The Genetic Landscape of a Metabolic Interaction

Thuy N. Nguyen[1,2,3,†], Christine Ingle[1,2,3], Samuel Thompson[4, ‡], Kimberly A. Reynolds[1,2,3]

[1]The Green Center for Systems Biology, [2]Department of Bioinformatics, and [3]Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, USA, 75390

[4]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA 94158

[†]Present address: Form Bio, Dallas, USA, 75226

[‡]Present address: Department of Bioengineering, Stanford University, Stanford, CA 94305

* Correspondence: kimberly.reynolds@utsouthwestern.edu (K.A. Reynolds)

Keywords: Dihydrofolate Reductase, DHFR, Thymidylate Synthase, TYMS, deep mutational scan, epistasis, genotype-to-phenotype, genetic interaction

Abstract: Enzyme abundance, catalytic activity, and ultimately sequence are all shaped by the need of growing cells to maintain metabolic flux while minimizing accumulation of deleterious intermediates. To quantify how variation in the activity of one enzyme constrains the biochemical parameters and sequence of another, we focused on dihydrofolate reductase (DHFR) and thymidylate synthase (TYMS), a pair of enzymes catalyzing consecutive reactions in folate metabolism. We used deep mutational scanning to quantify the growth rate effect of 2,696 DHFR single mutations in 3 TYMS backgrounds and show that our data are well-described by a relatively simple enzyme velocity to growth rate model. From the data and model we estimate the approximate effects of all single mutations on DHFR catalytic power. Together our data provide a comprehensive view of epistasis between mutations in a biochemically linked enzyme pair, reveal the structural distribution of positions tuning DHFR catalysis, and establish a foundation for the design of multi-enzyme systems.

INTRODUCTION

Enzymes function within biochemical pathways; exchanging substrates and products to generate useful metabolites. This metabolic context constrains enzyme velocity — the product of both catalytic activity and enzyme abundance. For example, the relative velocities of some enzymes must be coordinated to avoid accumulation of deleterious metabolic intermediates[1–3]. In other instances, optimal enzyme abundance is set by a tradeoff between the cost of protein synthesis and the benefit of efficient nutrient utilization[4–6]. Considered at the pathway scale, metabolic enzymes are often produced in evolutionarily conserved stoichiometric ratios across species[7], providing further indication that relative — not just absolute — enzyme velocity is under selection. More generally, the relationship between the velocity of a given enzyme, metabolic flux through a pathway, and cellular growth rate is non-linear and shaped by interactions amongst pathway enzymes (**Fig.1a**). Indeed, a key result of metabolic control theory is that the control coefficient of an enzyme — defined as the fractional change in pathway-level flux given a fractional change in enzyme velocity — depends on the starting (native) velocity of the enzyme, but *also* the velocity of all other enzymes in the pathway[8,9]. That is to say, given that enzymes act sequentially to produce metabolites, the effects of mutations on cellular phenotype can be buffered or amplified depending on which enzymatic reactions control metabolic flux. As a consequence, enzyme mutations that are neutral in one context may have profound consequences for metabolic flux and growth rate in the background of variation in another[10–13]. This context-dependence, or epistasis, amongst metabolic enzymes need not be mediated by direct physical binding, but emerges indirectly through shared metabolite pools and a need to maintain flux while avoiding the accumulation of deleterious intermediates[6,11,14].

While much prior work has explored the constraints on protein sequence and evolution induced by physical protein-protein interactions, the sequence-level constraints emerging from these sorts of non-binding functional interactions in metabolism remain unclear. How is this “indirect” epistasis organized in the protein structure and reflected in the sequence? A quantitative

66 understanding of how pathway context shapes sequence and activity would assist in the
67 interpretation of disease-associated mutations, the design of new enzymes, and directing the
68 laboratory evolution of metabolic pathways. To begin to address this, we examined the residue-
69 level epistatic interactions between a pair of enzymes that catalyze consecutive reactions in folate
70 metabolism: dihdyrofolate reductase (DHFR) and thymidylate synthase (TYMS). The activity of
71 these enzymes is strongly linked to *E. coli* growth rate, they are frequent targets of antibiotics
72 and chemotherapeutics, and our prior work showed that they co-evolve as a module both in the
73 laboratory and across thousands of bacterial genomes[1]. Taking this enzyme pair as a simplified
74 model system in which to examine a metabolic interaction, we created a mathematical model
75 relating variation in DHFR and TYMS catalytic parameters to growth rate using a focused set of
76 well-characterized point mutants. Then, to more deeply test this model and comprehensively map
77 the pattern of epistasis between these two enzymes, we measured the effect of nearly all possible
78 DHFR single mutations (2,696 in total) in the context of three TYMS variants selected to span a
79 range of catalytic activities. The model predicted – and the data showed – that TYMS background
80 profoundly changed both the sign (buffering vs. aggravating) and magnitude of DHFR epistasis.
81 Mapping the epistatic effects of mutation to the DHFR tertiary structure revealed that they are
82 organized into distinct clusters based on epistatic sign. Additionally, mutations with the largest
83 magnitude epistatic effect to TYMS centered around the DHFR active site, while more weakly
84 epistatic positions radiated outwards. Finally, we inferred approximate values for DHFR catalytic
85 power (k_{cat}/K_m) across all 2,696 mutations by using the growth rate measurements across TYMS
86 backgrounds to constrain the enzyme velocity to growth rate model. The residues linked to
87 catalysis form a structurally distributed network inside the enzyme and are highly evolutionarily
88 conserved. Taken together, our data demonstrates at single-residue resolution how epistasis
89 mediated through a metabolic interaction reshapes a mutational landscape. The enzyme velocity
90 to growth rate model opens the door to extracting quantitative biochemical parameters from
91 growth rate data, and invites one to consider new ideas for the design of multi-enzyme systems.

92

93 **RESULTS**

94 **An enzyme velocity to growth rate model for DHFR and TYMS**

95 Folate metabolism is a well-conserved biochemical pathway involved in the synthesis of purine
96 nucleotides, thymidine, glycine, and methionine[15] (**Fig. 1b**). Consequently, this pathway is
97 strongly linked to cell growth and a frequent target of antibiotics and chemotherapeutics. DHFR
98 and TYMS play a central role in folate metabolism. DHFR reduces dihydrofolate (DHF) to
99 tetrahydrofolate (THF) using NADPH as a cofactor. THF then serves as a carrier for activated
100 one-carbon units in downstream metabolic processes. TYMS catalyzes the oxidation of THF back
101 to DHF during deoxythymidine synthesis and is the sole enzyme responsible for recycling the
102 DHF pool[16,17]. Prior work from ourselves and others indicates that these two enzymes are strongly
103 functionally coupled to each other and less coupled to the remainder of the pathway: they co-
104 evolve in terms of synteny and gene presence-absence across bacterial species[1], inhibition of
105 DHFR with trimethoprim is rescued by suppressor mutations in TYMS in both the lab and the
106 clinic[1,18], and loss-of-function mutations in DHFR are rescued by loss-of-function mutations in
107 TYMS[1,19,20]. Metabolomics data indicated that loss of DHFR function results in accumulation of
108 DHF and depletion of reduced folates; compensatory loss of function mutations in TYMS help to
109 restore DHF and THF pools to more native-like levels[1,21,22]. Thus, DHFR and TYMS are a growth-
110 linked two-enzyme system where epistasis is driven by a biochemical interaction, with the added
111 simplification that they are relatively decoupled from surrounding metabolic context.

112 We sought to create a mathematical model relating changes in DHFR and TYMS catalytic
113 parameters to growth rate phenotype with the goals of (1) formalizing our previous empirical
114 findings, (2) quantifying the constraints on DHFR and TYMS catalytic activities, and (3) defining
115 the relationship between biochemical activity, epistatic sign, and epistatic magnitude. We
116 developed our model using a previously collected set of metabolomics and growth rate data for
117 five DHFR point mutants in the background of both WT TYMS and TYMS R166Q, a near-

118 catalytically-inactive point mutant (**Table S1, Table S2**). First, we considered the relationship
119 between intracellular THF abundance and growth rate as measured across all ten DHFR/TYMS
120 sequence combinations. THF limitation restricts the production of several growth-linked factors,
121 including thymidine, methionine, glycine, and the purine precursors inosine and AICAR. Under
122 the experimental conditions of our growth rate assays — M9 minimal media with 0.4% glucose,
123 0.2% amicase, and 50 μg/ml thymidine — thymidine is not growth limiting (TYMS R166Q is
124 rescued to WT-like growth) and amicase provides a source of free amino acids. Thus we reason
125 that growth rate is principally dependent upon purine production in our experiments. In any case,
126 we previously observed a hyperbolic dependence of growth rate on reduced folate abundance for
127 many THF species in these experimental conditions[1]. We selected 10-formyl THF with three
128 glutamates as a representative growth-linked reduced folate given it's clear relationship to growth
129 and proximity to purine biosynthesis. Following a similar approach as Rodrigues et al, we fit a
130 single four-parameter sigmoidal function relating growth rate to intracellular THF concentration[23]
131 (**Fig. 1c**).

132 Eqn. 1 $$g = \frac{g_{max} - g_{min}}{1 + (K/[THF])^n} + g_{min}$$

133 Here, g_{max} represents the maximal growth rate, g_{min} is the minimal growth rate, K is a constant
134 that captures the concentration of THF that yields 50% growth, and n is a Hill coefficient (**Table**
135 **S3**).

136 Next, we sought to connect variation in DHFR and TYMS enzyme velocity to intracellular
137 THF concentrations. To simplify our model, we reduced the pathway to a cycle in which DHFR
138 and TYMS catalyze opposing oxidation and reduction reactions (**Fig. 1d**). This abstraction
139 assumes that DHFR and TYMS dominate turnover of the DHF and THF pools, and that the
140 reduced folates are considered as a single THF pool. While this simplification clearly omits much
141 of folate metabolism, it allows us to write a rate equation that isolates the recycling of THF in
142 terms of a small number of measurable biochemical parameters:

Eqn. 2 $$\frac{d[\mathrm{THF}]}{dt} = \frac{[\mathrm{DHFR}] * k_{cat}^{\mathrm{DH}}}{1 + \mathrm{K}_m^{\mathrm{DH}}/([\mathrm{fol_{tot}}] - [\mathrm{THF}])} - \frac{[\mathrm{TYMS}] * k_{cat}^{\mathrm{TS}}}{1 + \mathrm{K}_m^{\mathrm{TS}}/([\mathrm{THF}])}$$

In this equation, DHFR and TYMS are treated as catalyzing opposing reactions with Michaelis Menten kinetics, providing a relationship between steady state kinetics parameters (k_{cat}^{DH}, $\mathrm{K}_m^{\mathrm{DH}}$, k_{cat}^{TS}, $\mathrm{K}_m^{\mathrm{TS}}$) and intracellular THF abundance. From this equation one can find an analytical solution for the steady state concentration of THF in the form of the Goldbeter-Koshland equation [24,25].

Eqn. 3 $$\frac{[\mathrm{THF}_{ss}]}{[\mathrm{fol_{tot}}]} = \frac{\frac{V_1}{V_2} * (1 - \hat{K}_{m2}) - \hat{K}_{m1} - 1 + \sqrt{4\hat{K}_{m2}\frac{V_1}{V_2}(\frac{V_1}{V_2} - 1) + (\frac{V_1}{V_2}(\hat{K}_{m2} - 1) + \hat{K}_{m1} + 1)^2}}{2 * (\frac{V_1}{V_2} - 1)}$$

Where:

$$V_1 = [\mathrm{DHFR}]k_{cat}^{\mathrm{DH}} \qquad \hat{K}_{m1} = \mathrm{K}_m^{\mathrm{DH}}/\mathrm{fol_{tot}}$$

$$V_2 = [\mathrm{TYMS}]k_{cat}^{\mathrm{TS}} \qquad \hat{K}_{m2} = \mathrm{K}_m^{\mathrm{TS}}/\mathrm{fol_{tot}}$$

In our initial model construction, the steady state catalytic parameters (k_{cat}^{DH}, $\mathrm{K}_m^{\mathrm{DH}}$, k_{cat}^{TS}, $\mathrm{K}_m^{\mathrm{TS}}$) were experimentally measured *in vitro* using purified samples of all mutants, with the exception of TYMS R166Q which is near-inactive and assigned an arbitrarily low k_{cat} and high $\mathrm{K_m}$ (**Table S1,S2**). Four fit parameters remain in Equation 3: (1) the concentration of the total folate pool ($[\mathrm{fol_{tot}}]$) (2) the intracellular concentration of DHFR ($[\mathrm{DHFR}]$), which we treated as identical across all variants (as our model will eventually describe thousands of DHFR mutations, and we wish to avoid overparameterization), (3) the intracellular concentration of WT TYMS ($[\mathrm{TYMS_{WT}}]$), and (4) the intracellular concentration of TYMS R166Q ($[\mathrm{TYMS_{R166Q}}]$). This relatively simplified model showed good correspondence to the data when fit (R^2 = 0.96, **Fig. 1e, Table S3**). Equations 1 and 3 were then combined to estimate growth rate as a function of both DHFR and TYMS activity,

163 by linking catalytic activity to THF abundance, and then THF abundance to growth rate. The
164 complete model worked well to predict growth rate on our initial training set (**Fig. 1f**).
165
166 **The growth rate effect of DHFR mutations changes magnitude and sign depending upon**
167 **TYMS background.**
168 To more rigorously test our model and understand its' predictions, we expanded our dataset to
169 include more DHFR and TYMS variants with experimentally characterized activities. As our initial
170 model was developed using only two extreme TYMS variants (wild-type and a near complete loss
171 of function variant, R166Q), we were particularly curious to evaluate model performance for TYMS
172 mutations with intermediate effects on catalysis and *E. coli* growth. We identified candidate TYMS
173 mutations by examining an earlier growth complementation study[26]. A handful of these mutants
174 were then cloned, screened for expression, and when possible, purified and characterized.
175 Through this mini-screen we selected two mutations that stably expressed, purified robustly, and
176 yielded intermediate activities: TYMS R127A and Q33S (**Fig. 2a**). The R127A mutation is located
177 in the TYMS active site and is one of four arginines that coordinate the substrate (dUMP)
178 phosphate group. The Q33S mutation is located at the TYMS dimer interface, distal to the active
179 site. We observed that R127A was more deleterious to catalytic function than Q33S, but that both
180 mutations were more active than R166Q (which shows almost no measurable activity *in vitro*, **Fig.**
181 **2b, Table S2**).
182 We measured growth rates for seven catalytically characterized DHFR variants (a set of
183 single and double mutants selected to span a range of catalytic activities) in the background of
184 these four TYMS mutants (WT, R127A, Q33S and R166Q) using a plate-reader-based assay (28
185 measurements total, each in triplicate; **Fig. 2c**, **Fig S1a,c**). We used this focused dataset to re-
186 parameterize the model equations, this time fitting five total parameters
187 ($[\mathrm{fol_{tot}}], [\mathrm{DHFR}], [\mathrm{TYMS_{WT}}], [\mathrm{TYMS_{Q33S}}], [\mathrm{TYMS_{R127A}}], [\mathrm{TYMS_{R166Q}}]$, **Table S3**). This second
188 round of fitting tested the ability of growth rate data only (in the absence of metabolomics

