A two-enzyme adaptive unit within bacterial folate metabolism

Andrew F. Schober[1,2], Andrew D. Mathis[1,2], Christine Ingle[1], Junyoung O. Park[3,4,‡], Li Chen[3,5], Joshua D. Rabinowitz[3,5], Ivan Junier[6], Olivier Rivoire[7] and Kimberly A. Reynolds[1,2*]

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

[3]*Lewis-Sigler Institute for Integrative Genomics,* [4]*Department of Chemical and Biological Engineering, and* [5]*Department of Chemistry, Princeton University, Princeton, NJ 08544, USA*

[6]*Centre National de la Recherche Scientifique (CNRS), Université Grenoble Alpes, TIMC-IMAG, F-38000 Grenoble, France*

[7]*Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, PSL Research University, F-75005 Paris, France*

‡*Current address: Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, Los Angeles, CA 90095, USA*

*Please address correspondence to:

Kimberly Reynolds
Green Center for Systems Biology
UT Southwestern Medical Center
6001 Forest Park Rd.
Dallas, TX 75390-8597

Ph: 214-645-6157
kimberly.reynolds@utsouthwestern.edu

Short title: Comparative genomics and adaptive interactions in folate metabolism

Keywords:

co-evolution, comparative genomics, folate metabolism, dihydrofolate reductase (DHFR), thymidylate synthase, experimental evolution

Metabolic enzyme function and evolution is influenced by the larger context of a biochemical pathway – deleterious mutations or perturbations in one enzyme can often be compensated by mutations to others. To explore strategies for mapping adaptive dependencies between enzymes, we used a combination of comparative genomics and experiments to examine interactions with the model metabolic enzyme Dihydrofolate Reductase (DHFR). Biochemically, DHFR shares a metabolic intermediate with numerous folate metabolic enzymes. In contrast, comparative genomics analyses of synteny and gene co-occurrence indicate a sparse pattern of evolutionary couplings in which DHFR is coupled to the enzyme thymidylate synthase (TYMS), but is relatively independent from the rest of folate metabolism. To test this apparent modularity, we used quantitative growth rate measurements and forward evolution in *E. coli* to demonstrate that the two enzymes are coupled to one another, and can adapt independently from the remainder of the genome. Mechanistically, the coupling between DHFR and TYMS is driven by a constraint wherein TYMS activity must not greatly exceed that of DHFR – both to avoid depletion of reduced folates and prevent accumulation of the metabolic intermediate dihydrofolate. Extending our comparative genomics analyses genome-wide reveals over 200 gene pairs with statistical signatures similar to DHFR/TYMS, suggesting the possibility that cellular pathways might be decomposed into small adaptive units.

Introduction

The collective action of enzymes in metabolism produces the basic materials for cells to grow and divide. Biochemical pathway maps provide a rich description of the reactions and intermediates needed for metabolism, yet it remains difficult to predict how metabolic systems will respond to perturbation. That is, if the activity or expression level of a particular enzyme were changed, which (if any) of the other enzymes in metabolism would require compensatory modification? Because the pattern of such adaptive interactions between proteins remains largely unknown, our ability to rationally engineer new systems [1-3], understand how the cell responds to perturbations [4, 5], and quantify the relationship between mutations and disease [6, 7] is limited. An ability to globally map dependencies between proteins and to identify groups of enzymes within a pathway that adapt as distinct subunits would help render cellular systems more tractable and predictable.

To begin to address this problem, we used a combination of comparative genomics and experiments to measure the pattern of adaptive interactions with the folate metabolic enzyme Dihydrofolate Reductase (DHFR). DHFR is an essential enzyme involved in the synthesis of purine nucleotides, thymidine and several amino acids [8]. As such, it is a common target of antibiotics, antimalarials, and chemotherapeutics [9, 10]. In recent years, it has become a prominent model system for understanding the evolution of drug resistance [11-15], evolution of protein conformational dynamics [16-18], and constraints on horizontal gene transfer [19, 20]. However, it remains unclear how changes to other enzymes in folate metabolism influence the relationship between DHFR function and cellular fitness. Understanding how mutations in other enzymes can compensate for loss of DHFR function is important to questions of antibiotic

resistance and the evolution of folate metabolism. More generally, DHFR provides a well-studied model system to test computational approaches for predicting adaptive interactions.

Examining patterns of gene neighborhood and gene loss/gain across species presents a practical strategy for inferring genes that share an adaptive constraint. We used analyses of both synteny (conservation of chromosomal proximity) and gene co-occurrence across 1445 bacterial genomes to create a statistical map of evolutionary couplings within folate metabolism. Though the folate biochemical pathway map is highly interconnected (many enzymes share a metabolic intermediate), our comparative genomics analyses suggest a sparse architecture of adaptive interactions in folate metabolism. Four small groups of genes appear as independent co-evolutionary units: the genes within a group co-evolve with each other, but each group appears to evolve relatively independently from the remainder of folate metabolism (and the genome). One of these gene groups contains the enzymes DHFR and Thymidylate Synthase (TYMS). Overall, these results suggest that small enzyme groups might adapt as relatively independent units inside the broader context of folate metabolism.

To test whether DHFR and TYMS are more coupled to each other than to the remainder of the biochemical pathway, we used quantitative measurements of growth rate, metabolomics and forward evolution in *E. coli* to map couplings to both DHFR and TYMS. We demonstrate that these two enzymes are (1) coupled through a shared metabolite, (2) are relatively less coupled to the remainder of folate metabolism and (3) can adapt independently from the remainder of the genome. In particular, we observe that mutations in these two enzymes are sufficient for *E. coli* to acquire resistance to the inhibition of DHFR by trimethoprim. Extending our statistical analyses of co-evolution genome wide reveals additional gene pairs that co-evolve with one another yet are relatively independent from the rest of the genome. This provides a set

of hypotheses for future experimental testing, and suggests that the finding of small adaptive units embedded within larger cellular systems may be a general feature of cellular networks.

Results

A comparative genomics-based map of evolutionary coupling in folate metabolism

Starting from DHFR (encoded by the *E. coli* gene *folA*), we selected 16 folate metabolic genes for study using information from the STRING database (v10.5 , [21]) and KEGG pathway maps [22]. STRINGdb is an online database of known and predicted protein interactions that uses a combination of information – including synteny, co-occurrence, online databases, co-expression, high-throughput experiments (e.g. yeast two hybrid data), and automated text mining – to produce an amalgamated confidence score [21]. This amalgamated score is benchmarked to predict the likelihood that two enzymes are in the same KEGG pathway. We selected all highest confidence interactions to DHFR (encoded by the *E. coli* gene *folA*). This results in fourteen genes, all of which encode an enzyme sharing a product or substrate with DHFR. We added two more genes (*metF*, *folD*) to this set to complete the folate cycle on the basis of KEGG pathway information.

This selection of genes encompasses core folate metabolism – a set of reactions that interconvert various folate species and produce methionine, glycine, thymidine and purine nucleotides in the process (Fig 1A and Table S1) [8]. The input of the pathway is 7,8-dihydrofolate (DHF), produced by the bifunctional enzyme dihydrofolate synthase/folylpolyglutamate synthetase (FPGS) through the addition of L-glutamate to dihydropteroate. Once DHF is formed, it is reduced to 5,6,7,8-tetrahydrofolate (THF) by the enzyme dihydrofolate reductase (DHFR) using NADPH as a co-factor. THF can then be

modified by a diversity of one-carbon groups at both the N-5 and N-10 positions, and subsequently serves as a one-carbon donor in several critical reactions, including the synthesis of a few amino acids and purine nucleotides (Fig 1A, bottom square portion of pathway). The only step that oxidizes reduced folate (5,10-Methylene THF) back to DHF is catalyzed by the enzyme thymidylate synthase (TYMS), which modifies uridine monophosphate (dUMP) to thymidine monophosphate (dTMP) in the process.

Figure 1. Biochemical and statistical representations of folate metabolism. **A**, Biochemical pathway map of folate metabolism. Blue font indicates metabolites, abbreviated enzyme (*gene*) names are in black or grey text. Black text and lines correspond to enzymes annotated as highest confidence interactions for DHFR (*folA*) in STRINGdb v10.5. See Table S1 for a more complete description of each enzyme. **B-C,** Heatmaps of evolutionary coupling between gene pairs in folate metabolism as evaluated by gene synteny and co-occurrence. Pixel color shows the extent of coupling between genes, indicated as a relative entropy, D_{ij}^{intra}. Enzyme names are indicated on the left and top of the matrix in black text, the corresponding gene names are given at right in grey italics. In *E. coli*, a single gene (*folD*) encodes a bifunctional enzyme that catalyzes both the methylene tetrahydrofolate dehydrogenase (MTD) and methenyltetrahydrofolate cyclohydrolase (MTCH) reactions in the biochemical pathway as shown in A. The majority of gene pairs show little co-evolution by either measure (dark purple pixels). See also Fig S1.

Consistent with the stated goals of the STRINGdb (to recover biochemical pathways), the STRING confidence scores indicate coupling between nearly all pairs of enzymes in folate metabolism. Both STRINGdb and the associated pathway map suggest that folate metabolism is

92 highly interconnected, with many enzymes interacting via a shared product or substrate (Fig 1A,
93 Fig S1A). To examine the extent to which these biochemical interactions lead to evolutionary
94 couplings, we focused our analysis on measures that report on evolutionary selection across
95 organisms. We computed two different statistical measures of evolutionary association across
96 1445 bacterial genomes: 1) synteny, the conservation of chromosomal proximity between genes
97 and 2) co-occurrence, the coordinated loss and gain of genes across species [23-25]. Both
98 approaches have been established as reliable indicators of protein functional relationships [24,
99 26-31], and technical variations on these methods are components of the amalgamated STRING
100 score. These approaches are expected to report on any interaction important to evolutionary
101 fitness, and can thus indicate gene pairs that interact through various mechanisms, including
102 physical binding or biochemical constraints.

103 In contrast to the amalgamated STRING score, analyzing these two signals individually
104 results in a sparse pattern of evolutionary couplings wherein most folate metabolic genes are
105 relatively independent from each other (Fig 1B-C, Fig S1B-C). Consistent with expectation, we
106 do observe evolutionary couplings between physically interacting gene products: the glycine
107 cleavage system proteins H, P and T (*gcvH*, *gcvP*, and *gcvT* in *E. coli*) form a macromolecular
108 complex [32]. We also see evolutionary couplings for three enzyme pairs: 1) DHFR/TYMS 2)
109 methionine synthase (MS) with methionine tetrahydrofolate reductase (MTHFR) and 3) the
110 purine biosynthesis proteins PGT and IMPS. While these enzyme pairs are not known to
111 physically bind, each pair of enzymes share a metabolic intermediate, suggesting that a
112 biochemical mechanism underlies their evolutionary coupling. Yet, most enzymes that catalyze
113 neighboring reactions do not show statistical correlation (Fig 1, Fig S1). Instead, synteny and co-

114 occurrence provide a different representation of folate metabolism in which most genes are
115 relatively independent and a few interact to form evolutionarily constrained units.

116 So what factors might contribute to the observed sparsity of evolutionary couplings? One
117 potential contributor is limited sensitivity of our methods. Indeed, false negatives – gene pairs
118 that are evolutionarily coupled but without a strong synteny or co-occurrence signal – are
119 expected due to limited statistical power. So perhaps the apparent modularity of the
120 DHFR/TYMS pair within folate metabolism just reflects an incomplete pattern of predicted
121 interactions. We find only six evolutionary couplings out of 120 possible pairwise interactions,
122 even though 83 enzyme pairs interact biochemically (by sharing a product or substrate). Thus, a
123 high false negative rate would be necessary to completely explain the difference between the
124 pattern of biochemical interactions and the pattern of evolutionary couplings.

125 More generally, the pattern of adaptive interactions between enzymes need not strictly
126 resemble the pattern of biochemical interactions, given the non-linear relationship between
127 enzyme activities, metabolite concentrations, and fitness. The observed sparsity might then
128 reflect a modular organization of adaptive constraints between enzymes, shaped by evolution.
129 The evolutionary couplings are a statistical inference across thousands of genomes, and are not
130 expected to capture idiosyncratic couplings specific to a particular organism or environmental
131 condition. Instead, we hypothesize that the evolutionary coupling map presents predictions of the
132 major adaptive couplings (and of the degree of independence) that do not strongly depend on
133 choice of organism or environment. If true, this would suggest a route to begin decomposing
134 metabolic pathways into meaningful, smaller multi-gene units.

