On the metabolic basis and predictability of global epistasis.

Pavithra Venkataraman[#], Supreet Saini

Department of Chemical Engineering, Indian Institute of Technology Bombay, India 400076

Corresponding Author: pavithrav@iitb.ac.in; +91 22 2576 4235

Running Title: *A mechanistic explanation of global epistasis*

Abstract

Fitness effects of mutations are epistatic, i.e., they depend on the genetic background in which they occur. Despite microscopic idiosyncrasies of mutational effects, laboratory evolution experiments showed consistent patterns of declining adaptability, indicating some underlying "rule" that dictates mutational effects. Recently, global epistasis showed that mutations become less beneficial as the fitness of the genetic background in which they occur increases, thus making epistasis predictable. However, there exists no mechanistic explanation of this fitness dependence of fitness effects.

In this work, we use a prokaryotic cell growth and division model based on metabolic sizing to study the mechanistic basis, and predict the likelihood of global epistasis. We report two important results. One, although the average extent of global epistasis does not change with genetic variation, the likelihood of strong global epistasis decreases as genetic variation increases. Two, the fitness effects of mutations in the slow reaction modules are well predicted using the global epistasis approach. Therefore, when the fitness of the cell is contingent on a rate-limiting step, mutations in that step are well-predicted by the fitness of the background in which they occur. Thus, declining adaptability is a consequence of evolution acting in a step-wise fashion to improve the slowest modules in a cell.

Overall, our work offers a first-principle explanation of global epistasis and declining adaptability.

Keywords: Epistasis; Declining Adaptability; Mutation Effects; Metabolism.

Introduction

Adaptation of living organisms to changing environments occurs via natural selection acting on randomly occurring mutations that could alter metabolic networks in cells. These adaptive changes have implications on biodiversity, genetics, human health, biogeochemical processes, and agriculture[1–5]. Prediction of these naturally occurring changes, which is a central goal of evolutionary biology, has been challenging because the effects of mutations on the phenotypes of living organisms is highly contingent on the genetic background in which they occur[6,7]. These genetic (G x G) interactions, which are both intra- and inter-genic, constitute the phenomenon of epistasis. Epistasis plays an important role in shaping evolutionary trajectories by altering both the sign and magnitude of fitness effects of mutations[7–12].

In the absence of information about how epistasis mechanistically operates at the cellular level, empirical studies have largely focused on building a statistical understanding of epistasis[13–15]. One approach is to characterize the fitness effects of individual mutations in different genetic backgrounds, and determine whether selection will favour the mutation or not[16–19]. An alternative approach to study epistasis empirically ignores individual mutations and focuses on drawing up a distribution of fitness effects (DFEs) of mutations for a particular genotype[20–26]. The challenge with these approaches is that there is an astronomical number of mutations and genotypes to study, and hence, impossible to fully trace experimentally[27,28].

Several empirical studies have shown that the adaptability of evolving populations decreases with time, indicating a generic "rule" affecting mutational effects, independent of molecular details[29,30]. This declining adaptability is explained by a pattern of global epistasis, called diminishing-returns epistasis, which suggests that the fitness effects of beneficial mutations decrease linearly with an increase in background fitness[31]. Correspondingly, the deleterious effects of deleterious mutations increase linearly with increasing background fitness, and this is called increasing costs epistasis[32]. These consistencies indicate that fitness gains saturate, and that highly fit genotypes are less robust to deleterious mutations. Global epistasis is the only convenient tool available to predict epistatic effects, but its underlying cellular origin is unknown[13].

Metabolic networks have been used to understand the non-additive effects of epistasis, by focusing on the microscopic epistatic patterns in metabolic pathways[33–36]. Epistatic effects have often been attributed to the structural and functional changes that mutations cause in proteins[37,38]. For example, epistasis can arise due to altered interactions at active sites of

enzymes, or due to rewiring of biochemical interactions between two or more protein molecules[39–43]. Although correlations between structural changes and mutational effects have been established in the past, there exists no generic first-principle explanation of epistasis. A recent work aimed at understanding the evolution of antibiotic resistance in β-lactamase gene showed that mutations which alter the rate-limiting step led to epistasis[44]. This important work shows how fundamental changes in the steps of enzyme catalysis could lead to altered evolutionary trajectories. However, much like constructing fitness landscapes and DFEs, it is impossible to identify all the underlying biochemical changes due to mutations in a cell. In fact, mutations in non-coding regions have also been shown to modulate epistasis[45–47]. In such a scenario, a mechanistic model of epistasis, that also captures global epistasis, remains elusive.

In this work, we attempt to explain the origin of epistasis and its globality feature, using a metabolic-sizing based cell growth model in which changes in coding and non-coding regions manifest as changes in the metabolic flux. Subsequently, we show how genetic variation modulates the strength and globality of epistasis, and offer a metabolic flux-based explanation of when global epistasis may be expected.

Methods

Cell growth and division model based on a metabolic adder and sizer

We model cell growth and division based on the production of three bottleneck metabolites. These bottleneck metabolites are produced via independent two-step enzymatic reactions, all of which are fed by a substrate drawn from the environment, as shown in **Figure 1**. When a cell accumulates predefined amounts (10^8 molecules in this work) of all the threshold metabolites, it divides, and each of the daughter cells inherits half of all the cellular constituents (mRNA, enzymes, and metabolites). The time taken for a cell to divide is t_{div}, and $\frac{1}{t_{div}}$, the growth rate, is considered a proxy for the fitness of the cell.

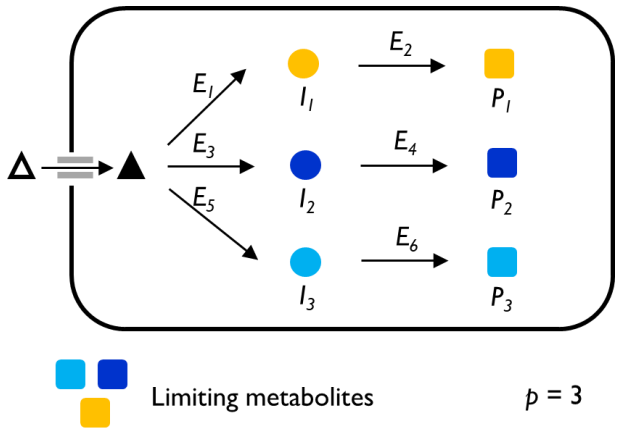

Figure 1. Essential metabolic pathways in the cell. The cells modelled in this work divide when they accumulate threshold amounts of limiting metabolites. A cell with three bottleneck metabolites (P_1, P_2, and P_3) is shown. Each terminal metabolite is produced via enzymatic cascades, feeding on the same substrate.

Stochastic gene expression and production of enzymes

Independent gene-promoter systems drive the production of the enzymes that drive the enzymatic reactions in the cell, as shown in **Figure 2**. The process of gene expression comprises of stochastic (burst-like) transcription (Ref), and deterministic translation (Ref). To implement stochastic-burst-like gene expression, we sample random "t_{ON}" and "t_{OFF}" durations of each of the promoters from two exponential distributions of set means. In this study, the means of "t_{ON}" and "t_{OFF}" lie between 0.5 and 1, and 0.5 and 5, respectively[48]. We assume no shortage of RNA polymerases, so the process of transcription proceeds in a deterministic fashion for "t_{ON}" duration. Each mRNA is assigned a lifetime duration drawn from an exponential distribution of a set mean (= 25 minutes)[49]. Only one ribosome is attached to an mRNA at a

given instant of time to translate it. We sample an exponential distribution of a set mean (= 20 hours) to assign lifetime durations for each protein molecule produced[50]. The length of the gene is 1029 bp[51], the rate of transcription is 10 nucleotides/s, and the rate of translation is 10 amino acids/s[52–54].

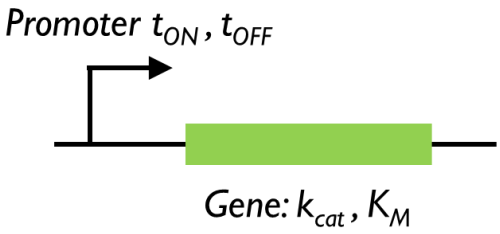

Figure 2. Gene and promoter characteristics. Every enzyme involved in the production of threshold metabolites is produced based on the characteristics of its gene and promoter. We model stochastic, burst-like transcription. The enzymes catalysis is quantified by k_{cat} and K_M.

Production of metabolites via enzymatic reactions

Every enzyme involved in the production of the intermediate and threshold metabolites is characterized by a k_{cat} and K_M, drawn from data reported in the literature[55]. Michaelis-Menten reaction laws govern the conversion of one metabolite to another. The amount of each metabolite that the cell must accumulate before it divides is set at 10^8, depending on division time distributions[56].