189 measurements) to constrain the model, and the capacity of the model to capture TYMS mutations
190 with intermediate effects on activity. The data were again well described by the model (**Fig. 2C,**
191 **Fig. S1b,d**). As a control for overfitting, we tested the ability of the model to predict growth rates
192 for arbitrary catalytic data. We randomly shuffled the catalytic parameters (k_{cat} and K_m) among
193 mutations for both DHFR and TYMS, refit all free model parameters, and calculated the RMSD
194 and R^2 values between the best fit model and the shuffled data. Importantly, the model was
195 generally unable to describe the experimental growth rate data when catalytic parameters were
196 shuffled across both DHFR and TYMS (**Fig S1e,f**). This indicated that the model provided a
197 specific description of our experiment and was not trivially overfit. The model was less sensitive
198 to shuffling TYMS catalytic parameters (presumably because we included fit parameters
199 describing the abundance of each TYMS mutation that can compensate for this shuffling, **Fig.**
200 **S1h**). However, it was strongly sensitive to shuffling DHFR parameters (**Fig. S1g**). Taken
201 together, this analysis indicated that the model provides a good description of the enzyme-
202 velocity-to-growth-rate relationship and can be used to predict and interpret how molecular
203 changes in DHFR and TYMS activity modulate growth rate phenotype.

204 As in previous work, we observed that decreasing DHFR activity was deleterious to growth
205 rate, and that loss-of-function mutations in DHFR can be partly or even entirely rescued by the
206 loss-of-function mutation TYMS R166Q (**Fig. S1a,c**). TYMS R127A, a less severe loss of function
207 mutation, showed a similar albeit more modest trend – this mutation was able to partly rescue
208 growth for some (though not all) DHFR mutations. As indicated by both our prior experimental
209 data and the model, reducing TYMS activity alongside DHFR loss-of-function prevents the
210 accumulation of DHF and depletion of THF pools, consequently preserving growth in thymidine-
211 supplemented media. Stated otherwise, loss-of-function mutations in TYMS help to preserve
212 reduced folate pools, allowing THF to shuttle one-carbon units in downstream biochemical
213 processes like purine biosynthesis. Thus, the TYMS R166Q and R127A variants show positive
214 (buffering) epistasis to low-activity DHFR mutations. In contrast to our expectation that a more

215 intermediate mutation would also demonstrate intermediate levels of buffering epistasis, TYMS
216 Q33S shows negative (or amplifying) epistasis to some DHFR mutations. This means that these
217 DHFR mutations are more deleterious in the background of TYMS Q33S than in the native TYMS
218 context. Our model accounted for this observation by increasing the intracellular concentration of
219 TYMS Q33S (a fit parameter, **Table S3**) such that the effective k_{cat} of TYMS Q33S is greater than
220 wildtype ($[\mathrm{TYMS_{Q33S}}]k_{cat}^{\mathrm{TYMS_Q33S}} > [\mathrm{TYMS_{WT}}]k_{cat}^{\mathrm{TYMS_WT}}$). This in turn increased the intracellular
221 requirement for DHFR activity, resulting in negative epistasis.

222 To further explore the pattern of epistasis across TYMS backgrounds, we simulated
223 growth rates over a range of DHFR k_{cat} and K_m values in each TYMS background (**Fig. 2d**). This
224 provided a comprehensive prediction of the TYMS-induced constraints on DHFR activity. In
225 particular, we obtained a regime of DHFR k_{cat} and K_m values that is sufficient to support growth
226 for each TYMS mutation. From these data we computed epistasis. These results indicated that
227 TYMS Q33S has negative epistasis to DHFR variants spanning a well-defined band of catalytic
228 parameters. The simulations also indicated that R127A has weak positive epistasis over a regime
229 of moderately impaired DHFR variants, but is insufficient to rescue growth for the strongest loss
230 of function variants. Finally, TYMS R166Q was observed to be broadly rescuing; DHFR variants
231 need only a negligible amount of activity to support growth in this context. Thus, our simulations
232 show that the sign and magnitude of DHFR epistasis are strongly tuned by TYMS background.
233 The model thus provides quantitative predictions of the catalytic regimes where epistasis will be
234 most apparent.

235

236 **The single-mutant landscape of DHFR is strongly modulated by TYMS context**

237 Next we wanted to examine if these observations — negative epistasis for Q33S and broadly
238 positive epistasis for R166Q — held true across a larger dataset. Additionally, we wanted to
239 characterize the structural pattern of biochemical epistasis at the residue level across DHFR. To

240 accomplish this, we created a plasmid-based saturation mutagenesis library of DHFR containing
241 all possible single mutations at every position (3002 total). This library was subcloned into all three
242 TYMS backgrounds; sequencing showed that these libraries are well-distributed and approach
243 full coverage of all single mutations (97.1% - WT TYMS, 94.6% - TYMS Q33S, 99.3% - TYMS
244 R166Q) (**Fig. S2**). We transformed these libraries into *E. coli* lacking the genes encoding DHFR
245 and thymidylate synthase (ER2566 *ΔfolA ΔthyA*). Transformants for each library were then grown
246 as a mixed population in selective media (M9 minimal media with 0.4% glucose, 0.2% amicase,
247 and 50 μg/ml thymidine) in a turbidostat to ensure maintenance of exponential growth and
248 constancy of media conditions. We took six time points over the course of 24 hours, and we
249 prepared these samples for next generation sequencing. By examining the change in the relative
250 frequency of individual mutant counts over time, we computed the growth rate difference relative
251 to WT DHFR for nearly all mutations in the library (**Fig. 3a**, **Table S4**, see methods for details).
252 All relative growth rate measurements were made in triplicate. We observed good concordance
253 among replicates (**Fig. S3**).

254 The entire dataset showed that the DHFR mutational landscape was strongly dependent
255 on TYMS background (**Fig. 3b-e**). In all three TYMS backgrounds, the distribution of growth rate
256 effects was bi-modal and reasonably well-described by a double gaussian containing one peak
257 of near-neutral mutations and another peak of mutations with highly deleterious growth rate
258 effects. This is the expected result for an enzyme that shows a sigmoidal relationship between
259 activity and growth. In the native TYMS context, the vast majority of mutations fall into the near-
260 neutral peak. However, there is a substantial fraction (12%, 343 total) that display growth rates at
261 or below that of "inactive", where "inactive" was defined as the average growth rate across
262 nonsense mutations in the first 120 residues of DHFR. Consistent with expectation, mutations at
263 known positions of functional importance tended to be deleterious in the WT TYMS context (W22,
264 D27, F31, T35, L54, R57, T113, G121, and D122)[27]. For example, both W22 and D27 are directly
265 in the active site and serve to coordinate substrate through a hydrogen bonding network[28], G121

266 and D122 are part of the βf-βg loop and stabilize conformational changes associated to
267 catalysis[29,30], and F31 contacts the substrate and is associated to the "network of promoting
268 motions"[31,32]. In the TYMS Q33S context, many of these deleterious mutations had even more
269 severe effects or were classified as "Null". Null mutations disappeared from our sequencing
270 counts within the first three time points (8 hours) of the selection experiment, preventing inference
271 of growth rate. For example, mutations at position 22 are deleterious in the WT TYMS context,
272 and appear as Null or very deleterious in the Q33S context. The same pattern can be readily
273 observed for positions 7,14,15, 22, 27, 31, 35, and 121. Again we saw that 12% of mutations have
274 growth rates at or below that of "inactive" variants. Finally, in the TYMS R166Q context, there are
275 very few deleterious mutations. Nearly all mutations are contained in the near-neutral peak,
276 including mutations at highly conserved active site positions like M20, W22, and L28. Stop codons
277 and mutations at the active site residue D27 continued to be deleterious, indicating that DHFR
278 activity was still under (very weak) selection in the TYMS R166Q background. Nonetheless, only
279 5% of mutations displayed growth rates at or below those of inactive mutations. Thus TYMS
280 R166Q is broadly buffering to DHFR variation.

281 To quantify the context dependence of mutational effects, we computed epistasis for all
282 DHFR mutations with measurable relative growth rates in each of the three TYMS backgrounds
283 (2,696 in total, see also methods) (**Fig. 4, Fig. S4, Table S5**). We assessed the statistical
284 significance of epistasis by unequal variance t-test under the null hypothesis that the mutations
285 have equal mean growth rates in both TYMS backgrounds. These p-values were compared to a
286 multiple-hypothesis testing adjusted p-value determined by Sequential Goodness of Fit (P = 0.035
287 for TYMS Q33S and P = 0.029 for TYMS R166Q, **Fig. 4a,b**) [33]. In the TYMS Q33S background,
288 95 mutations (3%) showed significant negative epistasis and 280 mutations (9%) showed
289 significant positive epistasis. Many of the DHFR mutations with positive epistasis to Q33S were
290 near-neutral in the WT context, and displayed small improvements in growth rate that were highly
291 significant due to the low experimental error for the best-growing mutations (**Fig. 4c**). In contrast,

292 the mutations with negative epistasis exhibited a range of growth rate effects in the WT context.
293 For the TYMS R166Q background the overall proportion of significant epistatic mutations was
294 larger: while only 41 mutations (1%) showed significant negative epistasis, 851 mutations (28%)
295 showed significant positive epistasis. Direct comparison of the relative growth rates of mutations
296 across the WT, Q33S, and R166Q TYMS backgrounds makes it very obvious that TYMS R166Q
297 was broadly rescuing, while TYMS Q33S had a more subtle effect that sometimes yielded
298 negative epistasis (**Fig. 4c,d**). These observations are consistent with our biochemical
299 understanding of the interplay between DHFR and TYMS relative velocities.
300

301 **The enzyme velocity to growth-rate model captures the observed fitness landscapes and**
302 **allows global estimation of mutational effects on catalysis**

303 Next we sought to further test our enzyme velocity to growth-rate model using the deep mutational
304 scanning data. We refit the model a third time, drawing upon a larger dataset of 34 DHFR single
305 mutants with previously reported k_{cat} and K_m values. We additionally characterized k_{cat} and K_m for
306 four new DHFR mutations (I5K, V13H, E17V and M20Q) that exhibited strong sign epistasis to
307 TYMS to more completely test our ability to predict epistasis. Together this yielded a set of 114
308 growth rate measurements with matched k_{cat} and K_m values for DHFR and TYMS (38 DHFR
309 mutations in 3 TYMS backgrounds, **Table S1**). We used these data to perform a bootstrap
310 analysis; iteratively subsampling the data and refitting the model 1000 times to obtain standard
311 deviations in our model fit and the eight associated parameters (**Fig. 5a**). The inferred parameters
312 for this large set of sequencing-based growth rate measurements were qualitatively similar to
313 those obtained for the smaller set of 28 plate-reader based growth rate measurements (7 DHFR
314 mutants in 4 TYMS backgrounds), but we observed some discrepancy in the estimated total folate
315 pool and intracellular concentrations of TYMS (**Table S3**). Overall both the predicted growth rates
316 and pattern of epistasis showed good agreement to our experimental observations (**Fig. 5a,b**).

317 Having established model performance on a subset of biochemically characterized DHFR
318 and TYMS sequences, we next examined consistency of the model with all growth rate
319 measurements (the total model fit). However the effect of most mutations on catalysis is unknown.
320 Thus, for each DHFR point mutant we used Monte Carlo sampling to identify a space of k_{cat} and
321 K_m values consistent with the three growth rate measurements (in the three TYMS backgrounds).
322 While three growth rate measurements were insufficient to uniquely constrain both k_{cat} and K_m
323 (the solution space is degenerate), this process did permit estimation of $\log_{10}$ catalytic power
324 (k_{cat}/K_m) for all 2,696 characterized point mutants with reasonable agreement to experiment (**Fig.**
325 **5c**). Once these catalytic parameters were estimated, we put them back into the model to assess
326 the correspondence between the predicted (modeled) growth rates, predicted epistasis, and our
327 experimental observations, yielding a global picture of model fit quality. Overall, we observed that
328 the model well-described the data with two exceptions. First, there was a small proportion of
329 DHFR mutations that were predicted to be rescued by TYMS R166Q but in actuality were not (70
330 total, 2% of all DHFR mutations, the horizontal stripe of red dots in **Fig. 5d**). It is possible that
331 these mutations caused a growth rate defect through DHFR mis-folding and aggregation, a factor
332 not captured by our model. Second, there was a proportion of DHFR mutations predicted to have
333 negative epistasis to TYMS R166Q but observed to exhibit mild positive epistasis (**Fig. 5e**). These
334 differences may be related to the fact that DHFR abundance is modeled with a single parameter
335 across all mutants, a factor which could be addressed in future work by including additional
336 experimental data. Nevertheless, the data indicated that our model can globally describe growth
337 rate phenotypes given variation in enzyme velocity. The resulting model and inferred catalytic
338 parameters now permit estimation of DHFR single mutant fitness in any TYMS background. We
339 computed the fraction of DHFR point mutants that are neutral (growth rate above 0.9) as a function
340 of variation in TYMS k_{cat} and K_m. These calculations highlighted that selection on DHFR activity
341 is strongly shaped by TYMS background, with low-activity TYMS variants increasing the
342 mutational tolerance of DHFR (**Fig. 5f,g**).