135 As a first step towards testing this broader hypothesis, we set out to experimentally test if
136 the statistical pattern of evolutionary coupling observed for DHFR and TYMS corresponds to the

137 pattern of adaptive coupling in an individual organism, namely *E. coli*. In particular, are DHFR
138 and TYMS indeed more coupled to each other than to other enzymes in the pathway, as our
139 statistical results suggest? And can perturbations in TYMS suffice to compensate for
140 perturbations in DHFR (without the need to modify other enzymes)? These proteins are not
141 known to physically interact, providing an opportunity to better understand how biochemical
142 constraints might drive co-evolution. Though the genes encoding DHFR and TYMS are
143 proximal along the chromosome of many bacterial species, they are approximately 2.9
144 megabases apart in *E. coli*. This means our experiments will determine whether adaptive
145 coupling between genes found in synteny persists even when they are distant on the
146 chromosome.
147
148 **The growth rate defect of DHFR knockdown is partially rescued by changes in the**
149 **expression of TYMS – but not other folate metabolic enzymes.**
150 As a first measurement of adaptive coupling between DHFR, TYMS and the rest of folate
151 metabolism, we performed targeted measurements of growth rate dependency. More specifically,
152 we decreased DHFR (or TYMS) expression, measured the effect on growth rate, and then
153 examined if the effect on growth rate was modified in the background of a second gene
154 knockdown. The null hypothesis is that the growth rate effect of reducing expression for one
155 gene should be the same regardless of whether or not the second gene is knocked down – unless
156 the two genes are coupled.
157 To make these measurements, we used CRISPR interference (CRISPRi) and a next-
158 generation sequencing based measurement of relative growth rate. In CRISPRi, a single guide
159 RNA (sgRNA) and a catalytically dead Cas9 enzyme are used to target and repress gene

160 expression [33]. Pairs of sgRNAs can be cloned into a single vector, permitting knockdown of
161 two genes at once (Fig S2A). Using this approach, we constructed pairwise knockdowns of
162 DHFR and TYMS with every other gene in the folate metabolic pathway as described in Fig 1,
163 with the exception of *lpdA* and *fmt*. We excluded *lpdA* and *fmt* from our experiments because
164 they are involved in other critical cellular processes and are expected to have pleiotropic effects
165 (*fmt* is necessary to produce the initiator methionine tRNA required for all protein translation;
166 *lpdA* binds to the GDC but is also part of two other physical complexes in the cell). We also
167 measured the effects of knocking down each gene individually by pairing each sgRNA with an
168 sgRNA lacking the homology region needed for targeting a gene (referred to as *none*). In total,
169 this results in a small library of 14 single and 25 double gene knockdowns (Table S2-S4). The
170 knockdown activity of individual sgRNAs was verified by comparing the growth rate effect of
171 knockdown to similar data for gene knockouts (data from the Keio collection, Fig S2C), and by
172 qPCR (Fig S2D, Table S5).

173 To measure growth rate effects for the entire library, we transformed the library into
174 MG1655 K12 *E. coli* and grew the cells in a turbidostat, a device for continuous culture that
175 ensures: 1) the cells remain in exponential growth phase for the duration of the experiment 2) the
176 media conditions are constant. We used M9 minimal media containing 0.4% glucose, 0.2%
177 amicase, and 5 μg/ml thymidine for growth. These conditions were selected to be similar to
178 previous forward evolution experiments ([15], with the addition of thymidine), and provide some
179 measure of buffering for variation in the expression of folate metabolic enzymes. We took time
180 points over the course of 12 hours, and then computed a relative growth rate for each knockdown
181 in the population by using next generation sequencing to "count" the number of each sgRNA
182 pair in the population (Fig S2B). The relative growth rate indicates the per hour decrease or

183 increase in the abundance of a particular knockdown relative to wild type (carrying the targetless
184 "*none*" sgRNA).

185 As expected, we observed that knocking down DHFR expression was highly deleterious.
186 This was true whether alone or in combination with any other gene – with the exception of
187 TYMS. Knocking down DHFR and TYMS together uniquely led to a partial rescue of growth
188 rate (Fig 2A). Knocking down TYMS expression alone was also deleterious (though the effect is
189 partly buffered by the presence of thymidine), and the deleterious effect was not improved by
190 knocking down any other gene (Fig 2B). This indicates asymmetry in the coupling between
191 DHFR and TYMS in this condition: knocking down TYMS buffers DHFR, but not vice versa.
192 These results are consistent with prior attempts to generate *ΔfolA* strains of *E. coli*; when
193 attempting to delete the gene encoding DHFR, a TYMS mutation is spontaneously acquired [34].
194 Overall the data show that among folate metabolic enzymes, decreasing TYMS expression
195 uniquely compensates for reductions in DHFR expression.

196

197 **Figure 2 CRISPRi-based measurements of growth rate dependency for DHFR and TYMS**.
198 **A-B,** Relative growth rates for CRISPRi knockdowns of folate genes paired with either DHFR
199 (**A**) or TYMS (**B**). Growth rates were measured in M9 minimal media supplemented with 0.2%
200 amicase and 5μg/ml thymidine; and are calculated relative to the fitness of a strain carrying an
201 sgRNA with no target homology region (*none*). Gray bars indicate the growth rate effect of a
202 single mutant; blue bars indicate the effect of the double mutant. Error bars correspond to a 95%
203 CI across at least three internal technical replicates (see also methods). The DHFR knockdown
204 is significantly rescued by TYMS knockdown (p-value < 0.005 by Student's t-test), but not by
205 other gene knockdown (p-values > 0.20). As a reference point, the absolute doubling time of the

206 *none* strain in turbidostat mono-culture is 0.83 ± 0.09 (95% CI) hours. From this, we estimate
207 that a relative growth rate of –0.2 in these mixed population measurements corresponds to a
208 doubling time of approximately 1.8 hours, and –0.4 is approximately zero growth rate (dead).
209 To the right of each bar graph, we also plot a histogram of the growth rate effects for all
210 knockdowns of DHFR (**A**) or TYMS (**B**).
211

212 **Forward evolution indicates adaptive independence of DHFR and TYMS from the rest of**
213 **the genome**

214 The above data are consistent with the idea that DHFR and TYMS are more adaptively
215 coupled to each other than to other genes in folate metabolism, but leave open the possibility that
216 DHFR and TYMS interact with other genes in the genome. Moreover, the comparative genomics
217 analyses of synteny and co-occurrence suggest that the coupling between DHFR and TYMS
218 leads them to co-evolve as a unit. To look for compensation by other enzymes genome-wide, and
219 test the idea that DHFR and TYMS can adapt as an independent unit, we conducted a global
220 suppressor screen. We inhibited DHFR with the common antibiotic trimethoprim and then
221 examined the pattern of compensatory mutations with whole genome sequencing. If DHFR and
222 TYMS indeed represent a quasi-independent adaptive unit, then suppressor mutations should be
223 found within these two genes (*E. coli folA* and *thyA*) with minimal contribution from other sites.

224 To facilitate forward evolution in the presence of trimethoprim, we used a specialized
225 device for continuous culture called a morbidostat [15, 35, 36]. The morbidostat dynamically
226 adjusts the trimethoprim concentration in response to bacterial growth rate and total optical
227 density, thereby providing steady selective pressure as resistance levels increase (Fig S3A-C).
228 The basic principle is that cells undergo regular dilutions with fresh media until they hit a target

229 optical density (OD = 0.15); once this density is reached, they are diluted with media containing
230 trimethoprim until growth rate is decreased. This approach makes it possible to obtain long
231 trajectories of adaptive mutations in the genome with sustained phenotypic adaptation [15]. For
232 example, in a single 13-day experiment, we observe resistance levels in our evolving bacterial
233 populations that approach the trimethoprim solubility limit in minimal (M9) media.

234 As for the CRISPRi-based growth rate measurements, we conducted the forward
235 evolution experiments in M9 minimal media supplemented with 0.2% amicase. We evolved
236 three populations in parallel: in each case the media was supplemented with a different
237 concentration of thymidine (5, 10, and 50 μg/ml) chosen to range from a condition in which
238 TYMS loss-of-function is deleterious to one which fully rescues of TYMS activity. By
239 alleviating selective pressure on the entire pathway, we sought to expose a larger range of
240 adaptive mutations without biasing the pathway towards a particular result; this is similar to the
241 common practice of conducting second site suppressor screens for essential genes under
242 relatively permissive conditions [37]. In this context, an absence of mutation outside of the two-
243 gene pair becomes more significant.

244 Over the 13 days of evolution, we observed that the three populations steadily increased
245 in trimethoprim resistance (Fig S3D, Fig S4). To identify the mutations underlying this
246 phenotype, we selected 10 single colonies from the endpoint of each of the three experimental
247 conditions for phenotypic and genotypic characterization (30 strains in total). For each strain, we
248 measured the trimethoprim IC50, growth rate dependence on thymidine, and conducted whole
249 genome sequencing (Fig 3, Fig S5 and Table S6-S8). All 30 strains were confirmed to be highly
250 trimethoprim resistant (Fig 3, Table S6), with the exception of strains 4, 5 and 10 from the 50
251 μg/ml thymidine condition: these three strains grew very slowly at all concentrations of

252 trimethoprim tested (up to 1500 μg/ml). Because the dependence of growth on trimethoprim did
253 not fit a sigmoidal function for these three strains, we do not report an IC50. Further, strains from
254 all three thymidine-supplemented conditions became dependent on exogenous thymidine for
255 growth, indicating a loss of function in the *thyA* gene that encodes TYMS (Fig 3 and Fig S5). We
256 confirmed that this loss of function is not a simple consequence of neutral genetic variation in the
257 presence of thymidine; cells grown in 50 μg/ml thymidine and 0.2% amicase in the absence of
258 trimethoprim retained TYMS function over similar time scales (Fig S6).
259
260 **Figure 3 Genotype and phenotype of selected strains from three evolved populations. A-C,**
261 Ten single colonies (strains) were selected at the endpoint of each forward evolution condition
262 for genotyping and phenotyping (30 in total). Each panel contains a map indicating the mutations
263 observed in the strains sampled from that evolution condition. Genes that were mutated in two or
264 fewer strains across all conditions are excluded, as are synonymous mutations (see Table S7 for
265 sequencing statistics, and Table S8 for a complete list of mutants). Gene names are labeled along
266 the left edge of the map, with the corresponding residue or nucleotide change(s) denoted along
267 the right. If a strain acquires any mutation in a particular gene, the column section corresponding
268 to that gene is shaded blue. Nearly all strains acquired mutations in both *folA* and *thyA*, encoding
269 DHFR and TYMS, with a few exceptions lacking a *folA* mutation. A small red arrow indicates
270 one strain with mutations in only DHFR and TYMS. To the right of each mutation map are
271 trimethoprim (TMP) IC50 and thymidine dependence measurements for each strain. Error bars
272 represent standard error over triplicate measurements. Evolved strains 4,5 and 10 grew very
273 slowly at all concentrations of trimethoprim measured, and we were unable to determine an IC50
274 by sigmoidal fit. Table S6 contains exact IC50 values and errors. Thymidine dependence is

275 represented as area under the log(OD_{600}) curve in 0 μg/ml thymidine over 10 hours. Evolved
276 strains are no longer viable in the absence of extracellular thymidine, indicating a loss-of-
277 function mutation in TYMS. See Fig S5 for growth rates across a range of thymidine
278 concentrations. Three 'reconstitution strains,' featuring representative *folA* and *thyA* mutations
279 recombined into a clean genetic background, have been included in panels A and C for
280 comparison (denoted **R2-4**). Red dots in the rows of reconstituted genotypes R2-4 indicate the
281 IC50 of a strain containing *only* the corresponding *folA* mutation. The *folA* single mutant strains
282 were described in [13] and were produced through a different recombination protocol than the
283 double mutants, and thus differ in the presence of additional chromosomal antibiotic markers but
284 are otherwise identical (see Experimental Procedures).
285
286 Whole genome sequencing for all 30 strains revealed that isolates from all three
287 conditions acquired coding-region mutations in both DHFR and TYMS, or even just in TYMS
288 (Fig 3). The sequencing data for strains 4,5,7 and 10 (in 50 μg/ml thymidine) did not indicate any
289 mutations or amplifications to DHFR or the promoter. For all other strains, the mutations
290 identified in DHFR reproduce those observed in an earlier morbidostat study of trimethoprim
291 resistance [15]. The mutations in TYMS – two insertion sequence elements, a frame shift
292 mutation, loss of two codons, and a non-synonymous active site mutation – are consistent with
293 loss of function. Thus, the observed pattern of mutation is consistent with our CRISPRi
294 experiment: reduced TYMS activity can buffer inhibition of DHFR.
295 But are mutations in DHFR and TYMS sufficient to explain the resistance phenotype of
296 the evolved strains? We did not observe mutations in any other folate metabolic genes during the
297 course of the experiment, but we did observe that a few other genes mutated under multiple

