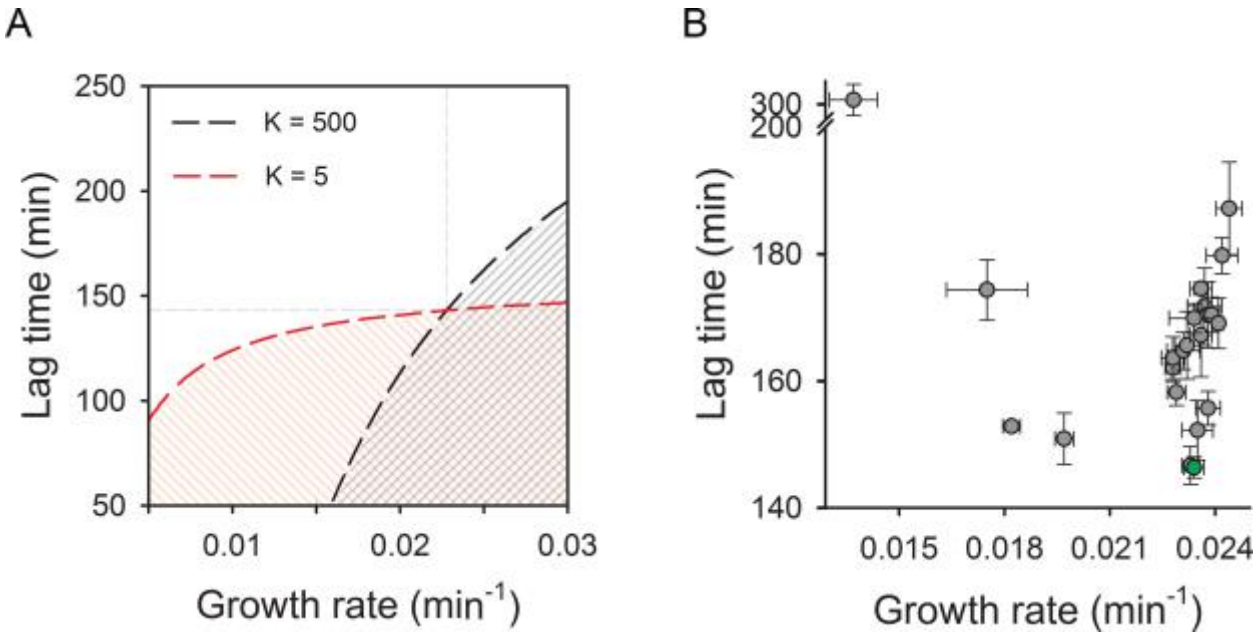


Fig 3: **Binary growth competition**. The growth of individual strains was modeled as per Gompertz equation (Eq. 2). The growth parameters for strain 1 were fixed to those obtained for WT Adk (dashed gray lines) while those for strain 2 were generated randomly over a wide range of growth rates (0.005 to 0.030 min^{-1}) and lag times (50 to 250 min). (A) Contour plot showing fraction of strain 1 (WT) at saturation when the competition is carried out under two different carrying capacities (red line indicates K=5 while the black line indicates K=500). The dashed lines indicate neutrality region where both strains have equal proportions at saturation. The areas below the neutrality line (filled with solid lines) represent the parameter space where strain 2 wins the competition (fraction of strain 2 > 0.5). (B) Scatter plot of growth rate (μ) versus lag time (λ). The data points represent the mean and error bars the standard deviation of 6 to 9 measurements (see Table S2). The growth rate and lag time appear to be statistically independent of each other across the Adk mutant strains (Spearman $\rho = 0.31$, $p = 0.15$).

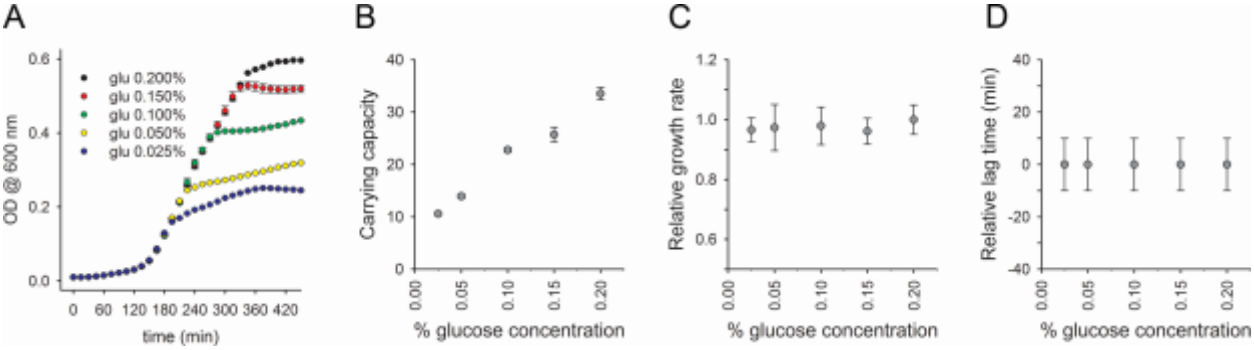


Fig 4: **Growth curves at various nutrient concentration.** (A) Growth curves of strains with WT Adk obtained under varying glucose concentrations in supplemented M9 medium. The fitted growth curve parameters are shown as functions of glucose concentration: (B) carrying capacity of ln(OD) as derived from Gompertz fitting, (C) growth rate, and (D) lag time. The growth rates and lag times are estimated from analysis of growth curve derivatives and are normalized relative to the respective values at 0.2% glucose concentration.

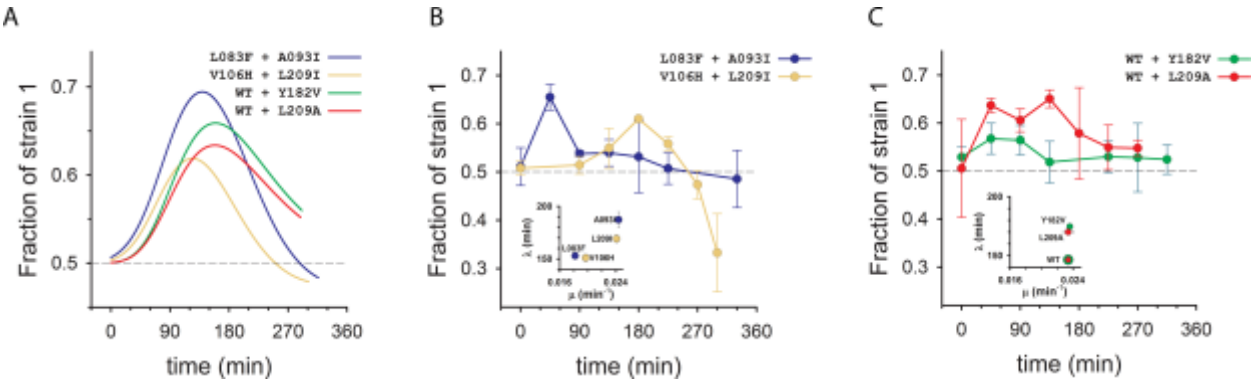


Fig 5: **Tradeoffs between lag and exponential growth in binary competitions.** (A) Fraction of the first strain as a function of time in simulated binary competitions. We modeled growth of each strain using the Gompertz 4-parameter equation (Eq. 2) with experimentally measured growth rate and lag time values. The initial OD for individual strains was assumed to be 0.006 at the start of competition, and growth was assumed to saturate at OD of 0.6. Despite having similar growth rates, the fraction of WT in WT + L209I and WT + Y182V competitions was always above 0.5 owing to the advantage it gained due to shorter lag time (scenario 2 in the text). L083F and V106H dominate at earlier time points (equivalent to low carrying capacities) due to their short lag times compared to their respective competitors. However, at longer times (high carrying capacities) the advantage due to lag is lost due to their lower growth rates. (B, C)

Experimental validations of the predictions in (A) using qPCR based mismatch amplification mutation assay (MAMA). The fraction of competing strains was estimated using Eq. 5. The data points are mean and error bars represent standard deviation of two measurements. See related Fig S11. The growth rates and lag times for the competing pairs are shown in insets.