343 **Epistasis between DHFR and TYMS is organized into structurally localized groups**

344 Next, we examined the structural pattern of DHFR positions with epistasis to TYMS Q33S and
345 TYMS R166Q. Given that mutations tend to have similar epistatic effects at a particular DHFR
346 position in our data set (**Fig. S4**), we used k-means clustering to sort positions into four categories
347 according to their pattern of epistasis: negative, insignificant, positive, and strong positive (**Fig.**
348 **6a, Table S6**). The strong positive category solely contained DHFR mutations in the TYMS
349 R166Q background, while the negative epistasis category was predominantly occupied by DHFR
350 mutations in the TYMS Q33S background. Mapping these positions to the DHFR structure
351 showed that epistasis is organized into spatially distinct regions of the tertiary structure (**Fig.**
352 **6b,c**). Mutations with negative epistasis to Q33S tended to be proximal to the DHFR active site,
353 particularly the folate binding pocket. The negative epistasis cluster included several key positions
354 near or in the Met-20 loop, which is known to undergo conformational fluctuations associated with
355 catalysis (residues A9, V13, E17 and M20) [27,30]. It also encompassed positions I5, L24, L28, and
356 F31 which surround the folate substrate. Several of these positions have known roles in catalysis;
357 mutations at position 31 promoted product release (while slowing hydride transfer), and dynamics
358 of the M20 loop (which includes V13,E17) are associated with substrate binding and product
359 release[31,34]. Additionally, specific mutations at positions 5, 20, and 28 result in trimethoprim
360 resistance by altering trimethoprim affinity[34]. These structural and biochemical observations are
361 consistent with the finding that mutations with negative epistasis tended to yield moderate to
362 severe growth rate defects. In contrast, positions with positive epistasis to Q33S often had very
363 little (or sometimes a beneficial) effect on growth rate, and were distributed around the DHFR
364 surface (**Fig. 4c, Fig. 6b**). In the context of TYMS R166Q only one position — C85 — was
365 included in the negative epistasis cluster (**Fig. 6c**). A large fraction of DHFR positions (53%, 84
366 total) displayed positive epistasis to TYMS R166Q; these positions were distributed throughout
367 the DHFR structure. The positions in the strong positive epistasis cluster included mutations with
368 some of the most severe effects on growth rate in the WT TYMS context. A number of these

369 positions were previously established as important to DHFR catalysis, including residues F31,
370 L54, G121, D122, and S148[27]. Mutations at these sites can be detrimental to k_{cat}, K_m, or both.
371

372 **Epistasis and the structural encoding of DHFR catalysis**

373 When the epistatic clusters are viewed together on the structure, one sees that they form
374 approximate distance-dependent shells around the active site (**Fig. 7a-d**). Considering the pattern
375 of epistasis to TYMS Q33S, positions with negative epistasis were closest to the active site,
376 surrounded by positions with insignificant epistasis, and finally positions in the positive epistasis
377 cluster form an outer shell (**Fig. 7a,b**). For TYMS R166Q, positions in the strong positive epistasis
378 cluster were closest to the active site, followed by positive epistasis positions, and finally those
379 with insignificant epistasis (**Fig. 7c,d**). For comparison, we also mapped the model-predicted
380 catalytic power averaged across all mutations at a position to the structure (**Fig. 7e**). Together,
381 these structural images paint a picture of the molecular encoding of catalysis and epistasis.
382 Mutations with predicted intermediate-to-large effects on catalysis were nestled near the active
383 site and showed negative epistasis to Q33S and strong positive to positive epistasis to R166Q,
384 while mutations with more mild effects on catalysis showed weaker positive to insignificant
385 epistasis to R166Q and Q33S. Though catalysis and epistasis showed an approximate distance-
386 dependent relationship to the DHFR active site, there a number of key positions distal to the active
387 site that exhibited large growth rate effects, strong positive epistasis to TYMS R166Q, and likely
388 act allosterically to tune catalytic activity (e.g. L110, G121, D122, W133, S148, and Y151). The
389 positions with the largest estimated effects on catalysis were highly evolutionarily conserved ($P <$
390 10^{-10} by Fisher's exact test, **Table S7**, **Fig. 7f**), indicating that our model and experimental data
391 are capturing features relevant to the fitness of DHFR.
392

393 **DISCUSSION**

394 It is well-appreciated that physical protein interactions place constraints on the individual
395 interacting monomers. Protein interfaces are organized to bind with appropriate affinity and avoid
396 non-specific interactions[35,36]. The individual components of physical complexes tend to be
397 expressed in similar ratios to avoid dosage related toxicity and aggregation[37,38]. However the
398 extent to which biochemical interactions constrain the function and sequence of individual
399 monomers has remained less clear. We have explicitly revealed these interactions at single-
400 residue resolution for one model system and coupled them with a mathematical model to quantify
401 the intracellular constraints on DHFR and TYMS relative catalytic activities.

402 Our mutagenesis data and modeling show that TYMS activity strongly modifies the
403 constraints on DHFR catalytic parameters; shaping both the range and relative importance of k_{cat}
404 and K_m in modulating growth. This biochemical interaction results in an approximately shell-like
405 pattern of mutational sensitivity to TYMS background (epistasis) in the DHFR tertiary structure.
406 Extreme loss-of-TYMS function buffered variation in some of the most conserved DHFR active
407 site positions, while moderate loss of function buffered variation at more peripheral surface
408 exposed sites. Given these data, we expect that inhibition or loss-of-function in TYMS will promote
409 the evolvability of DHFR, a finding with consequences for both laboratory and clinical evolution.
410 For example, inhibiting TYMS activity in the clinic may promote the evolution of drug resistance
411 in DHFR, while activating TYMS may restrict evolutionary accessible paths. In the laboratory,
412 strains with reduced TYMS activity could provide a less stringent context for testing designed
413 sequences or evolving new DHFR function.

414 The existence of an enzyme velocity to growth-rate mapping — by definition — allows us
415 to relate variation in DHFR and TYMS catalytic parameters to growth rate. It also allows one (in
416 principle) to do the inverse: infer *in vitro* catalytic parameters from growth rate measurements.
417 The intuition follows from classic steady-state Michaelis Menten experiments: to quantify steady
418 state kinetics *in vitro* one measures enzyme initial velocity as a function of substrate
419 concentration. In our sequencing-based experiments, variation in TYMS background effectively

420 titrates intracellular concentrations of DHF (substrate) while growth rate provides an estimate of
421 velocity. Though our current dataset of three TYMS backgrounds is insufficient to uniquely
422 constrain precise fits for k_{cat} and K_m, we anticipate that the addition of a few additional TYMS
423 backgrounds and/or the use of more sophisticated fitting approaches will permit more accurate
424 biochemical parameter inference. Indeed, recent work on peptide binding proteins (the PDZ and
425 SH3 domains) has shown how measuring the growth rate effect of mutations in different genetic
426 backgrounds and assay conditions can well-constrain biophysical parameters for binding affinity
427 and protein stability[39,40]. One might follow a conceptually similar strategy to learn quantitative
428 biochemical parameters from high throughput growth rate data. New microfluidics-based
429 approaches for high-throughput biochemistry could play a key role in refining and testing such
430 methodology[41].

431 Together our findings shape how we think about designing enzymes and metabolic
432 systems. Typical strategies for designing enzymes do not explicitly consider cellular context[42]. As
433 a result, a significant fraction of designs could fail simply because they are not properly "matched"
434 in terms of velocity to the surrounding pathway. The limited ability of homologs to complement
435 growth in another species has been observed for a number of enzymes[43–47], including DHFR[48,49].
436 Thus, even a well-designed catalytically active synthetic enzyme could fail to rescue growth if
437 placed in the wrong cellular context. Just as computational protein design considers entire
438 physical complexes to create binding interactions with altered affinity and specificity, one might
439 consider the joint design of biochemically-interacting enzymes to alter metabolic efficiency and
440 growth. Further study of enzyme rates and abundance across species, as well as
441 characterizations of enzyme velocity to growth rate mappings, will help shape our understanding
442 of the system level constraints placed on metabolic enzymes.

443

444 **MATERIALS AND METHODS**

445 **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

446 ***Escherichia coli expression and selection strains*:** ER2566 Δ*folA* Δ*thyA E. coli* were used for
447 all growth *in vivo* growth rate measurements; this strain was a kind gift from Dr. Steven Benkovic
448 and is the same used in Reynolds et al., 2011 and Thompson et al., 2020 [50,51]. XL1-Blue *E. coli*
449 (genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q}*ZΔM15* Tn*10*(Tet^{r})])
450 from Agilent Technologies were used for cloning, mutagenesis, and plasmid propagation.
451 BL21(DE3) *E. coli* (genotype*: fhuA2 [lon] ompT gal (λ DE3) [dcm]* Δ*hsdS*. *λ DE3 = λ sBamHIo*
452 Δ*EcoRI-B int::(lacI::PlacUV5::T7 gene1) i21* Δ*nin5*) from New England Biolabs were used for
453 protein expression.
454
455 ***Selection vector for DHFR constructs***: DHFR variants were cloned into a modified version of
456 the pACYC-Duet 1 vector (Novagen), which we refer to as pTet-Duet. pTet-Duet is a low-copy
457 number vector containing two multiple cloning sites; the first is under control of the T7 promoter
458 and the second was modified to be regulated by the tetracycline repressor (TetR). DHFR (*folA*) is
459 cloned into the first MCS; TYMS (*thyA*) is cloned into the second MCS. During selections we do
460 not induce expression of either gene but instead rely on leaky expression in ER2566 Δ*folA* Δ*thyA*
461 *E. coli*. The vector map for these constructs can be found on Addgene: 81596.
462
463 ***Expression vector for DHFR constructs***: *E.coli folA* (the gene encoding DHFR) was cloned into
464 the pHis8-3 expression vector using restriction sites NcoI and XhoI. DHFR was tagged in-frame
465 with an N-terminal 8X-Histidine tag separated from the *folA* reading frame by a thrombin cleavage
466 site. Individual point mutant clones were constructed using the Quikchange II site-directed
467 mutagenesis kit (Agilent).
468
469 ***Expression vector for TYMS constructs***: The *thyA* gene (encoding TYMS) was amplified by
470 PCR from *E. coli* MG1655 and cloned into the vector pET24A using XbaI/Xho restriction sites.

471 The point mutants of TYMS (Q33S, R127A, and Q33S) were made using the Agilent QuikChange
472 II site-directed mutagenesis kit.
473
474 **METHOD DETAILS**
475 **Plate-reader Based Growth Rate Assays**
476 DHFR and TYMS point mutant combinations in the selection vector were transformed into
477 ER2566 Δ*folA* Δ*thyA* chemically competent cells by heat shock. The cells were recovered for 60
478 minutes at 37°C with shaking at 220 rpm, spread on agar plates (Luria Broth (LB) containing 30
479 µg/ml chloramphenicol and 50 µg/ml thymidine), and grown at 37°C overnight. The next day,
480 liquid overnight cultures were inoculated from a streak over multiple colonies and grown overnight
481 at 37°C in LB supplemented with 30 µg/ml chloramphenicol and 50 µg/ml thymidine. These
482 overnight cultures were pelleted and washed with M9 minimal media, then resuspended in pre-
483 warmed M9 media supplemented with 0.4% glucose, 0.2% amicase, 2 mM $MgSO_4$, 0.1µM $CaCl_2$,
484 30 µg/ml chloramphenicol (henceforth referred to as M9 selection media). Next, OD_{600} for all
485 resuspended cultures was measured in a Perkin Elmer Victor X3 plate reader. Cultures were then
486 diluted to OD_{600}=0.1 in prewarmed M9 selection media and incubated for 4 hours at 30°C, shaking
487 at 220 rpm. After this period of adaptation and regrowth, cultures were back-diluted to OD_{600} =
488 0.1 in 1 ml prewarmed M9 selection media with 50 µg/ml thymidine. These cells were inoculated
489 into 96-well culture plate at OD_{600} = 0.005 (10 µl cells into 200 µl total well volume) containing
490 prewarmed M9 selection media with 50 µg/ml thymidine; plates were sealed with EasySeal
491 permeable covers (Sigma Aldrich). All growth rate measurements were made in triplicate. Plates
492 were shaken for 10 seconds before reading, and Readings of OD_{600} were taken every 6 minutes
493 over 24 hours using a BioTek Synergy Neo2 plate reader in a 30°C climate-controlled room.
494
495 **DHFR Saturation Mutagenesis Library Construction**

496 The DHFR saturation mutagenesis library was constructed as four sub-libraries in the DHFR
497 selection vector (see above for details) to ensure coverage of each mutated region with a 300
498 cycle Illumina sequencing kit. Each sublibrary covers 40 amino acid positions: 1-40 (sublibrary 1,
499 SL1), 41-80 (sublibrary 2, SL2), 81-120 (sublibrary 3, SL3), and 121-159 (sublibrary 4, SL4).
500 'Round the Horn' or inverse PCR (iPCR) with mutagenic NNS primers (N = A/T/G/C, S = G/C)
501 was used to introduce all 20 amino acid substitutions at a single amino acid position as described
502 in[51]. Library completeness was verified by deep sequencing. In our initial validation sequencing
503 run we found that mutations at positions W22 and L104 were systematically under-represented;
504 iPCR was repeated for these positions and they were supplemented into their respective
505 assembled sublibraries.

506 After sub-library assembly, restriction digest and ligation was used to subclone each
507 sublibrary into pTet-Duet plasmids containing different TYMS backgrounds (WT, R166Q, or
508 Q33S). The entire DHFR coding region containing restriction sites (NotI and EcoNI) was amplified
509 by PCR. PCR reaction was size-verified with agarose gel electrophoresis with an expected band
510 size of 627 bp. The library PCR products and target plasmids were double digested with NotI and
511 EcoNI for 3 hours at 37°C. To prevent re-circularization, the digested plasmid was treated with
512 Antarctic phosphatase for 1 hour at 37°C. The DHFR insert and treated plasmid were ligated with
513 T4 DNA ligase overnight at 16°C. The concentrated ligation product was then transformed into *E.*
514 *coli* XL1-blue by electroporation, and recovered in SOB for 1 hour at 37°C. 20 µL of the recovery
515 culture was serially diluted and plated on LB-agar with 50 µg/mL thymidine and 30 µg/mL
516 chloramphenicol, to permit quantification of transformants and estimate library coverage. The
517 minimum library coverage was 1000 CFU/mutant. The remaining recovery culture was grown in
518 a flask containing 12 ml LB with 30 µg/mL chloramphenicol and 50 µg/mL thymidine at 37°C, with
519 220 rpm shaking overnight. 10 ml of the overnight culture was miniprepped with the Gene-Jet
520 Mini-prep kit (Fisher Scientific, K0503) to obtain the plasmid library.