298 thymidine conditions (Table S9). These mutations might be neutral, contribute to trimethoprim
299 resistance, or be adaptive for growth in continuous culture. Of particular interest were recurring
300 mutations in *gadX*, a regulator of the acid response system, and the insertion sequences (IS1) in
301 *dusB*. Trimethoprim is known to induce an acid response in *E. coli* that results in up-regulation
302 of the *gadX* target genes *gadB/C* [38], and *dusB* is located in an operon upstream of the global E.
303 coli transcriptional regulator *fis* [39]. Thus, we wanted to test if mutations in DHFR and TYMS
304 alone were sufficient to explain the resistance phenotype. We introduced several of the observed
305 DHFR and TYMS mutations into a clean wild-type *E. coli* MG1655 background (referred to as
306 strains R1-R4) and measured the IC50. Strain R1 contained only a TYMS mutation (the deletion
307 **Δ**6(64-69)). As for strains 4,5 and 10 (in 50 μg/ml thymidine), strain R1 grew very slowly at all
308 concentrations of trimethoprim measured (up to 1800 μg/ml), preventing determination of a
309 reliable IC50. Strains containing mutations in both DHFR and TYMS (R2-R4) showed resistance
310 phenotypes comparable to the evolved strains (Fig 3). For example, the DHFR L28R/ TYMS
311 **Δ**6(64-69) double mutant in a clean genetic background (strain R4) was more resistant to
312 trimethoprim than all of the evolved strains containing these two mutations (Fig 3A). For
313 comparison, we also measured the IC50 for three previously constructed strains containing the
314 DHFR single mutants (P21L, L28R, and W30R) [13]. The DHFR mutations alone were
315 uniformly less trimethoprim resistant than the corresponding DHFR/TYMS double mutants.
316 Taken together, we conclude that the paired DHFR and TYMS mutations are both necessary and
317 sufficient to generate the observed resistance phenotype. Consistent with this, one of the evolved
318 strains contains only mutations in DHFR and TYMS (strain 1 in 50 μg/ml thymidine, Fig 3C).
319 Interestingly, TYMS loss-of-function mutations have been observed in trimethoprim
320 resistant clinical isolates for multiple genera of bacteria [40, 41]. This provides a degree of

321 confidence that the laboratory evolution experiment reflects natural conditions of adaptation.
322 Taken together the data support the idea that DHFR and TYMS represent a two-gene adaptive
323 unit, with compensatory mutations contained within the unit.
324

325 **DHFR and TYMS are coupled by constraints on metabolite concentration**

326 Given the above data, and no evidence for the physical association of DHFR and TYMS
327 in bacteria, what is the mechanism underlying their adaptive coupling? The CRISPRi
328 experiments and second site suppressor screen indicate that reductions in DHFR activity – either
329 from reducing expression, or inhibiting catalysis – are buffered by decreases in TYMS activity.
330 This suggests a simple hypothesis: their coupling arises from the need to balance the
331 concentration of key metabolites in the folate metabolic pathway. In particular, both the
332 depletion of reduced folates (THF species), and accumulation of DHF are expected to be
333 detrimental to cell growth [42, 43]. So, we wanted to test the relationship between the activity of
334 DHFR, the activity of TYMS, metabolite abundance and growth.

335 We started with a panel of ten DHFR mutations selected to span a range of *in vitro*
336 catalytic activities [44], and measured the effect of these mutations on *E. coli* growth in the
337 background of either WT TYMS or TYMS R166Q, a catalytically inactive variant. Again, we
338 used a next-generation sequencing based assay to measure the relative fitness of all possible
339 mutant combinations (20 total) [44]. In this system, DHFR and TYMS were expressed from a
340 single plasmid that contains two DNA barcodes – one associated with DHFR and one with
341 TYMS – that uniquely encode the identity of each mutant (Fig S7A). The full library of mutants
342 was transformed into the thymidine auxotroph strain *E. coli* ER2566 *ΔfolA ΔthyA*, and grown as

343 a mixed population in a turbidostat. Counting the frequency of the barcodes over time permits
344 estimation of a relative growth rate for each mutant in the population (Fig S7B, Fig S8).
345 Relative growth rates were measured across two thymidine supplementation conditions:
346 one in which TYMS R166Q produces a growth defect (5 μg/ml thymidine), and one that
347 constitutes a full rescue of TYMS activity (50 μg/ml). In the background of WT TYMS (grey
348 points, Fig 4A-B) growth rate was relatively insensitive to ~10-fold decrease in DHFR activity.
349 As DHFR activity was diminished further, the relationship between catalytic power and growth
350 rate became monotonic: decreases in DHFR activity result in slower growth. In the background
351 of TYMS R166Q, decreases in DHFR catalytic activity were buffered. More precisely, we
352 observed that the TYMS R166Q mutation is deleterious or neutral in the context of WT DHFR,
353 but is beneficial in the context of low activity DHFR mutants (e.g. F31Y.G121V). In high
354 thymidine conditions, this effect was very pronounced: the TYMS R166Q mutation was
355 sufficient to rescue growth to WT-like levels for DHFR mutants with orders of magnitude less
356 activity (Fig 4B).
357
358 **Figure 4. A loss-of-function mutation in TYMS buffers metabolic changes from decreased**
359 **DHFR activity**. **A, B**, Scatter plots of relative growth rate for DHFR mutants spanning a range
360 of catalytic specificities (k_{cat}/K_m), and either a wild-type (WT, gray points) or catalytically
361 dead (R166Q, red points) TYMS. Error bars correspond to standard error across triplicate
362 measurements. In both conditions, we observe a buffering behavior in which the cost of reducing
363 DHFR activity is mitigated by introducing a loss-of-function mutation in TYMS. **C,** Liquid
364 chromatography-mass spectrometry profiling of intracellular folate species in M9 media
365 supplemented with 0.1% amicase and 50 μg/ml thymidine. Rows reflect mutant DHFR/TYMS

366 combinations, columns correspond to metabolites. Data represent the mean of three replicates,
367 see also Fig S9 for associated errors. Each folate species can be modified by the addition of 1-5
368 glutamates. The color of each square denotes the $\log_2$ abundance of each species relative to wild-
369 type. The data show that mutations reducing DHFR activity (G121V, F31Y.L54I, M42F.G121V,
370 and F31Y.G121V) cause an accumulation of DHF and depletion of reduced folate species (THF)
371 (bottom four rows). This effect is partly compensated by an inactivating mutation in TYMS
372 (rows two-five). **D,** The corresponding doubling time for each mutant, as measured in batch
373 culture (conditions identical to panel A).
374
375 From this set of mutants, we selected ten DHFR/TYMS pairs for liquid chromatography-
376 mass spectrometry (LC-MS) profiling of folate pathway metabolites. The experiment was carried
377 out for log-phase cultures in M9 glucose media supplemented with 0.1% amicase and 50 μg/ml
378 thymidine, conditions in which the selected DHFR mutations displayed significant growth
379 defects individually, but in which the corresponding DHFR/TYMS double mutants are restored
380 to near wild-type growth. Current mass spectrometry methods allow discrimination between the
381 full diversity of folate species, which differ in oxidation, one-carbon modification, and
382 polyglutamylation states, permitting a broad metabolic study of the effects of mutations [45].
383 The data confirm that for DHFR loss-of-function mutants, intracellular DHF
384 concentration increases (Fig 4C, bottom four rows). In addition, we observe depletion of reduced
385 polyglutamated folates (Glu >= 3), while several mono- and di-glutamated THF species
386 accumulate (particularly for THF, Methylene THF and 5-Methyl THF). Prior work found that
387 DHF accumulation results in inhibition of the upstream enzyme FP-γ-GS [43]. FP-γ-GS
388 catalyzes the polyglutamylation of reduced folates, an important modification that increases

389 folate retention in the cell and promotes the use of reduced folates as substrates in a number of
390 downstream reactions [46]. The observed pattern of changes in the reduced folate pool (depletion
391 of polyglutamated THF forms) is consistent with inhibition of FP-γ-GS by DHF.

392 In the background of the corresponding TYMS loss-of-function mutant, the metabolite
393 profiles become more WT-like: the depletion of reduced folates is less severe, and the
394 accumulation of DHF more moderate (Fig 4C). Overall, the changes in metabolite concentration
395 are consistent with the observed growth rate defects in the DHFR loss-of-function mutants:
396 mutants with more severe DHF accumulation and THF depletion grow more slowly (Fig 4D, Fig
397 S9). Thus, coordinated decreases in the activity of DHFR and TYMS serve to maintain balance
398 in the DHF and THF metabolite pools, a condition associated with optimal growth. This provides
399 a plausible biochemical mechanism for the strong statistical association of DHFR and TYMS by
400 synteny and co-occurrence across thousands of bacteria. In this sense, the bifunctional fused
401 form of DHFR/TYMS found in protists and plants could be regarded as an extreme case that
402 guarantees stoichiometric expression [47, 48].
403

404 **Genome-wide analyses of co-evolution identifies additional adaptive pairs**

405 The comparative genomics analyses of folate metabolism indicate a sparse pattern of
406 adaptive interactions, in which DHFR and TYMS were more coupled to each other and less
407 coupled to the remainder of folate metabolism. The experimental data support the interpretation
408 that they form an adaptive unit embedded in folate metabolism: the two enzymes adapt together
409 in response to trimethoprim, and they are linked by a shared constraint on metabolite
410 concentration. Taken together, this suggests the hypothesis that analyses of synteny and gene co-
411 occurrence might be used to identify small adaptive units elsewhere in the cell.

412 With this goal, we extended our analyses of gene synteny and co-occurrence to include
413 all genes represented in *E. coli*. To ensure good statistics, we filtered the orthologs analyzed to
414 those that co-occur in a sufficiently large number of genomes (2095 COGs, ~500,000 pairs in
415 total) (see also Supplemental Experimental Procedures). We also compared our analysis to: (1)
416 metabolic annotations from KEGG [22] and (2) the set of high-confidence binding interactions in
417 *E. coli* reported by the STRINGdb [21]. Consistent with intuition and prior work, co-evolving
418 gene pairs show enrichment for physical complexes, enzymes in the same metabolic pathway,
419 and more specifically, enzymes with a shared metabolite (Fig 5A,B).
420
421 **Figure 5. Genome-wide analysis of co-evolution in *E. coli*. A,** Enrichment of physical and
422 metabolic interactions as a function of synteny coupling. **B**, Enrichment of physical and
423 metabolic interactions as a function of co-occurrence coupling. **C,** A scatterplot of synteny-based
424 coupling for all analyzed gene pairs. Each point represents a pair of orthologous genes (COGs);
425 coupling within the pair is shown on the x-axis, and the strongest coupling outside of the pair is
426 shown on the y-axis. Color-coding reflects annotations from the STRING database (physical
427 interactions) or KEGG database (metabolic pathways): green gene pairs bind, while pairs in dark
428 blue or light blue do not interact but are found in the same metabolic pathway. Dark blue gene
429 pairs share a metabolic intermediate. The DHFR/TYMS pair is highlighted in red. See Table S10
430 for an annotated list of gene pairs below the diagonal. **D**, Scatterplot of coupling by co-
431 occurrence for all analyzed gene pairs (same format as C).
432
433 To further identify pairs of genes that are strongly evolutionarily coupled to each other
434 and relatively decoupled from the remainder of the genome, we constructed scatterplots of all

435 gene pairs (Fig 5C,D). These plots indicate the strength of coupling within each pair (as a
436 relative entropy, along the x-axis) versus the strongest coupling outside of the pair (along the y-
437 axis). This analysis is limited to the identification of two gene units, and our analysis will need to
438 be extended to identify larger communities within the complete network of co-evolutionary
439 relationships [49]. It presents a simple graphical method for identifying gene pairs that are more
440 tightly coupled to each other than any other gene in the dataset: these are the set of points below
441 the diagonal (Table S10). Thus, these points represent predictions of two-gene coevolving units
442 (258 by synteny, 194 by co-occurrence), and include the DHFR/TYMS pair. Many of the
443 predicted adaptive units are two-protein physical complexes, while others share a metabolic
444 intermediate like DHFR/TYMS.