References

1 Sanjuan, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A* **101**, 8396-8401, doi:10.1073/pnas.0400146101

0400146101 [pii] (2004).

2 Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D. S. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**, 929-932, doi:10.1038/nature05385 (2006).

3 Bershtein, S., Mu, W. & Shakhnovich, E. I. Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc Natl Acad Sci U S A* **109**, 4857-4862, doi:1118157109 [pii]

10.1073/pnas.1118157109 (2012).

4 Bershtein, S., Mu, W., Serohijos, A. W., Zhou, J. & Shakhnovich, E. I. Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. *Molecular cell* **49**, 133-144, doi:10.1016/j.molcel.2012.11.004 (2013).

5 Bershtein, S. *et al.* Protein Homeostasis Imposes a Barrier on Functional Integration of Horizontally Transferred Genes in Bacteria. *PLoS genetics* **11**, e1005612, doi:10.1371/journal.pgen.1005612 (2015).

6 Rodrigues, J. V. *et al.* Biophysical principles predict fitness landscapes of drug resistance. *Proc Natl Acad Sci U S A* **113**, E1470-1478, doi:10.1073/pnas.1601441113 (2016).

7 Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631, doi:10.7554/eLife.00631 (2013).

8 Orr, H. A. Fitness and its role in evolutionary genetics. *Nat Rev Genet* **10**, 531-539, doi:10.1038/nrg2603 (2009).

9 Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature reviews. Genetics* **4**, 457-469, doi:10.1038/nrg1088 (2003).

10 Dykhuizen, D. E. & Hartl, D. L. Selection in chemostats. *Microbiological reviews* **47**, 150-168 (1983).

11 Biel, S. W. & Hartl, D. L. Evolution of transposons: natural selection for Tn5 in Escherichia coli K12. *Genetics* **103**, 581-592 (1983).

12 Fridman, O., Goldberg, A., Ronin, I., Shoresh, N. & Balaban, N. Q. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature* **513**, 418-421, doi:10.1038/nature13469 (2014).

13 Bajaj, K., Chakrabarti, P. & Varadarajan, R. Mutagenesis-based definitions and probes of residue burial in proteins. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16221-16226, doi:10.1073/pnas.0505089102 (2005).

14 Semisotnov, G. V. *et al.* Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31**, 119-128, doi:10.1002/bip.360310111 (1991).

15 Navarro, S. & Ventura, S. Fluorescent dye ProteoStat to detect and discriminate intracellular amyloid-like aggregates in Escherichia coli. *Biotechnology journal* **9**, 1259-1266, doi:10.1002/biot.201400291 (2014).

16 Dykhuizen, D. E., Dean, A. M. & Hartl, D. L. Metabolic flux and fitness. *Genetics* **115**, 25-31 (1987).

17 Peters, J. M. *et al.* A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165**, 1493-1506, doi:10.1016/j.cell.2016.05.003 (2016).

18 Hegreness, M., Shoresh, N., Hartl, D. & Kishony, R. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science* **311**, 1615-1617 (2006).

19 Cha, R. S., Zarbl, H., Keohavong, P. & Thilly, W. G. Mismatch amplification mutation assay (MAMA): application to the c-H-ras gene. *PCR methods and applications* **2**, 14-20 (1992).

20 Jiang, L., Mishra, P., Hietpas, R. T., Zeldovich, K. B. & Bolon, D. N. Latent effects of Hsp90 mutants revealed at reduced expression levels. *PLoS genetics* **9**, e1003600, doi:10.1371/journal.pgen.1003600 (2013).

21 Serohijos, A. W. & Shakhnovich, E. I. Merging molecular mechanism and evolution: theory and computation at the interface of biophysics and evolutionary population genetics. *Curr Opin Struct Biol* **26**, 84-91, doi:10.1016/j.sbi.2014.05.005 (2014).

22 Serohijos, A. W., Lee, S. Y. & Shakhnovich, E. I. Highly abundant proteins favor more stable 3D structures in yeast. *Biophys J* **104**, L1-3, doi:10.1016/j.bpj.2012.11.3838 (2013).

23 Hartl, D. L., Clark, A.G. *Principles of population genetics, 4th ed.*, (Sinauer, 2007).

24 Lambert, G., Liao, D., Vyawahare, S. & Austin, R. H. Anomalous spatial redistribution of competing bacteria under starvation conditions. *J Bacteriol* **193**, 1878-1883, doi:10.1128/JB.01430-10 (2011).

25 Palkova, Z. Multicellular microorganisms: laboratory versus nature. *EMBO reports* **5**, 470-476, doi:10.1038/sj.embor.7400145 (2004).

26 Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* **8**, 15-25, doi:10.1038/nrmicro2259 (2010).

27 Adkar, B. V. *et al.* Protein model discrimination using mutational sensitivity derived from deep sequencing. *Structure* **20**, 371-381, doi:10.1016/j.str.2011.11.021 (2012).

28 Muller, C. W., Schlauderer, G. J., Reinstein, J. & Schulz, G. E. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* **4**, 147-156 (1996).

Acknowledgements:

We thank Shimon Bershtein and Adrian Serohijos for helpful discussions.

Contributions:

BVA and EIS - designed research; BVA, SB, JT and MMu - performed experiments; BVA, MMa, SB and EIS - analyzed the data; BVA, MMa, SB and EIS - wrote the paper; All authors edited and approved the final version.

Supplementary Methods

Mutagenesis and protein purification

Adenylate kinase (Adk) is encoded by the *adk* gene, which was cloned under the T7-*lac* promoter in pET28a(+) vector (Invitrogen) between *Nde*I and *Xho*I restriction sites. We carried out mutagenesis with a pair of 30-35 bp long, partially-complementary primers and the inverse PCR technique using KOD hot-start DNA polymerase. The mutations were centered in the complementary regions of the primers. The mutagenic plasmids were transformed in *E. coli* DH5α cells for faithful propagation and storage, and in *E. coli* BL21(DE3) for protein overexpression and purification. The His-tagged proteins were purified by Ni-NTA affinity chromatography (Qiagen) and subsequently passed through a HiLoad Superdex 75pg column (GE). The monomeric peak was collected, concentrated and eventually stored in 10mM potassium phosphate buffer (pH 7.2). The concentration of the proteins was measured by BCA assay (ThermoScientific) with BSA as standard.

Biophysical characterization

Thermal denaturation: We assessed the thermal stability of WT and mutant proteins by differential scanning calorimetry (nanoDSC, TA instruments) using 20 μM of protein. The scans were carried out from 10 to 90 °C at a scan rate of 90 °C/hr. The thermodynamic parameters were derived by fitting the data to a two-state unfolding model using NanoAnalyze (TA instruments). We also carried out thermal denaturation using the melt-curve module of BioRad CFX96, with Sypro Orange dye as a probe for unfolding as described earlier[1]. The dye was added to the final concentration of 5× in a 25 μl reaction volume containing 4 μM of protein in 10 mM potassium phosphate buffer (pH 7.2). The data were fit to a standard four-parameter sigmoidal equation to obtain apparent melting temperatures.