521

522 **Growth Rate Measurements in the Turbidostat for all DHFR mutant libraries**

523 All sublibraries were inoculated, grown, and sampled in triplicate. Each plasmid sub-library was
524 transformed into the *E. coli* double knockout strain ER2566 Δ*folA* Δ*thyA* by electroporation and
525 recovered in SOB for one hour at 37°C. To estimate library coverage, 20 µL of the recovery culture
526 was serial diluted with SOB and plated on LB agar plates containing 30 µg/mL chloramphenicol
527 and 50 µg/mL of thymidine. The remainder of the recovery culture was inoculated into M9
528 selection media supplemented with 50 µg/mL thymidine and grown overnight at 37°C. The next
529 morning, library coverage was estimated from colony counts; all selection experiments in this
530 work had an estimated library coverage of 1000 CFU/mutant or greater. The overnight liquid
531 culture was washed and back-diluted to OD_{600}=0.1 in M9 selection media supplemented with 50
532 µg/mL thymidine, and incubated for four hours at 30°C to allow adaptation to selection
533 temperature and to return the culture to log-phase growth. Following adaptation, selection was
534 initiated by back-diluting these cultures to an OD_{600} of 0.1 into 17 mL of pre-warmed M9 selection
535 media supplemented with 50 µg/mL thymidine in continuous culture vials with stir bars. These
536 vials were then incubated in a turbidostat with a target OD_{600} of 0.15 at a temperature of 30°C.
537 The turbidostat maintained a set optical density by adding 2.8 mL dilutions of M9 selection media
538 supplemented with 50 µg/mL thymidine in response OD detection, and was built according to the
539 design of Toprak et al[52]. Culture samples (1 mL each) were taken at the beginning of selection (t
540 = 0 hours) and at 4, 8, 12, 20, and 24 hours into selection. Immediately after each time point,
541 these 1 mL samples were pelleted at 2,348 rcf in a benchtop microcentrifuge for 5 minutes at
542 room temperature. Supernatants were removed and the remaining pellet was stored at -20°C.

543

544 **Next Generation Sequencing Amplicon Sample Preparation**

545 Each turbidostat selection sample (representing a particular timepoint for a sub-library and
546 replicate) was prepared for sequencing as a PCR amplicon using Illumina TruSeq-HT i5 and i7
547 indexing barcodes. To generate these amplicons, each cell pellet from the growth rate assay was

548 thawed and lysed by resuspending the cells with 100 µL dH_2O and incubation at 95°C for 5
549 minutes. Lysates were then clarified by centrifugation at maximum speed for 10 minutes in a room
550 temperature bench top microcentrifuge. Supernatants containing plasmids were isolated from the
551 pellet. 1 µL of each supernatant was used as the template for an initial round of PCR with Q5-Hot
552 Start Polymerase (NEB) that amplified the DHFR coding region of the sublibrary (10 PCR cycles
553 total, standard Q5 reaction conditions). From this first PCR reaction, 1 µL was used in a second
554 round of PCR (22 cycles of denaturation/anneal/elongation) with primers that added Illumina
555 sequencing adaptors. Together, these two rounds of PCR yielded a final product of size: 315 bp
556 (SL1), 308 bp (SL2), 298 bp (SL3), 304 bp (SL4). The amplicons were size verified using 1%
557 agarose gel electrophoresis. In the case where a sample did not produce an amplicon, the first
558 round PCR was repeated with 2 µL of the supernatant rather than 1 µL, with the remaining
559 preparation identical. All amplicons were individually quantified using with Quant-iT™
560 PicoGreen™ dsDNA Assay Kit (ThermoFischer Scientific) and mixed in equimolar ratio, with a
561 final target amount greater than or equal to 2000 ng. Errors in pipetting volume were minimized
562 by ensuring that more than 2 µL was taken from each amplicon. This mixture was gel-purified and
563 then cleaned and concentrated using the Zymo Research DNA Clean & Concentrator-5 kit. To
564 assess purity, the A260/A80 and A260/A230 nm absorbance ratios of the sample library were
565 measured on a nanodrop. The sample library DNA concentration was measured using a Qubit
566 dsDNA HS Assay in a Qubit 3 Fluorometer (Invitrogen by Thermo Fischer Scientific). The sample
567 library was diluted to 30 nM in a volume of 50 µL of TE buffer (1 mM Tris-HCl (pH 8.5), 10 mM
568 EDTA (pH 4)). This mixed and quantified library was sequenced on an Illumina HiSeq (150 cycle
569 x 2 paired-end) by GeneWiz. Prior to sequencing, GeneWiz also provided quality control with
570 sample library quantification with Qubit dsDNA HS Assay and sample DNA fragment composition
571 with TapeStation (High Sensitivity D1000 ScreenTape, Agligent Technologies, Inc.). The NGS
572 sequencing run resulted in 251.61 GB of data, with 337,353,664 reads, and 101,209 Mbases.
573

574 **DHFR Expression and Purification**

575 DHFR mutant variants were expressed in BL21(DE3) *E. coli* grown at 30°C in 50 ml Terrific Broth
576 (TB) with 35 µg/ml Kanamycin (Kan) for selection. Expression was induced at an OD_{600} = 0.6-0.8
577 with 250 uM IPTG, and cells were grown at 18°C for 16-18 hours. Cultures were pelleted by
578 centrifugation for 10 minutes at 5000 x g, 4°C and supernatant removed; cell pellets were stored
579 at -80°C. Thawed cell pellets were lysed by sonication in 10ml lysis buffer (50 mM Tris, 500 mM
580 NaCl, 10 mM imidazole, pH 8.0 buffer containing 0.1 mM PMSF, 0.001 mg/ml pepstatin, 0.01
581 mg/ml leupeptin, 20 µg/ml DNAseI and 5 µg/mL lysozyme). The resulting lysate was clarified by
582 centrifugation and incubated with 0.1ml Ni-NTA agarose (Qiagen) slurry (0.05 ml column volume)
583 equilibrated in Nickel Binding Buffer (NiBB, 50 mM Tris pH 8.0, 500 mM NaCl, 10 mM imidazole)
584 for 15 minutes on a tube rocker at 4°C. The slurry was then transferred to a disposable
585 polypropylene column (BioRad). After washing with 10 column volumes (CV) of NiBB, DHFR was
586 eluted with 0.5 mL 50mM Tris pH 8.0, 500 mM NaCl, 400 mM imidazole. The eluted protein was
587 concentrated and buffer-exchanged to 50 mM Tris, pH 8.0 in a 10 kDa Amicon centrifugal
588 concentrator (Millipore) and centrifuged 15 min at 21,000 x g, 4°C to pellet any precipitates.
589 Following buffer exchange, the protein was purified by anion exchange chromatography (using a
590 BioRad HiTrapQ HP column on a BioRad NGC Quest FPLC). A linear gradient was run from 0-1
591 M NaCl in 50 mM Tris pH 8.0 over 30 ml (30 column volumes, CV) while collecting 0.5 ml fractions.
592 Fractions containing DHFR were combined, concentrated, flash-frozen in liquid nitrogen, and
593 stored at -80°C.

594

595 **TYMS Expression and Purification**

596 Individual TYMS mutants were expressed in BL21(DE3) *E. coli* grown at 37°C in 50 ml Terrific
597 Broth (TB) with 35 µg/ml Kanamycin (Kan) for selection. Expression was induced with 1mM IPTG
598 when the cells reached an OD_{600} = 0.6-0.8, and the cells were then grown at 18°C for 16-18 hours
599 before harvesting pellets for storage at -80°C. TYMS was purified from the frozen pellets following

600 a protocol adapted from Changchien et al [53]. Cell pellets were thawed and resuspended in TYMS
601 lysis buffer (20 mM Tris/10 mM MgCl2/0.1% DOC pH 7.5 with 5 mM DTT, 0.2 mg/ml lysozyme, 5
602 µg/ml DNAse I) and incubated at room temperature while rocking for 15 minutes. The resulting
603 supernatant was clarified by centrifugation. Next, streptomycin sulfate was added to a final
604 concentration of 0.75% to separate nucleic acids. The cells were incubated rocking at 4°C for 10
605 minutes and the supernatant was retained following centrifugation for 10 minutes at >10,000 x g.
606 Ammonium sulfate was then added at 50% saturation (0.3 g/ml), mixed for 10 minutes at 4°C,
607 then centrifuged as above, retaining supernatant. Additional ammonium sulfate was then added
608 to the supernatant at 80% saturation (an additional 0.2g/ml), mixed for 10 minutes at 4°C, and
609 centrifuged as above, retaining the pellet. The pellet was dissolved in 25mM potassium phosphate
610 pH 6.5 and dialyzed overnight at 4°C against 1L 25 mM potassium phosphate pH 6.5. Following
611 dialysis the protein was purified by anion exchange (HiTrap Q HP column, Cytiva) with a 25 CV
612 linear gradient from 0M NaCl to 1M NaCl in 25mM potassium phosphate pH 6.5. FPLC fractions
613 containing TYMS were combined and concentrated using a 10 kDa Amicon concentrator
614 (Millipore) and stored at 4°C for up to a week.
615
616 **DHFR Steady-state Michaelis Menten Kinetics**
617 DHFR k_{cat} and K_m were determined under Michaelis-Menten conditions with saturating
618 concentrations of NADPH as in prior work[54,55]. Briefly, DHFR protein concentration was
619 determined by measuring A_{280} (extinction coefficient = 33500 $M^{-1}cm^{-1}$). DHF (Sigma Aldrich) was
620 prepared in MTEN buffer (50 mM MES, 25 mM Tris base, 25 mM Ethanolamine, 100 mM NaCl,
621 pH 7.0) containing 5 mM DTT (Sigma Aldrich). 100 nM DHFR protein and 100 µM NADPH (Sigma
622 Aldrich) were combined in MTEN buffer with 5 mM DTT and pre-incubated for 1 hour at 25°C prior
623 to measurement. To initiate the reaction, the protein-NADPH solution was mixed with DHF in a
624 quartz cuvette (sampling DHF over a range of concentrations, tuned to the Km of the mutant).
625 The initial velocity of DHFR was measured spectrophotometrically by monitoring the consumption

626 of NADPH and DHF (decrease in absorbance at 340 nm, $\Delta\varepsilon_{340}$=13.2 mM^{-1} cm^{-1}). All
627 measurements were made in triplicate; analysis was performed using the Michaelis-Menten
628 nonlinear regression function of Graph Pad Prism.
629
630 **Preparation of TYMS substrate for assaying enzyme activity and steady state kinetics**
631 (6R)-methylenetetrahydrofolic acid (MTHF, CH_2H_4fol) was purchased from Merck & Cie
632 (Switzerland) and dissolved to 100 mM in nitrogen-sparged citrate-ascorbate buffer (10 mM
633 ascorbic acid, 8.5 mM citrate, pH 8.0). 30 µL aliquots were made in light-safe microcentrifuge
634 tubes, flash-frozen in liquid nitrogen, and stored at -80C. Before use, the stock was thawed and
635 diluted to 10 mM in TYMS kinetics reaction buffer (100 mM Tris base, 5 mM Formaldehyde, 1 mM
636 EDTA, pH 7.5) and quantified in an enzymatic assay: 50 µM MTHF, 200 µM dUMP and 1µM
637 TYMS protein were combined and A_{340} measured until steady-state reached. Actual concentration
638 was then calculated from the difference in A_{340} before and after the reaction using Beer's Law
639 (MTHF extinction coefficient: 6.4 $mM^{-1}cm^{-1}$).
640
641 **TYMS Steady-state Michaelis Menten Kinetics**
642 TYMS k_{cat} and K_m were determined for both dUMP and MTHF under Michaelis-Menten conditions
643 by varying one substrate and holding the other saturating as in prior work[56,57]. Briefly, TYMS
644 protein concentration was determined by measuring A_{280} (extinction coefficient = 53400 $M^{-1}cm^{-1}$).
645 TYMS protein was prepared in TYMS assay buffer (100 mM Tris base, 5 mM Formaldehyde, 1
646 mM EDTA, pH 7.5) containing 50 mM DTT (Sigma Aldrich). 50 nM TYMS protein and either 100
647 µM dUMP (Sigma Aldrich) or 150 µM MTHF (Merck & Cie) were combined with varying
648 concentrations of the other substrate to initiate the reaction. The production of DHF was monitored
649 spectrophotometrically (increase in absorbance at 340nm, $\Delta\varepsilon_{340}$=6.4 $mM^{-1}cm^{-1}$) for 2 minutes per

650 reaction. All measurements were made in triplicate; analysis was performed using the Michaelis-
651 Menten nonlinear regression function of Graph Pad Prism.
652

653 **QUANTIFICATION AND STATISTICAL ANALYSIS**

654 **Enzyme Velocity to Growth Rate Model Construction and Parameterization**

655 For the purposes of modeling, we approximated DHFR and TYMS as a two-enzyme cycle in which
656 DHFR produces THF and consumes DHF, and TYMS produces DHF and consumes THF. This
657 abstraction ignores the different carbon-carrying THF species, instead collapsing them into a
658 single "reduced folate" pool. This simplification allows us to construct an analytically solvable
659 model for steady state THF concentration that we can then relate to growth (Equation 3).

