Title:

Deep mutational scanning of *Pneumocystis jirovecii* dihydrofolate reductase reveals allosteric mechanism of resistance to an antifolate

Authors:

Francois D. Rouleau[1,2,3,4,ǂ], Alexandre K Dubé[1,2,3,5], Isabelle Gagnon-Arsenault[1,2,3,5], Soham Dibyachintan[1,2,3,4], Alicia Pageau[1,2,3,4], Philippe C. Després[1,2,3,4], Patrick Lagüe[1,2,3,4] and Christian R Landry[1,2,3,4,5]

Affiliations:

1) Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada
2) Department of Biochemistry, Microbiology and Bioinformatics, Université Laval, Québec, QC, Canada
3) Quebec Network for Research on Protein Function, Engineering, and Applications (PROTEO), Université du Québec à Montréal, Montréal, QC, Canada
4) Université Laval Big Data Research Center (BDRC_UL), Québec, QC, Canada
5) Department of Biology, Université Laval, Québec, QC, Canada

ǂAuthor for correspondence:

Francois D. Rouleau: Francois.rouleau.2@ulaval.ca

Author contributions:

Contributor Role	Role Definition
Conceptualization	Francois D. Rouleau, Christian R. Landry
Data Curation	Francois D. Rouleau
Formal Analysis	Francois D. Rouleau, Soham Dibyachintan, Alicia Pageau
Funding Acquisition	Christian R Landry, Francois D. Rouleau
Investigation	Francois D. Rouleau, Alexandre K Dubé, Isabelle Gagnon-Arsenault, Christian R Landry
Methodology	Francois D. Rouleau, Alexandre K Dubé, Soham Dibyachintan, Isabelle Gagnon-Arsenault, Patrick Lagüe, Christian R Landry
Project Administration	Francois D. Rouleau, Alexandre K Dubé, Isabelle Gagnon-Arsenault, Christian R Landry
Resources	Francois D. Rouleau, Alexandre K Dubé, Isabelle Gagnon-Arsenault, Patrick Lagüe, Christian R Landry
Software	Francois D. Rouleau, Alicia Pageau, Philippe C Després, Patrick Lagüe
Supervision	Alexandre K Dubé, Isabelle Gagnon-Arsenault, Christian R Landry
Validation	Francois D. Rouleau, Alexandre K Dubé, Soham Dibyachintan, Patrick Lagüe, Christian R Landry
Visualization	Francois D. Rouleau
Writing – Original Draft Preparation	Francois D. Rouleau
Writing – Review and Editing	Francois D. Rouleau, Alexandre K Dubé, Isabelle Gagnon-Arsenault, Alicia Pageau, Philippe C Després, Soham Dibyachintan, Patrick Lagüe, Christian R Landry

Abstract

Pneumocystis jirovecii is a fungal pathogen that causes pneumocystis pneumonia, a disease that mainly affects immunocompromised individuals. This fungus has historically been hard to study because of our inability to grow it *in vitro*. One of the main drug targets in *P. jirovecii* is its dihydrofolate reductase (PjDHFR). Here, by using functional complementation of the baker's yeast ortholog, we show that PjDHFR can be inhibited by the antifolate methotrexate in a dose-dependent manner. Using deep mutational scanning of PjDHFR, we identify mutations conferring resistance to methotrexate. Thirty-one sites spanning the protein have at least one mutation that leads to resistance, for a total of 355 high-confidence resistance mutations. Most resistance-inducing mutations are found inside the active site, and many are structurally equivalent to mutations known to lead to resistance to different antifolates in other organisms. Some sites show specific resistance mutations, where only a single substitution confers resistance, whereas others are more permissive, as several substitutions at these sites confer resistance. Surprisingly, one of the permissive sites (F199) is without direct contact to either ligand or cofactor, suggesting that it acts through an allosteric mechanism. Modeling changes in binding energy between F199 mutants and drug shows that most mutations destabilize interactions between the protein and the drug. This evidence points towards a more important role of this position in resistance than previously estimated and highlights potential unknown allosteric mechanisms of resistance to antifolate in DHFRs. Our results offer unprecedented resources for the interpretation of mutation effects in the main drug target of an uncultivable fungal pathogen.

Introduction

Some essential proteins are highly conserved across species, and therefore, they serve as targets of broad-spectrum drugs [1]. One such enzyme is the dihydrofolate reductase (DHFR), which catalyzes the reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF) in an NADPH-dependent manner [2]. THF is an essential cofactor involved in nucleotide and amino acid biosynthesis. Because of these properties, DHFRs are among the best-studied proteins, with several hundreds of orthologs and variants sequenced, characterized, and crystalized [3–10]. Despite their amino acid sequence diversity across orthologs, the core function and structure of DHFRs are highly conserved, and functional replacement is possible between several organisms [11,12].

Because of their essential nature, DHFRs have long been valuable targets for drug development [13]. For instance, in *Escherichia coli*, in *Plasmodium falciparum* and in humans, this enzyme has been thoroughly studied as therapeutic targets using molecules generally referred to as antifolates. Some drugs appear to inhibit many orthologous DHFRs (e.g. methotrexate (MTX) and pemetrexed) while others appear to be more specific (e.g. cycloguanil and pyrimethamine) [14,15]. Antifolates have been broadly used to treat various infections and diseases, from bacterial and fungal infections to rheumatoid arthritis and some cancers [10,12,14,16].

For many orthologs, mutations that confer resistance to antifolates have been identified. Resistance mutations act through different mechanisms depending on the specific protein and drug involved [17,18]. Indeed, antifolates have a wide range of structures, with antifolates having different binding affinities and forming different contacts with the enzyme. Recently, DHFRs have been identified as potential targets for the treatment of fungal infections, where treatment options are often limited given the low number of antifungal classes [19]. Despite this, fungal DHFRs have rarely been the focus of systematic studies and therefore, their ability to evolve towards antifolate resistance remains to be elucidated. While resistance to antifolates has been rigorously investigated for some organisms of clinical interest, especially in bacteria and intracellular parasites, fungal resistance has been much less studied, even in pathogens against which antifolates have historically been administered [14].

Such is the case of *Pneumocystis jirovecii*, an opportunistic fungal pathogen and the etiological agent of pneumocystis pneumonia in humans. *P. jirovecii* was originally classified as a protozoan and named *Pneumocystis carinii*, but was officially reclassified as a fungus in 1988 with the advent of sequence-based identification [20], and renamed *P. jirovecii* in 1999 to avoid confusion with the rat-specific *P. carinii,* another member of this genus [21–23]. Members of the *Pneumocystis* genus are host-specific, and it has been reported that even amongst primates, transmission of the *Pneumocystis* associated with one species of primate to another is very unlikely, and has not been reported as of yet [24].

P. jirovecii infection is most often observed in HIV patients. It is one of the most prevalent opportunistic pathogens in this population, exhibiting a high organismal load and a mortality rate of up to 20% [25–28]. In recent years, its occurrence in immunocompromised non-HIV patients, such as organ transplant recipients, has been increasing. In this group, the disease caused by *P. jirovecii* has a mortality rate close to

60% despite the use of prophylaxis. This increase raises major concerns about the future of *Pneumocystis* pneumonia in vulnerable populations [29–31]. Furthermore, *P. jirovecii* has been detected in healthy, non-immunocompromised groups, and in up to 50% of the general population, implying a higher prevalence than previously estimated [26,32–35]. The potential for this fungus to cause human infections is therefore likely to be underestimated compared to other pathogenic fungi, and warrants the development of tools for its monitoring, as well as for functional analyses.

Furthermore, *P. jirovecii* appears to be insensitive to currently used antifungal drugs such as azoles, polyenes, allylamines and echinocandins [36]. Trimethoprim-sulfamethoxazole (TMP-SMX, Bactrim) is the main drug combination used for both prophylaxis and treatment of *Pneumocystis* pneumonia [37]. In recent years, many clinical cases of *Pneumocystis* pneumonia have shown resistance to treatment, with many reporting mutations in the dihydropteroate synthase (DHPS), the target of SMX, and more recently, in DHFR, the target of TMP [38–41]. With other classes of DHPS inhibitors having acute side effects [42,43], and the toxicity associated with high doses of TMP [44], there is an urgent need to investigate the mechanisms through which antifolate resistance evolves in this fungal pathogen.