Urea denaturation: We carried out isothermal urea denaturation with WT and mutant proteins to assess the stability of the proteins to chemical denaturants. We incubated 5 μM of protein for ~4 hrs at 25 °C with varying concentrations of urea (0-8 M). The urea concentrations were estimated by refractive index measurements. The denaturation was monitored by measuring the ellipticity at 222 nm using a CD spectrometer (Jasco). The melt data was fitted assuming a model

of two-state unfolding with linear free energy as described earlier[2,3]. The m-value was fixed to 3300 cal/mol/M for fitting.

Gel filtration: We assessed the oligomeric status of purified proteins by gel filtration using 50 µg of protein on sephadex 75 analytical columns.

ANS and proteostat binding: We used 12 µM of bisANS for assessing binding to 2 µM of protein in 10 mM potassium phosphate buffer (pH 7.2). The excitation and emission wavelengths were set to 395 nm and 490 nm, respectively. 2 µM of protein was incubated with 3.5 mM of the proteostat dye in 1X assay buffer (Enzo LifeSciences). For this the excitation and emission wavelengths were set to 550 and 600 nm, respectively.

Enzyme activity: We measured the activity of Adk in terms of ADP formation by an end-point assay as described earlier[4]. Briefly, the concentration of AMP was fixed to 500 µM and ATP concentration was varied from 0 to 500 µM in an enzymatic reaction. 5 nM of Adk was used to initiate the reaction and 500 µM of Ap5A was used for quenching at 20, 40, and 60 second time points. The amount of ADP formed was measured by LDH-Pyruvate kinase-coupled reaction and the kinetic parameters were derived by fitting the data to the Michaelis-Menten equation.

Adk overexpression: The *adk* gene was cloned in a pBAD plasmid and transformed in the *E. coli* BW27783 strain (CGSC#12119). This strain constitutively expresses the arabinose transporter (araE) which enables uniform uptake of arabinose. The cells were induced with increasing concentrations of arabinose from 0 to 0.05%.

Intracellular protein abundance: Cells were grown in supplemented M9 medium for 4 hours at 37°C, harvested and subsequently lysed with 1× BugBuster (Novagen) and 25 units/ml of Benzonase. Total amount of proteins in cell lysate was estimated by BCA assay. The specific fraction of Adk was determined by SDS-PAGE followed by western blot using rabbit anti-Adk polyclonal antibodies (custom- raised by Pacific Immunology).

Estimation of viable cells in saturating culture: The overnight culture was grown in supplemented M9 medium for 16 hours at 30 °C and the proportion of live:dead cells was measured using Live/Dead BacLight Bacterial Viability Kits (Molecular Probes) according to the manufacturer's instructions. Briefly, 1×10^8 cells (in a volume of 1ml) were mixed with 3ul of a

1:1 proportion of Syto9 dye and Propidium Iodide (PI). The mixture was incubated in the dark for 15 minutes, following which the fluorescence was measured at 530nm and 630nm. Syto9 dye stains live cells and emits fluorescence at 530nm (green), while PI stains dead cells and can be detected at 630nm (red). The ratio of fluorescence values at 530nm:630nm corresponds to the proportion of live:dead cells in that sample which was eventually used to estimate the percentage of live cells in a sample, according to the manufacturer's instructions. An exponentially growing culture (considered as 100% live) and cells treated with 70% ethanol for 1 hour (considered 100% dead) were mixed in different known proportions, and their 530:630nm ratio was used to generate a standard curve.

1 Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212-2221, doi:10.1038/nprot.2007.321 (2007).

2 Chen, B. L. & Schellman, J. A. Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies. *Biochemistry* **28**, 685-691, doi:10.1021/bi00428a041 (1989).

3 Schellman, J. A. The thermodynamic stability of proteins. *Annu Rev Biophys Biophys Chem* **16**, 115-137, doi:10.1146/annurev.bb.16.060187.000555 (1987).

4 Pena, M. I., Davlieva, M., Bennett, M. R., Olson, J. S. & Shamoo, Y. Evolutionary fates within a microbial population highlight an essential role for protein folding during natural selection. *Mol Syst Biol* **6**, 387, doi:10.1038/msb.2010.43 (2010).

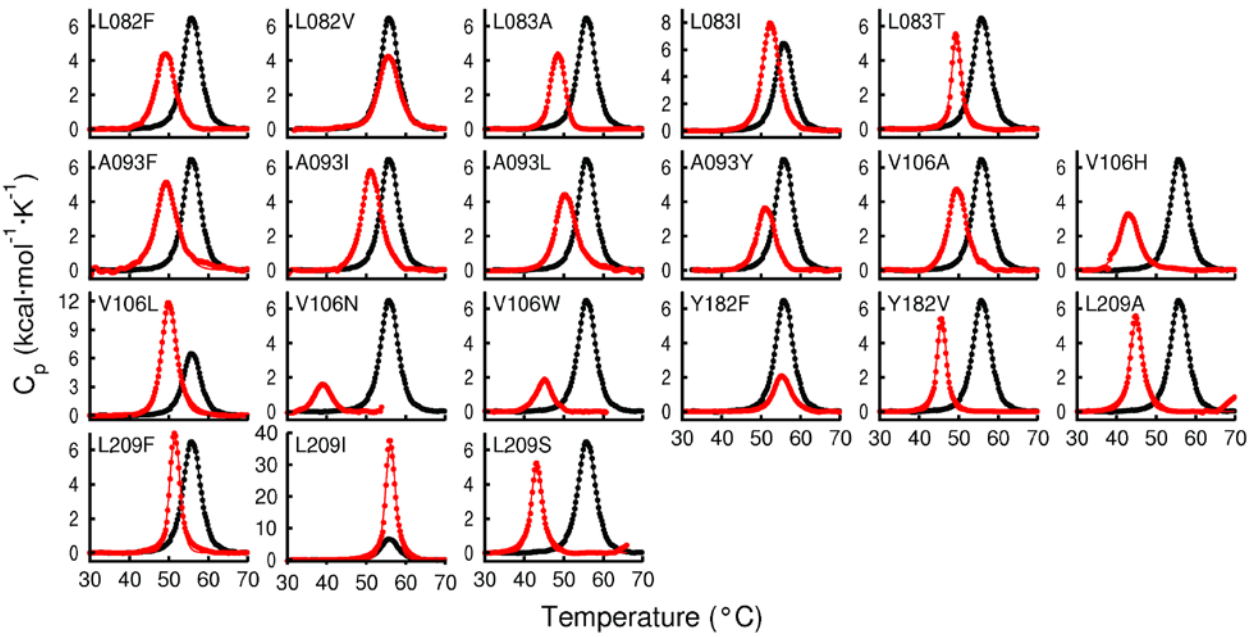


Fig S1: Thermal unfolding monitored by Differential Scanning Calorimetry (DSC) for WT (black trace) and 20 different Adk mutant proteins (red trace). The molar heat capacity (C_p) is shown as a function of temperature. The scan rate was 90 °C/hr. The data was fitted to a two-state thermal unfolding model to derive the thermodynamic parameters.