660 First, we fit the free parameters in the Goldbeter-Koshland equation ($[DHFR]$, $[TYMS_{WT}]$,
661 $[TYMS_{R166Q}]$, fol_{tot}) using a set of ten metabolomics measurements for the relative abundance of
662 3-glutamate form of formyl THF as obtained in prior work [1]. These measurements were made for
663 DHFR mutations G121V, F31Y/L54I, M42F/G121V, F31Y/G121V and the WT in the background
664 of WT TYMS and TYMS R166Q. So why this particular folate species? We noticed that the relative
665 abundance of many of the reduced THF species in our data set was correlated, and chose formyl
666 THF to model because the experimental data were less variable and showed a strong, monotonic
667 relationship with cell growth. Then, we fit the free parameters in equation two (g_{max}, g_{min}, K, n)
668 using a set of ten growth rate measurements for the same DHFR/TYMS mutation pairs. This fitting
669 process gave rise to the fits shown in figure 1. When assessing model performance against the
670 larger set of TYMS variants (as in figure 2) we refit all parameters (g_{max}, g_{min}, K, n, $[DHFR]$,
671 $[TYMS_{WT}]$, $[TYMS_{R166Q}]$, $[TYMS_{Q33S}]$, $[TYMS_{R127A}]$, fol_{tot}) to the growth rate data only since we did
672 not have metabolomics data for this larger set. All parameter fits were made in python using
673 scipy.optimize least_squares; the complete fitting process is documented in Jupyter notebook
674 1_KGmodel.ipynb in the associated github repository.

675 We assessed the model sensitivity to shuffling the data (Figure 2 – supplement 2) by
676 randomly shuffling all catalytic parameters (k_{cat}, K_m) 50 times across DHFR and TYMS and
677 computing an R^2 value. We also assessed model sensitivity to subsampling the data; error bars
678 in Figures 1, 2, and Figure 2 – supplement 2 correspond to SEM across "jackknife" re-samplings
679 of the data wherein one DHFR/TYMS combination was left out for each re-sampling. Finally, to
680 assess the global model fit to the data (as in Figure 6 and Figure 6 – supplement 1) we first fit the
681 9 model parameters (g_{max}, g_{min}, K, n, [DHFR], [$TYMS_{WT}$], [$TYMS_{R166Q}$], [$TYMS_{Q33S}$], fol_{tot}) using the
682 growth rate measurements of 16 DHFR mutations for which experimental k_{cat} and K_m were known
683 (48 total observations given the three TYMS backgrounds). Then, fixing these parameters, we fit
684 k_{cat} and K_m values for all 2696 mutations with growth rate measurements in all three TYMS
685 backgrounds to the complete data set of 8,088 sequencing-based growth rate observations. This
686 process is documented in Jupyter notebook 4_ModelAndDMSData.ipynb in the associated github
687 repository.
688

689 **Next Generation Sequencing Data Processing and Read Counting**

690 All PCR amplicons (corresponding to individual replicates, timepoints and sublibraries) were
691 sequenced on an Illumina HiSeq using 2 x 150 paired end reads. The resulting fastq files were
692 processed and filtered prior to read counting. Briefly, the forward and reverse reads were merged
693 using USEARCH. Each read was quality score filtered (Q-Score ≥ 20) and identified as a WT or
694 mutant of DHFR using a custom python script. This python script filtered for full length reads and
695 base call quality scores greater than 20 (error rate ≤ 1:100). The reads passing these quality
696 control criteria were compared against the wild-type reference sequence to determine mutation
697 identify. Reads that contained multiple point mutations or mutations outside the sublibrary of
698 interest were removed from analysis. This process resulted in counts for the WT and each mutant
699 at each time point and replicate. These counts were further corrected given the expected error in

700 the data (q-score) and Hamming distance from the WT codon to account for potential hopping of
701 WT reads to mutations; a process that was detailed in McCormick et al [55].
702
703 **Relative Growth Rate Calculations**
704 We calculated relative growth rates for individual mutations and the WT over time from the
705 sequencing-based counts (N_t^{mut}, N_t^{WT}). Mutants with fewer than 10 counts were considered
706 absent from the data set and were set to zero to reduce noise. From these thresholded counts,
707 we calculated a log normalized relative frequency of each mutation over time:
708 $$log_2(f(t)) = log_2(N_t^{mut}/N_t^{WT}) - log_2(N_{t=0}^{mut}/N_{t=0}^{WT})$$
709 We then calculated relative growth rate (m_{DHmut}^{TS}) as the slope of the log relative frequency over
710 time by linear regression. Linear regression was performed using scikit Learn, and individual
711 points were weighted by the number of counts (in order to down weight less-sampled mutants at
712 later time points). Relative growth rate or a mutant was only calculated if the mutant was present
713 over at least the first three time points, otherwise it was classified as a "Null" mutant. Finally, all
714 relative growth rates were normalized such that WT has a relative growth rate of 1. Growth rates
715 were additionally normalized by the bulk culture growth rate (estimated from the turbidostat, in
716 units of generations per hour) to account for small vial-to-vial variations culture doublings across
717 the experiment. All calculations are shown in Jupyter notebook 2_DMSGrowthRates.ipynb in the
718 associated github repository.
719
720 **Epistasis Analysis**
721 Epistasis was calculated according to an additive model:
722 $$\varepsilon_{DHmut,TSmut} = m_{DHmut}^{TS_mut} - m_{DHmut}^{TS_WT}$$
723 In our experiments TYMS R127A, Q33S and R166Q have no growth rate effect in the WT DHFR
724 context due to thymidine supplementation. Under this formalism, mutations that show improved

725 growth in the mutated TYMS background have positive epistasis, while mutations with reduced
726 growth in the mutated TYMS background have negative epistasis. We assessed the statistical
727 significance of epistasis by unequal variance t-test under the null hypothesis that the mutations
728 have equal mean growth rates in both TYMS backgrounds. These p-values were compared to a
729 multiple-hypothesis testing adjusted p-value determined by Sequential Goodness of Fit (P = 0.035
730 for TYMS Q33S and P = 0.029 for TYMS R166Q) [33]. K-means clustering of epistatic positions
731 was performed using a custom script based on that described in Thompson et al [51]. All epistasis
732 calculations are shown in Jupyter notebook 3_Epistasis.ipynb in the associated github repository.
733

734 **ACKNOWLEDGEMENTS**
735 The authors thank Olivier Rivoire for early conversations regarding our data. We also thank the
736 Reynolds lab for feedback on experimental design, data analysis, and the manuscript.
737

738 **FUNDING**
739 Research reported in this publication was supported by the National Institute of General Medical
740 Sciences of the National Institutes of Health under Award Number R01GM136842. The content
741 is solely the responsibility of the authors and does not necessarily represent the official views of
742 the National Institutes of Health. This work was also partly supported in its early stages by the
743 Gordon and Betty Moore Foundation's Data Driven Discovery Initiative through grant GBMF4557
744 to KAR.
745

746 **AUTHOR CONTRIBUTIONS**
747 TNN and KAR conceptualized the work and designed experiments. KAR created the
748 mathematical model. TNN collected all deep mutational scanning data. TNN and KAR analyzed
749 the data. CI collected plate-reader based growth rate measurements (used in model
750 development) and performed all Michaelis-Menten enzyme kinetics assays. SMT constructed the

751 deep mutational scanning library. KAR wrote the original draft. TNN, CI, SMT and KAR revised

752 the paper. KAR provided supervision and funding acquisition.

753

754 **COMPETING INTERESTS**

755 The authors have no competing interests to declare.

756

FIGURES AND LEGENDS

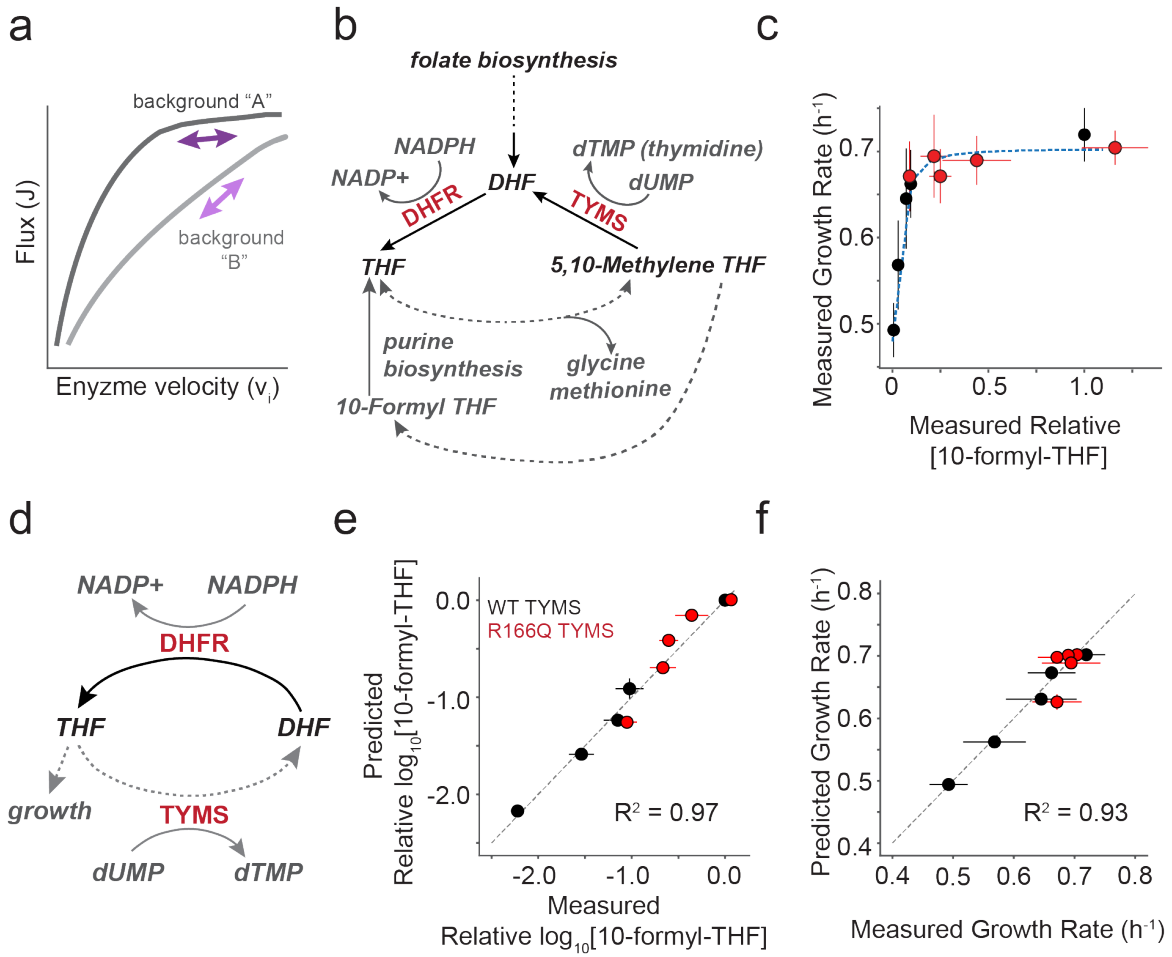


Figure 1. Constructing a biochemistry-to-growth model for DHFR and TYMS.

a) Schematic describing the relationship between metabolic pathway flux and enzyme velocity. Many enzymes show a hyperbolic relationship between velocity and flux; the enzyme control coefficient describes the fractional change in flux given a fractional change in velocity. Control coefficients vary with the starting enzyme velocity (purple and green arrows, background "A") and can change with genetic background (violet arrow, background "B").

b) The role of DHFR and TYMS in folate metabolism. Metabolites are labeled in grey or black italic text. Dotted lines indicate multiple intermediate reactions that are summarized with a single line.

c) The relationship between the experimentally measured relative abundance of [10-formyl-THF] and *E. coli* growth rate. Red points indicate five DHFR variants in the background of TYMS R166Q (a near catalytically inactive variant) and black indicates the same DHFR variants in the context of WT TYMS. Error bars indicate the standard deviation across N=3 replicates for both growth rate (y-axis) and 10-formyl-THF abundance (x-axis). The blue dotted line indicates the best fit for a hyperbolic model (Equation 1) relating THF abundance to growth.

d) A simplified, abstracted version of the DHFR and TYMS system. Again dotted lines indicate multiple intermediate reactions that are summarized with a single line.

776 **e)** The correlation between experimentally measured $\log_{10}$[10-formyl-THF] relative abundance
777 and the model prediction (as computed with Equation 3). The grey dotted line indicates x=y.
778 Color coding is identical to **c**.
779 **f)** The correlation between experimentally measured and predicted growth rates for five DHFR
780 point mutations in two different TYMS backgrounds (same mutants as in **c**,**e**). The grey dotted
781 line indicates x=y. Color coding is identical to **c**.
782

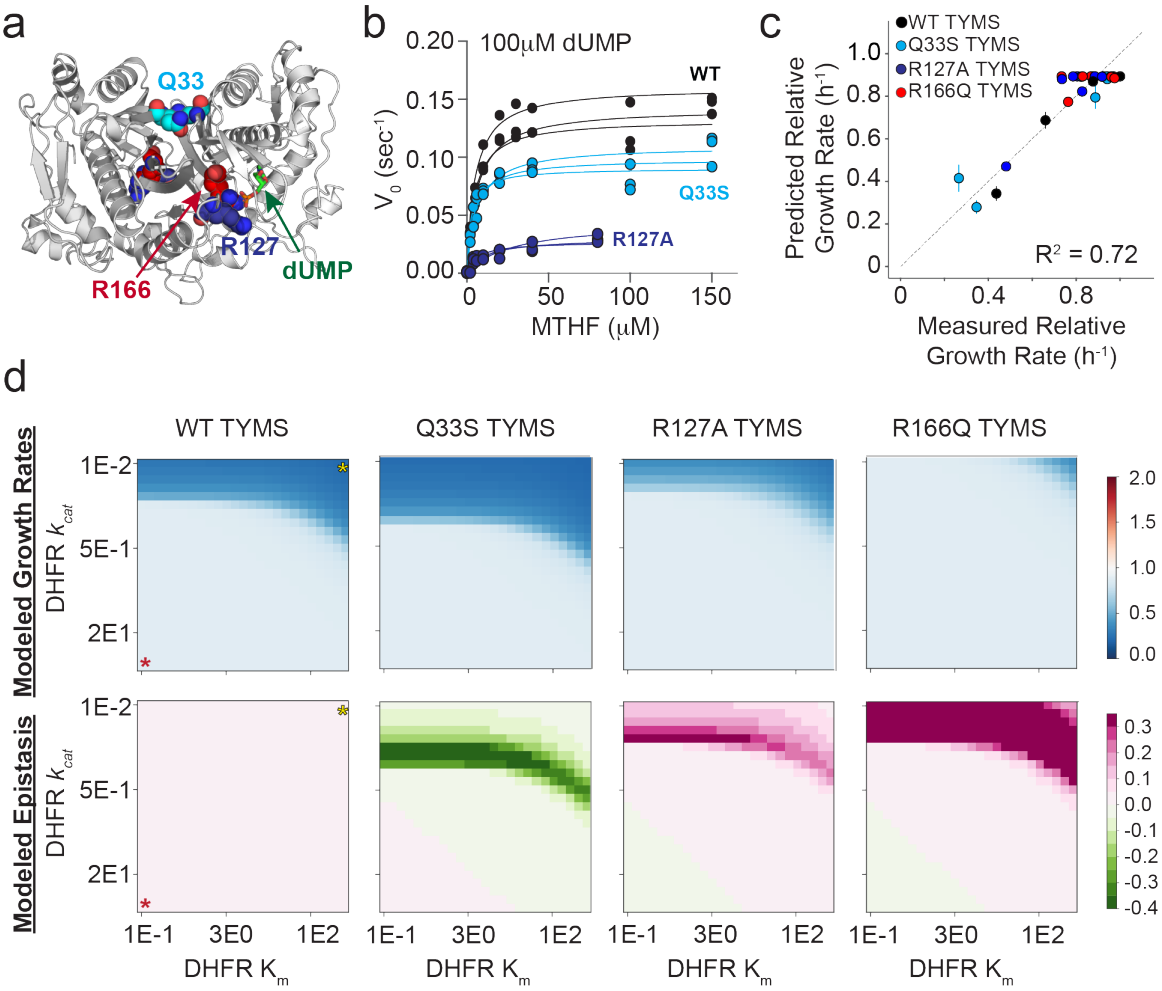


Figure 2. **Evaluating biochemistry-to-growth model performance across additional TYMS variants**.