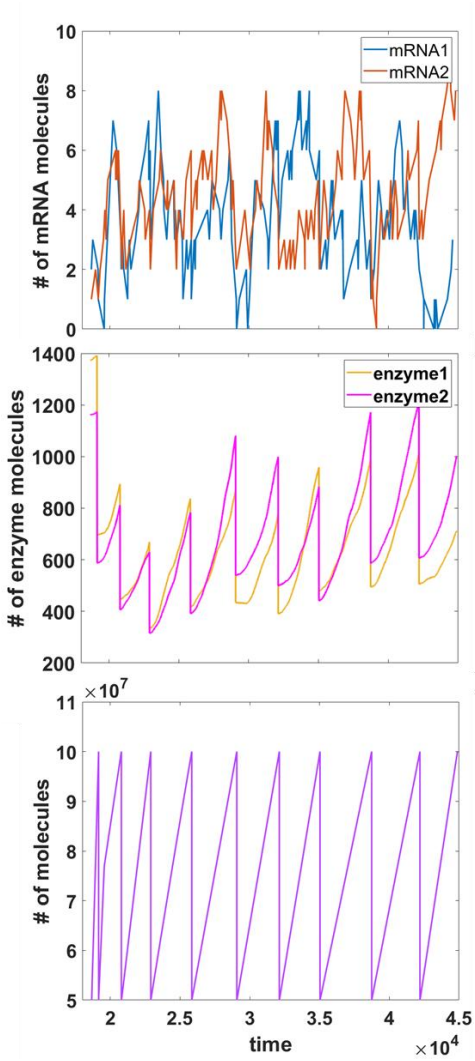

Figure 3. Temporal profiles of mRNAs, enzymes, and threshold metabolite. We model a cell with stochastic transcription, deterministic translation, and division based on accumulation of threshold metabolites. In this scenario, for one of the three enzymatic cascades, the temporal profiles of the two mRNAs (top panel), two enzymes (middle panel), and the threshold metabolite (bottom panel) are shown.

Creation of background variation

As shown above, the cells we model have six enzymes (*3 threshold metabolites x 2 reactions to produce each of them*). Each of these enzymes' amounts and catalytic properties are governed by four parameters – t_{ON}, t_{OFF}, k_{cat}, and K_M. Therefore, the fitness of a cell is a non-linear function of twenty-four parameters. To create variations in the cell, one could vary any of these twenty-four parameters. However, in a module, varying a gene or a promoter has similar distribution of fitness effects (**Supplementary Figure 1**).

In this study, unless specified otherwise, we study the effect of a mutation iin t_{ON1}, which changes its value from 0.5 to 1. We call this mutation the focal mutation. For a focal mutation in one of the modules, we could have background mutations in - (a) the same module, (b) a different module in the same path, (c) a module in the same position but different path, and (d) a module in a different path and position. Different combinations of these classes yield fifteen ($^4C_1 + {}^4C_2 + {}^4C_3 + {}^4C_4$) types of background variations that could be created. The parameters that we vary corresponding to cases (a), (b), (c), and (d) as explained above, are k_{cat1}, t_{ON2}, k_{cat3}, and t_{OFF6}, and they are denoted by *1, 2, 3,* and *4*, respectively, in this paper.

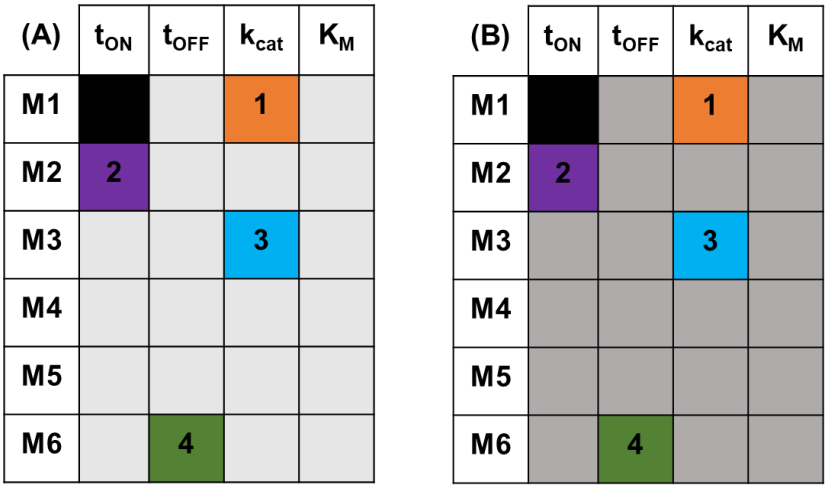

Figure 4. Variations in the parameters to create different backgrounds. In this figure, the rows indicate the six modules in the cell, and the columns indicate the four parameters in each of these modules. In this study, a mutation in t_{ON1} (0.5 to 1, cell highlighted in black in **(A)** and **(B)**) has been studied in different "genetic" backgrounds. In order to create these different backgrounds, four parameters are varied in different combinations, while keeping the others constant. The four parameters are k_{cat1} (indicated by 1 in this study, highlighted in orange in **(A)** and **(B)**), t_{ON2} (indicated by 2 in this study, highlighted in violet in **(A)** and **(B)**), k_{cat3} (indicated by 3 in this study, highlighted in yellow in **(A)** and **(B)**), and t_{OFF6} (indicated by 4 in this study, highlighted in green in **(A)** and **(B)**). Changes to the cells highlighted in different shades of grey indicate changes in the macro interactions, and we equate such changes with differences between strains.

Consider the case where there is one varying parameter, say t_{ON2}. In order to study the epistatic effects of microscopic variation, we create a pool of twenty-five cells which differ in t_{ON2}, identical to each other in twenty-two parameters, and the twenty-fourth parameter is the focal mutation. This group of cells can be thought of as those belonging to the same strain. (Similarly,

for cases where there are two, three, and four varying parameters, twenty-one, twenty, and nineteen parameters constant, respectively).

To study the epistatic effects of macroscopic variation, we create twenty such distinct pools (say, twenty different strains), and each pool has a unique set of twenty-two parameters. We then introduce the focal mutation and study its fitness effects in these resulting five hundred cells. We study the effect of the focal mutation using this approach in all the fifteen possible background variations.

Creation of cells with predefined metabolic speeds

In the cells that we model, there are six enzymatic reaction steps. We create cells with predefined metabolic speeds to study global epistasis. The speed of a reaction is either "slow", "same" or "fast" relative to the other reactions. In this study, we create toy cells in which the t_{OFF} and K_M of all the six modules are identical and equal to 5 and 60220 respectively. t_{ON} of all six modules are identical to each other, but is a value between 0.5 and 1. This variation in t_{ON} gives cells of variable fitness. A reaction step's speed, relative to the others, is decided by the k_{cat}, which is 10 for "slow" reactions, and 100 for "fast" reactions.