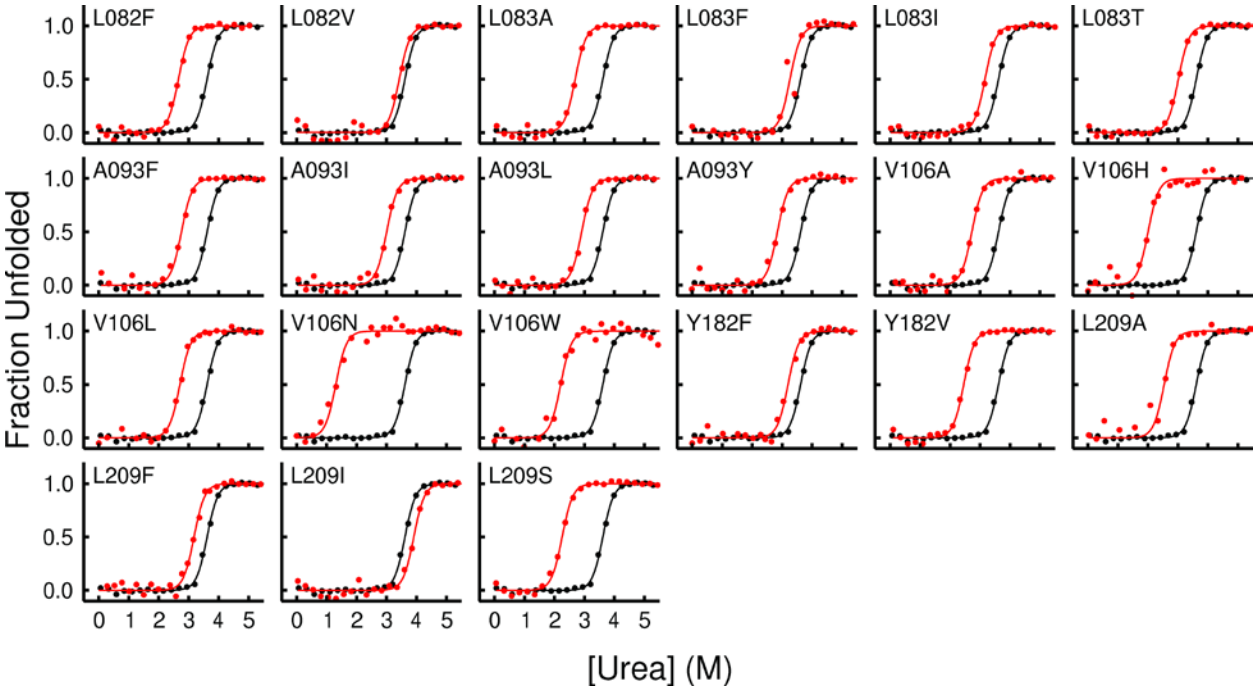


Fig S2: Isothermal urea denaturation curves at 25 °C for WT (black dots) and mutant Adk proteins (red dots). The fraction unfolded (Fu) is plotted as a function of denaturant concentration. Protein denaturation was monitored by recording the CD signal at 222 nm. The data was fit to a two-state unfolding model.

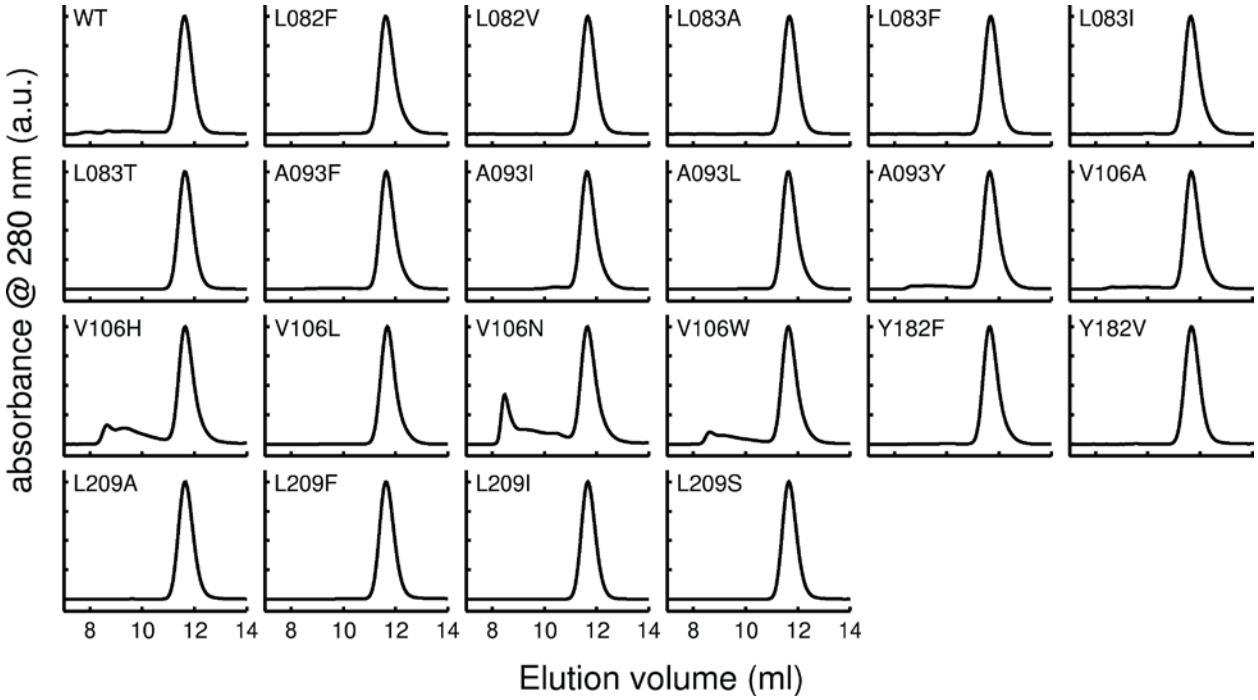


Fig S3: Analytical gel-filtration profile of WT and 20 mutant Adk proteins on a Superdex-75 column at room-temperature. The absorbance at 280 nm is shown as a function of elution volume. For comparison all the monomeric peaks were normalized to 1. WT Adk along with most other mutant proteins elutes at the expected position for a monomer. Exceptions were V106H, V106N and V106W, where additional peaks appear at much higher molecular weights.

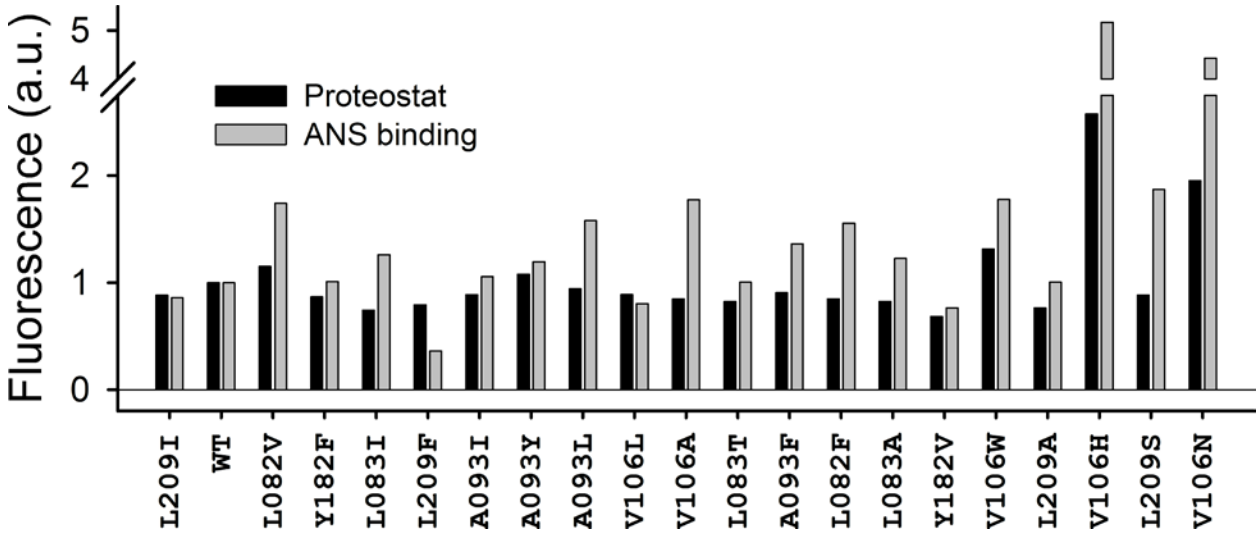


Fig S4: Aggregation propensity and molten-globule states of mutant proteins. Bar plots represent the extent of ProteoStat and ANS binding to WT and mutant Adk proteins. The proteins on the x-axis are arranged in decreasing order of stability from left to right.