However, as there are currently no available methods to culture *P. jirovecii in vitro* or in animal models, let alone edit its genome, the systematic study of this fungal pathogen and of its drug resistance mechanisms remains a challenge. Here, we develop an experimental system and a resource for the study of resistance mutations to antifolates in the DHFR of *P. jirovecii* (PjDHFR). We constructed yeast strains and plasmids that allow for the complementation of the *Saccharomyces cerevisiae* DHFR with PjDHFR. We then built a library of mutants of the entire protein to comprehensively identify mutations that confer resistance to MTX, an antifolate with well-known resistance mutations and mechanisms in other eukaryotic DHFRs, including yeast, mouse, and human. We chose MTX as our antifolate of choice, as *S. cerevisiae* has been reported as being insensitive to TMP through mechanisms not directly involving the DHFR [45]. We use the library of mutants to conduct pooled competition assays in the presence of two different concentrations of MTX, as well as in the absence of the drug. We identify sites along the protein where resistance mutations are the most prevalent and show that many mutations associated with MTX resistance in PjDHFR are found within the active site. We show that some are structurally conserved with MTX resistance mutations reported in orthologous eukaryotic DHFRs, and to other antifolates in *E. coli*. We also identify MTX resistance-conferring mutations at a position previously unreported in eukaryotes, position F199. F199 is not in direct contact with either drug, substrate, or cofactor, revealing that resistance to antifolates can evolve through allosteric mechanisms. Finally, we investigate mutations at this position to assess their impact on protein stability and binding energy with ligands using modeling. We show that mutations at F199 destabilize interaction between PjDHFR and MTX more than with DHF, and that destabilization of interaction with MTX correlates with resistance to this drug. These results imply a more complex landscape of resistance to MTX than previously suspected in eukaryotic DHFRs.

Methods

Strains, plasmids, and culture media

All strains used in this study were constructed from the reference strain BY4741 as an initial parent strain [46]. All transformations were conducted using the standard lithium acetate transformation protocol [47]. All PCR reactions were conducted using KAPA HiFi Hotstart DNA polymerase (Sigma-Aldrich, catalog #BK1000) unless stated otherwise. Details for all strains, plasmids, oligonucleotides, and media recipes are available in their respective supplementary tables.

To know if PjDHFR is functional in *S. cerevisiae*, we had to inactivate or delete the yeast's endogenous DHFR, *DFR1*. Since it is an essential gene, to be able to grow a *dfr1* deletion strain, we had to first insert another DHFR (*dfrB1*) which would be expressed under the control of a b-estradiol induced promoter (*GAL1pr*). *GAL1pr* allows for the expression of the downstream gene in the presence of β-estradiol (Sigma-Aldrich, catalog #E2758-5G). Construction of strain IGA130 (*his3 ura3 met15 LEU2::GEM, GAL1pr::dfrB1-NATMX*) was done by inserting the *dfrB1* gene (previously known as R67 - yeast codon optimized sequence) at the *GAL1* locus in strain AKD0679 (*LEU2::GEM, GAL1pr*::Stuffer-*NATMX*) by CRISPR-Cas9-mediated integration following the protocol detailed in (Ryan et al., 2016) and using pCas plasmid with a gRNA sequence targeting the Stuffer from (Dionne et al., 2021). Transformants were selected on YPD + Geneticin 200 µg/mL (G418) + Nourseothricin 100 µg/mL (NAT). Following loss of pCAS plasmid, insertion of *dfrB1* in the genome was verified by colony PCR using oligos OP151-E5 and ADHterm_R with BioShop Taq DNA Polymerase and validated by Sanger sequencing. Construction of strain FDR0001 (*his3 ura3 met15 LEU2::GEM, GAL1pr::dfrB1-NATMX dfr1::Stuffer*) was done using a similar CRISPR-Cas9-mediated protocol as described above, but using a Stuffer sequence and a gRNA targeting *DFR1*. Oligos DFR1_KO_F and DFR1_KO_R were used to amplify the Stuffer sequence used to knockout *DFR1*. Transformants were selected on YPD + G418 + NAT + 100 nM β-estradiol. Following the loss of pCAS plasmid, deletions were confirmed by PCR using oligos DFR1_KO_valid_F and DFR1_KO_valid_R and Sanger sequencing (BioShop Taq DNA Polymerase). Detailed PCR protocols are available in Supplementary Table PCR. Details for media composition are available in Supplementary Table Media.

All yeast experiments were conducted at 30 °C for 48 h unless stated otherwise. All experiments involving bacteria were conducted at 37 °C using 2YT with appropriate antibiotics unless stated otherwise. All liquid cultures were incubated with 250 rpm orbital agitation. All Sanger sequencing was performed by the Sequencing Platform of the Centre Hospitalier de l'Université Laval (Université Laval, Québec, Canada), and all next-generation sequencing was performed using the MiSeq Reagent Kit v3 on an Illumina MiSeq for 600 cycles at the Sequencing Platform of the Institut de Biologie Intégrative et des Systèmes (Université Laval, Québec, Canada).

Plasmid pAG416-GPD-PjDHFR was constructed from pAG416-GPD-ccdB and *Pneumocystis jirovecii* DHFR (PjDHFR) (UniPROT: A0EPZ9), codon-optimized using GenScript Biotech's GenSmart™ Codon Optimization tool, and ordered from Twist Bioscience (San Francisco, California, USA) as a Gene Fragment. pAG416-GPD-ccdB was digested using XbaI and HindIII-HF in CutSmart® buffer (New England Biolabs,

catalog #R0145 and #R0104), according to manufacturer's instructions, as the backbone. PjDHFR was amplified by PCR using oligos Pj_to_pAG_F and Pj_to_pAG_R under parameters specified in supplementary table PCR, as the insert. Backbone and Insert were assembled using an in-house Gibson assembly mix (1:2 vector:insert ratio) and cloned into *Escherichia coli* MC1061 ([araD139]B/r ∆(araA-leu)7697 ∆lacX74 galK16 galE15(GalS) λ-e14-mcrA0 relA1 rpsL150(strR) spoT1 mcrB1 hsdR2) using a standard heat shock transformation method and plated onto ampicillin-containing 2YT media (100 µg/mL). Clones were confirmed by colony PCR using oligos pAG_insert_F and pAG_insert_R, and the amplicon sequence was confirmed by Sanger sequencing. The same protocol was used to construct pAG416-GPD-ScDHFR (*DFR1* insert amplified from BY4741 genomic DNA using oligos Sc_to_pAG_F and Sc_to_pAG_R) and pAG416-GPD-mDHFR (mDHFR, synthetic, L22F and F31S, insert amplified from pGEST-linker-fulllength-DHFR-HYG_B plasmid using oligos Murine_to_pAG_F and Murine_to_pAG_R.)

Functional complementation assay

The complementation assay was conducted by transforming the DHFR-containing plasmids into strain FDR0001 and plating transformations on solid SC-URA+MSG+100 nM β-estradiol media to ensure growth of transformants even in the absence of complementation. Four individual colonies were harvested from each transformation and streaked on solid SC-URA+MSG+100 nM β-estradiol media, and colony PCR was conducted to confirm transformation success. For both spot-dilution assays, transformants were grown in liquid SC-URA_MSG+100 nM β-estradiol media overnight. Cultures were diluted to 0.1 OD_{600} in the same media and grown to 0.3 OD_{600} to ensure that cells were in exponential phase. 1.0 OD_{600} unit worth of cells were harvested from the cultures, washed, and resuspended in ddH_20, and 5 µL was spotted on MTX media (See Supplementary Table Media for detailed recipe) with and without 440 µM methotrexate (BioShop, catalog #MTX440) and with and without 100 nM β-estradiol. The following spots were done by five-fold serial dilution in the same media.

Dose response curves

Dose response curves were conducted in triplicate using a Tecan Infinite M Nano (Tecan Life Sciences) plate reader. DHFR-containing plasmids were transformed into strain IGA130. Three individual colonies were sampled from transformation plates and grown individually in 3 mL SC-URA+MSG in V-bottom 24-well deep well plates overnight. The next morning, 1.0 OD_{600} unit of cells was harvested and spun down at ~500 rcf for 4 minutes. The supernatant was removed, and the pellet washed with sterile deionized water, and then resuspended in 1 mL of sterile deionized water. 20 µL of this resuspension was added in a flat bottom Greiner 96 well plate, to 180 µL of liquid MTX media prepared the same day, for a final OD_{600} of 0.1. Liquid MTX media was prepared for growth curves using a ten-fold dilution of MTX media at 440 µM MTX to 44 µM, and then subsequent two-fold dilutions to attain 0.086 µM of MTX. Plates' edges were sealed with Parafilm to limit media evaporation and placed into plate readers at 30 °C for 48h without agitation. OD_{600} was read every 15 minutes at four points within the well, and mean OD_{600} was used as a reading.

223 The growth rate for strain IGA130 with a plasmid-encoded DHFR was calculated
224 from the slope at every time point before 40h, and taking the median of the five largest
225 values for a more robust estimate. This was done for every replicate. Growth coefficient
226 was measured for each replicate using the following equation (Eq1), with the derivative
227 growth rate of strain IGA130 with plasmid pAG416-GPD-Empty (Empty) without MTX as
228 reference for maximum growth rate:

229 **Eq1**: $Growth\ coefficient = 1 - \frac{(Reference\ growth\ rate - Sample\ growth\ rate)}{Reference\ growth\ rate}$

230 Inhibitory concentrations (ICs) were computed using the standard Hill equation, using
231 Python's SciPy optimize.curve_fit method (SciPy version 1.7.3).