a) Location of the TYMS point mutations (PDBID: 1BID[58]). TYMS functions as an obligate domain-swapped homodimer; active sites include residues from both monomers (white and grey cartoon). Positions mutated in this study are in colored spheres, and indicated with arrows (Q33 – cyan, R127 – navy, R166 – red). The dUMP substrate is in sticks, labeled and colored green.

b) Michaelis Menten enzyme kinetics for WT TYMS (black), TYMS Q33S (cyan), and TYMS R127A (navy). Individual replicates (3 total) are plotted. Points indicate experimental data and lines the best fit steady state model.

c) Correlation between experimentally measured and model-predicted relative growth rates for seven DHFR variants in four TYMS backgrounds. Each point represents one DHFR/TYMS combination. Error bars in the x direction are SEM across triplicate growth rate measurements, error bars in y are the SEM estimated from jackknife (leave-one-out) sub-sampling the data and refitting the model.

d) Heatmaps of simulated growth rates (top row) and epistasis (bottom row) as computed over a range of DHFR kinetic parameters in four TYMS backgrounds. In the left-most column of heatmaps a red star marks the highest activity enzyme (low K_m, high k_{cat}), while a yellow star marks the lowest activity enzyme. Growth rates are indicated with a blue-white-red color map, where a relative growth rate of one (white) is equivalent to WT. Epistasis values are indicated with a green-white-pink color map, where zero epistasis is shown in white.

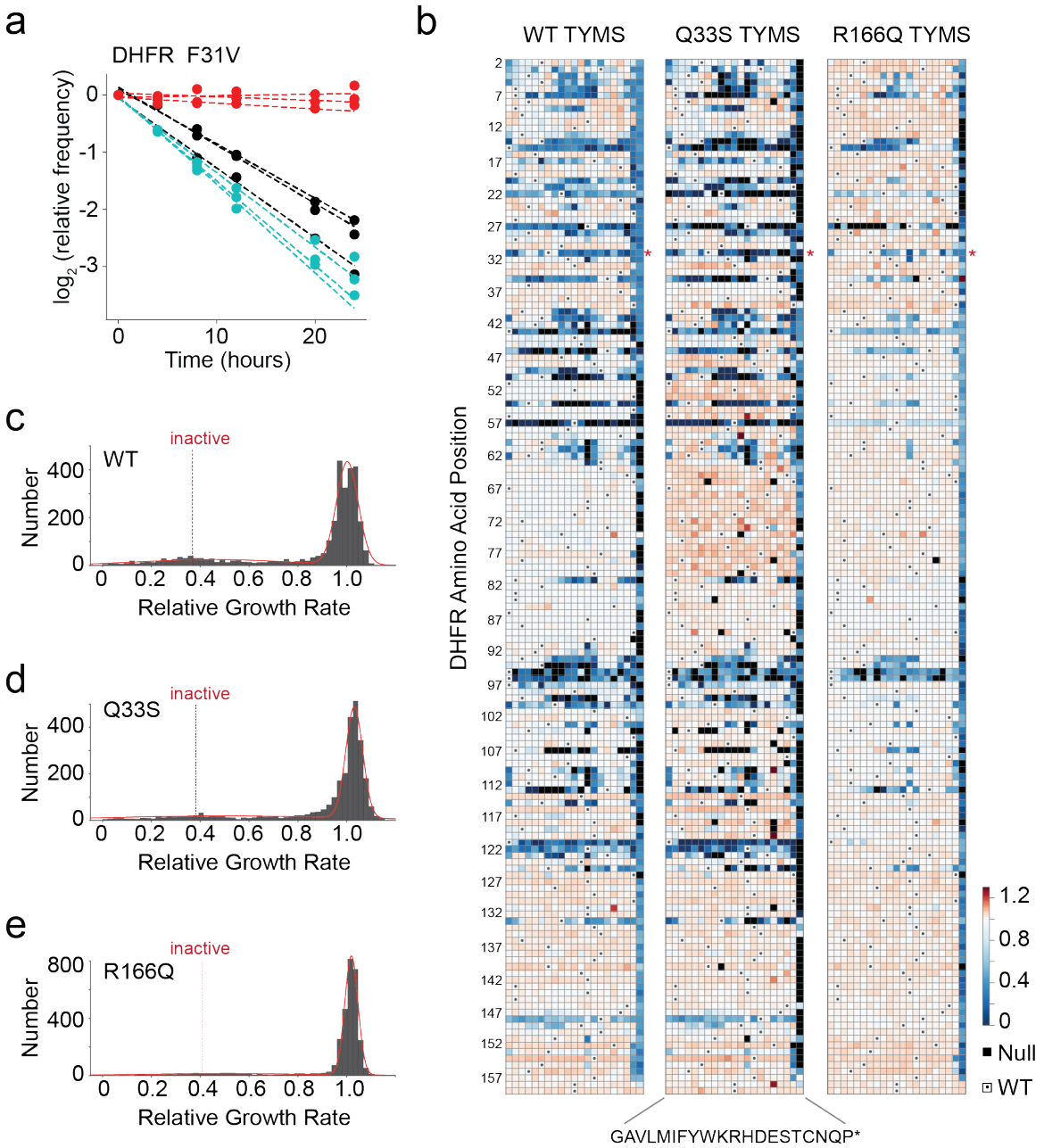


Figure 3. The effects of DHFR mutation on growth rate in three TYMS backgrounds.

a) Sequencing-based growth rate measurements for DHFR F31V in three TYMS backgrounds: R166Q (red), Q33S (cyan), and WT (black). Each point represents one triplicate experimental measurement. Dotted lines indicate linear regression fits to each replicate, the slope of each line is the inferred growth rate (relative to WT) for that DHFR/TYMS mutant combination.

b) Heatmaps of the growth rate effect for all DHFR single mutations. DHFR positions are along the horizontal axis; amino acid residues (along the vertical axis) are organized by physiochemical similarity. The displayed relative growth rate is an average across three replicates, and is normalized such that the WT DHFR is equal to one. Red indicates mutations that increase growth rate, white indicates mutations with wild-type like growth, and blue indicates mutations that decrease growth rate. Null mutations (black squares) were not

observed by sequencing after the first two time points, and thus there was insufficient data for growth rate inference. Small dots mark the WT residue identity in each column.

c) The distribution of DHFR mutational effects in the WT TYMS background. The red line indicates a best-fit double gaussian, grey bars are the data. The red, dashed “inactive” line marks the average relative growth rate for nonsense mutations (stop codons) in the first 120 positions of DHFR. The WT DHFR growth rate is equal to one.

d) The distribution of DHFR mutational effects in the TYMS Q33S background, color coding identical to (**c**)

e) The distribution of DHFR mutational effects in the TYMS R166Q background, color coding identical to (**c**). Note that the y-axis for (**e**) is distinct from (**c**) and (**d**).

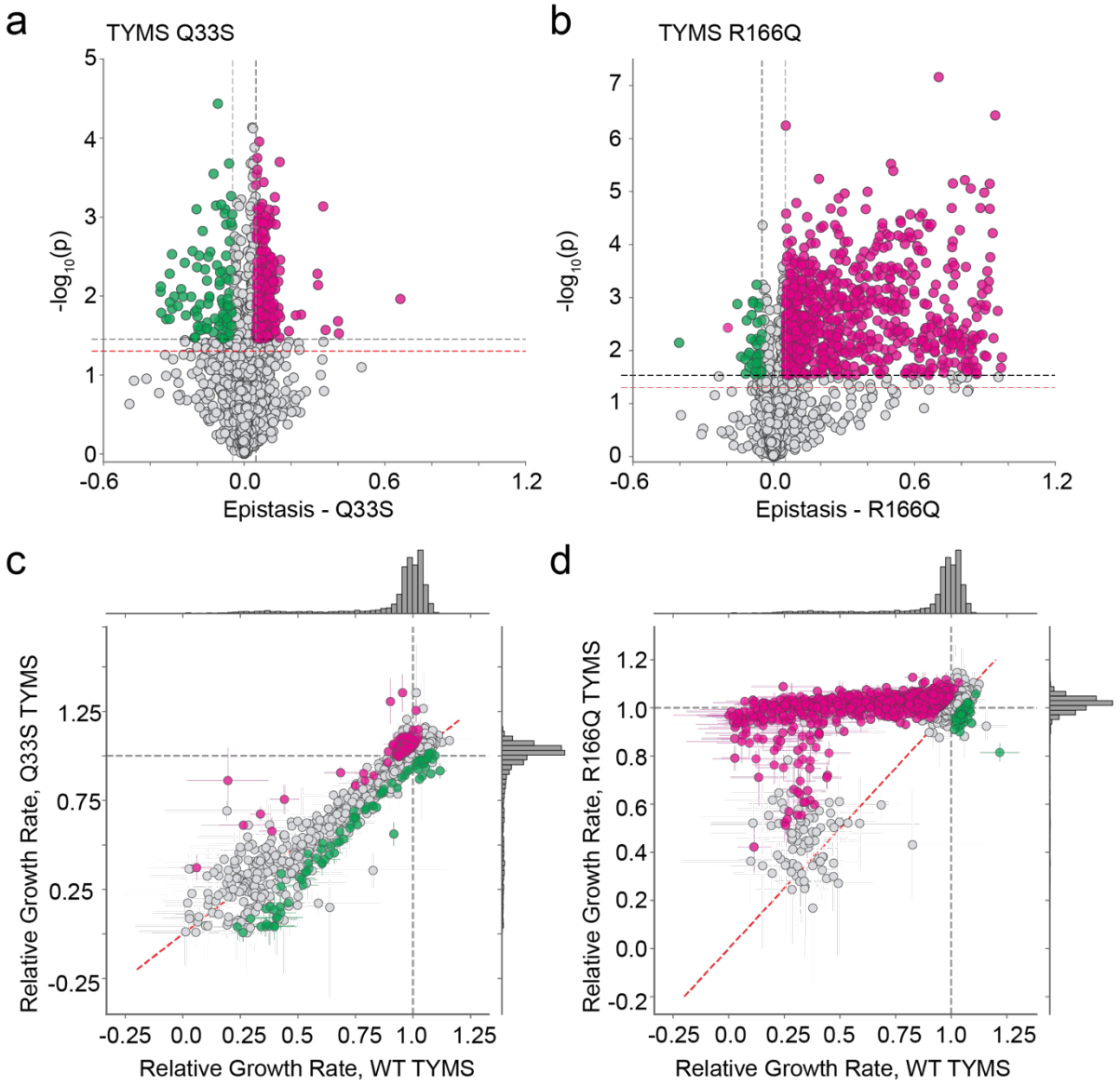
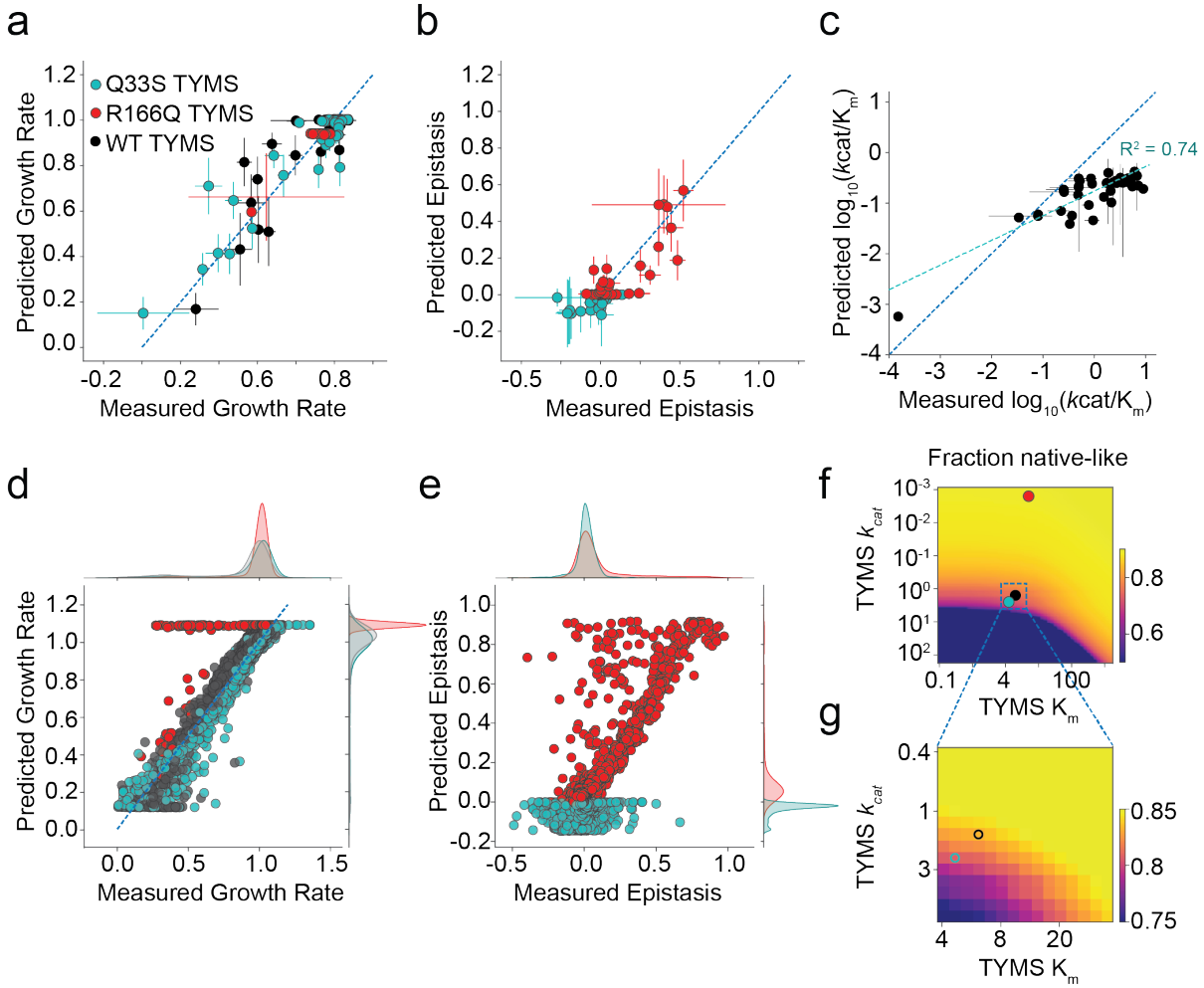