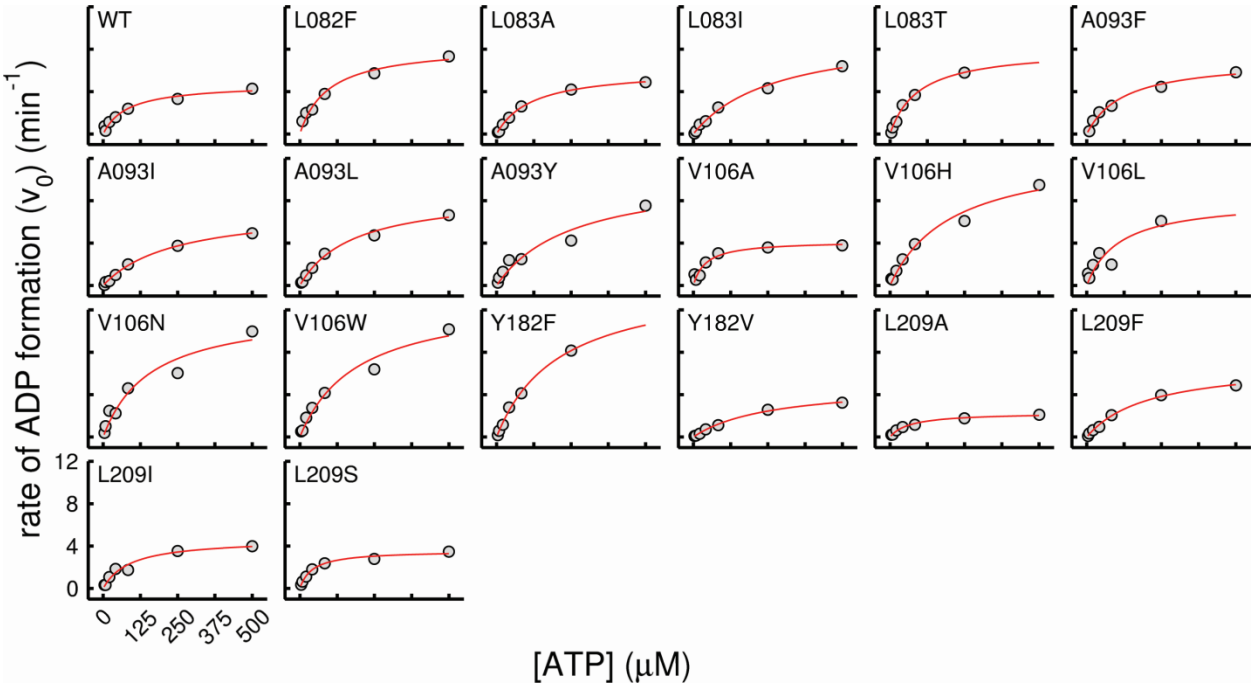


Fig S5: Enzyme activity of Adk mutants at 25 °C measured as described in SI Text. The initial velocity, shown as a function of ATP concentration, was calculated as the amount of ADP produced per minute by 1 nmol of Adenylate Kinase. The concentration of AMP in all experiments was fixed to 500 μM. The data (gray circles) was fitted using the Michaelis-Menten equation of enzyme activity to extract relevant parameters (fitted line in red).

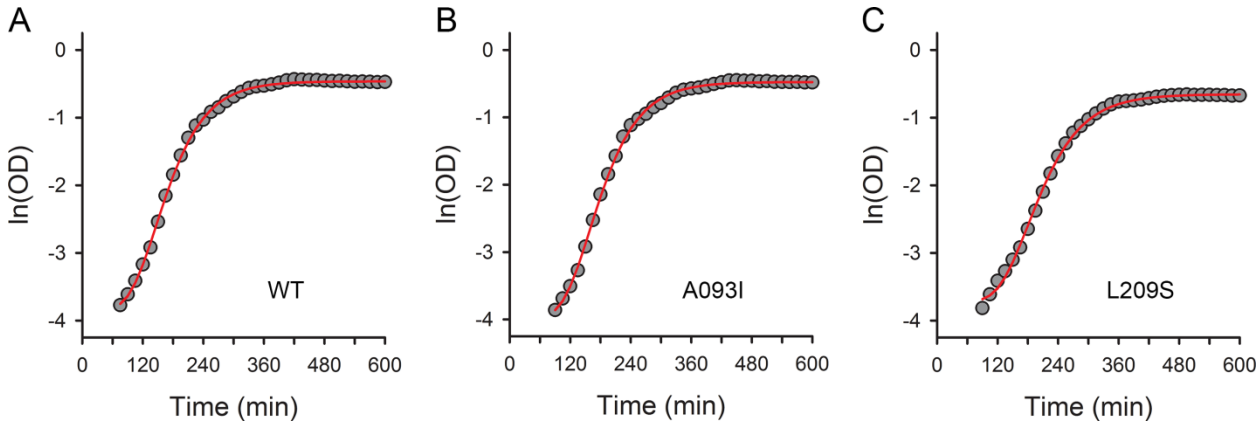


Fig S6: Representative growth curves of (A) WT, (B) A093I, and (C) L209S strains. Each growth curve is shown as ln(OD) vs time plot. The experimental data is shown in gray circles and the Gompertz fit is shown in solid red line. The strains were chosen to illustrate the quality of the fit across different range of growth rates and lag times (see Table S2 for growth parameters).