232 **Deep mutational scanning library preparation**

233 A deep mutational scanning library of PjDHFR was constructed using the
234 previously described megaprimer method [48,49]. Primers with NNK degenerate
235 sequence at the codon of interest were ordered from Eurofins Genomics for all positions
236 along PjDHFR, except for the start and stop codons. By using these primers with reverse
237 primers Mega_1_R, Mega_2_R and Mega_3_R, for fragments 1, 2 and 3, respectively,
238 megaprimers were generated by PCR using KAPA Hi-Fi polymerase and pAG416-GPD-
239 PjDHFR as template on a per-position basis. This PCR product was diluted 1/2500 with
240 two serial 50-fold dilutions, and 2 µL of these dilutions were used as megaprimers for the
241 amplification of pAG416-GDP-PjDHFR using KAPA Hi-Fi polymerase. PCR products
242 were digested using 0.2 µL of DpnI restriction enzyme (New England Biolabs, catalog
243 #R0176L) (37 °C for 2 h) to remove parental DNA. This product was then transformed in
244 *E. coli* MC1061, and colonies were counted. Biomass from petri dishes with >400 colonies
245 was harvested (>10x coverage for each codon), and plasmid libraries extracted per
246 position using the FroggaBio Presto Mini Plasmid Kit (catalog #PDH300.) Each plasmid
247 positional library was quantified using a ThermoScientific Nanodrop 2000
248 spectrophotometer and library content was assessed for diversity using the RC-PCR
249 method from [50] with primers Lib_seq_F1_F/R, Lib_seq_F2_F/R and Lib_seq_F3_F/R
250 for the different fragments. Positions where diversity was assessed as being too low were
251 redone using the same protocol. All oligos used for library generation are available in
252 Supplementary Table Oligos.

253 Once diversity was confirmed at every position, plasmid libraries were transformed
254 using the same method as above into IGA130 and plated on SC-URA+MSG media.
255 Biomass from Petri dishes with >1000 colonies were harvested, and glycerol stock
256 cultures were prepared for each position. Once every position was transformed in yeast
257 and appropriately frozen as glycerol stocks, 100 µL of each glycerol stock was added to
258 individual 5 mL culture tubes of SCc-URA+MSG liquid media and grown overnight. OD_{600}
259 was measured for each position, and an equal number of cells from each position was
260 mixed as a master pool for each fragment, and a glycerol stock was prepared for each
261 fragment.

262 **Pooled competition assay**

263 Starting from these master stocks, 100 µL were added to 5 mL of SC-URA+MSG
264 media and grown overnight in triplicate. These cultures were considered as T_0 samples.

265 OD_{600} was measured, and 1.5 OD_{600} units were spun down at 224 rcf for 3 minutes and
266 resuspended in 1 mL of sterile water. 14 mL of 1.07x MTX media was inoculated with 1
267 mL of 1.5 OD_{600} culture, to yield 1.0x media at 0.1 OD_{600}, in triplicate. Depending on the
268 condition, these cultures had either 0 μM (DMSO), 8.8 μM (IC 75) or 44 μM (IC90) of
269 MTX. DMSO was used as the drugless control, as it is the solvent for MTX. Once cultures
270 reached either 1.2 OD_{600} or were grown for 72 h, cultures were diluted back to 0.1 OD_{600}
271 in a fresh batch of the same media using the same method. Once grown, these culturse
272 were considered as T_{final}, and cells were harvested once cultures reached either 1.2 OD_{600}
273 or 72 h of growth in this second passage. At every passage and for every culture, triplicate
274 glycerol stocks were made, and 5.0 OD_{600} units were harvested and spun down at 224
275 rcf for 3 minutes. Supernatant was removed and pellets were frozen for downstream DNA
276 extraction. DNA extractions were conducted using Zymoprep Yeast Plasmid Miniprep II
277 (Zymo Research, catalog #D2004). The coding sequence of PjDHFR was split into three
278 fragments of roughly equal length (F1, positions 2 to 73, F2, 74 to 145, and F3, 146 to
279 206) to allow Illumina MiSeq 300 bp paired-end sequencing. Extracted libraries were
280 amplified by fragments using Lib_seq_F/R primers using a cycle described in
281 Supplementary Table PCR, purified using magnetic beads (AMPureXP from Beckman
282 Coulter Life Sciences, catalog #A63882) and sent for sequencing for a total of 15M reads
283 on an Illumina MiSeq.

284 **Next-generation sequencing read processing**

285 After sequencing, reads were analyzed using custom Python and Bash scripts.
286 Briefly, demultiplexed R1 and R2 reads were trimmed to remove primer sequences and
287 merged using PANDAseq [51]. Merged reads were aggregated and counted using
288 VSEARCH [52]. Using custom Python scripts, expected codon mutations were matched
289 with read counts, and the frequency of each mutant was assessed in each individual
290 culture. To minimize sampling bias, we ensured that every retained mutant had high initial
291 coverage in T_0 to minimize bias in downstream analysis and discarded all mutants that
292 had an initial coverage of less than 15 high quality reads. Synonymous codons to wild-
293 type PjDHFR were used as internal control and reference for normalization between
294 fragments. The selection coefficient was calculated according to Eq2, where the
295 normalized log2-fold change of the mutants between T_0 and T_{final} was divided by the
296 median normalized log2-fold change of the synonymous mutants and normalized by the
297 number of mitotic generations measured for each replicate (g). The final selection
298 coefficient value for a mutant was computed using the median selection coefficient
299 between all synonymous mutants within each individual culture.

300 **Eq2**: $$Selection\ coefficient = \frac{Log2\left(\frac{Freq\ mutant\ T_{final}}{Median\ freq\ silent\ T_{final}}\right) - Log2\left(\frac{Freq\ mutant\ T_0}{Median\ freq\ silent\ T_{final}}\right)}{g}$$

301

302 This normalization allowed us to generate a standardized score where 0 is the
303 median score of synonymous mutants, which gives an accurate estimate of the wild-type
304 fitness in the tested conditions and of the relative performance of all mutants in
305 comparison to wild-type PjDHFR in every condition. The resulting distribution of selection
306 coefficients was used in downstream analyses.

Statistical analysis of selection coefficients and false-discovery rate

To assess resistance thresholds using statistical means, we used the sklearn mixture package (version 1.0.2) [53] to estimate the fraction of silent mutations, resistant and very resistant mutations by fitting the distribution of selection coefficients under MTX selection to a mixture model of underlying Gaussian distributions. Using custom scripts, we estimated how many parameters recapitulated the overlying function the best by minimizing Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for the gaussian mixture models using an expectation-maximization model. AIC and BIC are metrics to minimize to favor the simplest model to maximize the fitting. Low BIC favors simpler models, and low AIC favors more accurate models. By using the best fit model, we separated our distributions of selection coefficients into their underlying Gaussian mixture models, and set our thresholds based on parameters from the two rightmost distributions (most resistant mutants and second-to most resistant mutants). To minimize overlap effects between underlying Gaussian distributions, we chose the median values of the second-to-rightmost underlying distribution as the resistant threshold, and the intersection second-to-rightmost and rightmost as the very resistant threshold in each condition where the drug was present (S1 Fig.).

To correct for experimental noise and false discovery rate, we conducted a one sided Welch's *t*-test (scipy.stats version 1.7.3) [54]. We generated a matrix for each mutation at each position, which regrouped values for all replicates and synonymous mutations for a specific amino acid substitution and compared it to the distribution of silent mutations to assess if the selection coefficient of mutants with this substitution was significantly greater (i.e. more resistant) than the distribution of silent mutations. We then used the Benjamini-Hochberg method to control for the false discovery rate of significantly resistant mutants at 5% (statsmodels version 0.13.5) [55]. This allowed us to identify a subset of mutants that have selection coefficients significantly greater than the wild-type, which we classify as being resistant, and corrected for false discovery rate. All custom code is available at https://github.com/Landrylab/Rouleau_et_al_2023.git.

Construction of validation mutants

To validate results from the pooled competition assay, we selected different mutants based on their scores, patterns of resistance, and reproducibility across replicates and conditions, and reconstructed these manually (Supplementary Table Validations). To generate these mutants, identified by their specific reconstructed mutation in Supplementary Table Plasmids, we used fusion PCR followed by Gibson assembly. First, by using oligonucleotides containing the mutation of interest that are a reverse complement to each other, we generated the two fusion PCR fragments using Valid_master_F and any R_valid oligo, and Valid_master_R and any F_valid oligo. pAG416-GPD-PjDHFR was used as the template vector. Then, PCR products were mixed and diluted 1/2500, and used as a template with Valid_master_F and Valid_master_R oligos for the generation of the final product. The fusion PCR product was then used in Gibson cloning, using the same method and backbone as described above, and clones were validated using Sanger sequencing. Constructions were transformed into the relevant strain (IGA130 or FDR0001) and plated on SC-URA+MSG +/- 100 nM β-estradiol depending on the strain. Individual growth curves and growth rates

351 were conducted and measured as described above. Subsequent spot-dilution assays
352 were conducted as described above, but with 44 µM of MTX instead of 440 µM.