445 The distance from the diagonal indicates the degree of modularity of a pair. By
446 construction, points near the diagonal have an interaction with at least one other gene that is
447 similar to the degree of coupling within the pair. The most modular pairs occupy the space in the
448 lower right corner of the graph – by this criteria, DHFR and TYMS show an exceptionally strong
449 signature of modularity. Considering the regime where $D_{ij}^{\mathrm{intra}} > 1.0$ and $D_{ij}^{\mathrm{exter}} < 0.5$ on the
450 synteny plot, we observe a few other pairs with experimental evidence for adaptive coupling. For
451 example the gene pair *accB*/*accC* which encodes two of the four subunits of acetyl-CoA
452 carboxylase, the first enzymatic step in fatty acid biosynthesis. Overexpression of either *accB* or
453 *accC* individually causes reductions in fatty acid biosynthesis, but overexpressing the two genes
454 in stoichiometric amounts rescues this defect [50, 51]. Constraints on relative expression have
455 also been noted for the *selA*/*selB* and *tatB*/*tatC* gene pairs [52, 53]. The *tatB/tatC* genes encode
456 components of the TatABCE twin-arginine translocation complex, while *selA/selB* are not
457 known to bind but are both involved in selenoprotein biosynthesis. The full set of gene pairs

below the diagonal now serve as a starting place for more deeply understanding the hierarchical pattern of evolutionary couplings inside of cellular pathways, and testing the prevalence of adaptive modules more generally (Table S10).

Discussion and Conclusions

Comparative genomics analyses of synteny and co-occurrence are well-established tools for the prediction of interactions between genes [24, 26-30]. However, they are not typically used to infer independence. Here we effectively ask whether, for evolutionary couplings between genes, the absence of evidence can be evidence for absence. Conceptually, this is analogous to recent work in proteins, in which co-evolution was used to map quasi-independent adaptive and functional residue groups within a single protein [54]. In our case study of folate metabolism, comparative genomics suggested a sparse pattern of evolutionary coupling between folate metabolic enzymes, with DHFR and TYMS showing significant coupling to each other but less coupling to the remainder of the folate metabolic pathway. Experimental analyses support the interpretation that these two enzymes form a relatively independent adaptive unit: they are more coupled to each other than any other enzyme in the pathway, they adapt together in response trimethoprim, and their activities are coupled by a shared constraint on metabolite concentrations. So, in this case, comparative genomics identified a meaningful adaptive unit much smaller than the scale of the entire biochemical pathway. In light of these findings, it becomes less obvious that shared membership in a KEGG metabolic pathway should be a default gold standard for interaction prediction approaches; instead the goals of the method and type of interaction being predicted (shared biochemical intermediate, physical complex or adaptive interaction) should be carefully considered.

481 DHFR and TYMS represent a first case study, and comparative genomics approaches
482 will not predict every adaptive interaction. Analyses of synteny and co-occurrence are an
483 inference across thousands of species, encapsulating hundreds of millions of years of
484 evolutionary divergence. In the context of specific environmental conditions (including
485 conditions never encountered in past evolutionary history), or other model organisms, additional
486 adaptive interactions likely exist. We expect that the adaptive units predicted by synteny and co-
487 occurrence represent well-conserved, core, adaptive interactions that can sometimes be
488 elaborated on under particular environmental conditions or in particular species. To illustrate,
489 consider recent work in which the authors introduced a loss-of-function mutation to DHFR, and
490 then evolved the cells to grow in M9 media while gradually reducing the exogenous
491 supplementation of several key folate metabolites (including adenine, inosine, methionine, and
492 thymine) (xx). In this case, adaptation to DHFR loss-of-function was commonly mediated by
493 two loss-of-function mutations in other genes, consistent with a sparse pattern of adaptive
494 couplings to DHFR. A loss-of-function mutation in TYMS was acquired in 12/15 trajectories,
495 and 11/15 trajectories acquired a second loss-of-function mutation in the gene *deoB*. The *deoB*
496 mutation appears after TYMS mutation in the trajectories, thus the first response to DHFR
497 perturbation in these experiments is TYMS loss of function. Mutations were not systematically
498 observed in any other genes. The *deoB* gene encodes phosphopentomutase, a metabolic enzyme
499 involved in deoxyribose degradation. Deoxyribose is used to convert exogenously supplied
500 thymine to thymidine; the authors suggest that inactivating *deoB* provides a strategy for cells to
501 make efficient use of the small amounts of thymine supplied in the media. In contrast, in our
502 experiment, we directly supplement with thymidine rather than thymine and do not observe the
503 *deoB* mutation. Thus, while the core adaptive unit of DHFR and TYMS is maintained across

504 both experiments, an additional adaptive mutation (*deoB*) emerges in one environment to permit
505 efficient use of imported nutrients. Taken together, these results underscore the importance of
506 further developing computational frameworks that can robustly identify core adaptive
507 interactions that persist across environments and even genetic background.

508 If it is generally possible to decompose metabolic pathways into smaller, relatively
509 independent, pieces, this would suggest a route to identify self-contained units for biosynthetic
510 engineering, and provide fundamental insights into how cells maintain homeostasis and adapt in
511 the face of changing environments. For example, thymidine synthesis is the rate-limiting step for
512 DNA synthesis in eukaryotic cells, and transcription of the TYMS and DHFR genes is greatly
513 upregulated (via a common transcription factor) at the G_1/S cell cycle transition [55]. In
514 computational models of eukaryotic folate metabolism, computationally increasing the activities
515 of DHFR and TYMS 100-fold results in increased thymidine synthesis but only modestly
516 changes the concentration of other folates. Thus decoupling the DHFR/TYMS pair from the
517 remainder of folate metabolism may be important for ensuring modularity in different metabolite
518 pools. Substantial further work is needed to go beyond the DHFR/TYMS pair – to
519 comprehensively test the relationship between the units identified by comparative genomics,
520 patterns of adaptive dependencies, and adaptation to environmental changes. The data presented
521 here provide the computational hypotheses, technical approaches, and motivation for now doing
522 so.
523

524 **Materials and Methods**

525 **Comparative genomics analyses.** Synteny analysis was conducted using a simplified version of
526 the methods described in [23, 24]. See the Supplemental Experimental Procedures for a detailed
527 description of the synteny and co-occurrence calculations.
528
529 **CRISPRi-based growth rate measurements.** We used next-generation sequencing based
530 measurements of relative growth rate to quantify the effects of CRISPRi knockdowns on *E. coli*
531 growth. The methods are described in detail in the Supplemental Information; here we provide a
532 brief summary of the approach. Following from Qi et al., pairs of single guide RNAs (sgRNAs)
533 were ligated into expression vectors using a single golden gate cloning reaction [33, 56]. We
534 modified the sgRNA-containing plasmid to contain unique barcodes that allow for replicate
535 measurements and error correction within a single experiment (see SI for more details). The
536 resulting pairwise sgRNA library was transformed into MG1655 *E. coli* engineered to contain a
537 chromosomally encoded, catalytically dead Cas9 (dCas9) under control of an inducible Tet
538 promoter. The resulting transformants were grown in a continuous culture device (turbidostat) in
539 M9 minimal media with 0.4% glucose, 35 μg/mL Kanamycin, 5 μg/mL thymidine, and 0.2%
540 amicase for selection. After adaptation to the turbidostat, dCas9 was induced by the addition of
541 anhydrous tetracycline (aTC) and culture time points were taken over 12 hours. From each time
542 point, the sgRNA-containing region of each vector was deep sequenced to find the relative
543 frequency of each sgRNA pair compared to a negative control sgRNA (lacking a homology
544 targeting region). These data were used to calculate relative growth rates.
545
546 **Forward evolution of trimethoprim resistance in the morbidostat.**
547 The morbidostat/turbidostat apparatus was constructed as described by Toprak and colleagues
548 [36]. The founder strain for the forward evolution experiment was *E. coli* MG1655 modified by

549 phage transduction to encode green fluorescent protein (*egfp*) and chloramphenicol resistance
550 (*cat*) at the P21 attachment site. The goal of this modification was to prevent and detect
551 contamination with other strains. Throughout the forward evolution experiment, cells were
552 grown at 30°C in M9 media supplemented with 0.4% glucose and 0.2% amicase (Sigma); 30
553 μg/ml of chloramphenicol (Cam) was added for positive selection.

554 To begin the experiment, the founder strain was cultured overnight at 37°C in Luria
555 Broth (LB) + 30 μg/ml Cam. This culture was washed twice with M9, and back diluted into M9
556 + 30 μg/ml Cam supplemented with 5, 10, or 50 μg/ml thymidine (thy) for overnight adaptation
557 in culture tubes at 30°C. The next day (henceforth referred to as day 0; day 1 is the end of the
558 first day of adaptation), these overnight cultures were streaked onto LB agar plates: two colonies
559 per condition were chosen for whole genome sequencing (WGS) in order to obtain an accurate
560 sequence for the founder strain. The remainder of the overnight cultures was used to inoculate
561 three morbidostat tubes at containing M9 media with varying thymidine supplementation (5, 10,
562 and 50μg/ml thy). The starting optical density was approximately 0.005. Initial antibiotic
563 concentrations were 0, 11.5 and 57.5 μg/ml trimethoprim for media stocks A, B, and C
564 respectively. Each culture grew unperturbed until it surpassed an OD_{600} of 0.06, at which point it
565 underwent periodic dilutions with fresh media. The dilution rate is given by the formula
566 $r_{dil} = f \ln \frac{V}{V+\Delta V}$, where $V = 15ml$ is the culture volume, and $\Delta V = 3ml$ is volume added. We
567 chose a dilution frequency $f = 3\ h^{-1}$, to give $r_{dil} = 0.55$. Above the $OD_{600} = 0.15$, these dilutions
568 are used to introduce TMP into the culture (see also Fig S3). This allows controlled inhibition of
569 DHFR activity in response to growth rate. Cycles of growth and dilution continued for a period
570 of ~22 hours, at which point the run was paused to make glycerol stocks, replenish media, and
571 update TMP stock concentrations. Culture vials for the next day of evolution were filled with

572 fresh media and inoculated using 300µl from the previous culture. Complete trajectories of
573 OD_{600} and drug concentration are shown in Fig S4. Endpoint cultures were streaked onto LB
574 agar plates supplemented with 30 µg/ml of Cam and 50 µg/ml thymidine to obtain isolated
575 colonies for whole genome sequencing.
576
577 **Whole genome sequencing.** Two isolates were selected from each adapted day 0 culture, and
578 ten clonal isolates (colonies) were randomly selected from the endpoint of each evolution
579 condition, totaling 36 strains. Isolation of genomic DNA was performed using the QIAamp DNA
580 Mini Kit (Qiagen). The Nextera XT DNA Library Prep Kit (Illumina) was used to fragment and
581 label each genome for paired-end sequencing using a v2 300-cycle MiSeq kit (Illumina).
582 Average read length and coverage can be found in Table S7. Genome assembly and mutation
583 prediction was performed using *breseq* [57]. The reference sequence was a modification of the *E.*
584 *coli* MG1655 complete genome (accession no. NC_000193), edited to include the GFP marker
585 and chloramphenicol resistance cassette in our founder strain. The modified reference sequence
586 and all complete genome sequences from the beginning and endpoint of forward evolution are
587 available in the NCBI BioProject database (accession number: PRJNA378892).
588
589 **Measurements of thymidine dependence.** All strains were grown overnight in LB + 5µg/ml
590 thy, with the exception of the strains evolved in the 50µg/ml thy, which were supplemented with
591 50µg/ml thy to ensure viability. Cultures were then washed twice in M9 media without
592 thymidine, and inoculated at an OD_{600}=0.005 in 96-well plates containing M9 media
593 supplemented with 10-fold serial dilutions of thymidine, ranging from 0.005 µg/ml to 50 µg/ml
594 (in singlicate). OD_{600} was monitored in a Victor X3 plate reader at 30°C over a period of 20