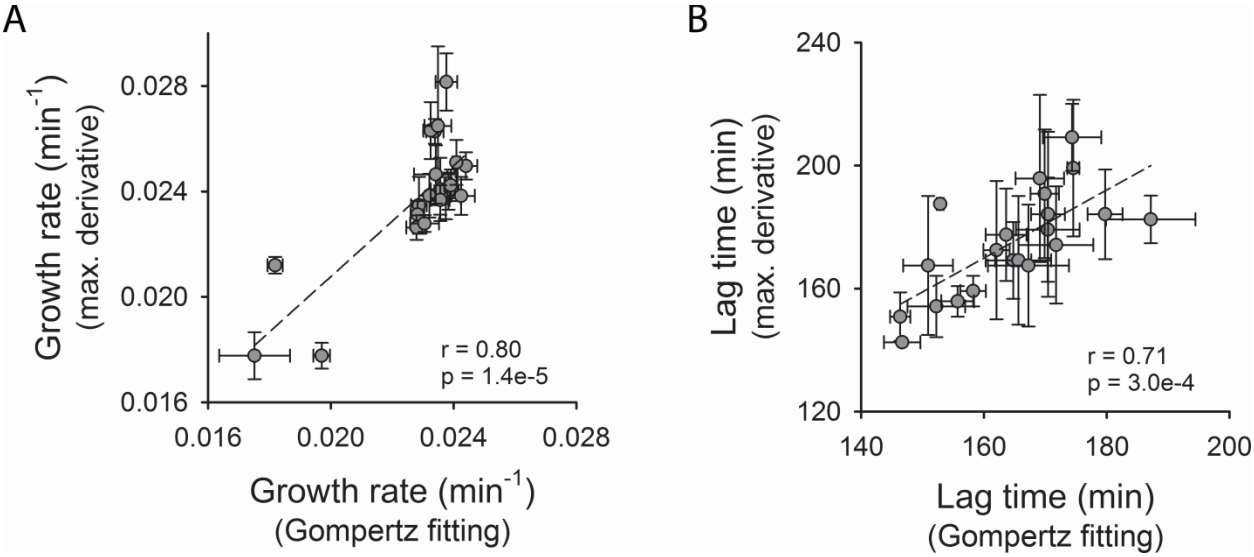


Fig S7: Correlation between growth parameters derived from Gompertz fitting and maximum-derivative method. The parameters derived from both the methods correlate very well as indicated by Pearson's correlation parameters (r and p-values). The data points represent mean and error bars are standard deviation of 6 or 9 measurements (see Table S2).

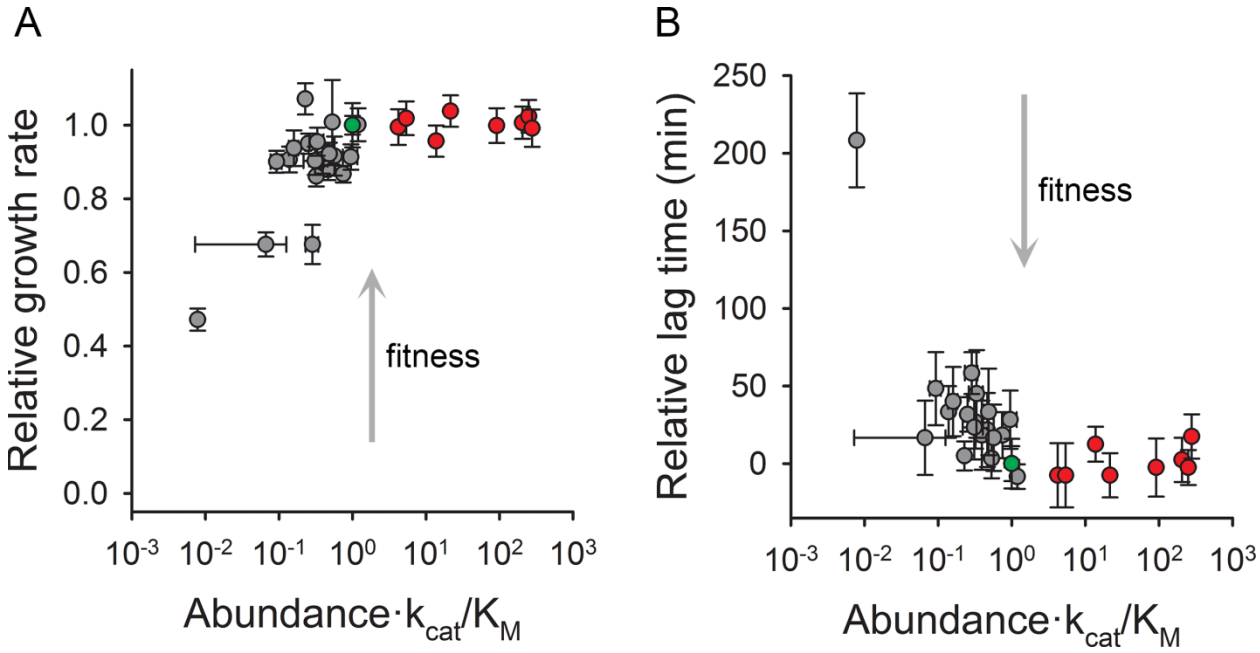


Fig S8: Traits of population growth. (A) Growth rates and (B) lag time obtained from analysis of growth curve derivatives shown as a function of catalytic capacity which is defined as abundance $\times k_{cat}/K_M$. The mutant data is shown in gray circles, whereas red circles represent the BW27783 strain with varying degrees of overexpression of WT Adk from a pBAD plasmid. Data for WT is shown in green. Fig 2 is an equivalent figure with growth rate and lag times obtained after fitting the raw data with Gompertz equation (Eq. 2).

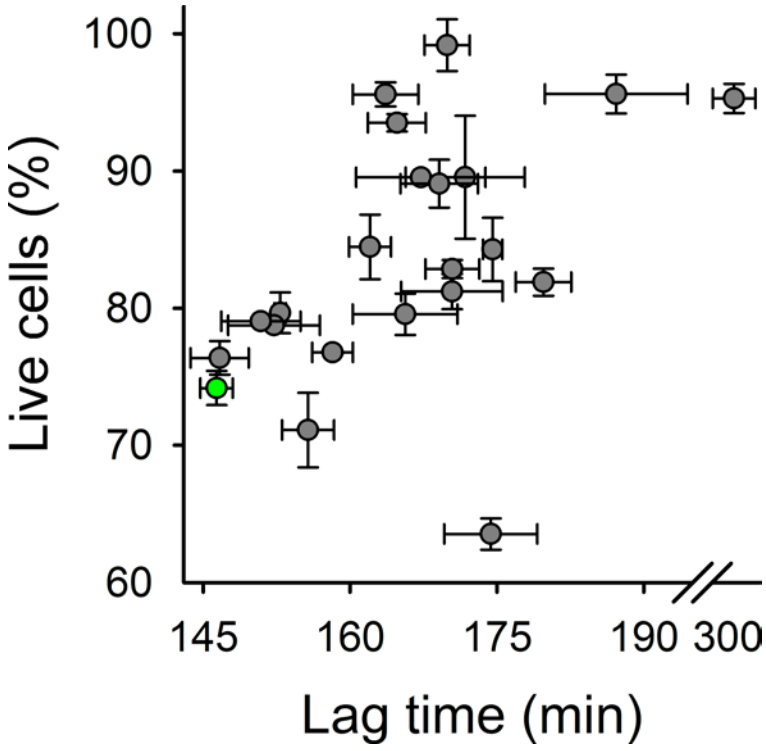
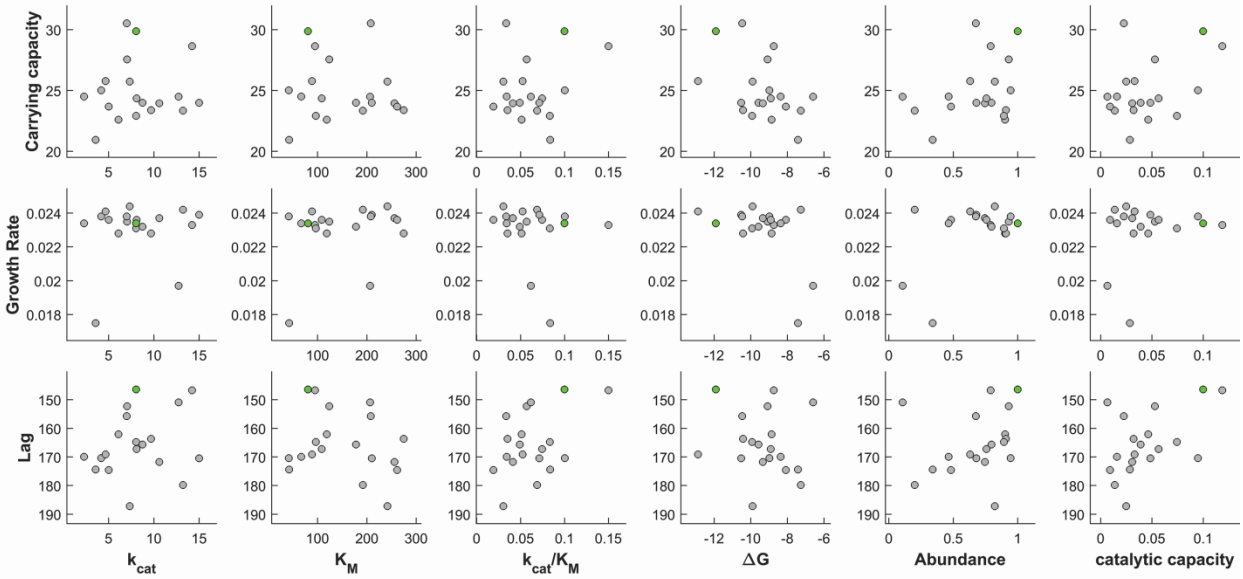