353 **Western blot analysis**

354 For western blot analysis, different variants of PjDHFR and ScDHFR were tagged
355 using either 3xFLAG tag in C-terminal (3') or 1xFLAG tag in N-terminal (5'). For N-terminal
356 constructions, pAG416-GPD-PjDHFR (wild-type, validation mutants pAG416-GPD-
357 PjDHFR-M33P and pAG416-GPD-PjDHFR-F199P) or pAG416-GPD-ScDHFR were
358 digested using ClaI enzyme in rCutSmart® buffer (New England Biolabs, catalog #R0197)
359 according to manufacturer's instructions, and purified using magnetic beads as described
360 above. 3xFLAG cassettes were amplified using oligos C3xFLAG_F and _R from p3-
361 3xFLAG-KanMX. The PCR product was purified using magnetic beads. Gibson assembly
362 was conducted as described above and confirmed by Sanger sequencing using
363 pAG_insert_F. For C-terminal tagging, 1xFLAG cassette was amplified from p3-3xFLAG-
364 KanMX using N1xFLAG_F1_F and N1xFLAG_F1_R_nomet or _met, to remove or keep
365 the PjDHFR start codon, respectively. Then, PjDHFR was amplified from pAG416-GPD-
366 PjDHFR using N1xFLAG_F2_F_nomet or _met and N1xFLAG_F2_R, to remove or keep
367 PjDHFR start codon, respectively. These fragments were assembled using fusion PCR,
368 as described above, and assembled into a digested (HindIII, rCutSmart®, New England
369 Biolabs, catalog #R0104) and purified pAG416-GPD-ccdB using in-house Gibson
370 assembly, as described above. All constructions were confirmed by Sanger sequencing
371 using pAG_insert_F. All constructions were transformed in IGA130 as described above,
372 and transformants were grown overnight in SC-URA+MSG+100 nM β-estradiol in
373 triplicate. A schematic for all constructions is available in S2A Fig.

374 Yeast proteins were extracted by resuspending 12.5 OD_{600} units in 250 µL of lysis
375 buffer (Complete Mini Protease inhibitor cocktail, Millipore-Sigma 11836153001). Glass
376 beads were added to the cell suspension and the tubes were vortexed for 5 min on a
377 Turbomix. Following the vortex step, 25 µL SDS 10% were added to each tube before
378 boiling them for 10 min. Once the tubes were boiled, they were centrifuged for 5 min at
379 16,000 rcf to clear the supernatant. Samples for migration were prepared by mixing 17.5
380 µL of cell extract with 2.5 µL DTT 1 M and 5 µL loading buffer 5X (250 mM Tris-Cl pH 6.8,
381 10% SDS, 0.5% Bromophenol Blue, 50% glycerol). Samples were migrated on a 15%
382 SDS-PAGE acrylamide gel for 55 min, then transferred on a nitrocellulose membrane
383 using a semi-dry system (TE 77 PWR, Amersham Bioscience) for 1h at 0.8 mA/cm^2.
384 Following the transfer, membranes were soaked in a Ponceau red solution for 5 min,
385 destained quickly with water and imaged. To remove remaining Ponceau staining, the
386 membranes were shaken with PBS (10 mM sodium phosphate dibasic, 135 mM NaCl, 2
387 mM KCl, 1.5 mM monopotassium phosphate) solution for around 45 min. Membranes
388 were then blocked overnight with agitation in blocking solution (Intercept® (PBS) Blocking
389 Buffer, Mandel Scientific, catalog #927-70003) at room temperature. The following
390 morning, membranes were incubated with primary antibody 0.15 µg/mL (Anti-Flag M2,
391 F3165-1MG, Millipore-Sigma) in blocking solution + 0.02% Tween 20 for 30 min.
392 Membranes were washed 3 times for 5 min with PBS-T (PBS with 0.1% Tween 20). The
393 secondary antibody was then added at 0.075 µg/mL (Anti-mouse 800, LIC-926-32210,
394 Mandel Scientific) in blocking solution + 0.02% Tween 20 for 30 min. Membranes were

washed again 3 times for 5 min with PBS+T, followed by imaging with an Odyssey Fc instrument (Licor, Mandel Scientific) in channel 700 (30s) and 800 (120s).

Structural analysis, alignments, and allostery analysis

The structure of PjDHFR was recovered from AlphaFoldDB (UniProt: A0EPZ9). Crystal structures of DHFR ortholog from *P. carinii*, in complex with MTX/NADPH (PDB: 3CD2) and DHF/NADPH (PDB: 4CD2), were recovered from UniProt. Using ChimeraX's (version 1.6) [56] matchmaker algorithm, the structures were aligned with each other, and ligands from crystal structures were extracted and added to PjDHFR AlphaFold structure. The resulting complexed structures were then relaxed using the FoldX (version 5.0) optimize command to remove van der Waals clashes. Using the MutateX protocol on the PjDHFR complex with MTX/NADPH and DHF/NADPH, changes in protein free energy were computed.

For visual representation of selection coefficient on protein structure, chains were colored using the ChimeraX color *byattribute* command, and attributes were manually defined within custom .defattr files. Data for coloring *E. coli* DHFR according to allostery was recovered from supplementary S1 table from [57], with darker colors corresponding to more stringent confidence intervals for allostery. Alignments between *E. coli* DHFR and PjDHFR were done using ChimeraX's matchmaker.

To measure distance between protein residues and ligands, a custom script was used, based on tools from BIO.PDB (version 1.78) [58]. Briefly, by using a .pdb file coordinate parser, distance was estimated between α carbons from all residues in PjDHFR and all atoms from the ligand of interest, except hydrogen atoms. Then, the minimal distance between the residue α carbon and any atoms in the ligand of interest was considered for establishing contact. Any distance shorter than 8 Å was considered a contact.

Modeling of binding affinity changes

To model changes in binding affinity to ligands caused by mutations, we used the Rosetta FlexddG method [59]. Briefly, binding molecules were parameterized using Rosetta's molfile_to_params.py script. *In silico* saturation mutagenesis was conducted using a modified version of the provided script from the FlexddG GitHub tutorial, and the generated params files as extra param files [59]. 100 structures with 35000 backrub steps were used as parameters, based on recommendations from [60], and all individual resulting structures were analyzed using a modified version of the provided script from the FlexddG GitHub tutorial. This resulted in a distribution of ddG values for 100 individual replicates. To minimize extreme effects caused by potential clashes or other artifacts, despite the backrub steps, the median values from these distributions were used as the final ddG for downstream analyses.

Results

PjDHFR can functionally complement the deletion of native DHFR, *DFR1*, in *S. cerevisiae* and is sensitive to methotrexate

In the budding yeast *S. cerevisiae*, the DHFR is encoded by the gene *DFR1* (ScDHFR). *DFR1* is an essential gene [61], which means that complementation assays

437 require the use of an engineered strain. Some strains with *DFR1* deletion or
438 downregulation have been constructed, but they require multiple mutations, such as
439 inactivation of *TUP1,* and specific growth media to be handled [12,62,63]. Here, we
440 performed the full deletion of *DFR1* in *S. cerevisiae* for functional complementation with
441 PjDHFR. We engineered a strain in which the function of *DFR1* is complemented by
442 introducing another DHFR-coding gene (*dfrB1*) from bacteria, that is nonhomologous to
443 *DFR1* and that is insensitive to antifolates [64]. Upon induction of a β-estradiol inducible
444 promoter, expression of this *dfrB1* masks the deletion of *DFR1,* and allows for growth in
445 the presence of MTX (Fig 1A and 1B) [65]. This strain, FDR0001 (*dfr1Δ*), is therefore non-
446 viable in the absence of β-estradiol, but viable and MTX-resistant in the presence of β-
447 estradiol.

448 By transforming FDR0001 with a plasmid harboring PjDHFR under the regulation
449 of the constitutive GPD promoter and by growing it in the absence of β-estradiol, we found
450 that PjDHFR functionally complements the deletion of *DFR1* in *S. cerevisiae* and rescues
451 growth at a near wild-type level (Fig 1B). In the absence of β-estradiol, the strain without
452 DHFR on the plasmid (Empty, negative control) was unable to grow, confirming that the
453 phenotype observed depends on the presence of a complementing DHFR on the plasmid.
454 Using media supplemented with MTX at a concentration known to fully inhibit *DFR1* at its
455 endogenous expression level (440 μM) [66], we used this same system to assess
456 PjDHFR sensitivity to this antifolate. The MTX-resistant mouse DHFR (mDHFR) showed
457 complementation and resistance to MTX, as expected [67]. Complementation of the
458 inhibition of *DFR1* by this mDHFR in *S. cerevisiae* and its associated resistance to MTX
459 have been reported in several studies [66,67]. Interestingly, when using a plasmid
460 harboring ScDHFR under the regulation of the GPD promoter as a positive control for
461 complementation, we observed resistance to 440 μM of MTX on solid media. Since *DFR1*,
462 at its native expression level and without vector-expressed DHFR (Empty) is sensitive to
463 MTX, this experiment shows that high expression of this protein is sufficient to achieve
464 resistance to MTX at these concentrations. Resistance caused by overexpression has
465 previously been reported in *E. coli* for the antifolate TMP [68]. Despite being under the
466 regulation of the same promoter, the PjDHFR-harboring strain was sensitive to MTX.
467 Western blot results (S2B Fig), alongside complementation assays, show that PjDHFR is
468 expressed and functional in *S. cerevisiae,* and that PjDHFR is more sensitive to this
469 antifolate than wild-type *DFR1*, at least in this context. This highlights the potential for
470 PjDHFR to evolve toward resistance to MTX and the usefulness of our system to identify
471 resistance mutations.