595 hours. Growth was quantified using the positive integral of OD_{600} over time. This measure
596 captures mutational or drug-induced changes in the duration of lag phase as well as perturbations
597 in growth rate [15]. For each strain, we identified a start-time (t_0) at the end of lag-phase for the
598 fully-rescued 50μg/ml thy condition. We chose each t_0 computationally as the last point before
599 monotonic growth above the limit of detection. The log(OD_{600}) versus time curves for all
600 conditions are then vertically shifted ('background-subtracted'), such that the function value at
601 this start-time is zero. This curve is then numerically integrated from t_0 to t_0+15 hours using the
602 trapezoid method.
603
604 **Measurements of trimethoprim resistance (IC50).** All strains were grown overnight in LB +
605 5μg/ml thy, with the exception of the strains evolved in the 50μg/ml thy, which were
606 supplemented with 50μg/ml thy to ensure viability. Each strain was then washed into media
607 conditions corresponding to the strain's forward evolution condition, and adapted for 5.5 hours at
608 30°C. The recovery cultures were used to inoculate 96-well plates containing M9 media
609 sampling serial dilutions of TMP (in triplicate), with a starting OD_{600} = 0.005. OD_{600} was
610 monitored using a Tecan Infinite M200 Pro microplate reader and Freedom Evo robot at 30°C
611 over a period of at least 12 hours. The trimethoprim resistance of each strain was quantified by
612 its absolute IC50, the drug concentration (μg/ml) at which growth is half-maximal. The
613 relationship between growth and trimethoprim inhibition is modeled using the four parameter
614 logistic function:

$$Y = \frac{a - d}{1 + (X/c)^b} + d$$

615 where Y is growth, X is TMP concentration, a is the asymptote for uninhibited growth, d is the
616 limit for inhibited growth, c provides the concentration midway between a and d, and b captures

617 sensitivity [58]. Growth was quantified using the positive integral of OD_{600} data over a 10h
618 period of growth (see also the methods for measurement of thymidine dependence). For each
619 strain, we identify a start-time (t_0) at the end of lag-phase for the uninhibited 0μg/ml TMP
620 condition. Growth versus TMP concentration was fit to the above model using MATLAB. IC50
621 was calculated as the concentration X^* for which growth $Y(X^*) = a/2$.
622
623 **Growth without trimethoprim selection in 50μg/ml thymidine using the turbidostat**
624 The founder strain for this experiment was identical to that used for evolution of trimethoprim
625 resistance. Throughout the experiment, cells were grown at 30°C in M9 media supplemented
626 with 0.4% glucose and 0.2% amicase (Sigma); 30 μg/ml of chloramphenicol (CAM) was added
627 for positive selection. To begin the experiment, the founder strain was cultured overnight at 37°C
628 in Luria Broth (LB) + 30 μg/ml Cam. This culture was washed twice with M9, and back diluted
629 into M9 supplemented with 50μg/ml thymidine (thy) for overnight adaptation in culture tubes at
630 30°C. The next day (henceforth referred to as day 0; day 1 is the end of the first day of
631 continuous culture), the overnight culture was used to inoculate three turbidostat tubes
632 containing 17ml of M9 supplemented with 50 thy. The starting optical density was
633 approximately 0.005. Each culture grew unperturbed until it reached an OD_{600} of 0.15, at which
634 point it was diluted with 2.4 ml of fresh media. These cycles of growth and dilution persisted for
635 a period of ~22 hours, at which point the run was paused to make glycerol stocks and replenish
636 media. Culture vials for each following day of evolution were filled with fresh media and
637 inoculated using 300μl from the previous culture.
638

Growth Rate Measurements for DHFR and TYMS point mutants. All relative growth rate measurements were performed in the *E. coli* folate auxotroph strain ER2566 *ΔfolA ΔthyA* [59]. DHFR (*folA*) and TYMS (*thyA*) are provided on the plasmid pACYC-Duet1 (in MCS1 and MCS2, respectively) and are each under control of a T7 promoter. For these experiments, we use leaky expression (no IPTG induction). Each mutant plasmid (20 in total) is marked with a genetic barcode in a non-coding region between the two genes. Plasmids were transformed into the auxotroph strain, and each mutant was grown overnight in separate LB +30μg/ml Cam +50μg/ml thy cultures. Then, cultures were washed 2x in M9 media supplemented with 0.4% glucose and 0.2% amicase and 30μg/ml Cam, and adapted overnight at 30°C. All mutants were mixed in equal ratios based on OD_{600} and inoculated at a starting OD_{600} = 0.1 in the turbidostat. Growth rates were measured under two conditions: 5 thy and 50 thy, with three replicates each. The turbidostat clamps the culture to a fixed OD_{600} = 0.15 by adding fresh dilutions of media. Every 2 hours over the course of 12 hours a 1ml sample was removed, pelleted and frozen for next-generation sequencing. Amplicons containing the barcoded region with appropriate sequencing adaptors (350 basepairs in total size) were generated by two sequential rounds of PCR with Q5 polymerase. The barcoded region was sequenced with a single-end MiSeq run using a v2 50 cycle kit (Illumina). We obtained 14,348,937 reads. Data analysis was performed using a series of custom python scripts to count barcodes, and MATLAB to fit relative growth rates.

Constructing DHFR/TYMS mutants in a clean genetic background. We followed the protocol for scarless genome integration using the modified λ-red system developed by Tas et al. [60]. In this method, a tetracycline (Tet) resistance cassette ("landing pad") is first integrated at

662 the site targeted for mutagenesis. Then, the landing pad is excised by the endonuclease I-SceI,
663 and replaced with the desired mutation by λ-red mediated recombination. $NiCl_2$ is used to
664 counterselect against cells that retain the tetracycline cassette. Tas et al. provides a detailed
665 protocol; here we give the specifics necessary for our experiments. For the λ-red machinery, we
666 transformed the plasmid pTKRED (Genbank accession number GU327533) into
667 electrocompetent *E. coli* MG1655 with a genomic *egfp/cat* resistance cassette (the forward
668 evolution founder strain). For the Δ25-26 TYMS mutation, we introduced the *tetA* landing pad
669 between genome positions 2,964,900 and 2,965,201 (genome NC000913) corresponding to the
670 N-terminus of the *thyA* gene. For the DHFR mutations (L28R, W30R, and P21L), the landing
671 pad was recombined between genome positions 49,684 and 49,990 (genome NC000913). In
672 order to replace the Tet cassette, cells were induced with 2mM IPTG and 0.4% arabinose, and
673 then transformed with 100ng of dsDNA PCR product containing the mutation of interest (with
674 appropriate homology arms). This reaction experienced 3 days of outgrowth at 30°C in rich
675 defined media (RDM, Teknova) with glucose substituted for 0.5% v/v glycerol. The media was
676 supplemented with 6 mM or 4mM $NiCl_2$ for counterselection against *tetA* at the *thyA* locus or
677 *folA* locus respectively. The outgrowth culture was streaked onto agar plates and screened daily
678 for the mutant of interest using LB supplemented with 50 μg/ml thy, 30 μg/ml spectinomycin,
679 and +/- 5-10 μg/ml Tet. All mutations were confirmed by Sanger sequencing of the complete
680 *folA* and *thyA* open reading frame; for *folA* the promoter region was also sequenced.
681
682 **LC-MS Metabolite Measurements**. Cells were cultured in M9 0.2% glucose media containing
683 0.1% amicase, 50 ug/ml thy, and 30 ug/ml Cam at 30°C for metabolite analysis. In mid-log
684 phase at OD_{600} ~0.2, *E. coli* culture (3 ml for nucleotide measurement and 7 ml for folate

685 measurement) was filtered on a nylon membrane (0.2 μm), and the residual medium was quickly
686 washed away by filtering warm saline solution (200 mM NaCl at 30'C) over the membrane
687 loaded with cells to exclude non-desirable extracellular metabolites from LC-MS analysis. The
688 membrane was immediately transferred to a 6 cm Petri dish containing 1 ml cold extraction
689 solvent (-20°C 40:40:20 methanol/acetonitrile/water; for folate stability, 2.5 mM sodium
690 ascorbate and 25 mM ammonium acetate in folate extraction solvent [45]) to quench metabolism.
691 After washing the membrane, the cell extract solution was transferred to a microcentrifuge tube
692 and centrifuged at 13000 rcf for 10 min. The supernatant was transferred to a new
693 microcentrifuge tube. Folate samples were prepared with an additional extraction: the pellet was
694 resuspended in the cold extraction solvent and sonicated for 10 min in an ice bath. After the
695 second extraction and centrifugation, the supernatant was combined with the initial supernatant.
696 The metabolite extracts were dried under nitrogen flow and reconstituted in HPLC-grade water
697 for LC-MS analysis. Metabolites were measured using stand-alone orbitrap mass spectrometers
698 (ThermoFisher Exactive and Q-Exactive) operating in negative ion mode with reverse-phase
699 liquid chromatography [61]. Exactive chromatographic separation was achieved on a Synergy
700 Hydro-RP column (100 mm×2 mm, 2.5 μm particle size, Phenomenex) with a flow rate of 200
701 μL/min. Solvent A was 97:3 H_2O/MeOH with 10 mM tributylamine and 15 mM acetic acid;
702 solvent B was methanol. The gradient was 0 min, 5% B; 5 min, 5% B; 7 min, 20% B; 17 min,
703 95% B; 20 min, 100% B; 24 min, 5% B; 30 min, 5% B. Q-Exactive chromatographic separation
704 was achieved on an Poroshell 120 Bonus-RP column (150 x 2.1 mm, 2.7 μm particle size,
705 Agilent) with a flow rate of 200 μL/min. Solvent A is 10mM ammonium acetate + 0.1% acetic
706 acid in 98:2 water:acetonitrile and solvent B is acetonitrile. The gradient was 0 min, 2% B; 4

707 min, 0% B; 6 min, 30% B; 11 min, 100% B; 15 min, 100% B; 16 min, 2% B; 20 min, 2% B. LC-
708 MS data were analyzed using the MAVEN software package [62].
709
710 **Acknowledgements** We thank members of the Reynolds lab for review of the manuscript, E.
711 Toprak for extensive advice on morbidostat construction and operation, S. Benkovic for the
712 ER2566 *ΔfolA ΔthyA* strain, D. Bikard for the *E. coli* MG1655 strain containing dCas9, T.
713 Kuhlman for molecular biology reagents used in genome editing, S. Thompson for the *folA*
714 supplementation mix recipe, and T. Bergmiller for the GFP/Chloramphenicol resistance marker
715 incorporated into our founder strain.

References

1. Kim J, Copley SD. Inhibitory cross-talk upon introduction of a new metabolic pathway into an existing metabolic network. Proceedings of the National Academy of Sciences of the United States of America. 2012 Oct 16;109(42):E2856-64. PubMed PMID: 22984162. Pubmed Central PMCID: 3479479.

2. Michener JK, Camargo Neves AA, Vuilleumier S, Bringel F, Marx CJ. Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires post-transfer refinement. eLife. 2014;3. PubMed PMID: 25418043. Pubmed Central PMCID: 4271186.

3. Michener JK, Vuilleumier S, Bringel F, Marx CJ. Phylogeny poorly predicts the utility of a challenging horizontally transferred gene in Methylobacterium strains. Journal of bacteriology. 2014 Jun;196(11):2101-7. PubMed PMID: 24682326. Pubmed Central PMCID: Pmc4010993. Epub 2014/04/01. eng.

4. Kim J, Kershner JP, Novikov Y, Shoemaker RK, Copley SD. Three serendipitous pathways in E. coli can bypass a block in pyridoxal-5'-phosphate synthesis. Molecular systems biology. 2010 Nov 30;6:436. PubMed PMID: 21119630. Pubmed Central PMCID: 3010111.

5. Long CP, Gonzalez JE, Feist AM, Palsson BO, Antoniewicz MR. Dissecting the genetic and metabolic mechanisms of adaptation to the knockout of a major metabolic enzyme in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America. 2018 Jan 2;115(1):222-7. PubMed PMID: 29255023. Pubmed Central PMCID: 5776819.

6. Kondrashov AS, Sunyaev S, Kondrashov FA. Dobzhansky-Muller incompatibilities in protein evolution. Proceedings of the National Academy of Sciences of the United States of America. 2002 Nov 12;99(23):14878-83. PubMed PMID: 12403824. Pubmed Central PMCID: 137512.

7. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences of the United States of America. 2012 Jan 24;109(4):1193-8. PubMed PMID: 22223662. Pubmed Central PMCID: 3268279.