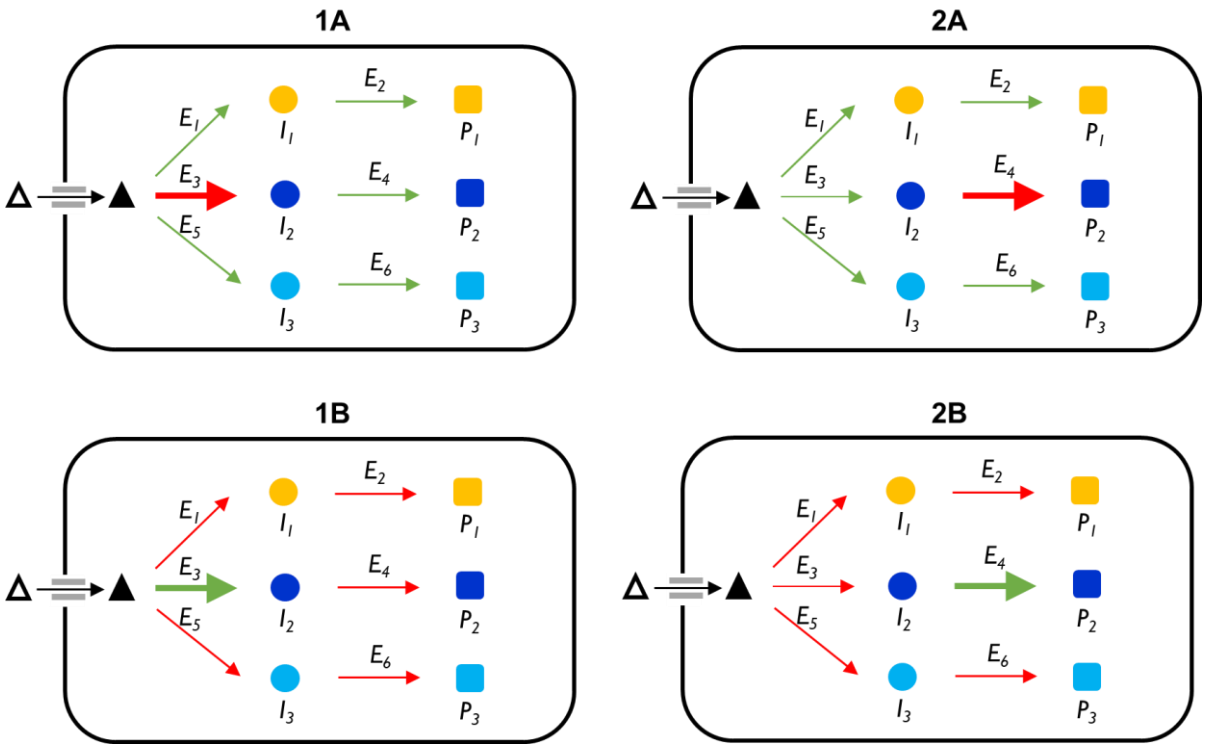

Figure 5. Toy cells with predefined metabolic speeds. The cells that we model have three reaction arms which produce the three threshold metabolites. Each of the reaction arms have

two reaction steps. In total, there are six reaction modules. We create toy cells in which the speeds of each of these modules is predefined as either slow or fast, relative to the others. Arrows in red and green indicate slow and fast reactions, respectively. In **1A** type of cells, one of the first reaction steps is slow, while the others are all equally faster. In **1B**, one of the first step reactions is fast, and the other five are all equally slow. In **2A**, one of the second step reactions is slow and the other five are equally fast, and in **2B**, one of the second step reactions is fast and the other five modules are equally slow.

Results

- **Fitness effects of mutations are dependent on the genetic background in which they occur.**

We modelled cell growth and division following a metabolic adder-sizer principle, and introduced a "mutation" by changing the value of one of the parameters controlling one of the six enzymatic reactions in the cell, i.e., in module I (denoted as M1 henceforth). We picked this module because it exerts direct control over the other two reaction arms in the cell, as opposed to M2 (module II). The mutation is introduced in cells with different types and extents of background variation, and we had fifteen types of variants (See Methods for more details).

In all the fifteen cases, we observed that the fitness effect of this mutation was non-uniform, and was dependent on the genetic background in which it occurred, as shown in Figure 6**.** This result was applicable with changes in the number of threshold metabolites (See **Supplementary Figure 2**). These results, predictably, showed that our model captures epistasis.

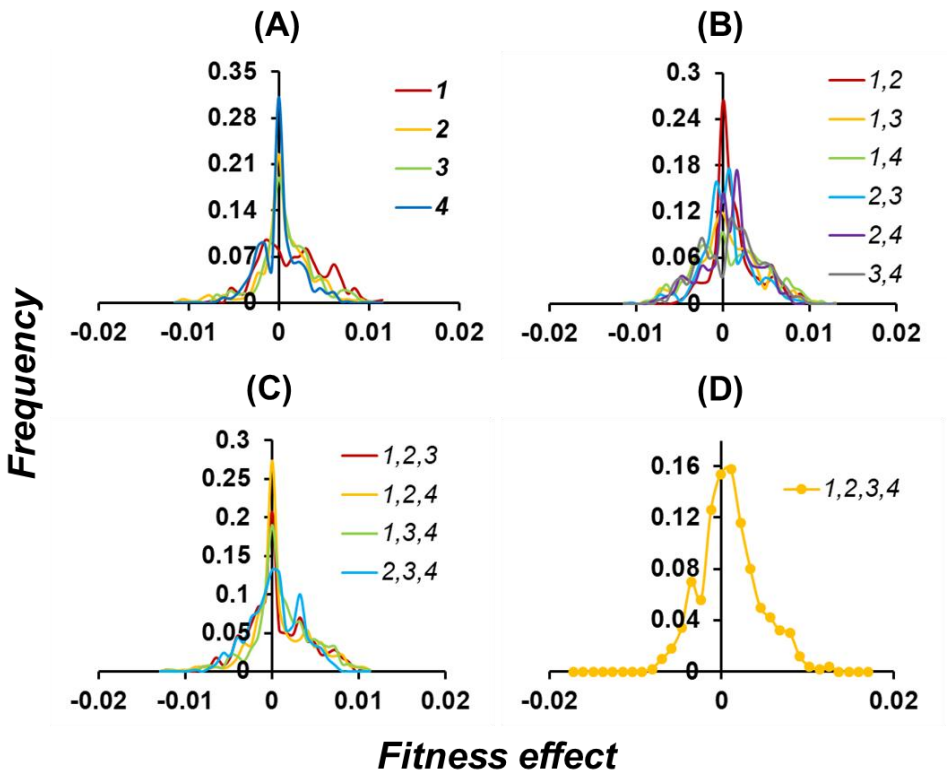

Figure 6. The fitness effects of mutations are background-dependent. We test if our model can capture background-dependent fitness effects of mutations, and ask how this

depends on the extent and type of background variation. The figures above correlate the frequency of mutations with fitness effects. **(A), (B), (C)** and **(D)** show the fitness effect of the focal mutation in different combinations of one, two, three, and four background variations, respectively. The legends indicate the background parameters that are varied to create different genotypes. Refer to methods for more details.

- **The fitness effects of mutations are predictable based on the fitness of the background in which they occur.**

One of the most widely studied method to predict effect of epistasis is the global epistasis approach, in which the fitness of the background in which the mutation occurs is used as a predictor of the mutation's fitness effects. More specifically, experiments have shown that the fitness effect of a mutation linearly decreases with an increase in background fitness. Greater the R^2 of this linear fit, greater is the globality of epistasis[57]. Epistasis of different globalities has been observed in several laboratory evolution experiments[31,32,57,58]. We tested if our model captured these experimental observations of this statistical pattern of mutational effects.

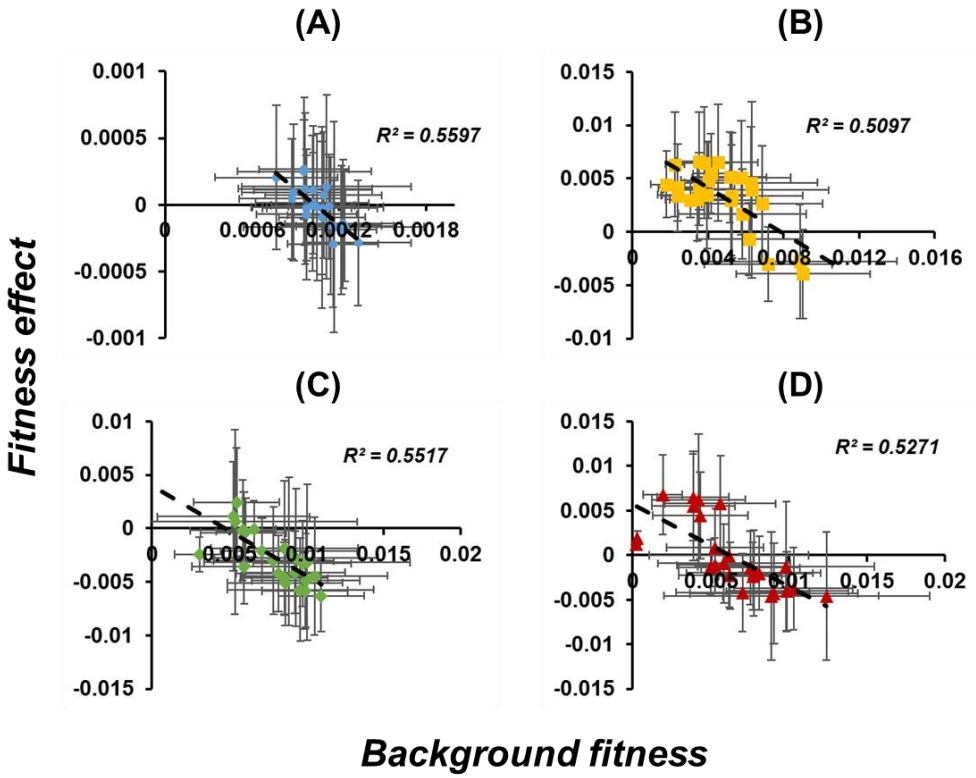

Figure 7. The fitness effects of mutations exhibit global epistasis. We test if our model captures empirical observations of global epistasis. **(A), (B), (C)** and **(D)** depict the fitness effect of changing t_{ON1} from 0.5 to 1 in genetic backgrounds that differ by

one (k_{cat3}), two (k_{cat3} & t_{OFF6}), three (t_{ON2}, k_{cat3}, t_{OFF6}) and four sites (k_{cat1}, t_{ON2}, k_{cat3}, t_{OFF6}), respectively. Clearly, the fitness effect of a mutation decreases linearly with an increase in background fitness.