472

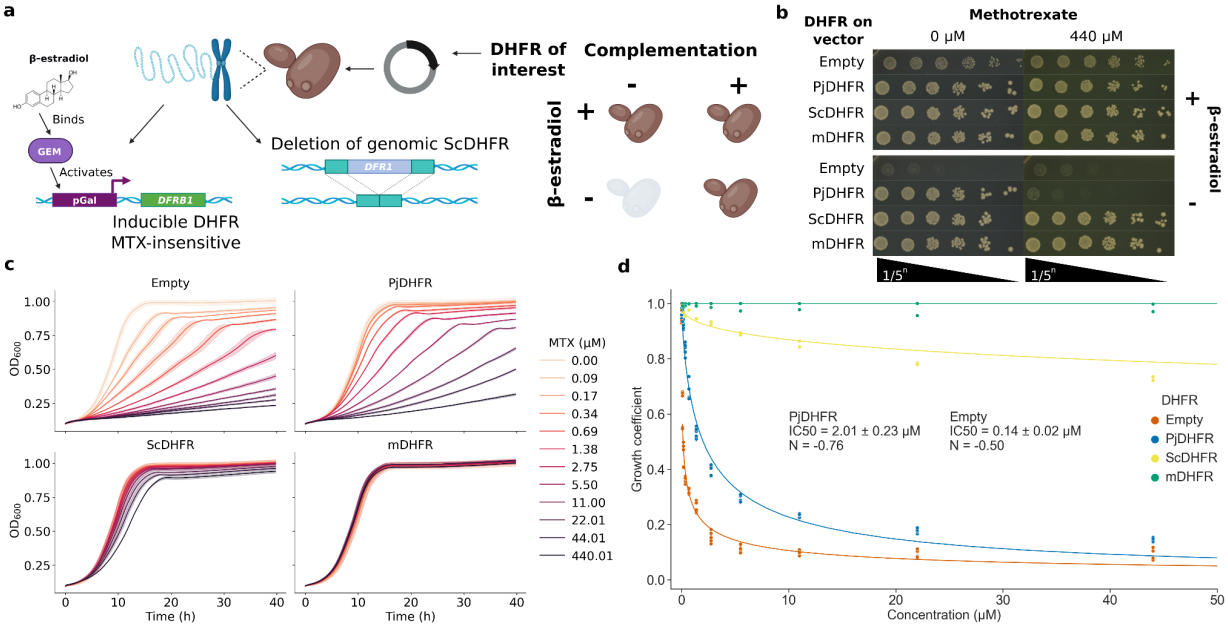


Fig 1. Experimental setup for complementation assay using PjDHFR and dose response curves. a) Experimental setup and strain construction. The yeast strain, FDR0001, harbors the *dfrB1* gene under the regulation of a β-estradiol inducible promoter (*Galpr*) [65], and a *dfr1*Δ deletion. *DfrB1* is insensitive to MTX, therefore all strains can grow on MTX-containing media in the presence of β-estradiol. This panel was made with BioRender. **b)** Spot-dilution assays to test for DHFR complementation and for MTX sensitivity. PjDHFR, ScDHFR and mDHFR are from *P. jirovecii* (GenBank: ABB84736.1, codon optimized for *S. cerevisiae*), *S. cerevisiae BY4741 DFR1*, and *M. musculus* mDHFR (synthetic, L22F and F31S, from [67]), respectively. ScDHFR was used as a complementation control, and mDHFR as a MTX-resistant control. When β-estradiol is added to the media, *dfrB1* is expressed, which is necessary for cell growth in a *dfr1*Δ *S. cerevisiae* background without DHFR on the vector. **c)** Growth curves of yeast strain IGA130 (wild-type for *DFR1*) each expressing a different DHFR. Shaded areas represent confidence intervals across biological triplicates. Empty vector shows sensitivity to low concentrations of MTX. Strains with vector-expressed DHFRs show increased resistance, with PjDHFR being the most sensitive, followed by ScDHFR, and mDHFR(L22F/F31S) being fully resistant. **d)** Growth coefficient (measure from growth rate relative to Empty without selection) of strain IGA130 with different DHFRs. R^2 for Hill equation fit: Empty =0.96, PjDHFR = 0.98, ScDHFR = 0.91 and mDHFR = 0.

To quantitatively assess the extent of inhibition and calculate inhibitory concentrations (IC) of PjDHFR in *S. cerevisiae*, we measured dose-response curves in the presence of decreasing concentrations of MTX. By comparing growth rates measured in the presence of different concentrations of the drugs to the growth from the Empty vector in the absence of the drug as growth coefficient, we estimated the sensitivity of strain IGA130, which still has wild-type *DFR1* at the genomic locus. Here, we switched from strain FDR0001 (*dfr1Δ*) to the parent strain IGA130 (wild-type for *DFR1*) for downstream screening because FDR0001 had insufficient transformation efficiency for the construction of DMS libraries.

In strain IGA130, mDHFR and Empty vector were used as positive and negative controls for MTX resistance, respectively (Fig 1C). None of the tested concentrations had a detectable effect on the growth rate of strains harboring MTX-resistant mDHFR. Empty plasmid showed the highest sensitivity, as only the natively expressed *DFR1* was present in this strain. Using a Hill-equation curve fit, we calculated ICs for these different DHFRs

505 (Fig 1D). IC50, when expressing no exogenous DHFR, was over 10 times lower than
506 when expressing PjDHFR (0.14 μM and 2.01 μM, respectively), highlighting the sensitivity
507 of the genomic *DFR1* and the relatively limited impact of genomic *DFR1* on resistance to
508 MTX. The conditions that were selected for downstream competition assays correspond
509 to roughly 75% and 90% inhibition for PjDHFR (IC75 = 8.8 μM MTX and IC90 = 44 μM
510 MTX).

511 **Pooled competition assays of PjDHFR mutant library identify mutations**
512 **conferring resistance to methotrexate**

513 Deep mutational scanning (DMS) was conducted on PjDHFR using the
514 megaprimer method [49]. We generated all possible single amino acid changes of
515 PjDHFR on the plasmid described above. To assess the fitness effect of mutations, we
516 measured the change in frequency for every mutant between the initial culture (T_0) and
517 the second passage in the presence of MTX (T_{final}). The pooled competition assay in
518 DMSO was included to identify any mutations that could have dominant negative effects
519 on growth and that could thus confound resistant or sensitive phenotypes (Fig 2A). At
520 IC75 and IC90 (Fig 2B and 2C) selection coefficients were used to identify mutations that
521 cause resistance to MTX. The final count of mutants covered 97.2% of all possible amino
522 acid replacements (4166/4284). We calculated a selection coefficient for each mutant in
523 each condition, which quantifies the effect on growth of an amino acid substitution relative
524 to the wild-type PjDHFR, where a positive score represents a growth advantage, and a
525 negative score a growth disadvantage. Replicates within conditions were well correlated
526 (S3A, S3B and S3C Fig).