8. Green JM, Matthews RG. Folate Biosynthesis, Reduction, and Polyglutamylation and the Interconversion of Folate Derivatives. EcoSal Plus. 2013.

9. Ducker GS, Rabinowitz JD. One-Carbon Metabolism in Health and Disease. Cell metabolism. 2016 Sep 8. PubMed PMID: 27641100.

10. Gangjee A, Jain HD, Kurup S. Recent advances in classical and non-classical antifolates as antitumor and antiopportunistic infection agents: part I. Anti-cancer agents in medicinal chemistry. 2007 Sep;7(5):524-42. PubMed PMID: 17896913.

11. Costanzo MS, Hartl DL. The evolutionary landscape of antifolate resistance in Plasmodium falciparum. Journal of genetics. 2011 Aug;90(2):187-90. PubMed PMID: 21869466. Pubmed Central PMCID: 3212943.

12. Ogbunugafor CB, Wylie CS, Diakite I, Weinreich DM, Hartl DL. Adaptive Landscape by Environment Interactions Dictate Evolutionary Dynamics in Models of Drug Resistance. PLoS computational biology. 2016 Jan;12(1):e1004710. PubMed PMID: 26808374. Pubmed Central PMCID: 4726534.

13. Palmer AC, Toprak E, Baym M, Kim S, Veres A, Bershtein S, et al. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. Nature communications. 2015 Jun 10;6:7385. PubMed PMID: 26060115. Pubmed Central PMCID: 4548896.

14. Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI. Biophysical principles predict fitness landscapes of drug resistance. Proceedings of the National Academy of Sciences of the United States of America. 2016 Mar 15;113(11):E1470-8. PubMed PMID: 26929328. Pubmed Central PMCID: 4801265.

15. Toprak E, Veres A, Michel JB, Chait R, Hartl DL, Kishony R. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. Nature genetics. 2012 Jan;44(1):101-5. PubMed PMID: 22179135. Pubmed Central PMCID: Pmc3534735. Epub 2011/12/20. eng.

16. Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM, Kroon G, et al. Divergent evolution of protein conformational dynamics in dihydrofolate reductase. Nature structural & molecular biology. 2013 Nov;20(11):1243-9. PubMed PMID: 24077226. Pubmed Central PMCID: 3823643.

17. Francis K, Stojkovic V, Kohen A. Preservation of protein dynamics in dihydrofolate reductase evolution. The Journal of biological chemistry. 2013 Dec 13;288(50):35961-8. PubMed PMID: 24158440. Pubmed Central PMCID: 3861645.

18. Liu CT, Hanoian P, French JB, Pringle TH, Hammes-Schiffer S, Benkovic SJ. Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. Proceedings of the National Academy of Sciences of the United States of America. 2013 Jun 18;110(25):10159-64. PubMed PMID: 23733948. Pubmed Central PMCID: 3690862.

19. Bershtein S, Serohijos AW, Bhattacharyya S, Manhart M, Choi JM, Mu W, et al. Protein Homeostasis Imposes a Barrier on Functional Integration of Horizontally Transferred Genes in Bacteria. PLoS genetics. 2015 Oct;11(10):e1005612. PubMed PMID: 26484862. Pubmed Central PMCID: 4618355.

20. Bhattacharyya S, Bershtein S, Yan J, Argun T, Gilson AI, Trauger SA, et al. Transient protein-protein interactions perturb E. coli metabolome and cause gene dosage toxicity. eLife. 2016 Dec 10;5. PubMed PMID: 27938662. Pubmed Central PMCID: 5176355.

21. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research. 2015 Jan;43(Database issue):D447-52. PubMed PMID: 25352553. Pubmed Central PMCID: 4383874.

22. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic acids research. 2012 Jan;40(Database issue):D109-14. PubMed PMID: 22080510. Pubmed Central PMCID: 3245020.

23. Junier I, Rivoire O. Synteny in Bacterial Genomes: Inference, Organization and Evolution. arXiv:13074291. 2013.

24. Junier I, Rivoire O. Conserved Units of Co-Expression in Bacterial Genomes: An Evolutionary Insight into Transcriptional Regulation. PloS one. 2016;11(5):e0155740. PubMed PMID: 27195891. Pubmed Central PMCID: 4873041.

25. Rivoire O. Elements of coevolution in biological sequences. Physical review letters. 2013 Apr 26;110(17):178102. PubMed PMID: 23679784.

26. Huynen M, Snel B, Lathe W, 3rd, Bork P. Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome research. 2000 Aug;10(8):1204-10. PubMed PMID: 10958638. Pubmed Central PMCID: 310926.

27. Janga SC, Collado-Vides J, Moreno-Hagelsieb G. Nebulon: a system for the inference of functional relationships of gene products from the rearrangement of predicted operons. Nucleic acids research. 2005;33(8):2521-30. PubMed PMID: 15867197. Pubmed Central PMCID: 1088069.

28. Kim PJ, Price ND. Genetic co-occurrence network across sequenced microbes. PLoS computational biology. 2011 Dec;7(12):e1002340. PubMed PMID: 22219725. Pubmed Central PMCID: 3248385.

29. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proceedings of the National Academy of Sciences of the United States of America. 1999 Apr 13;96(8):4285-8. PubMed PMID: 10200254. Pubmed Central PMCID: 16324.

30. Snel B, Bork P, Huynen MA. The identification of functional modules from the genomic association of genes. Proceedings of the National Academy of Sciences of the United States of America. 2002 Apr 30;99(9):5890-5. PubMed PMID: 11983890. Pubmed Central PMCID: 122872.

31. Beck C, Knoop H, Steuer R. Modules of co-occurrence in the cyanobacterial pan-genome reveal functional associations between groups of ortholog genes. PLoS genetics. 2018 Mar;14(3):e1007239. PubMed PMID: 29522508. Pubmed Central PMCID: 5862535.

32. Okamura-Ikeda K, Ohmura Y, Fujiwara K, Motokawa Y. Cloning and nucleotide sequence of the gcv operon encoding the Escherichia coli glycine-cleavage system. European journal of biochemistry / FEBS. 1993 Sep 1;216(2):539-48. PubMed PMID: 8375392.

33. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013 Feb 28;152(5):1173-83. PubMed PMID: 23452860. Pubmed Central PMCID: 3664290.

34. Howell EE, Foster PG, Foster LM. Construction of a dihydrofolate reductase-deficient mutant of Escherichia coli by gene replacement. Journal of bacteriology. 1988 Jul;170(7):3040-5. PubMed PMID: 2838456. Pubmed Central PMCID: 211246.

35. Chevereau G, Dravecka M, Batur T, Guvenek A, Ayhan DH, Toprak E, et al. Quantifying the Determinants of Evolutionary Dynamics Leading to Drug Resistance. PLoS biology. 2015;13(11):e1002299. PubMed PMID: 26581035. Pubmed Central PMCID: 4651364.

36. Toprak E, Veres A, Yildiz S, Pedraza JM, Chait R, Paulsson J, et al. Building a morbidostat: an automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition. Nature protocols. 2013 Mar;8(3):555-67. PubMed PMID: 23429717. Pubmed Central PMCID: Pmc3708598. Epub 2013/02/23. eng.

37. Forsburg SL. The art and design of genetic screens: yeast. Nature reviews Genetics. 2001 Sep;2(9):659-68. PubMed PMID: 11533715.

38. Mitosch K, Rieckh G, Bollenbach T. Noisy Response to Antibiotic Stress Predicts Subsequent Single-Cell Survival in an Acidic Environment. Cell systems. 2017 Apr 26;4(4):393-403 e5. PubMed PMID: 28342718.

39. Bradley MD, Beach MB, de Koning AP, Pratt TS, Osuna R. Effects of Fis on Escherichia coli gene expression during different growth stages. Microbiology. 2007 Sep;153(Pt 9):2922-40. PubMed PMID: 17768236.

40. King CH, Shlaes DM, Dul MJ. Infection caused by thymidine-requiring, trimethoprim-resistant bacteria. Journal of clinical microbiology. 1983 Jul;18(1):79-83. PubMed PMID: 6604070. Pubmed Central PMCID: 270747.

41. Kriegeskorte A, Block D, Drescher M, Windmuller N, Mellmann A, Baum C, et al. Inactivation of thyA in Staphylococcus aureus attenuates virulence and has a strong impact on metabolism and virulence gene expression. mBio. 2014 Jul 29;5(4):e01447-14. PubMed PMID: 25073642. Pubmed Central PMCID: 4128360.

42. Kwon YK, Higgins MB, Rabinowitz JD. Antifolate-induced depletion of intracellular glycine and purines inhibits thymineless death in E. coli. ACS chemical biology. 2010 Aug 20;5(8):787-95. PubMed PMID: 20553049. Pubmed Central PMCID: 2945287.

43. Kwon YK, Lu W, Melamud E, Khanam N, Bognar A, Rabinowitz JD. A domino effect in antifolate drug action in Escherichia coli. Nature chemical biology. 2008 Oct;4(10):602-8. PubMed PMID: 18724364. Pubmed Central PMCID: 2699209.

44. Reynolds KA, McLaughlin RN, Ranganathan R. Hotspots for allosteric regulation on protein surfaces. Cell. 2011 Dec 23;147(7):1564-75. PubMed PMID: 22196731. Pubmed Central PMCID: Pmc3414429. eng.

45. Lu W, Kwon YK, Rabinowitz JD. Isotope ratio-based profiling of microbial folates. Journal of the American Society for Mass Spectrometry. 2007 May;18(5):898-909. PubMed PMID: 17360194. Pubmed Central PMCID: 1909916.

46. McGuire JJ, Bertino JR. Enzymatic synthesis and function of folylpolyglutamates. Molecular and cellular biochemistry. 1981 Aug 11;38 Spec No(Pt 1):19-48. PubMed PMID: 7027025.

47. Beverley SM, Ellenberger TE, Cordingley JS. Primary structure of the gene encoding the bifunctional dihydrofolate reductase-thymidylate synthase of Leishmania major. Proceedings of the National Academy of Sciences of the United States of America. 1986 Apr;83(8):2584-8. PubMed PMID: 3458220. Pubmed Central PMCID: 323343.

48. Lazar G, Zhang H, Goodman HM. The origin of the bifunctional dihydrofolate reductase-thymidylate synthase isogenes of Arabidopsis thaliana. The Plant journal : for cell and molecular biology. 1993 May;3(5):657-68. PubMed PMID: 8374616.

49. Newman MEJ. Networks : an introduction. Oxford ; New York: Oxford University Press; 2010. xi, 772 p. p.

50. Abdel-Hamid AM, Cronan JE. Coordinate expression of the acetyl coenzyme A carboxylase genes, accB and accC, is necessary for normal regulation of biotin synthesis in Escherichia coli. Journal of bacteriology. 2007 Jan;189(2):369-76. PubMed PMID: 17056747. Pubmed Central PMCID: 1797400.

51. Janssen HJ, Steinbuchel A. Fatty acid synthesis in Escherichia coli and its applications towards the production of fatty acid based biofuels. Biotechnology for biofuels. 2014 Jan 09;7(1):7. PubMed PMID: 24405789. Pubmed Central PMCID: 3896788.

52. Bolhuis A, Mathers JE, Thomas JD, Barrett CM, Robinson C. TatB and TatC form a functional and structural unit of the twin-arginine translocase from Escherichia coli. The Journal of biological chemistry. 2001 Jun 08;276(23):20213-9. PubMed PMID: 11279240.

53. Rengby O, Johansson L, Carlson LA, Serini E, Vlamis-Gardikas A, Karsnas P, et al. Assessment of production conditions for efficient use of Escherichia coli in high-yield heterologous recombinant selenoprotein synthesis. Applied and environmental microbiology. 2004 Sep;70(9):5159-67. PubMed PMID: 15345395. Pubmed Central PMCID: 520894.

54. Halabi N, Rivoire O, Leibler S, Ranganathan R. Protein sectors: evolutionary units of three-dimensional structure. Cell. 2009 Aug 21;138(4):774-86. PubMed PMID: 19703402. Pubmed Central PMCID: 3210731.

55. Bjarnason GA, Jordan RC, Wood PA, Li Q, Lincoln DW, Sothern RB, et al. Circadian expression of clock genes in human oral mucosa and skin: association with specific cell-

cycle phases. The American journal of pathology. 2001 May;158(5):1793-801. PubMed PMID: 11337377. Pubmed Central PMCID: 1891949.