A fixed mutation in M1, described above, was studied in cells belonging to fifteen types of background genetic variations. As shown in Figure **2A-2D**, our model captures global epistasis as the extent of background variation increases. Refer to **Supplementary Figure 3** for all the 300 regressions (15 x 20). Interestingly, epistasis of a wide range of globalities is observed in all types of genetic backgrounds, i.e., the extent of background variability among the cells did not qualitatively dictate the globality of epistasis. The average globality of epistasis did not change significantly with a change in background variation extent or type, as shown **Figure 8.**

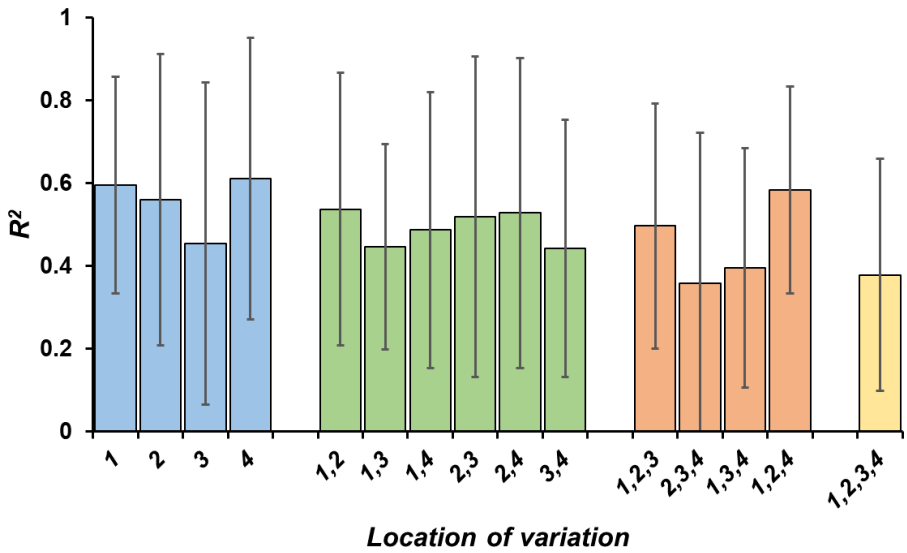

Figure 8. The average globality of epistasis for a mutation does not change with change in genetic variability. Based on Supplementary Figure 3, we measure the average globality of epistasis in the fifteen different types of cells in our study. For the given mutation, the average (indicated by the bars) does not change with changes in the position or extent of background variation. However, because of a large variance (the error bars indicate standard deviations), the globality of epistasis it is not predictable based on genetic variability.

- **The likelihood of occurrence of strong global epistasis changes with extent of background variation.**

The shape of global epistasis is inferred from the slope of the regression between fitness effects and background fitness. Negative slope indicates diminishing-returns and

increasing-costs epistasis, while positive slope indicates increasing-returns and diminishing-costs epistasis. Based on the slopes of the regression lines in **Supplementary Figure 3**, we identified if global epistasis was negative or positive. In most cases (84.28% of 299 cases, 1 of the cases was discarded), negative global epistasis was observed, as shown in **Figure 9**.

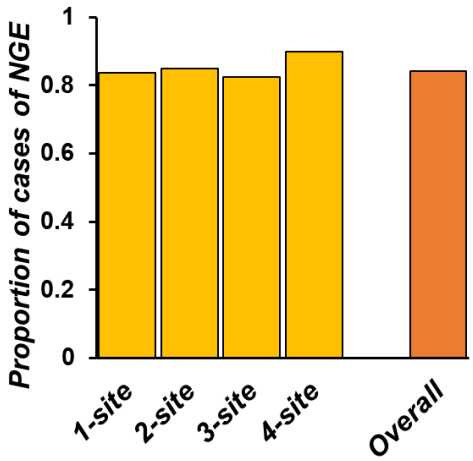

Figure 9. Negative global epistasis is ubiquitous. Global epistasis is negative when the fitness effect of a mutation decreases linearly with an increase in background fitness. In the 299 regressions in our study, negative global epistasis was common in all types of background variations.

We considered the R^2 values of the negative global epistasis cases, in each of the fifteen types of background variation. The expectation of R^2 for a mutation is the same, regardless of whether it occurs in a group of cells that are very different from each other, or in a group of cells which are almost identical, as shown in **Figure 10**. Similarly, for the average R^2 of negative global epistasis, it also does not matter where in the metabolic network the background variation is.

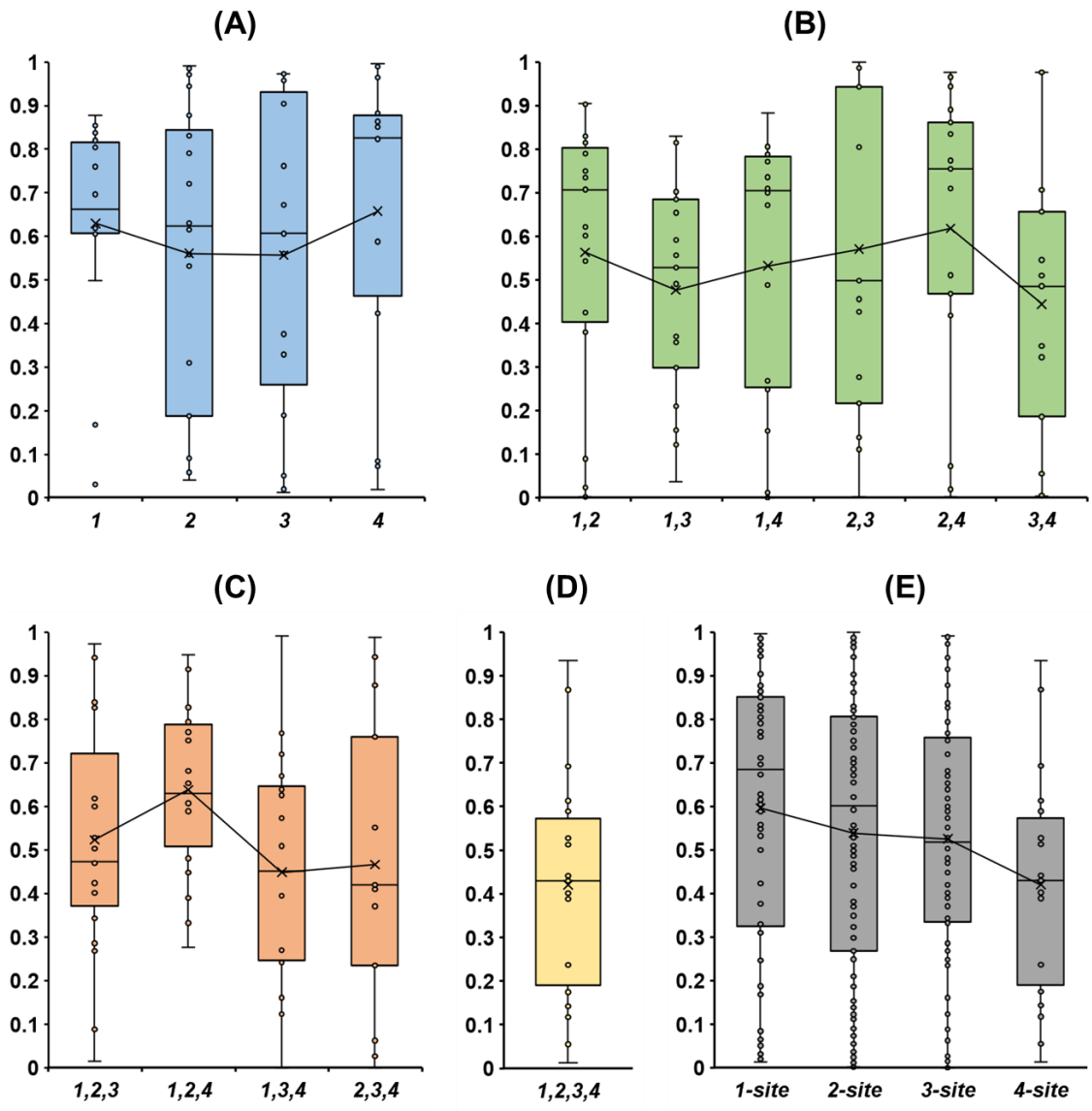

Figure 10. The average globality of negative epistasis for a mutation does not change with change in genetic variability. In each of the fifteen types of background variations, we identify the R^2 values corresponding to negative global epistasis cases in Supplementary Figure 3. **(A)** shows the distribution of R^2 in the four one-background-site variation conditions. Similarly, **(B)**, **(C)**, and **(D)** show R^2 values for the cases in which there are two, three and four background variable parameters, respectively. Clearly, it does not matter where the variation exists, but the average globality of negative epistasis does not change significantly. **(E)** shows the overall distribution of the globalities of epistasis in each of these four extents of variability. From this figure, it is clear that increasing the extent of background variation does not change the average globality of epistasis of a mutation.