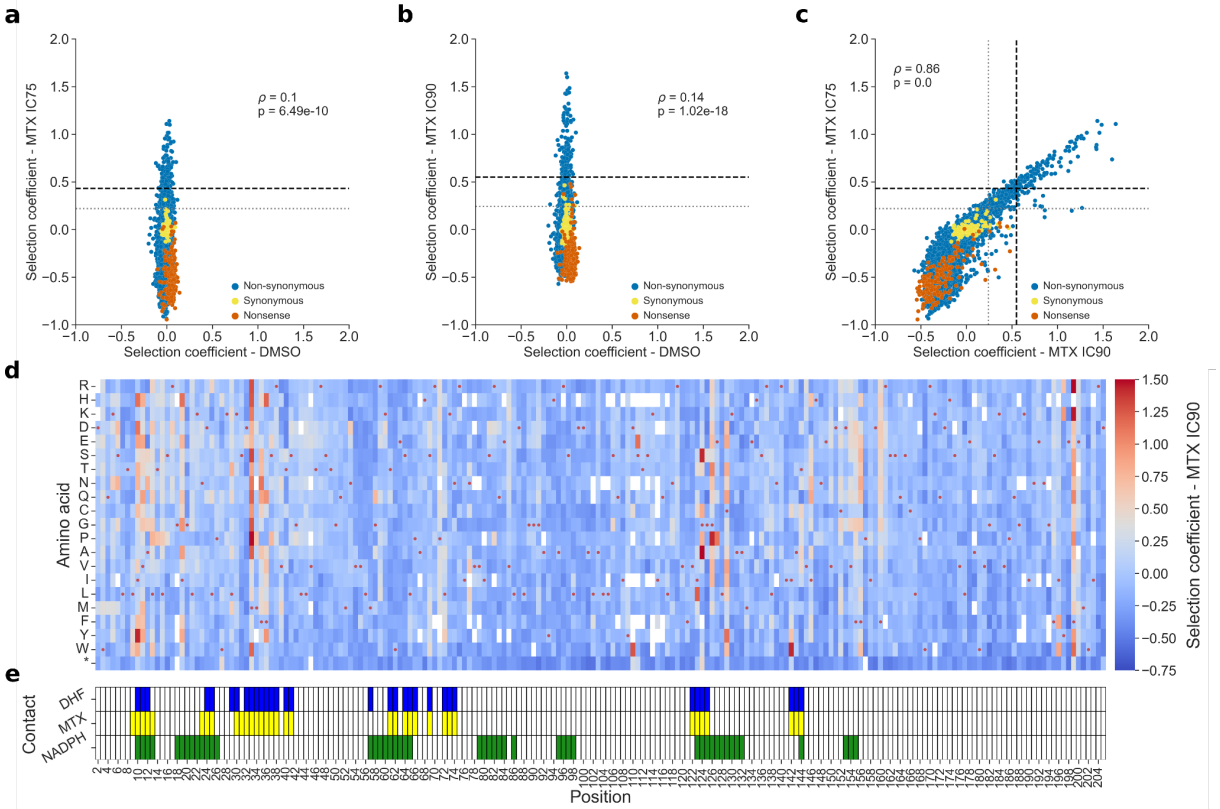


Fig 2. Effects of amino acid substitutions on the resistance of PjDHFR to methotrexate. a) Comparison of selection coefficients between control condition (DMSO) and at IC75 (Spearman's ρ = 0.1, p = 6.49e-10), **b)** DMSO and IC90 (Spearman's ρ = 0.14, p = 1.02e-18), and **c)** IC75 and IC90 (Spearman's ρ = 0.86, p < 1e-100). Different types of variants found in the library are marked. Nonsense mutants have premature stop codons. Synonymous are mutants that have different codon sequences but that have the same amino acid sequences as the wild-type sequence. Non-synonymous are amino acid changes relative to wild-type PjDHFR. Resistant (grey/dotted line) and very resistant (black/dashed line) thresholds are shown for both MTX conditions (See *Statistical analysis of selection coefficients* in methods for threshold determination, S1 Fig). **d)** Selection coefficient of a given amino acid (y-axis) at a given position (x-axis) at a MTX concentration corresponding to 90% of growth inhibition (IC90). Red points show the wild-type sequence of PjDHFR. **e)** Positions of contacting residues along PjDHFR modeled on structural alignments between PjDHFR and orthologous *P. carinii* DHFR (PDB: 3CD2 (MTX and NADPH) and 4CD2 (DHF)). Contact was established as amino acids with an α carbon located less than 8 Å from MTX, DHF or NADPH. Detailed heatmaps for DMSO and IC75 are available in S4 Fig.

By using results from the competition assay with MTX, we identified mutations PjDHFR that confer resistance to this antifolate. To classify mutants as being resistant or not, we first used a Gaussian mixture model to separate the distribution of fitness effects into the number of Gaussian components that best recapitulate the overlying distribution of observed fitness effects at IC75 (five components) and IC90 (four components) (S1 Fig). By using the estimated parameters from the underlying functions, we set the thresholds for resistance using quantiles of the distributions centered on a selection coefficient of 0. For IC75, the thresholds for a resistant mutant and a very resistant mutant were set at selection coefficients of over 0.220 (404/4284 mutants, 9.43%) and 0.432 (129/4284, 3.01%), respectively. For IC90, 0.243 (437/4284, 10.20%) and 0.550 (134/4284, 3.13%), were set for resistant and very resistant, respectively. Because of the

difference in the number of modelled Gaussian function components in the two conditions, potentially caused by the difference in strength of the selection pressures, slight differences in the number of mutants that pass the different thresholds are expected.

Overall, 31 positions, or about 15% of the positions along the protein, had at least one mutant that passed the very resistant threshold in both concentrations of MTX. These cutoffs do not consider the various numbers of replicates per amino acid in these experiments, so mutations above these thresholds do not have the same levels of confidence. We therefore also conducted a one-sided Welch's t-test and corrected for false discovery rates using the Benjamini–Hochberg procedure to control for the false discovery rate of significantly resistant mutants at 5%. We tested whether selection coefficients for amino acid variants were significantly greater than that of synonymous mutants of the wild-type PjDHFR. For IC75, 836 mutants with selection coefficients >0 were considered as significant, and 694 for IC90, with an intersection between the two concentrations of 603 mutations (S1C and S1D Fig, green line, and S5 Fig). Finally, analyzing intersecting datasets from thresholds and statistical significance, we identified 355 mutants as being resistant with a high level of confidence, and 116 mutants that we consider very resistant with a high level of confidence.

Individual growth assays validate results from the pooled competition assay

To validate the results of the pooled competition assays, we manually reconstructed 44 mutations in PjDHFR (Fig 3A, 3B and 3C). These mutations were chosen because they had various selection coefficients along the gradient of values observed. We also included positions that showed discrepancies between the IC75 and IC90 conditions, such as G58Q, D110W, S111E, I196W and I196F (Fig 3A, highlighted green). These mutants exhibited high resistance at IC90 but behaved differently at IC75, not passing the threshold for very resistant. Because the results strongly correlate between these two conditions, we reasoned that these could be experimental outliers, as we had observed increased noise between replicates at IC90 compared to IC75 (S3B an S3C Fig). These outlier mutants are enriched for amino acids encoded by a single codon in our experiment, which could partly explain their behavior. Validation mutants were tested using growth curves and spot-dilution assays at IC90 in the same conditions as the competition assay, as well as in DMSO. Results of the growth assays confirmed that these outliers were indeed sensitive to MTX at IC90 (Fig 3B). Growth rate for all other mutants individually tested correlated strongly with their selection coefficient from the competition assay based on Spearman's rank correlation (Fig 3B and D). All mutants grew similarly in DMSO (Fig 3C and 3D).