56. Hawkins JS, Wong S, Peters JM, Almeida R, Qi LS. Targeted Transcriptional Repression in Bacteria Using CRISPR Interference (CRISPRi). Methods in molecular biology. 2015;1311:349-62. PubMed PMID: 25981485. Pubmed Central PMCID: 4845913.

57. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods in molecular biology. 2014;1151:165-88. PubMed PMID: 24838886. Pubmed Central PMCID: 4239701.

58. Sebaugh JL. Guidelines for accurate EC50/IC50 estimation. Pharmaceutical statistics. 2011 Mar-Apr;10(2):128-34. PubMed PMID: 22328315. Epub 2012/02/14. eng.

59. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, et al. Surface sites for engineering allosteric control in proteins. Science (New York, NY). 2008 Oct 17;322(5900):438-42. PubMed PMID: 18927392. Pubmed Central PMCID: 3071530.

60. Tas H, Nguyen CT, Patel R, Kim NH, Kuhlman TE. An Integrated System for Precise Genome Modification in Escherichia coli. PloS one. 2015;10(9):e0136963. PubMed PMID: 26332675. Pubmed Central PMCID: 4558010.

61. Lu W, Clasquin MF, Melamud E, Amador-Noguez D, Caudy AA, Rabinowitz JD. Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. Analytical chemistry. 2010 Apr 15;82(8):3212-21. PubMed PMID: 20349993. Pubmed Central PMCID: 2863137.

62. Clasquin MF, Melamud E, Rabinowitz JD. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]. 2012 Mar;Chapter 14:Unit14 1. PubMed PMID: 22389014. Pubmed Central PMCID: 4055029.

Supplemental Legends

Figure S1 Predicted interactions in folate metabolism from STRINGdb. All data were obtained from the file COG.links.v10.5.txt, as downloaded from STRINGdb. **A**, The combined interaction, or confidence score between genes in central folate metabolism. The confidence score is benchmarked to approximate the probability that two enzymes share a pathway in the KEGG database. **B**, The neighborhood component of the interaction score, which captures genome context and chromosomal proximity. **C**, The co-occurrence component of the interaction score, based on correlations in COG presence/absence across species.

Figure S2 Methods and validation for CRISPRi inhibition of expression. A, Schematic of the vector containing sgRNAs for CRISPRi gene knockdowns. Each plasmid in the library contains two sgRNAs. For single gene knockdowns, an sgRNA that targets a gene is paired with an sgRNA lacking a targeting region (none). **B**, A representative sampling of frequency data and growth rate fits for the CRISPRi NGS-fit assay. Frequency is computed relative to the none:none sgRNA pair. **C**, The correlation between CRISPRi effects and gene knockout effects on growth suggests CRISPRi is working as expected. CRISPRi measurements are normalized to wildtype (none:none sgRNA), and were performed in M9 minimal media with glucose. Knockout growth data (x-axis) are published numbers associated with the KEIO collection (Baba et al, 2006 Mol Syst Biol), and report the normalized increase in OD600 over 24 hours of growth in MOPS minimal media with 0.4% glucose. **D**, RT-qPCR mRNA quantitation of CRISPRi effects on target gene mRNA abundance. All qPCR reactions were performed in technical triplicate except glyA, which was performed in technical duplicate. The $\Delta\Delta C_T$ method was used to quantity changes in gene expression. Standard error of the mean associated with each average C_T value

was propagated using a first order Taylor expansion to produce the error bars associated with each gene knockdown.

Figure S3 Evolution of trimethoprim (TMP) resistance in MG1655 cells using the morbidostat. A, Schematic of a continuous culture tube. Dilutions were made through inlet tubes labeled 'A', 'B', and 'C'. A constant volume of 15 ml was maintained by aspirating extra medium through the waste line after mixing. **B**, Control strategy for the addition of trimethoprim. Once the culture exceeded an OD_{600} of 0.06 periodic dilutions of 3 ml were made every 20 minutes. For OD_{600} between 0.06 and 0.15, media 'A,' containing no TMP, was added. While the OD_{600} was above a threshold of 0.15, media 'B' was used to introduce TMP. Media 'B' was added periodically until the culture tube reached a TMP concentration greater than 60% of the stock, at which point media 'C,' containing 5-fold more TMP was added. If a culture's growth rate diminished to below that of the dilution rate, then media 'A' would be used in the next dilution to incrementally reduce the TMP concentration. At the end of a given day, if media 'C' had been utilized, then all stock concentrations of TMP were increased by a factor of 5X to enable continued adaptation in the following day. **C**, A representative growth trajectory, color-coded by TMP concentration (day 7, 50 μg/ml thymidine). See Figure S4 for full growth trajectories over 13 days. **D**, The trajectory of estimated TMP resistance (as measured by the median TMP concentration on each day) versus number of generations for each experimental condition (5, 10 and 50 μg/ml thymidine).

Figure S4 OD600 measurements and trimethoprim concentration over 13 days of forward evolution. Each plot corresponds to a different thymidine concentration, denoted in the upper

left hand corner. The x-axis displays the number of days in real time. Discontinuities at day 5 for all three conditions are the result of a technical interruption; cultures were restarted from the previous day's glycerol stock. An enhanced view of the 50 μg/ml trajectory day 7 (boxed in a dashed line) is shown in Fig S3B.

Figure S5 Thymidine dependence of the 30 evolved strains. The y-axis denotes the positive integral of log(OD_{600}) evaluated over 20 hours of growth (see Experimental Procedures). Strains evolved for trimethoprim resistance in the presence of thymidine become auxotrophs.

Figure S6 Lack of thymidine dependence after growth in the absence of TMP. Loss of function in TYMS due to selection with TMP. **A**, Ten colonies from day 6 of the morbidostat TMP selection (50 μg/ml thymidine condition). Replica plating on 0 and 50 μg/ml thymidine indicates that all ten strains are thymidine auxotrophs. **B**, Ten colonies from three replicate growths in 50 μg/ml thymidine without TMP selection (turbidostat). These cultures were grown until biofilm formation became prohibitive. Replica plating on 0 and 50 μg/ml thymidine indicates that all strains retain TYMS activity.

Figure S7 Strategy for measuring DHFR/TYMS mutant growth rates. A, Genetic barcoding scheme for deep sequencing. Each plasmid contains two barcodes uniquely encoding the identity of its corresponding *folA* and *thyA* genotypes. **B**, Genotype frequency over time as determined by barcode sequencing. The relative fitness of each allele pair is given by the linear slope *m*. See Fig S8 for all growth rate fits.

Figure S8 Relative growth rate measurements for DHFR/TYMS mutants. Points represent the normalized relative frequency (log scale) of DHFR mutants during turbidostat growth in either 5 or 50 μg/ml thymidine. The y-axis indicates the genetic background of TYMS: either WT or R166Q. Relative growth rate fits are shown by the solid lines.

Figure S9 The relationship between intracellular folate species and doubling time. The y-axis depicts the log2 abundance of each metabolite normalized to WT. For each folate species, the five glutamylation states are shown in different colors. Doubling times were measured in M9 minimal media supplemented with 50 μg/ml thymidine. Error bars denote standard error of the mean across triplicate measurements.

Figure 1
Schober et al.

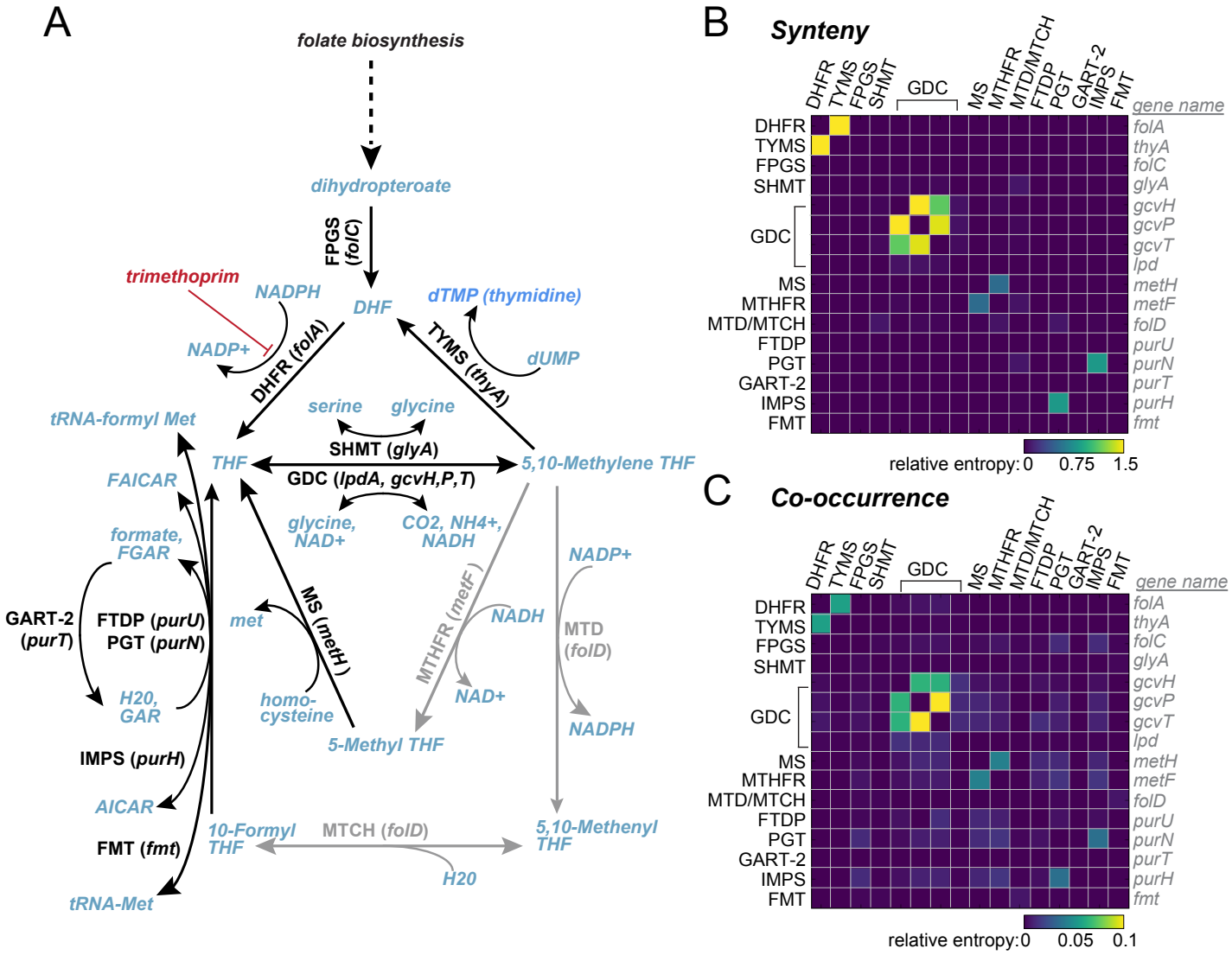

Figure 2
Schober et al.