Epistasis of variable globality has been observed in experimental studies. Can we predict when a mutation would exhibit negative global epistasis? Is it when it occurs in

a group of cells that are not *very* different from each other genetically, or when it occurs in a group of cells that are different from each other? Does it matter in what way the cells in this group differ from each other? We asked these questions using the R^2 of all cases corresponding to negative global epistasis, in all the fifteen types of cell types that we created.

Although the average strength and globality of epistasis are non-variable for a mutation, the distributions R^2 are not the same in all fifteen cases in our study. We set $R^2 = 0.5$ as the threshold for weak negative global epistasis. Therefore, any value of R^2 greater than 0.5 indicates strong negative global epistasis.

Using this threshold, we counted the number of instances of strong globality for different extents of background variation. As shown in **Figure 11**, as the extent of background variation increases, the likelihood of observing strong negative global epistasis decreases. This means that when a mutation occurs in a population of cells that are almost like each other, global interactions shape epistatic effects. But if the same mutation were to occur in a population of cells that harbour a large amount of variation, epistatic effects are more likely to be driven by local interactions that are unpredictable using macroscopic traits like fitness.

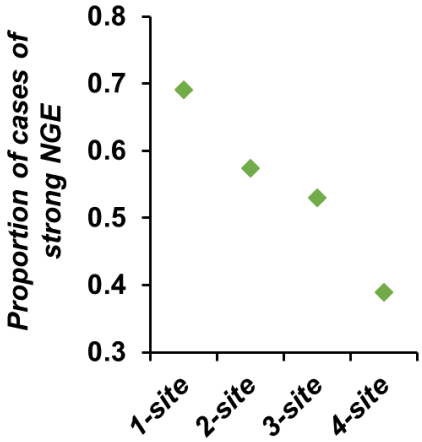

Figure 11. Increasing background variability decreases the likelihood of strong negative global epistasis. Although the average globality of negative epistasis did not change with background variation, it varying distribution indicates that as the extent of variability in a group of cells increases, the likelihood of observing strong negative global epistasis *($R^2 > 0.5$)* decreases.

It must be noted that in the 4-site variation case, all three reaction arms have variations, almost behaving like different strains (See Methods). That is clearly not the case with other extents of variation.

- **Mutations in the slow reaction steps exhibit strong global epistasis.**

Our mechanistic model captures the experimental observations of global epistasis. However, these results do nont allow us to predict from first principles the likelihood of global epistasis. Therefore, we start by asking if the globality of epistasis is predictable based on the speeds of enzymatic reactions in the cell.

We create toy cells in which the relative speeds of the six enzymatic reactions are pre-decided as fast or slow (see Methods). Broadly, we create four types of toy cells in which (a) one of the first step reactions is slow, while all the other steps are fast (1A) (b) one of the first step reactions is fast, while all other steps are slow (1B), (c) one of the second step reactions is slow, while all other steps are fast (2A), and (d) one of the second step reactions is fast, while all other steps are slow (2B). As a result, we have four cell types, in which there is one reaction step that is the odd one out in comparison to the others. We simulate several cells (twenty-five) of each of the four types, and then introduce a hundred mutations in these "odd" steps of each cell, and ask if and how the globality of epistasis varies.

Our results, as shown in **Figure 12**, indicate that mutations occurring in the second step reaction when it is slow (2A), are likely to exhibit strong global epistasis. Fitness effects of mutations in the other types of reaction steps (1A, 1B and 2B) show, on average, a weaker correlation with background fitness.

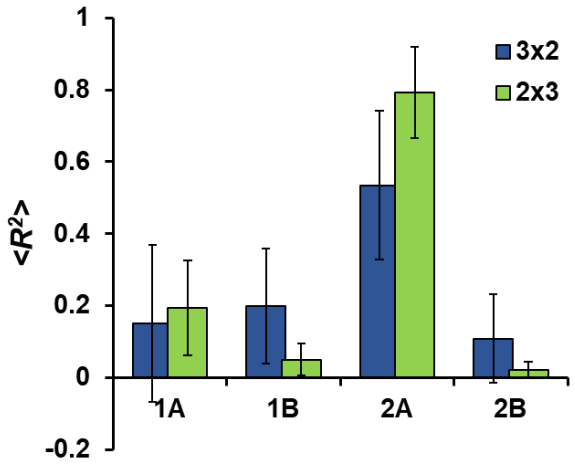

Figure 12. Metabolic speeds alter the extent of global epistasis of mutations. Using toy cells with predefined speeds of reaction modules as described above, we seek to identify the metabolic basis of global epistasis. In cells that have 3 reaction arms with 2 steps in each (3 x 2), the fitness effects of mutations in the first reaction step, when it is slow (1A) or fast (1B), do not exhibit a very strong correlation with background fitness. However, when mutations occur in the second step reaction when it is slow (2A), their fitness effects are much more predictable using background fitness. But background fitness is a poor predictor of the fitness effects of mutations in the second reaction step when it is fast (2A). These results hold when the metabolic architecture is changed to 2 x 3, i.e., 2 reaction arms with 3 steps in each. Here, 1A and 1B correspond to the first reaction steps, and 2A and 2B to the last.

To ensure that this result was not unique to the metabolic structure (3 x 2, 3 threshold reaction arms and 2 enzymatic reactions in each arm), we simulated the four cases – 1A, 1B, 2A and 2B – in a cell with 2 x 3 (2 threshold reaction arms, and 3 enzymatic reactions in each arm) architecture. As shown in **Figure 12** by the green bars, the results obtained for the 3x2 case hold even when the number of "interacting" reaction arms changes.

A recent study on global epistasis predicted that the availability of beneficial mutations should decrease with increase in background fitness, and deleterious mutations would increase correspondingly, indicating that global epistasis is a strong determinant of the sign of mutational effects[15]. This prediction was proven to be right experimentally in a later study[58]. We sought to identify the metabolic basis of these observations.

Using the dataset generated by introducing a hundred mutations in twenty-five genotypes of the four cell types, we obtained a hundred (25 x 4) DFEs. We draw up the mean and variance of these DFEs. As shown in **Figure 13**, although the means of the DFEs decreased linearly with an increase in the background fitness in all cases, the correlation was the strongest for the case 2A. The variance, on the other hand, showed no consistent trends, as depicted in **Supplementary Figure 4**.

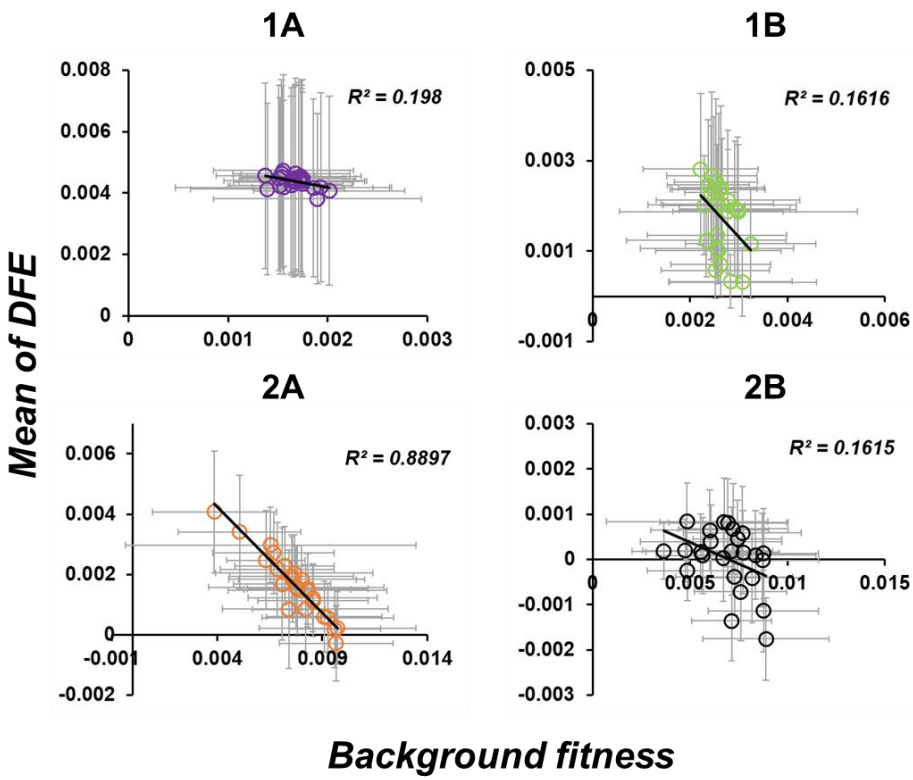

Figure 13. Strong global epistasis helps predict DFEs. Using cells with predefined metabolic fluxes, we ask how effective a predictor of the distribution of fitness effects (DFEs) of mutations is background fitness. As shown in above, the mean of the DFE is best predicted in the 2A case, i.e., when mutations occur in the slow second step reaction. In combination with the results shown in Figure 12, it is clear that strong global epistasis makes DFEs predictable.