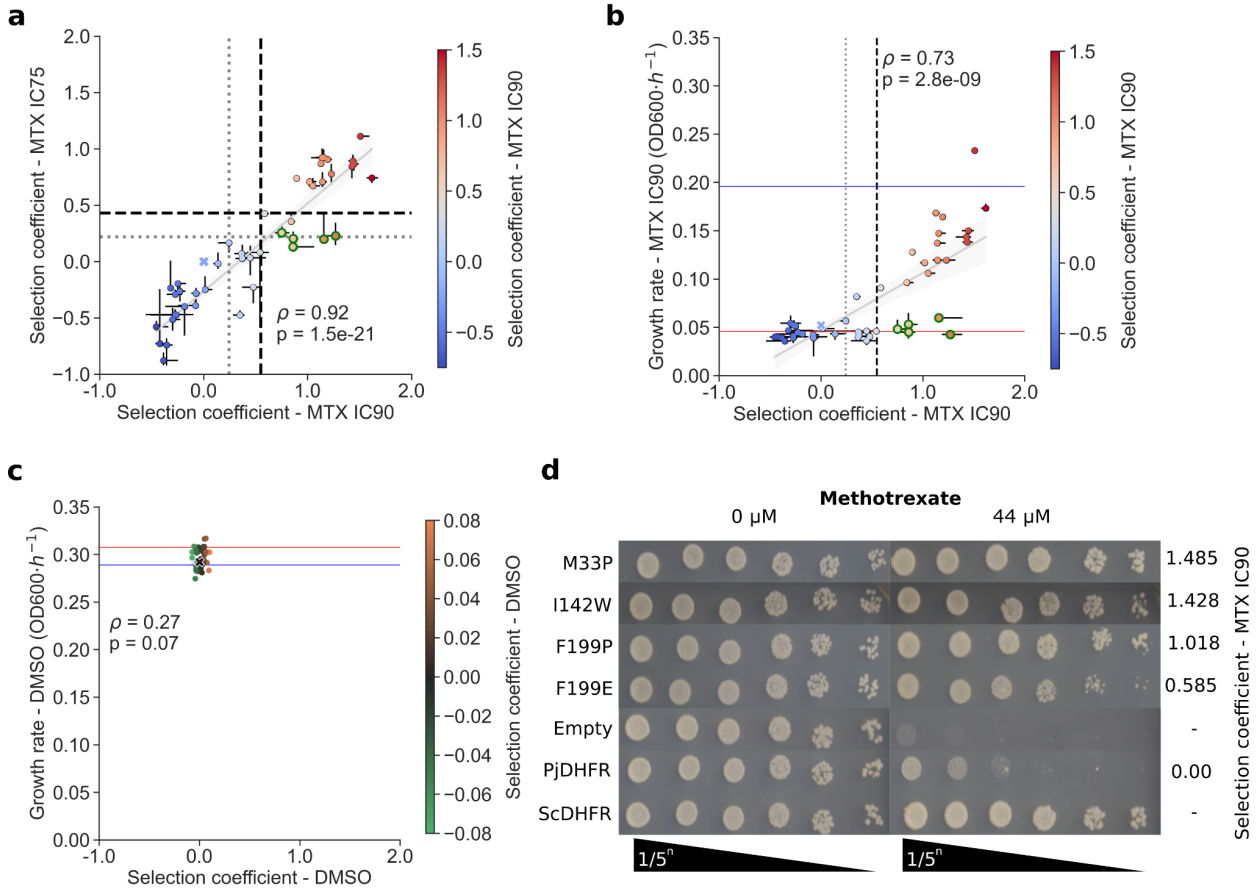


Fig 3. Validation of pooled competition assay with individually reconstructed mutants. a) Selection coefficient at IC75 and IC90 for mutants that were selected. Variants at positions showing large deviations between IC75 and IC90 are visible in the bottom right quadrant (gray lines, resistant threshold, black lines, very resistant thresholds, from gaussian mixture models, S3 Fig). Wild-type PjDHFR (blue X) is at 0.0 selection coefficient. Error bars in x represent mutant selection coefficients across synonymous codons and replicates at IC90 MTX and in y, at IC75 MTX. **b)** Growth rate ($OD_{600} \cdot h^{-1}$) compared to selection coefficient in IC90 for validation mutants. The blue line represents the derivative growth rate for the strain with ScDHFR on the plasmid and the red line represents the growth rate for the strain with the Empty plasmid (negative control). Error bars in x represent mutant selection coefficients across synonymous codons and replicates, and error bars in y represent growth rate measurements in triplicate. Suspected outliers are outlined in green. **c**) Growth rate compared to selection coefficient in DMSO for selected mutants. All mutants grew at similar rates in this condition. All selection coefficient axis limits are scaled to ease comparison between panels a), b) and c). Individual growth curves are shown in S6 Fig. Statistical test in a), b) and c) is Spearman's rank correlation. **d)** Spot-dilution assay for controls and some resistant mutants with their respective selection coefficients in IC90. ScDHFR and Empty are used as positive and negative controls, respectively.

To further validate our results and assess the presence of tradeoff between function and resistance, we reintroduced some of the reconstructed validation mutants into strain FDR0001 to assess the mutant protein's potential for functional complementation. In the absence of β-estradiol, all mutants identified as resistant could complement the deletion of *DFR1* in this strain, as expected. Interestingly, most mutants which were identified as being sensitive lost their ability to complement the deletion (5/7), implying that these mutations strongly negatively impacted protein function (S7A Fig). In

613 the presence of β-estradiol, all mutants grew similarly in the presence or absence of MTX
614 (S7B Fig). Growth experiments in strain FDR0001 exhibited higher levels of noise than
615 similar experiments in IGA130 (Fig 3A vs S7A Fig error bars).

616 **Premature stop codons reduce gene dosage toxicity**

617 In the absence of MTX, the distribution of selection coefficients showed little
618 variance around 0. As the endogenous *DFR1* was present in strain IGA130, the effect of
619 PjDHFR mutants on growth without MTX was expected to be minimal. However, we
620 observed an enrichment of positive selection coefficients for nonsense mutations (Fig
621 S4A, row marked with *, S8 Fig.) in this condition. Gene dosage toxicity has previously
622 been reported for DHFRs at expression levels higher than wild-type, suggesting that this
623 enrichment is caused by a fitness increase caused by the interruption of functional
624 PjDHFR production [67,69]. Because early stop codons are more likely to prevent the
625 formation of a functional protein than downstream ones, this effect is mainly observable
626 at the beginning of the sequence and is lost by the end of the protein, with selection
627 coefficients becoming slightly negative at around position 160 (S8 Fig). Other mutations,
628 especially before position 50, with similar fitness effects are observable across the
629 protein, which could hint at mutations that affect the function or folding of PjDHFR,
630 preventing gene dosage toxicity. However, the effect of most such mutations (mean
631 across all positive mutations in DMSO = 0.023, median = 0.016) is generally smaller than
632 the mean of stop codons before position 160 (0.060, median = 0.059), with only 0.26%
633 (11/4284) of mutations scoring significatively higher than this mean across the entire
634 protein. When using FoldX to predict the effect of mutations on the stability of PjDHFR,
635 we did not find a significant correlation between the effect of mutations and the selection
636 coefficient in DMSO (S9 Fig). Here, we expected highly destabilizing mutations to be
637 enriched in DMSO, as they would have potentially affected protein folding and thus
638 catalytic activity. The absence of clear positive correlation implies that a higher selection
639 coefficient in DMSO cannot be explained solely by changes in protein free energy.
640 Spearman's rank correlation shows a slight negative correlation (S9 Fig). While the
641 observed effects were small, the detection of the effect of stop codons highlighted the
642 sensitivity of this type of competition assay. Indeed, when manually validating mutants
643 from the competition assay using growth curves and spot-dilution assays, we did not
644 detect this phenotype, although the strain with the Empty plasmid had slightly higher
645 growth rate than any other wild-type DHFRs in this assay (in DMSO, Empty has a growth
646 rate of 0.308 $OD_{600} \cdot h^{-1}$, and ScDHFR of 0.289 $OD_{600} \cdot h^{-1}$, Fig 3C red and blue horizontal
647 lines, respectively). Most importantly, the experiments in the absence of MTX show that
648 there are no major dominant negative effects of amino acid substitutions that would
649 confound the measures of sensitivity to MTX.

650 **Resistance mutations most often appear at position within ligand binding**
651 **pocket**

652 Since there is no available crystal structure of PjDHFR, the predicted protein
653 structure was recovered from AlphaFoldDB (UniProt: A0EPZ9). Crystal structures of the
654 DHFR ortholog from *P. carinii* (61% identity, 75% positives), in complex with MTX/NADPH
655 (PDB: 3CD2) and DHF/NADPH (PDB: 4CD2), were recovered from the Protein Data Bank
656 (PDB) and aligned to generate the holoprotein structure of PjDHFR with MTX/NADPH

657 and DHF/NADPH. Despite sequence identity of only 61% between the orthologs, the
658 structures were highly conserved (root-mean-square deviation (RMSD) of 0.644 Å across
659 the protein backbone heavy atoms, 2.094 Å across all heavy atom pairs), allowing for
660 high-confidence structural alignments of ligands without the use of *ab initio* docking. We
661 then mapped the selection coefficients on the structure of PjDHFR (Fig 4A & 4B).

662 Most resistance positions are located within the PjDHFR substrate binding pocket,
663 or are in contact with NADPH, as we would expect for a competitive inhibitor such as
664 MTX. These sites overlap with residues that bind MTX. Not all mutations at these
665 positions lead to resistance. Indeed, resistance implies that the enzyme is no longer
666 inhibited by MTX but maintains proper interactions with the substrate so that it can still
667 perform its catalytic reaction that is required for growth. This leaves very few sites that
668 can be modified. At some of these key positions, such as M33 (4.8 Å from MTX),
669 resistance appears to also correlate with amino acid properties, where hydrophobic amino
670 acids, such as the wild-type methionine, have generally lower selection coefficients in
671 comparison to other polar residues, such as hydrophobic M33F (0.369 in IC90, p-value =
672 1.5e-8, from one-sided >0 Welch t-test), M33Y (0.243, p-value = 0.01) and M33V (0.441,
673 p-value = 1.1e-5) compared to charged M33R (1.287, p-value = 8.8e-25), M33K (1.088,
674 p-value = 4.7e-7) and short M33G (1.262, p-value = 1.3e-13). At other positions, fewer
675 amino acids had positive selection coefficients. At I142 (7.1 Å from MTX), only two amino
676 acid substitutions, I142W (1.428, p-value = 4.7e-4) and I142Y (0.668, p-value = 2.5e-3),
677 caused resistance, implying that other position-dependent properties, such as side chain
678 bulkiness, are at play. I142W, as well as a sensitive mutation at this same position, I142G
679 (-0.167, p-value = 0.97), were manually reconstructed as part of the validations by growth
680 curves and spot-dilution assays.