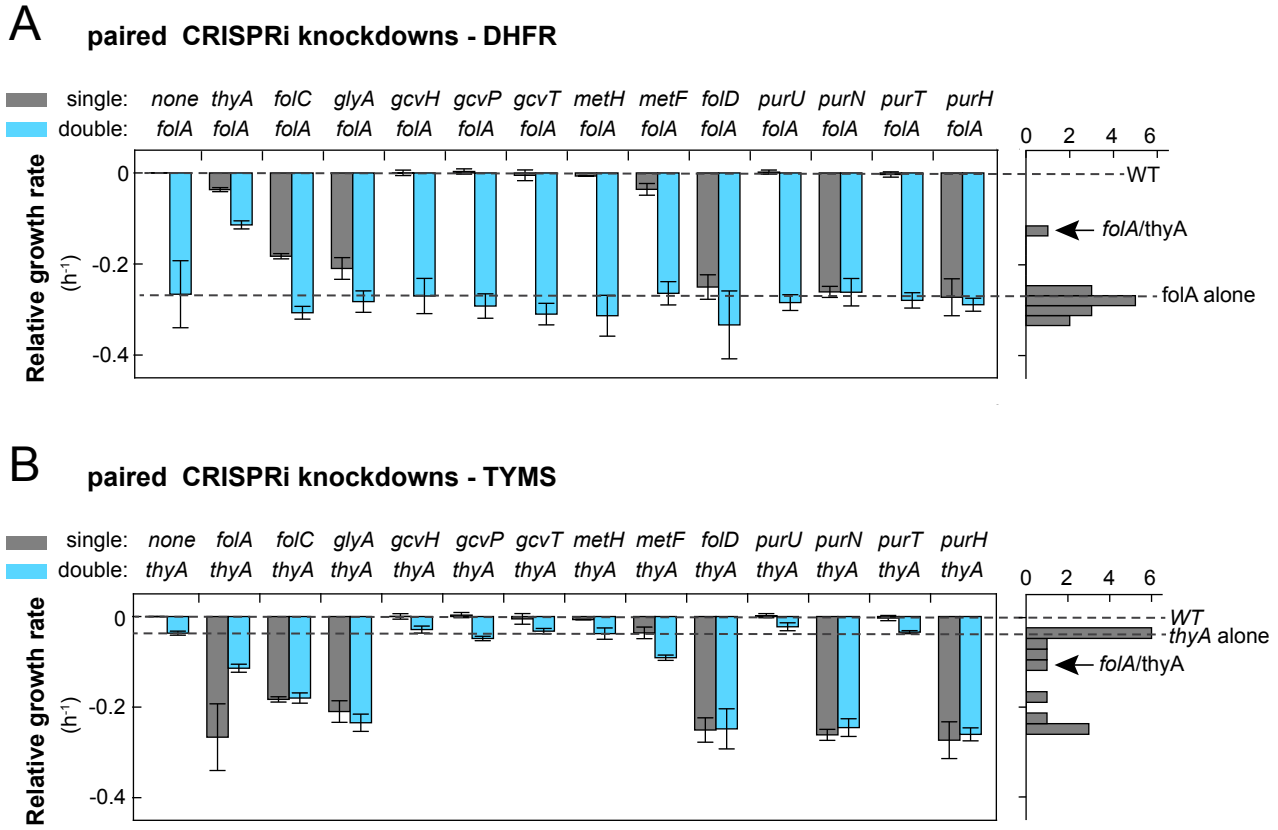

Figure 3
Schober et al.

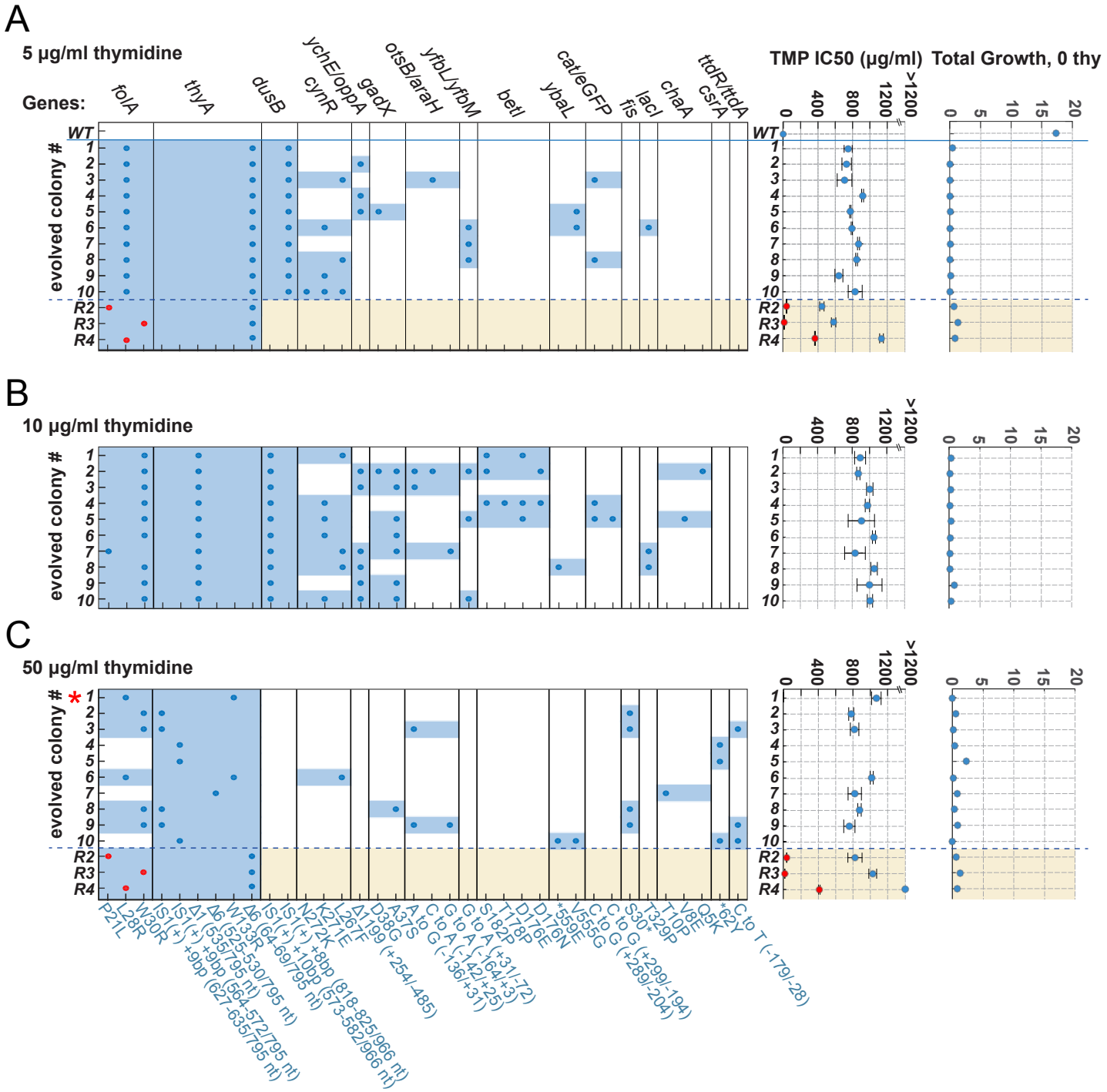

Figure 4
Schober et al.

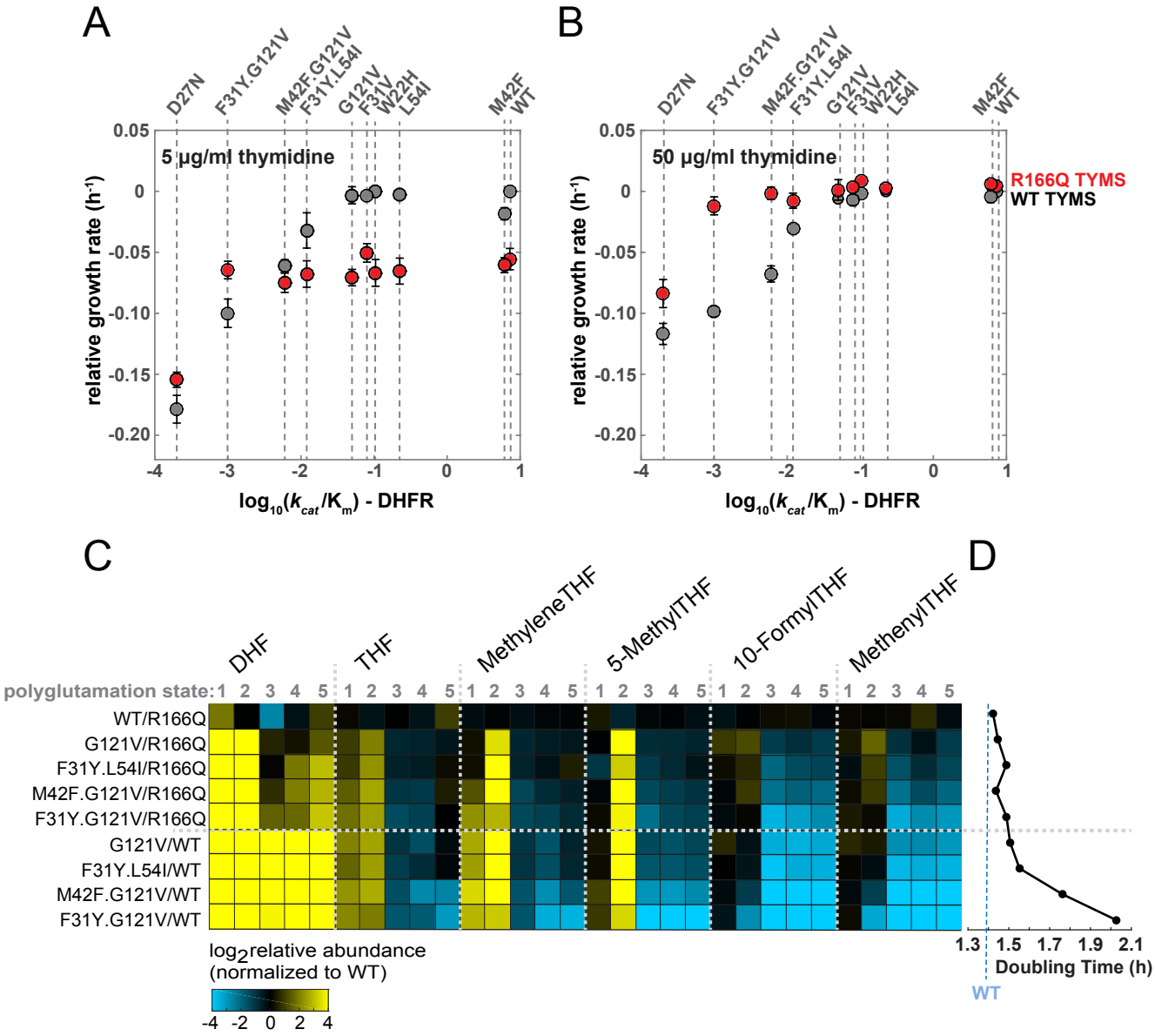
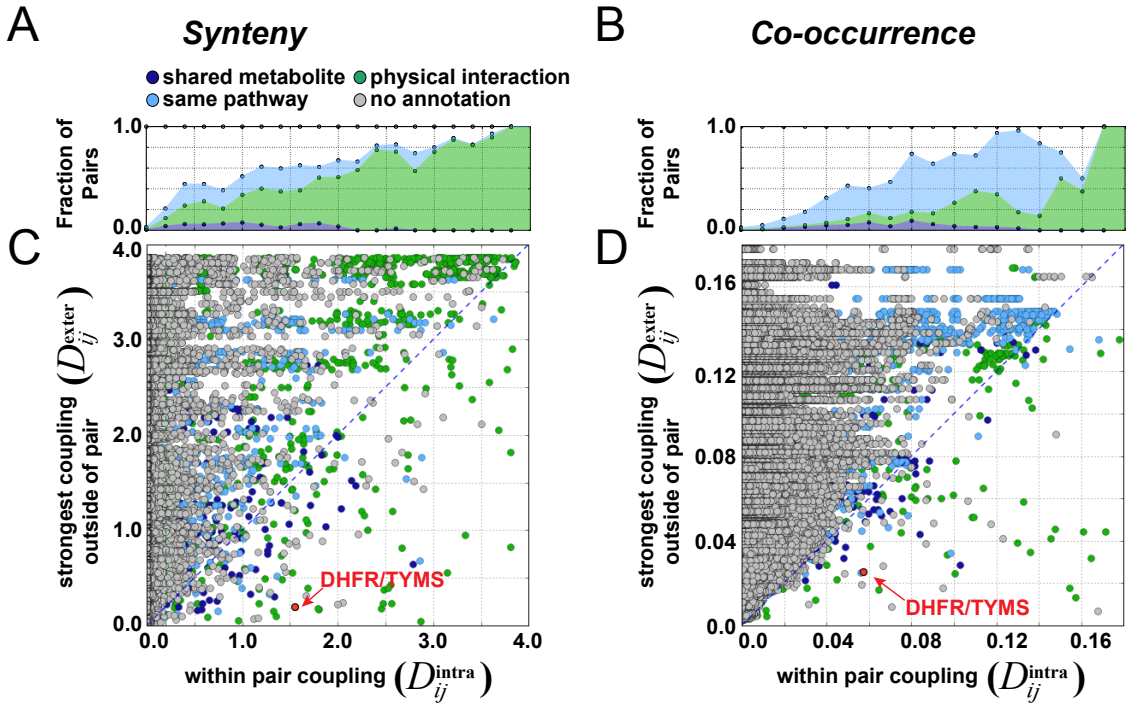
Figure 5
Schober et al.