Linear regressions of the fitness effects of each of these mutations with the background fitness have an x-intercept, slope, and y-intercept. The x-intercept is the background fitness at which the sign of the mutation changes. This background fitness value has been called the "pivot" growth rate in a recent empirical study carried out by Ardell *et. al.*[58]. This study also showed that a regression line between these slopes and intercepts passes through the origin.

We identified the pivot growth rates and the slope-intercept relationship for the hundred mutations in twenty-five genotypes of the four different cell types. As shown in **Figure 14A**, the linear regression between y-intercepts and slopes passes through the origin in three cases – case 1B, 2A and 2B. The pivot growth rate is predictable, according to these trends, for mutations in 1B, 2A and 2B cases, as shown in **Figure 14B**. In case 1A, the R^2 of the regression between intercepts and slopes is the lowest among the four

cases, and it indicates background-independent fitness effects of mutations. Additionally, there is a significant number of instances of positive global epistasis (indicated by the positive slopes) compared to the other three cases. All these factors result in a non-predictable pivot growth rate, as shown in **Figure 14B**.

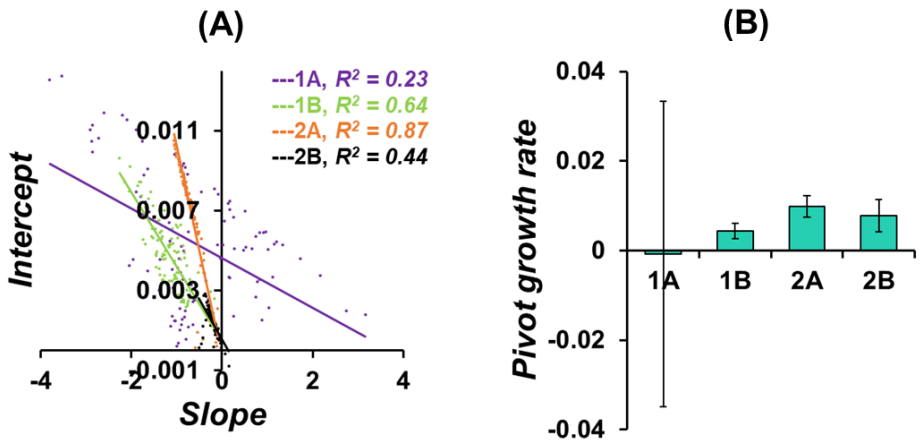

Figure 14. Metabolic flux alters the statistical properties of global epistasis. We asked using cells with predefined metabolic speeds how intercept-slope relationships, and pivot growth rate of mutations exhibiting global epistasis is modulated mechanistically. As shown in **(A)**, intercept-slope regressions pass through the origin in all cases, except 1A, where mutational fitness effects appear to be background-independent. From **(B)**, we see that the pivot growth rate

Discussion

Predicting the fitness effect of mutations is a long-standing goal in biology. Mutations in coding regions could manifest as altered mRNAs or proteins. On the other hand, mutations in the non-coding regions are expected to alter the dosage of proteins and mRNAs. Identifying the microscopic effect of every mutation is impossible, so empirical studies stick to measuring the effects of mutations on macroscopic yet evolutionarily important traits like fitness. Due to epistasis, the fitness effects of mutations depend on the genetic background in which they occur, and often exhibit a linear relationship with the background fitness in which they occur. This phenomenon, called global epistasis, has been observed in several microbial evolution studies, predicted using statistical models, without a robust mechanistic idea of why epistasis manifests this way[8,14].

In this work, we report a mechanistic cell growth and division model that captures epistasis and its globality feature. As explained in the methods, this cell grows until it accumulates enough limiting metabolites, and then divides. The thresholds for the three metabolites are set based on division time distributions that match with realistic division time distributions of prokaryotic cells[56]. In living organisms, metabolic networks are complex and reaction arms are mostly interconnected, except terminal steps lead to the production of metabolites[59]. While the cell that we model does not encode intricate networks, we have two types of reaction modules – ones that control the flux through the other paths, and ones that do not exert control over other paths. This metabolic structure in a cell allows us to comment on the macroscopic effects of mutations in both these types of modules.

In our study, we use the R^2 of the linear fit between fitness effects and background fitness as a measure of the globality of epistasis. The shape of global epistasis (diminishing-returns or increasing-returns) is inferred from slope of the linear regression. In most cases (84.28%) the slope is negative, indicating diminishing-returns and increasing-costs, or, *negative global epistasis*. An empirical manifestation of diminishing returns epistasis is declining adaptability trends in laboratory evolution experiments[29,30,60]. A classic example is the *E. coli* long term evolution experiment, in which the rate of fitness increase decreased with time[21]. This trend has been observed in other studies, when the founder populations differed by a handful, to a thousand mutations[31,61].

However, some exceptions to this trend, where the ancestor fitness did not effectively predict evolutionary fitness gains, are worth noting. When replicates of genetically diverged *E. coli*

strains were evolved, Smith *et. al.* observed declining adaptability among populations that descended from the same strain, but not across strains[62]. Similarly, a study carried out using yeast deletion strains by Rojas Echenique *et. al.* identified that the functional module in which the mutation occurred played an important role in adaptation[63]. Baheti *et. al.* analysed a high-throughput fitness landscape of the *folA* gene in *E. coli* and found that only a few mutations exhibited global epistasis[64]. Therefore, although a negative correlation between fitness effects and background fitness is common, it is not an inevitable occurrence.

Our model captures this feature of epistasis, and provides an insight into the predictability of negative global epistasis. If a mutation occurs in a set of very similar genetic backgrounds, it is highly likely that the fitness effects of this mutation exhibit a strong negative correlation with background fitness. This likelihood of observing strong negative global epistasis decreases as the variability among the genetic backgrounds, in which the mutation occurs, increases. However, it must be noted that negative global epistasis is observed independent of the extent of variability in a group of cells. Therefore, the first important result of our work is that declining adaptability is a highly likely outcome of evolution in an environment. But if the starting populations differed qualitatively in the profiles of their essential metabolites, the likelihood of observing negative global epistasis decreases, there could be idiosyncratic epistatic interactions, and consequently, the likelihood of declining adaptability also decreases.

After we could qualitatively comment on the likelihood of observing global epistasis, we sought a mechanistic explanation of global epistasis. In our cell growth and division model, the effects of mutations manifest as changes in the metabolic fluxes. We ask if the likelihood of the occurrence of global epistasis can be explained by the flux through a module, and the extent of metabolic control a module exerts over the others in the cell. We do this by creating toy cells with predefined speeds of metabolic steps. Given the metabolic architecture of the cells that we model, we see that fitness effects of mutations in the slowest reaction step, which has minimal control over other fluxes (2A condition), correlate strongly with the background fitness. This result was applicable even when mutations were introduced in a cell with altered metabolic architecture (Figure 12). Therefore, an important finding of this work is that metabolic control dictates the likelihood of epistasis.

A potential explanation for this observation is, because the substrate is common to the three (or two) first step reactions, mutations in any first step reaction alters the fluxes through all the three (or two) reaction branches in the cell that we model. That is not the case with second step

452 reactions. Therefore, mutations in the rate-limiting step that has little control over other
453 reactions, which is the case 2A, exhibit strong global epistasis[65]. We verified this hypothesis
454 by simulating cells with feedback mechanisms – once a threshold metabolite is produced in
455 sufficient quantities, the corresponding metabolic pathway was forced to shut down. This way,
456 there is reduced control of a first step reaction on the flux through the other modules. As shown
457 in **Supplementary Figure 5**, mutations in 1A and 2A exhibit strong global epistasis, when
458 there is feedback in the cell. Therefore, incorporating feedback in the cells that we model does
459 not change the results by a whole lot – it is still mutations in the less controlled, slow, reaction
460 steps which exhibit strong global epistasis.