Figure 4. Epistatic coupling of DHFR to two TYMS backgrounds.

a) Volcano plot examining the statistical significance of epistasis across all DHFR point mutations in the Q33S background. P-values were calculated by unequal variance t-test under the null hypothesis that the mutations have equal mean growth rates in both TYMS backgrounds. The red horizontal dashed line marks the standard significance cutoff of P=0.05, the black horizontal dashed line indicates a multiple-hypothesis testing adjusted p-value (P=0.035). The grey vertical dashed lines indicate an empirical threshold for epistasis. Pink and green indicate statistically significant positive and negative epistasis respectively.

b) Volcano plot examining the statistical significance of epistasis across all DHFR point mutations in the R166Q background. P-values were calculated by unequal variance t-test under the null hypothesis that the mutations have equal mean growth rates in both TYMS backgrounds. The red horizontal dashed line marks the standard significance cutoff of P=0.05, the black horizontal dashed line indicates a multiple-hypothesis testing adjusted p-value (P=0.029). The grey vertical dashed lines indicate an empirical threshold for epistasis. Pink and green indicate statistically significant positive and negative epistasis respectively.

843 **c)** Comparison of the relative growth rate effects for DHFR single mutants in the WT and TYMS
844 Q33S backgrounds. The marginal distribution of growth rate effects is shown along each axis.
845 Mutations with statistically significant positive and negative epistasis are indicated in pink and
846 green respectively. The WT relative growth rate equals one, and is indicated with a dashed
847 grey line across each axis. The dashed red line marks x=y.
848 **d)** Comparison of the relative growth rate effects for DHFR single mutants in the WT and TYMS
849 R166Q backgrounds. Plot layout and color coding is identical to (**c**).
850
851



852 **Figure 5. Global comparison of the biochemistry-to-growth model and deep mutational**
853 **scanning data set.**
854 **a)** Correlation between the experimentally measured and predicted growth rates of 114
855 DHFR/TYMS mutant combinations (circles colored according to TYMS background).
856 Horizontal error bars indicate standard deviation in experimentally measured growth rates
857 across three replicate measurements, vertical error bars are the standard deviation in the
858 predicted growth rates estimated by performing 1000 bootstrap re-samplings (and model fits)
859 of the data. The dashed grey line indicates y=x.
860 **b)** Correlation between the experimentally measured and model-predicted epistasis, as
861 computed from the growth rate data in (**a**). Again, color coding indicates TYMS background
862 (identical to **a**). The dashed grey line indicates y=x.
863 **c)** Correlation between the experimentally measured and computationally inferred $\log_{10}(k_{cat}/K_m)$
864 values for 38 mutants of DHFR. Horizontal error bars describe the standard deviation across
865 triplicate experimental measurements, vertical error bars indicate the standard deviation
866 across 50 iterations of stochastic (Monte-Carlo based) model inference.
867 **d)** Correlation between experimentally measured and predicted growth rates across the entire
868 deep mutational scanning dataset. The marginal distribution of growth rate effects is shown
869 along each axis.
870 **e)** Correlation between experimentally measured and predicted epistasis across the entire deep
871 mutational scanning dataset. The marginal distribution of epistatic effects is shown along
872 each axis.

f) Mutational tolerance of DHFR as a function of TYMS background. The heatmap shows the fraction of DHFR mutations with growth rates of 0.9 or better as TYMS k_{cat} and K_m are discretely varied. The values for TYMS R166Q, Q33S and WT are marked with red, cyan and black circles respectively.

g) A zoomed-in version of (**f**), focusing on the mutational tolerance of DHFR for TYMS backgrounds similar in velocity to WT and Q33S TYMS.

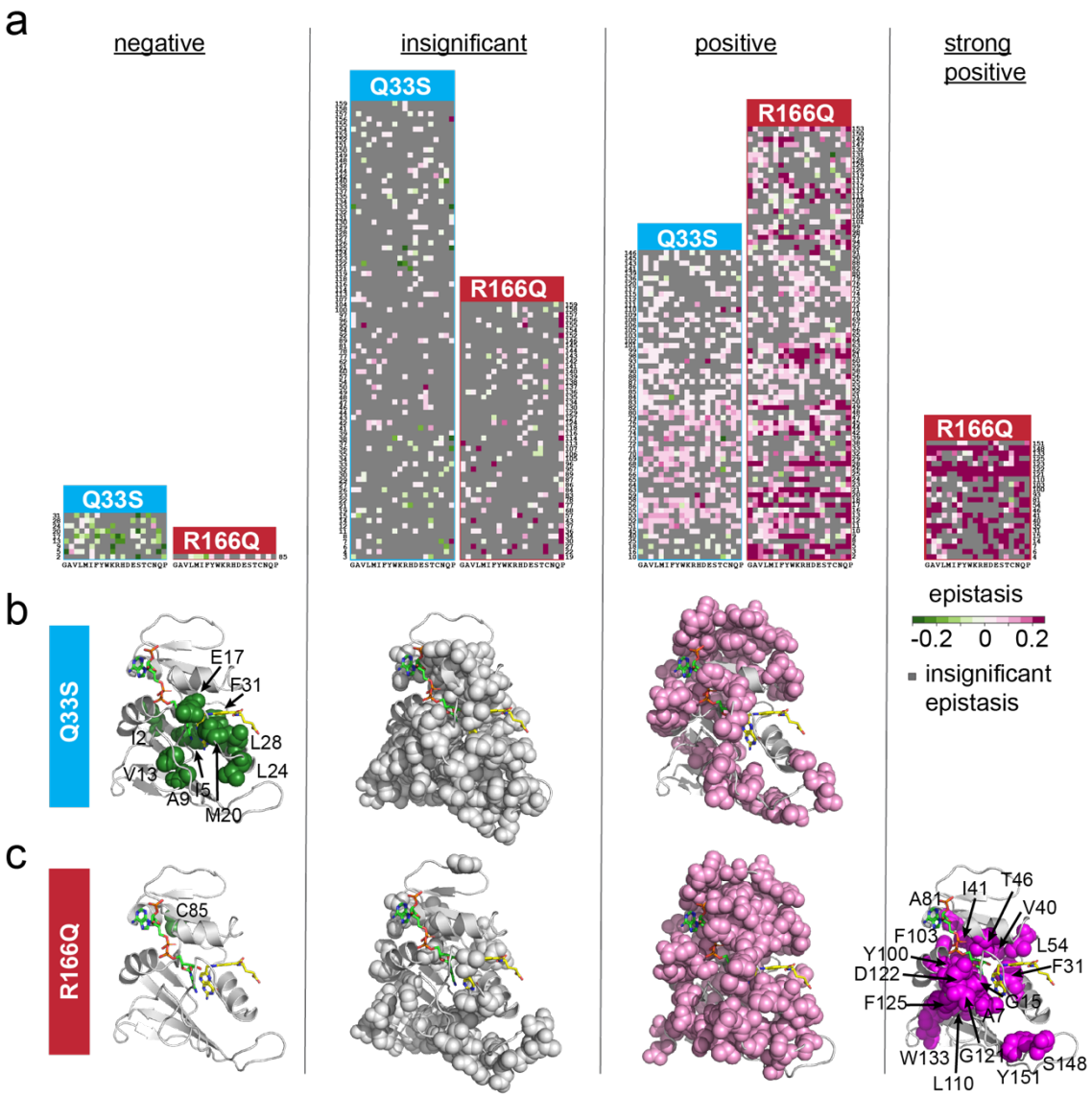


Figure 6. DHFR positions clustered by epistatic mutational effect.

a) Clusters of DHFR positions organized by predominant epistasis type. In each heat map DHFR positions are ordered along the vertical axis; amino acid residues are organized by physiochemical similarity along the horizontal axis. As in earlier plots, green indicates negative epistasis, and pink indicates positive epistasis. Grey pixels mark mutations with statistically insignificant epistasis.

b) Structural location of epistatic clusters for DHFR to TYMS Q33S. The DHFR backbone is in grey cartoon (PDBID: 1RX2[30]). Folate, the DHFR substrate is indicated with yellow sticks. The NADP+ cofactor is in green sticks.

c) Structural location of epistatic clusters for DHFR to TYMS R166Q. Color coding is identical to panel (**b**).

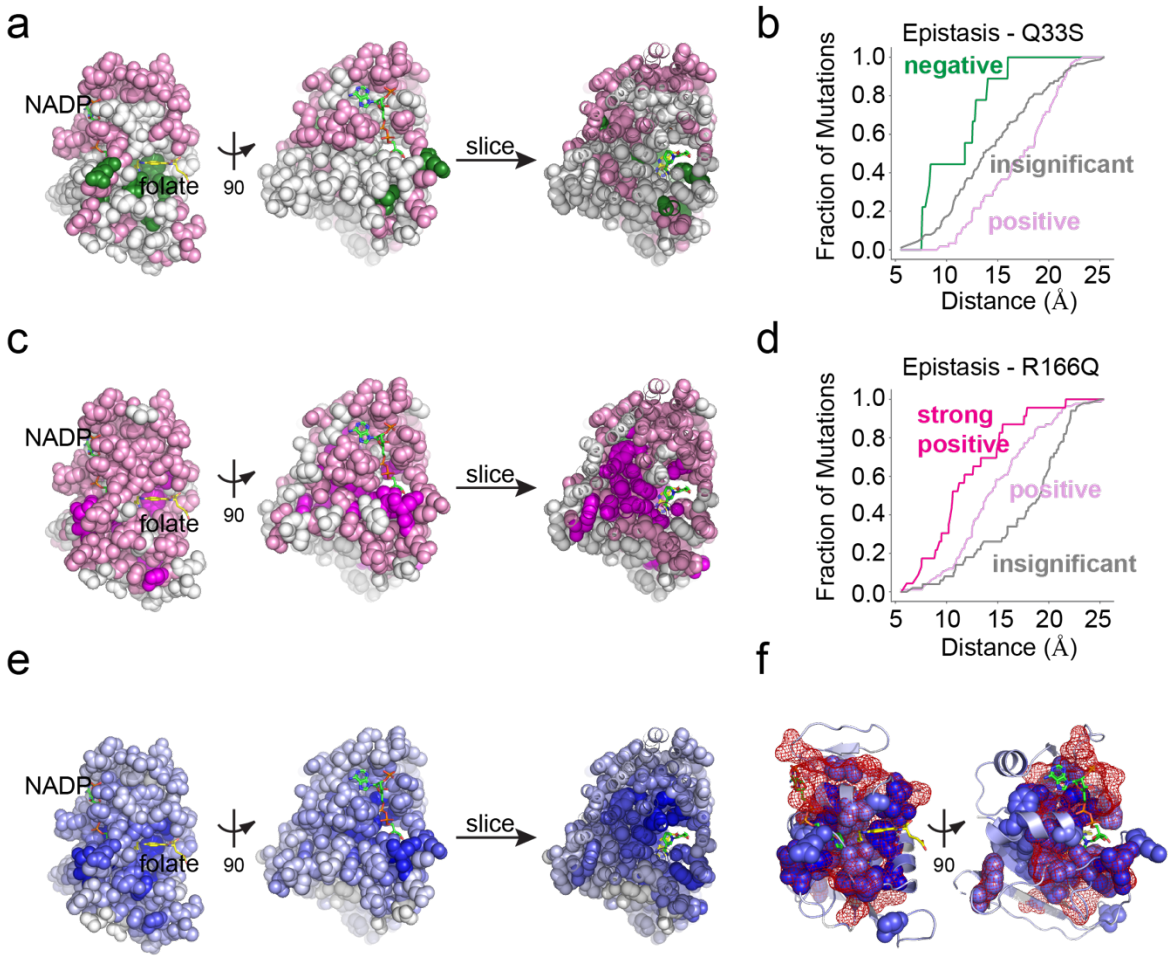


Figure 7. The structural organization of epistasis in DHFR.

a) Epistasis of individual DHFR positions to TYMS Q33S. The DHFR structure is shown in space filling spheres (PDBID: 1RX2), with the NADP co-factor in green sticks, and folate in yellow sticks. A slice through the structure shows the interior arrangement of epistasis. Positions in the negative epistasis cluster are colored green, positions in the positive epistasis cluster are colored pink. Grey spheres indicate positions in the insignificant epistasis cluster.

b) Cumulative distribution of positions in each epistatic cluster by distance to the DHFR active site for the TYMS Q33S background. In this case, active site was defined as the C6 atom of the folate substrate. Color coding follows from (**a**)

c) Epistasis of individual DHFR positions to TYMS R166Q. The DHFR structure is shown in space filling spheres (PDBID: 1RX2), with the NADP co-factor in green sticks, and folate in yellow sticks. A slice through the structure shows the interior arrangement of epistasis. Positions with strong positive epistasis are colored magenta, positions in the positive epistasis cluster are colored pink. Grey spheres indicate positions in the insignificant epistasis cluster.

d) Cumulative distribution of positions in each epistatic cluster by distance to the DHFR active site for the TYMS R166Q background. In this case, active site was defined as the C6 atom of the folate substrate. Color coding follows from (**c**)

e) The average effect of mutations on $\log_{10}$ catalytic power. All residues are indicated in space filling and color coded by the average mutational effect. Blue indicates positions where mutations have a deleterious effect on catalytic power (on average), while white indicates mutations that have little to no effect on catalytic power. Again, the NADP co-factor is shown

in green sticks, and folate in yellow sticks. A slice through the structure shows the interior distribution of mutational effects on catalysis.

f) The structural overlap between positions associated to catalysis and evolutionary conservation. The DHFR backbone is shown in grey cartoon, the NADP co-factor in green sticks, and folate in yellow sticks. Positions where mutations have (on average) a deleterious effect on catalysis are shown in blue space filling (color coding identical to **c**). Evolutionarily conserved positions (as computed by the Kullback-Leibler relative entropy in a large alignment of DHFR sequences) are outlined in red mesh.