Fig S9: Percentage of live or viable cells of WT and mutant Adk strains at saturation (16 hours of growth) versus their population lag time. The cultures were grown overnight at 30 °C, and then stained using fluorescent dyes Syto9 (specific for live cells) and propidium iodide (specific for dead cells). The data points are mean and error bars represent standard deviation of 2 biological replicates. WT Adk strain is shown in green.

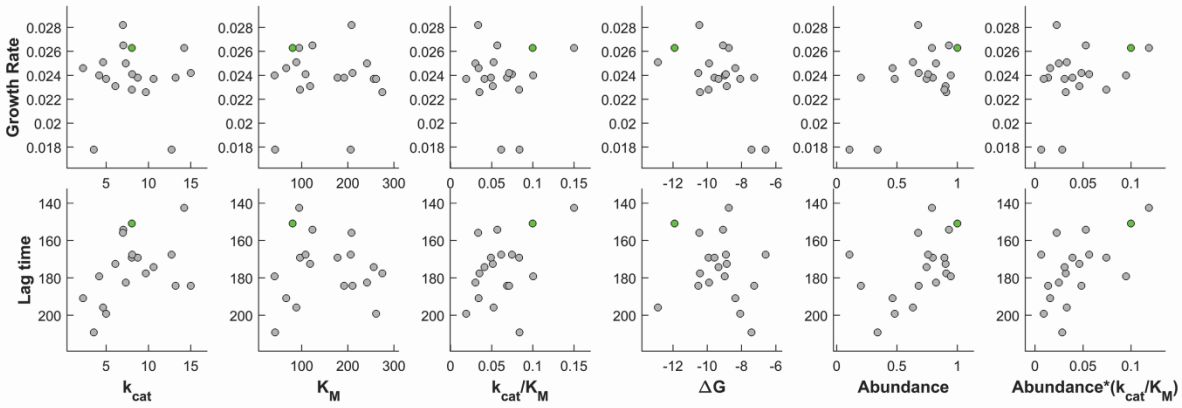
A



B

		k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($\mu M^{-1} \cdot min^{-1}$)	ΔG ($kcal \cdot mol^{-1}$)	Abundance	Abundance × (k_{cat}/K_M)
Carrying capacity	ρ	-0.02	-0.17	0.04	-0.38	0.22	0.24
	p-value	0.943	0.495	0.870	0.108	0.375	0.314
Growth rate (min^{-1})	ρ	0.04	0.20	-0.21	-0.32	-0.07	-0.08
	p-value	0.861	0.424	0.394	0.179	0.791	0.737
Lag time (min)	ρ	-0.19	0.20	-0.30	0.20	-0.39	-0.44
	p-value	0.442	0.412	0.212	0.420	0.103	0.057

C



D

		k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($\mu M^{-1} \cdot min^{-1}$)	ΔG ($kcal \cdot mol^{-1}$)	Abundance	Abundance × (k_{cat}/K_M)
Growth rate (min^{-1})	ρ	-0.06	-0.17	0.00	-0.44	0.24	0.32
	p-value	0.822	0.483	0.991	0.057	0.313	0.183
Lag time (min)	ρ	-0.42	0.03	-0.30	0.17	-0.48	-0.51
	p-value	0.077	0.903	0.209	0.486	0.038	0.027

Fig S10: Scatter plots of growth parameters (carrying capacity, growth rate and lag times) and molecular and cellular properties of Adk. Parameters were obtained using (A) Gompertz fit and (C) analysis of growth curve derivatives. Panels (B) and (D) show Spearman correlation coefficients (ρ) and *p*-values for each of the sub-plots in panels (A) and (C) respectively. The highest correlation values in each panel are highlighted in yellow.

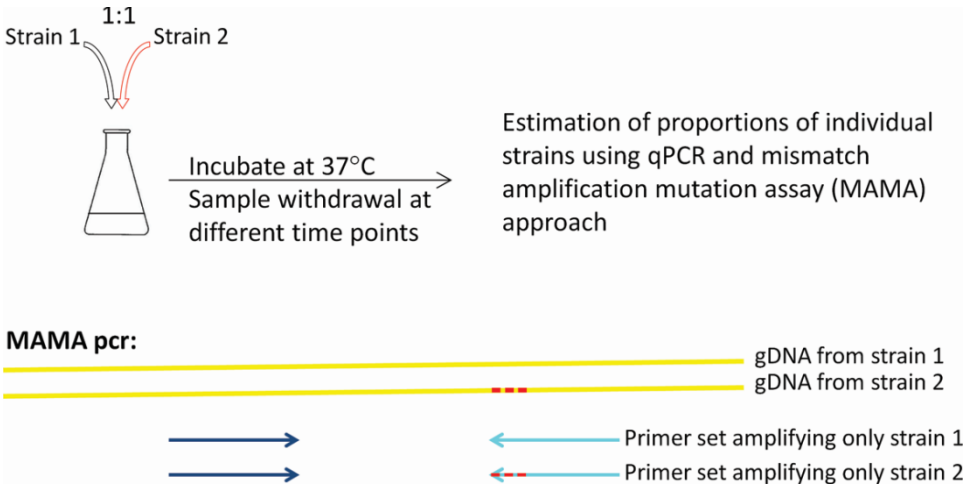


Fig S11: Schematic representation of binary growth competition experiments and estimation of relative proportion of competing strains. The strains (1) and (2) are mixed in 1:1 proportion and were grown at 37 °C. Samples were drawn at different time points, normalized for OD, and genomic DNA was extracted. The proportions of individual strains were estimated by a qPCR method employing mismatch amplification mutation assay method (see Methods). We designed a set of primers to differentially amplify the strains by matching the 3'-end of one of the primers to the site of mutation and using Taq DNA polymerase for amplification.