461 The distribution of fitness effects of mutations is a measure of how evolvable a genotype is[66].
462 The mean of the DFE is indicative of the average fitness effect of a mutation on a genetic
463 background. Higher the mean of the DFE, greater is the proportion of beneficial mutations.
464 Like we were able to predict the strength of global epistasis using the metabolic flux, we asked
465 if the DFE also shows predictable patterns based on the flux. In all the four types of toy cells
466 that we simulated (1A, 1B, 2A and 2B), the mean of the DFE decreased linearly with an
467 increase in background fitness, as observed in empirical studies in the past. The R^2 of the linear
468 fit between means of the DFE and background fitness, however, varied with the type of module
469 in which the mutations occurred. The mean of the DFE is well-predicted by the background
470 fitness in the module in which strong global epistasis was observed, i.e., case 2A. Incorporating
471 feedback manifests as greater predictability of the distribution of mutational effects in case 1A,
472 as shown in **Supplementary Figure 6**. However, the fitness effects of mutations in the fast
473 reaction steps do not exhibit strong global epistasis.

474 The fitness of a cell depends on the rate of the reactions, and is dictated by how metabolic
475 networks are wired. When there is only one slow reaction in the cell, its division time depends
476 on this slow reaction, as opposed to the rates of all reactions in the cell, i.e., fitness is directly
477 proportional to one of the reaction rates. This results in a decrease in the "interactions". This
478 linear relationship results in a negative dependence of the fitness effect of a mutation with
479 background fitness, which is negative global epistasis. Evolution acts in a step-wise manner to
480 improve each of the slow reactions in the cell, starting with the slowest one[67]. As a result,
481 declining adaptability is a common observation in laboratory evolution experiments.

482 Positive global epistasis, or increasing-returns epistasis, occurs when a beneficial mutation's
483 fitness effects increase with an increase in background fitness, indicating synergistic effects

484 between the mutations[68–70]. Our work indicates that such a scenario is most likely when there
485 is a mutation in a highly interconnected module, such as a first step reaction in our model,
486 allowing for a significant fitness gain in one evolutionary step. For instance, if one of the three
487 first step reactions is much faster than the others, mutations that slow does this reaction to
488 match the speeds of the other first step reactions would be beneficial. In this scenario, there is
489 an inverse relationship of the fitness with reaction rate of the fast step. However, fitness gains
490 depend non-linearly on the changes caused to the other reaction arms due to mutations in the
491 first reaction. Therefore, the second important result from our work, is the metabolic basis of
492 global epistasis, and its modulation of the predictability of the distribution of mutational effects.

493 The findings of our work are easily experimentally verifiable using synthetically engineered
494 prokaryotic cells. For example, when *E. coli* is grown in anaerobic conditions, it uses acetyl
495 CoA as a substrate for Krebs cycle, fatty acid biosynthesis, and mixed acid fermentation[71–73].
496 By carefully introducing mutations in the enzymatic pathways involved in these three reaction
497 chains, it is possible to empirically characterize fitness effects of mutations and study their
498 predictability. It was also recently shown that global epistasis helps predict function of
499 microbial communities[74]. This model of cell growth and division has been used to understand
500 metabolic trade and auxotrophy in naturally occurring microbial communities[56]. Therefore, it
501 would be an interesting exercise to extend this model to understand the metabolic basis of
502 predictability in community interactions.

503 Overall, this work is a novel attempt at modelling cell growth and division based on first
504 principles of biology and explaining statistical observations regarding global epistasis. Further
505 work is required to understand how environment and altered metabolic structures modulate
506 global epistasis at the cellular level.

References

1. Wang, L. & Wang, H. Marine Biogeochemical Processes and Ecosystem Evolution: Observational and Predictive Approaches. *ijms* (2024) doi:10.5376/ijms.2024.14.0034.
2. Omenn, G. S. Evolution and public health. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1702–1709 (2010).
3. Thrall, P. H. *et al.* Evolution in agriculture: the application of evolutionary approaches to the management of biotic interactions in agro-ecosystems. *Evolutionary Applications* **4**, 200–215 (2011).
4. Mackay, T. F. & Moore, J. H. Why epistasis is important for tackling complex human disease genetics. *Genome Med* **6**, 42 (2014).
5. Tang, D., Freudenberg, J. & Dahl, A. Factorizing polygenic epistasis improves prediction and uncovers biological pathways in complex traits. *The American Journal of Human Genetics* **110**, 1875–1887 (2023).
6. Bateson, W., Mendel, G. & Leighton, A. G. *Mendel's Principles of Heredity, by W. Bateson*. (University Press, Cambridge [Eng.], 1909). doi:10.5962/bhl.title.1057.
7. Østman, B., Hintze, A. & Adami, C. Impact of epistasis and pleiotropy on evolutionary adaptation. *Proceedings of the Royal Society B: Biological Sciences* **279**, 247–256 (2012).
8. Lehner, B. Molecular mechanisms of epistasis within and between genes. *Trends in Genetics* **27**, 323–331 (2011).
9. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**, 855–867 (2008).
10. Fenster, C. B., Galloway, L. F. & Chao, L. Epistasis and its consequences for the evolution of natural populations. *Trends in Ecology & Evolution* **12**, 282–286 (1997).
11. Kojima, K. & Lewontin, R. C. Evolutionary Significance of Linkage and Epistasis. in *Mathematical Topics in Population Genetics* (ed. Kojima, K.) 367–388 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1970). doi:10.1007/978-3-642-46244-3_12.
12. Weinreich, D. M., Watson, R. A. & Chao, L. PERSPECTIVE: SIGN EPISTASIS AND GENETIC COSTRAINT ON EVOLUTIONARY TRAJECTORIES. *Evolution* **59**, 1165–1174 (2005).

13. Diaz-Colunga, J. *et al.* Global epistasis on fitness landscapes. *Philosophical Transactions of the Royal Society B* **378**, 20220053 (2023).
14. Johnson, M. S., Reddy, G. & Desai, M. M. Epistasis and evolution: recent advances and an outlook for prediction. *BMC Biol* **21**, 120 (2023).
15. Reddy, G. & Desai, M. M. Global epistasis emerges from a generic model of a complex trait. *eLife* **10**, e64740 (2021).
16. Bank, C. Epistasis and Adaptation on Fitness Landscapes. *Annu. Rev. Ecol. Evol. Syst.* **53**, 457–479 (2022).
17. Weinreich, D. M., Delaney, N. F., DePristo, M. A. & Hartl, D. L. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science* **312**, 111–114 (2006).
18. Jayaraman, V., Toledo-Patiño, S., Noda-García, L. & Laurino, P. Mechanisms of protein evolution. *Protein Sci* **31**, e4362 (2022).
19. Pokusaeva, V. O. *et al.* An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. *PLoS Genet* **15**, e1008079 (2019).
20. Good, B. H. & Desai, M. M. The Impact of Macroscopic Epistasis on Long-Term Evolutionary Dynamics. *Genetics* **199**, 177–190 (2015).
21. Wiser, M. J., Ribeck, N. & Lenski, R. E. Long-Term Dynamics of Adaptation in Asexual Populations. *Science* **342**, 1364–1367 (2013).
22. Kassen, R. & Bataillon, T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat Genet* **38**, 484–488 (2006).
23. Levy, S. F. *et al.* Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).
24. Domingo-Calap, P., Cuevas, J. M. & Sanjuán, R. The Fitness Effects of Random Mutations in Single-Stranded DNA and RNA Bacteriophages. *PLoS Genet* **5**, e1000742 (2009).
25. Aggeli, D., Li, Y. & Sherlock, G. Changes in the distribution of fitness effects and adaptive mutational spectra following a single first step towards adaptation. *Nat Commun* **12**, 5193 (2021).
26. Sane, M., Diwan, G. D., Bhat, B. A., Wahl, L. M. & Agashe, D. Shifts in mutation spectra enhance access to beneficial mutations. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2207355120 (2023).