927 **DATA AND MATERIALS AVAILABILITY**

928 The DHFR deep mutational scanning libraries (in all three TYMS backgrounds) have been
929 deposited at addgene under **deposit number 81596**.
930
931 Code for the enzyme velocity to growth rate model, and analysis of all deep mutational scanning
932 data is available on github: https://github.com/reynoldsk/dhfr-tyms-epistasis
933
934 The raw sequencing data is available in FASTQ format through the NCBI sequencing read
935 archive, under **BioProject ID PRJNA791680**.
936

937 **REFERENCES**

938 1. Schober, A. F. *et al.* A two-enzyme adaptive unit within bacterial folate metabolism. *Cell*
939 *Reports* **27**, 3359-3370.e7 (2019).
940 2. Kemble, H. E. *et al. Flux, toxicity and protein expression costs shape genetic interaction in a*
941 *metabolic pathway*. http://biorxiv.org/lookup/doi/10.1101/362327 (2018)
942 doi:10.1101/362327.
943 3. Lilja, E. E. & Johnson, D. R. Metabolite toxicity determines the pace of molecular evolution
944 within microbial populations. *BMC Evol Biol* **17**, 52 (2017).
945 4. Noor, E. *et al.* The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws
946 and Cost Minimization. *PLOS Comput. Biol.* **12**, e1005167 (2016).
947 5. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein.
948 *Nature* **436**, 588–592 (2005).
949 6. Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical
950 Trends Shaping Enzyme Parameters. *Biochemistry* **50**, 4402–4410 (2011).
951 7. Lalanne, J.-B. *et al.* Evolutionary Convergence of Pathway-Specific Enzyme Expression
952 Stoichiometry. *Cell* **173**, 749-761.e38 (2018).
953 8. Kacser, H. & Burns, J. A. The control of flux. *Biochem Soc Trans* **23**, 341–366 (1995).
954 9. Dykhuizen, D. E., Dean, A. M. & Hartl, D. L. Metabolic Flux and Fitness. *Genetics* **115**, 25–
955 31 (1987).
956 10. Kacser, H. & Burns, J. A. The Molecular Basis of Dominance. *Genetics* **97**, 639–666 (1981).

957 11. Domingo, J., Baeza-Centurion, P. & Lehner, B. The Causes and Consequences of Genetic
958 Interactions (Epistasis). *Annu. Rev. Genom. Hum. Genet.* **20**, 433–460 (2019).
959 12. Pokusaeva, V. O. *et al.* An experimental assay of the interactions of amino acids from
960 orthologous sequences shaping a complex fitness landscape. *PLoS Genet.* **15**, e1008079
961 (2019).
962 13. Kondrashov, A. S., Sunyaev, S. & Kondrashov, F. A. Dobzhansky-Muller incompatibilities in
963 protein evolution. *Proc Natl Acad Sci U S A* **99**, 14878–14883 (2002).
964 14. Noda-Garcia, L., Liebermeister, W. & Tawfik, D. S. Metabolite-Enzyme Coevolution: From
965 Single Enzymes to Metabolic Pathways and Networks. *Annu Rev Biochem* **87**, 187–216
966 (2018).
967 15. Ducker, G. S. & Rabinowitz, J. D. One-Carbon Metabolism in Health and Disease. *Cell*
968 *Metabolism* **25**, 27–42 (2017).
969 16. Nijhout, H. F., Reed, M. C., Budu, P. & Ulrich, C. M. A Mathematical Model of the Folate
970 Cycle: NEW INSIGHTS INTO FOLATE HOMEOSTASIS. *J. Biol. Chem.* **279**, 55008–55016
971 (2004).
972 17. Leduc, D. *et al.* Flavin-Dependent Thymidylate Synthase ThyX Activity: Implications for the
973 Folate Cycle in Bacteria. *Journal of Bacteriology* **189**, 8537–8545 (2007).
974 18. King, C. H., Shlaes, D. M. & Dul, M. J. Infection caused by thymidine-requiring,
975 trimethoprim-resistant bacteria. *Journal of Clinical Microbiology* **18**, 79–83 (1983).
976 19. Howell, E. E., Foster, P. G. & Foster, L. M. Construction of a dihydrofolate reductase-
977 deficient mutant of Escherichia coli by gene replacement. *J Bacteriol* **170**, 3040–3045
978 (1988).
979 20. Rodrigues, J. V. & Shakhnovich, E. I. Adaptation to mutational inactivation of an essential
980 gene converges to an accessible suboptimal fitness peak. *Elife* **8**, e50509 (2019).

981 21. Kwon, Y. K., Higgins, M. B. & Rabinowitz, J. D. Antifolate-Induced Depletion of Intracellular
982 Glycine and Purines Inhibits Thymineless Death in *E. coli*. *ACS Chem. Biol.* **5**, 787–795
983 (2010).
984 22. Kwon, Y. K. *et al.* A domino effect in antifolate drug action in Escherichia coli. *Nat Chem Biol*
985 **4**, 602–608 (2008).
986 23. Rodrigues, J. V. *et al.* Biophysical principles predict fitness landscapes of drug resistance.
987 *PNAS* **113**, E1470–E1478 (2016).
988 24. Ferrell, J. E. & Ha, S. H. Ultrasensitivity part I: Michaelian responses and zero-order
989 ultrasensitivity. *Trends in Biochemical Sciences* **39**, 496–503 (2014).
990 25. Goldbeter, A. & Koshland, D. E. An amplified sensitivity arising from covalent modification in
991 biological systems. *Proceedings of the National Academy of Sciences* **78**, 6840–6844
992 (1981).
993 26. Michaels, M. L., Kim, C. W., Matthews, D. A. & Miller, J. H. Escherichia coli thymidylate
994 synthase: amino acid substitutions by suppression of amber nonsense mutations. *PNAS* **87**,
995 3957–3961 (1990).
996 27. Schnell, J. R., Dyson, H. J. & Wright, P. E. Structure, Dynamics, and Catalytic Function of
997 Dihydrofolate Reductase. *Annual Review of Biophysics and Biomolecular Structure* **33**,
998 119–140 (2004).
999 28. Shrimpton, P. & Allemann, R. K. Role of water in the catalytic cycle of E. coli dihydrofolate
1000 reductase. *Protein Sci* **11**, 1442–1451 (2002).
1001 29. Gekko, K., Kunori, Y., Takeuchi, H., Ichihara, S. & Kodama, M. Point Mutations at Glycine-
1002 121 of Escherichia coli Dihydrofolate Reductase: Important Roles of a Flexible Loop in the
1003 Stability and Function. *J Biochem* **116**, 34–41 (1994).
1004 30. Sawaya, M. R. & Kraut, J. Loop and subdomain movements in the mechanism of
1005 *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry* **36**, 586–
1006 603 (1997).

1007 31. Chen, J. T., Taira, K., Tu, C. P. D. & Benkovic, S. J. Probing the functional role of
1008 phenylalanine-31 of Escherichia coli dihydrofolate reductase by site-directed mutagenesis.
1009 *Biochemistry* **26**, 4093–4100 (1987).
1010 32. Watney, J. B., Agarwal, P. K. & Hammes-Schiffer, S. Effect of Mutation on Enzyme Motion
1011 in Dihydrofolate Reductase. *J. Am. Chem. Soc.* **125**, 3745–3750 (2003).
1012 33. Carvajal-Rodríguez, A., de Uña-Alvarez, J. & Rolán-Alvarez, E. A new multitest correction
1013 (SGoF) that increases its statistical power when increasing the number of tests. *BMC*
1014 *Bioinformatics* **10**, 209 (2009).
1015 34. Tamer, Y. T. *et al.* High-Order Epistasis in Catalytic Power of Dihydrofolate Reductase
1016 Gives Rise to a Rugged Fitness Landscape in the Presence of Trimethoprim Selection.
1017 *Molecular Biology and Evolution* **36**, 1533–1550 (2019).
1018 35. Karanicolas, J. & Kuhlman, B. Computational design of affinity and specificity at protein–
1019 protein interfaces. *Current Opinion in Structural Biology* **19**, 458–463 (2009).
1020 36. McClune, C. J., Alvarez-Buylla, A., Voigt, C. A. & Laub, M. T. Engineering orthogonal
1021 signalling pathways reveals the sparse occupancy of sequence space. *Nature* **574**, 702–
1022 706 (2019).
1023 37. Li, G.-W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying Absolute Protein
1024 Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell* **157**,
1025 624–635 (2014).
1026 38. Papp, B., Pál, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in
1027 yeast. *Nature* **424**, 194–197 (2003).
1028 39. Faure, A. J. *et al.* Mapping the energetic and allosteric landscapes of protein binding
1029 domains. *Nature* **604**, 175–183 (2022).
1030 40. Otwinowski, J. Biophysical Inference of Epistasis and the Effects of Mutations on Protein
1031 Stability and Function. *Molecular Biology and Evolution* **35**, 2345–2354 (2018).

1032 41. Markin, C. J. *et al.* Revealing enzyme functional architecture via high-throughput microfluidic
1033 enzyme kinetics. *Science* **373**, eabf8761 (2021).
1034 42. Lovelock, S. L. *et al.* The road to fully programmable protein catalysis. *Nature* **606**, 49–58
1035 (2022).
1036 43. Michener, J. K., Camargo Neves, A. A., Vuilleumier, S., Bringel, F. & Marx, C. J. Effective
1037 use of a horizontally-transferred pathway for dichloromethane catabolism requires post–
1038 transfer refinement. *eLife* **3**, e04279 (2014).
1039 44. Michener, J. K., Vuilleumier, S., Bringel, F. & Marx, C. J. Phylogeny poorly predicts the utility
1040 of a challenging horizontally transferred gene in Methylobacterium strains. *J Bacteriol* **196**,
1041 2101–2107 (2014).
1042 45. Kim, J. & Copley, S. D. Inhibitory cross-talk upon introduction of a new metabolic pathway
1043 into an existing metabolic network. *Proc Natl Acad Sci U S A* **109**, E2856-2864 (2012).
1044 46. Kachroo, A. H. *et al.* Systematic humanization of yeast genes reveals conserved functions
1045 and genetic modularity. *Science* **348**, 921–925 (2015).
1046 47. Russ, W. P. *et al.* An evolution-based model for designing chorismate mutase enzymes.
1047 *Science* **369**, 440–445 (2020).
1048 48. Bhattacharyya, S. *et al.* Transient protein-protein interactions perturb E. coli metabolome
1049 and cause gene dosage toxicity. *Elife* **5**, e20309 (2016).
1050 49. Bershtein, S. *et al.* Protein homeostasis imposes a barrier on functional integration of
1051 horizontally transferred genes in bacteria. *PLoS Genet* **11**, e1005612 (2015).
1052 50. Reynolds, K. A., McLaughlin, R. N. & Ranganathan, R. Hot Spots for Allosteric Regulation
1053 on Protein Surfaces. *Cell* **147**, 1564–1575 (2011).
1054 51. Thompson, S., Zhang, Y., Ingle, C., Reynolds, K. A. & Kortemme, T. Altered expression of a
1055 quality control protease in E. coli reshapes the in vivo mutational landscape of a model
1056 enzyme. *eLife* **9**, e53476 (2020).

1057 52. Toprak, E. *et al.* Building a morbidostat: an automated continuous-culture device for
1058 studying bacterial drug resistance under dynamically sustained drug inhibition. *Nature*
1059 *Protocols* **8**, 555–567 (2013).
1060 53. Changchien, L.-M. *et al.* High-Level Expression of Escherichia coli and Bacillus subtilis
1061 Thymidylate Synthases. *Protein Expression and Purification* **19**, 265–270 (2000).
1062 54. Reynolds, K. A., McLaughlin, R. N. & Ranganathan, R. Hot Spots for Allosteric Regulation
1063 on Protein Surfaces. *Cell* **147**, 1564–1575 (2011).
1064 55. McCormick, J. W., Russo, M. A., Thompson, S., Blevins, A. & Reynolds, K. A. Structurally
1065 distributed surface sites tune allosteric regulation. *Elife* **10**, e68346 (2021).
1066 56. Agrawal, N., Hong, B., Mihai, C. & Kohen, A. Vibrationally enhanced hydrogen tunneling in
1067 the Escherichia coli thymidylate synthase catalyzed reaction. *Biochemistry* **43**, 1998–2006
1068 (2004).
1069 57. Wang, Z. *et al.* Mg2+ Binds to the Surface of Thymidylate Synthase and Affects Hydride
1070 Transfer at the Interior Active Site. *J. Am. Chem. Soc.* **135**, 7583–7592 (2013).
1071 58. Stout, T. J., Sage, C. R. & Stroud, R. M. The additivity of substrate fragments in enzyme–
1072 ligand binding. *Structure* **6**, 839–848 (1998).
1073