Table S1: Structural and biophysical parameters

adk strain	SC acc[a] (%)	Residue depth (Å)	Fraction of WT in MSA	Fraction of mutant in MSA[b]	T_m (DSC) (°C)	T_m (TFA)[c] (°C)	ΔG (kcal/mol)	k_{cat}[d] (min^{-1})	K_M[e] (μM)	k_{cat}/K_M ($\mu M^{-1}min^{-1}$)
WT	--	--	--	--	55.9	53.8	-11.9	8.05	80.58	9.99E-02
L082F	0.0	8.8	0.57	0.00	49.2	47.4	-8.7	14.23	94.86	1.50E-01
L082V	0.0	8.8	0.57	0.14	55.7	53.2	-11.3	n.d.[f]	n.d.[f]	n.d.[f]
L083A	0.1	9.0	0.70	0.00	48.5	46.9	-8.9	6.10	118.73	5.13E-02
L083F	0.1	9.0	0.70	0.60	n.d.[f]	54.2	-10.8	n.d.[f]	n.d.[f]	n.d.[f]
L083I	0.1	9.0	0.70	0.23	52.4	50.7	-10.4	9.70	274.80	3.53E-02
L083T	0.1	9.0	0.70	0.00	49.4	49.3	-9.9	8.05	96.55	8.34E-02
A093F	0.0	5.9	0.85	0.00	49.4	47.2	-9.1	7.05	123.83	5.69E-02
A093I	0.0	5.9	0.85	0.00	51.3	49.7	-9.9	7.34	241.75	3.04E-02
A093L	0.0	5.9	0.85	0.07	50.6	47.9	-9.6	8.75	178.05	4.91E-02
A093Y	0.0	5.9	0.85	0.00	51.2	49.4	-9.3	10.61	256.23	4.14E-02
V106A	0.0	7.4	0.74	0.38	49.7	47.3	-9.0	4.19	41.78	1.00E-01
V106H	0.0	7.4	0.74	0.00	43.2	41.2	-6.6	12.74	206.54	6.17E-02
V106L	0.0	7.4	0.74	0.04	50.1	48.5	-8.9	8.10	108.60	7.46E-02
V106N	0.0	7.4	0.74	0.00	39.0	37.9	-4.8	12.07	158.06	7.64E-02
V106W	0.0	7.4	0.74	0.04	45.0	43.0	-7.3	13.22	192.05	6.88E-02
Y182F	7.2	6.3	0.86	0.14	55.4	53.8	-10.5	15.00	210.25	7.13E-02
Y182V	7.2	6.3	0.86	0.00	45.7	46.2	-8.1	5.01	261.53	1.92E-02
L209A	0.0	7.0	0.23	0.08	45.0	44.3	-8.4	2.30	66.94	3.43E-02
L209F	0.0	7.0	0.23	0.01	51.5	51.0	-10.5	7.00	208.17	3.36E-02
L209I[g]	0.0	7.0	0.23	0.58	56.1	55.5	-12.9	4.66	88.82	5.25E-02
L209S	0.0	7.0	0.23	0.00	43.2	42.4	-7.4	3.57	42.55	8.38E-02

[a] % sidechain accessibility calculated using coordinates of pdb 4ake

[b] fraction in multiple sequence alignment when WT amino acid is excluded

[c] melting temperture from thermofluor assay

[d] kcat for ADP formation

[e] KM for ATP

[f] not determined

[g] the only case in this dataset where fraction of mutant amino acid was greater that WT amino acid in MSA

Table S2: Intracellular abundance and growth parameters of adk mutants

adk strain	Carrying capacity, K[a]	s.d. in K[b]	Growth rate, μ[a] (min^{-1})	s.d. in μ[b]	Lag time, λ[a,c] (min)	s.d. in λ[b]	Abundance[d]	s.d. in abundance[e]
WT	29.9	3.0	0.0234	2.85E-04	146.4	1.7	1.00	0.00
L082F	28.7	2.9	0.0233	2.40E-04	146.7	3.0	0.79	0.05
L082V	24.4	0.8	0.0229	2.60E-04	158.2	2.1	n.d.[f]	n.d.[f]
L083A	22.6	0.8	0.0228	2.22E-04	162.0	2.1	0.90	0.08
L083F[g]	27.3	1.3	0.0182	2.48E-04	152.9	0.6	1.19	0.10
L083I	23.4	0.9	0.0228	3.30E-04	163.6	3.3	0.91	0.02
L083T	22.9	1.0	0.0231	4.72E-04	164.8	2.9	0.89	0.03
A093F	27.6	4.3	0.0235	4.37E-04	152.2	4.7	0.93	0.03
A093I[g]	25.7	0.5	0.0244	3.69E-04	187.2	7.3	0.82	0.07
A093L	24.0	1.0	0.0232	3.44E-04	165.6	5.3	0.80	0.13
A093Y	24.0	1.1	0.0237	4.90E-04	171.7	6.1	0.75	0.23
V106A	25.0	0.5	0.0238	3.28E-04	170.4	5.2	0.95	0.21
V106H[g]	24.5	2.2	0.0197	2.68E-04	150.9	4.1	0.11	0.10
V106L	24.4	2.2	0.0236	3.08E-04	167.2	6.6	0.76	0.05
V106N	5.9	0.3	0.0137	6.85E-04	305.8	20.0	0.01	0.00
V106W	23.4	1.0	0.0242	4.48E-04	179.7	2.8	0.20	0.01
Y182F	24.0	1.3	0.0239	2.15E-04	170.4	2.8	0.68	0.09
Y182V	23.7	1.1	0.0236	3.08E-04	174.6	1.0	0.48	0.08
L209A	24.5	1.3	0.0234	7.12E-04	169.9	2.3	0.46	0.07
L209F	30.5	1.8	0.0238	3.54E-04	155.7	2.7	0.68	0.05
L209I	25.8	1.2	0.0241	1.27E-04	169.1	4.0	0.63	0.14
L209S	20.9	1.5	0.0175	1.15E-03	174.4	4.8	0.34	0.07

[a] parameters derived by fitting Gompertz equation to ln(OD) vs time at 37 C

[b] standard deviation derived from 9 replicates (triplicates of 3 biological replicates)

[c] time required to achieve maximum growth rate

[d] abundance measured after 4h of growth at 37 C

[e] standard deviation derived 2 biological replicates

[f] not determined

[g] growth parameters derived for 6 replicates (triplicates of 2 biological replicates)

Table S3: Intracellular abundance and growth parameters of WT adk overexpression from pBAD plasmid in *E. coli* BW27783 strain

arabinose concentration (%)	Carrying capacity, K[a]	s.d. in K[b]	Growth rate, μ[a] (min^{-1})	s.d. in μ[b]	Lag time, λ[a,c] (min)	s.d. in λ[b]	Abundance[d]	s.d. in abundance[e]
no plasmid	29.5	2.5	0.0195	5.10E-04	126.7	1.0	1.00	0.10
0.00E+00	32.0	0.5	0.0195	2.00E-04	127.2	0.7	4.20	0.42
3.05E-06	30.9	1.8	0.0191	2.00E-04	127.2	2.0	13.73	1.37
1.22E-05	31.3	1.0	0.0197	2.00E-04	125.0	0.4	5.39	0.54
4.88E-05	27.1	1.0	0.0199	2.65E-04	135.7	1.1	21.55	2.15
1.95E-04	32.7	0.9	0.0191	1.00E-04	127.5	1.2	91.27	9.12
7.81E-04	31.6	2.5	0.0192	2.52E-04	127.0	2.1	205.31	20.51
3.13E-03	32.6	1.6	0.0192	4.73E-04	126.1	1.4	250.39	25.02
5.00E-02	28.4	3.1	0.0195	3.61E-04	127.1	2.5	278.69	27.84

[a] parameters derived by fitting Gompertz equation to ln(OD) vs time at 37 C

[b] standard deviation of 3 replicates

[c] time required to achieve maximum growth rate

[d] abundance measured after 4h of growth at 37 C

[e] standard deviation of 2 biological replicates