27. Bank, C., Matuszewski, S., Hietpas, R. T. & Jensen, J. D. On the (un)predictability of a large intragenic fitness landscape. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14085–14090 (2016).
28. Lässig, M., Mustonen, V. & Walczak, A. M. Predicting evolution. *Nat Ecol Evol* **1**, 0077 (2017).
29. Couce, A. & Tenaillon, O. A. The rule of declining adaptability in microbial evolution experiments. *Front. Genet.* **6**, 99 (2015).
30. Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D. & Marx, C. J. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011).
31. Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522 (2014).
32. Johnson, M. S., Martsul, A., Kryazhimskiy, S. & Desai, M. M. Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science* **366**, 490–493 (2019).
33. Segrè, A. V., Murray, A. W. & Leu, J.-Y. High-Resolution Mutation Mapping Reveals Parallel Experimental Evolution in Yeast. *PLoS Biol* **4**, e256 (2006).
34. Szathmáry, E. Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. *Genetics* **133**, 127–132 (1993).
35. Kryazhimskiy, S. Emergence and propagation of epistasis in metabolic networks. *eLife* **10**, e60200 (2021).
36. Maclean, R. C. Predicting epistasis: an experimental test of metabolic control theory with bacterial transcription and translation. *Journal of Evolutionary Biology* **23**, 488–493 (2010).
37. Dean, A. M., Dykhuizen, D. E. & Hartl, D. L. Fitness as a function of beta-galactosidase activity in Escherichia coli. *Genet Res* **48**, 1–8 (1986).
38. Lagator, M., Paixão, T., Barton, N. H., Bollback, J. P. & Guet, C. C. On the mechanistic nature of epistasis in a canonical cis-regulatory element. *eLife* **6**, e25192 (2017).
39. Ortlund, E. A., Bridgham, J. T., Redinbo, M. R. & Thornton, J. W. Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis. *Science* **317**, 1544–1548 (2007).
40. Otten, R. *et al.* How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science* **370**, 1442–1446 (2020).

590 41. Sethi, A., Eargle, J., Black, A. A. & Luthey-Schulten, Z. Dynamical networks in tRNA:protein
591 complexes. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6620–6625 (2009).
592 42. Campbell, E. C. *et al.* Laboratory evolution of protein conformational dynamics. *Current Opinion*
593 *in Structural Biology* **50**, 49–57 (2018).
594 43. Jiménez-Osés, G. *et al.* The role of distant mutations and allosteric regulation on LovD active site
595 dynamics. *Nat Chem Biol* **10**, 431–436 (2014).
596 44. Fröhlich, C. *et al.* Epistasis arises from shifting the rate-limiting step during enzyme evolution of
597 a β-lactamase. *Nat Catal* **7**, 499–509 (2024).
598 45. Cisneros, A. F. *et al.* Epistasis between promoter activity and coding mutations shapes gene
599 evolvability. *Sci Adv* **9**, eadd9109 (2023).
600 46. González-González, A., Batarseh, T. N., Rodríguez-Verdugo, A. & Gaut, B. S. Patterns of Fitness
601 and Gene Expression Epistasis Generated by Beneficial Mutations in the *rho* and *rpoB* Genes of
602 *Escherichia coli* during High-Temperature Adaptation. *Molecular Biology and Evolution* **41**,
603 msae187 (2024).
604 47. Chou, H.-H., Delaney, N. F., Draghi, J. A. & Marx, C. J. Mapping the Fitness Landscape of Gene
605 Expression Uncovers the Cause of Antagonism and Sign Epistasis between Adaptive Mutations.
606 *PLoS Genet* **10**, e1004149 (2014).
607 48. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-Time Kinetics of Gene Activity in
608 Individual Bacteria. *Cell* **123**, 1025–1036 (2005).
609 49. Chen, H., Shiroguchi, K., Ge, H. & Xie, X. S. Genome-wide study of mRNA degradation and
610 transcript elongation in Escherichia coli. *Mol Syst Biol* **11**, 781 (2015).
611 50. Moran, M. A. *et al.* Sizing up metatranscriptomics. *ISME J* **7**, 237–243 (2013).
612 51. Gong, X. *et al.* Comparative analysis of essential genes and nonessential genes in Escherichia coli
613 K12. *Mol Genet Genomics* **279**, 87–94 (2008).
614 52. Dennis, P. P. & Nomura, M. Stringent control of ribosomal protein gene expression in Escherichia
615 coli. *Proc Natl Acad Sci U S A* **71**, 3819–3823 (1974).
616 53. Dennis, P. P. & Bremer, H. Differential rate of ribosomal protein synthesis in Escherichia coli B/r.
617 *J Mol Biol* **84**, 407–422 (1974).

618 54. Young, R. & Bremer, H. Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of
619 growth rate. *Biochem J* **160**, 185–194 (1976).
620 55. Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends
621 Shaping Enzyme Parameters. *Biochemistry* **50**, 4402–4410 (2011).
622 56. Dutta, D. & Saini, S. Cell Growth Model with Stochastic Gene Expression Helps Understand the
623 Growth Advantage of Metabolic Exchange and Auxotrophy. *mSystems* **6**,
624 10.1128/msystems.00448-21 (2021).
625 57. Diaz-Colunga, J., Sanchez, A. & Ogbunugafor, C. B. Environmental modulation of global
626 epistasis in a drug resistance fitness landscape. *Nat Commun* **14**, 8055 (2023).
627 58. Ardell, S., Martsul, A., Johnson, M. S. & Kryazhimskiy, S. Environment-independent distribution
628 of mutational effects emerges from microscopic epistasis. *bioRxiv* 2023.11.18.567655 (2024)
629 doi:10.1101/2023.11.18.567655.
630 59. Judge, A. & Dodd, M. S. Metabolism. *Essays Biochem* **64**, 607–647 (2020).
631 60. Wünsche, A. *et al.* Diminishing-returns epistasis decreases adaptability along an evolutionary
632 trajectory. *Nat Ecol Evol* **1**, 0061 (2017).
633 61. Jerison, E. R. *et al.* Genetic variation in adaptability and pleiotropy in budding yeast. *eLife* **6**,
634 e27167 (2017).
635 62. Smith, C. E., Smith, A. N. H., Cooper, T. F. & Moore, F. B.-G. Fitness of evolving bacterial
636 populations is contingent on deep and shallow history but only shallow history creates predictable
637 patterns. *Proc Biol Sci* **289**, 20221292 (2022).
638 63. Rojas Echenique, J. I., Kryazhimskiy, S., Nguyen Ba, A. N. & Desai, M. M. Modular epistasis
639 and the compensatory evolution of gene deletion mutants. *PLoS Genet* **15**, e1007958 (2019).
640 64. Baheti, S., Raj, N. & Saini, S. Variations and predictability of epistasis on an intragenic fitness
641 landscape. Preprint at https://doi.org/10.7554/eLife.104848.1 (2025).
642 65. Bruggeman, F. J. *et al.* Whole-cell metabolic control analysis. *Biosystems* **234**, 105067 (2023).
643 66. Payne, J. L. & Wagner, A. The causes of evolvability and their evolution. *Nat Rev Genet* **20**, 24–
644 38 (2019).

645 67. Venkataram, S., Monasky, R., Sikaroodi, S. H., Kryazhimskiy, S. & Kacar, B. Evolutionary
646 stalling and a limit on the power of natural selection to improve a cellular module. *Proc. Natl.*
647 *Acad. Sci. U.S.A.* **117**, 18582–18590 (2020).
648 68. Chou, H.-H., Berthet, J. & Marx, C. J. Fast growth increases the selective advantage of a
649 mutation arising recurrently during evolution under metal limitation. *PLoS Genet* **5**, e1000652
650 (2009).
651 69. Hsieh, Y.-Y. P., Makrantoni, V., Robertson, D., Marston, A. L. & Murray, A. W. Evolutionary
652 repair: Changes in multiple functional modules allow meiotic cohesin to support mitosis. *PLoS*
653 *Biol* **18**, e3000635 (2020).
654 70. Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key
655 innovation in an experimental population of Escherichia coli. *Proc Natl Acad Sci U S A* **105**,
656 7899–7906 (2008).
657 71. Krivoruchko, A., Zhang, Y., Siewers, V., Chen, Y. & Nielsen, J. Microbial acetyl-CoA metabolism
658 and metabolic engineering. *Metabolic Engineering* **28**, 28–42 (2015).
659 72. Janßen, H. & Steinbüchel, A. Fatty acid synthesis in Escherichia coli and its applications towards
660 the production of fatty acid based biofuels. *Biotechnol Biofuels* **7**, 7 (2014).
661 73. El-Mansi, E. M. T. & Holms, W. H. Control of Carbon Flux to Acetate Excretion During Growth
662 of Escherichia coli in Batch and Continuous Cultures. *Microbiology* **135**, 2875–2883 (1989).
663 74. Diaz-Colunga, J., Skwara, A., Vila, J. C. C., Bajic, D. & Sanchez, A. Global epistasis and the
664 emergence of function in microbial consortia. *Cell* **187**, 3108-3119.e30 (2024).
665