681 Another region where strong effects were observed was the glycine triplet at
682 positions 124 to 126, which is conserved across species (3/3 in humans, *S. cerevisiae*
683 and *E. coli*, 2/3 (positions 124 and 125) in mice). All three glycines from this triplet are in
684 contact either with both MTX/DHF and NADPH (124 and 125) or with NADPH only (126),
685 which could imply a role in protein activity and/or ligand binding. Interestingly, at IC75 and
686 IC90, mutants at position 124 with a positive selection coefficient generally have short
687 chain amino acids (G124S/A/P 0.865/0.736/0.244 in IC75, p-values = 8.5e-13/4.8e-
688 8/3.6e-7). Position 126 is less constrained, with 12/20 substitutions having a selection
689 coefficient above the resistance threshold. However, at IC75, position 125 was one of the
690 least permissive positions observed in the competition assay, with the lowest median
691 score by position (-0.735), and no substitution with a selection coefficient higher than 0,
692 implying that the glycine at this position is under strong selection in this protein.

693 Mechanisms of DHFR resistance to MTX have previously been investigated in
694 eukaryotes, mainly in baker's yeast, mice, and humans [67,70,71]. We examined how
695 conserved the mechanisms were for PjDHFR (Supplementary table Annotation Card).
696 Positions with mutations that are most often reported as conferring resistance to MTX,
697 namely M35 in yeast, F31, F34 and Q35 in mice and in humans, are structurally
698 equivalent to positions M33 (maximum selection coefficient at this position in IC90 =
699 1.485, p-value = 2.8e-11), F36 (maximum = 1.202, p-value = 2.1e-5) and S37 (maximum
700 = 0.760, p-value = 3.1e-8) in PjDHFR, all being <5 Å from MTX. In PjDHFR, some
701 mutations have been reported as conferring resistance to TMP in clinical isolates, with

position F36 being reported as conferring the highest resistance to this drug [72]. This reveals that resistance mechanisms are highly conserved among DHFRs. As antifolates are competitive inhibitors, we expect most mutations that confer resistance to come with a tradeoff in protein activity. In mice and in humans, biochemical properties were measured for some mutants at position F31, structurally equivalent to position M33 in PjDHFR. Resistance was associated with an increased K_m for DHF, which translates to a less active enzyme, up to four-fold for some mutants at this position, consistent with a tradeoff between resistance to MTX and activity [67,71,73].

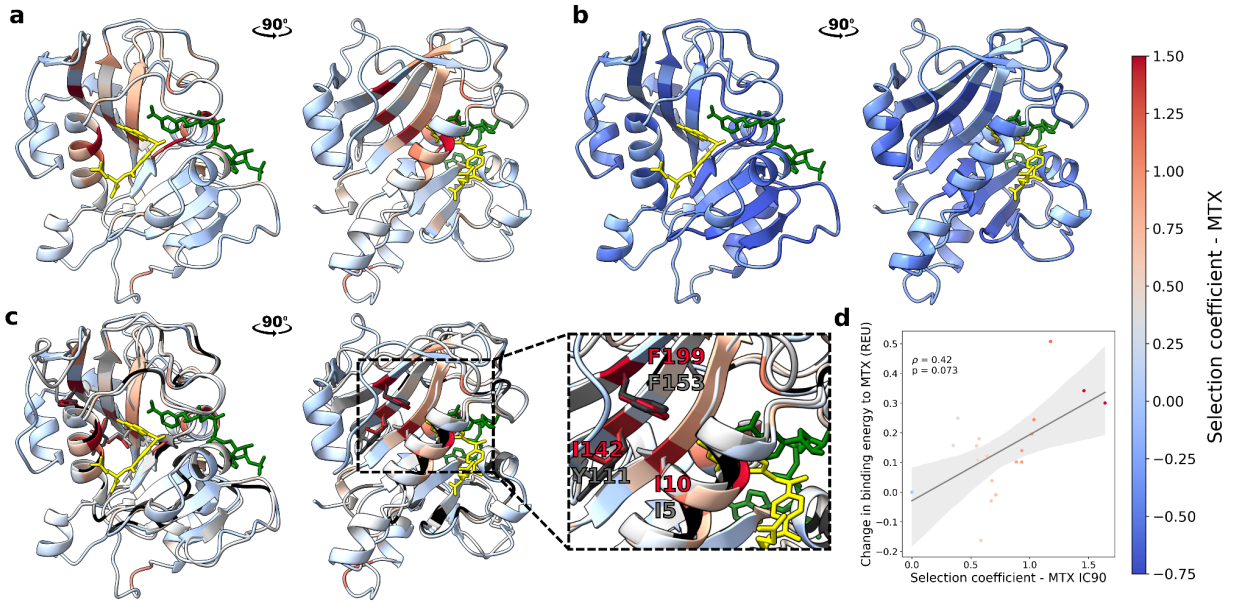


Fig 4. Selection coefficients mapped onto PjDHFR structure in complex with methotrexate and NADPH. a) Maximum selection coefficient from IC90 and **b)** minimum selection coefficient from IC75 observed for amino acid substitutions mapped on the PjDHFR structure by position. MTX (yellow) and NADPH (green) are visible. Nonsense mutations were not considered when identifying minimum selection coefficients. **c)** Structure superposition of PjDHFR (blue/red scale) and *E. coli* FolA (gray/black scale)**.** FolA is colored according to predicted allostery from [57], with darker shades of gray representing increased confidence in predicted allosteric score. PjDHFR is colored by maximum selection coefficient in IC90. Orientations of superimposed structures are the same as panel A. Positions 199 and 142 are predicted to have allosteric effects on the substrate and cofactor binding pocket [57]. I10 is not predicted to have an allosteric effect. Side chains for these positions, and corresponding in *E. coli.*, are shown and labeled. **d)** Rosetta FlexddG predicts that mutations at position 199 would destabilize interactions between PjDHFR and MTX, and this effect correlates with the measured selection coefficient. Spearman's rank coefficient is used to measure correlation. Y axis units are REU (Rosetta Energy Units).

Potential mechanism of resistance through allostery

By mapping the maximum selection coefficient values at IC90 (Fig 4A) and minimum values at IC75 (Fig 4B) by position on the structure, we can visualize structural determinants of resistance. Most positions where at least one mutant is highly resistant are proximal to MTX/DHF and NADPH, or close to other positions where resistance is observed (Fig 4A). Inversely, almost every position (202/204) has at least one mutant that performed worse than the wild-type PjDHFR at IC75, except for positions M33 and F199.

732 This is consistent with the expectation that most positions along a protein can have a
733 deleterious effect if mutated, especially in the presence of strong selection like MTX [74].

734 While some mutations were consistent with the literature and structural properties
735 of eukaryotic DHFRs, other resistant mutations were unexpected and may thus reveal
736 new mechanisms of resistance. F199, a position located in a distal β-strand, and which
737 has no contacts with drug, substrate, or NADPH, showed high levels of resistance to MTX
738 for many different substitutions, and consistently so across both conditions. The
739 maximum selection coefficients are 1.11 at IC75 and 1.64 at IC90, the second and first
740 highest in their respective assays. Two mutants from this position, F199E (0.585 at IC90)
741 and F199P (1.018 at IC90), were manually reconstructed by site-directed mutagenesis,
742 and were validated for resistance to MTX (S6 Fig). While this position had not been
743 reported in eukaryotic resistance studies, a structurally equivalent position (F153) had
744 been predicted in *E. coli* as being part of the allosteric "sector" of FolA (PDB: 1RX2),
745 which could hint at a yet unreported mechanism of resistance to MTX [57]. One mutation
746 at this position, F153S, has also been identified in *E. coli* as slightly increasing resistance
747 to TMP [68]. Interestingly, positions I10 (I5 in *E. coli*, I11 in baker's yeast, I7 in mouse and
748 humans) and I142 (Y111 in *E. coli*, L139 in yeast, F134 in mouse and human) are two
749 other positions where high levels of resistance were observed. Both positions are part of
750 the same β-sheet and physically close to F199, forming a "line/groove" towards the α-
751 helix containing M33, another position where several mutants had increased resistance
752 to MTX, and which is in direct contact with MTX and DHF. Y111 in FolA, the structurally
753 equivalent position to I142, has also been identified as being part of the allosteric "sector"
754 of FolA [57]. All three positions, I10, I142 and F199, are relatively buried, and are
755 hydrophobic amino acids, a property that is conserved across *E. coli*, baker's yeast,
756 mouse, and humans. In the case of F199, amino acids with similar properties have lower
757 selection coefficients in MTX, and polar/longer amino acids have higher selection
758 coefficients, with F199R and F199K being respectively the first and fourth mutants with
759 the highest selection coefficients at IC90. Position 199 also displays unique properties in
760 DMSO, being the only position where a mutation, F199P (0.102, p-value = 0.22), is
761 enriched, but the stop codon is not, implying that F199P is potentially more disruptive for
762 protein function or folding than a nonsense mutation at this position. However, this is
763 based on a single datapoint with three replicates and could therefore be due to
764 experimental noise. Overall, this evidence points towards a more important role of this
765 position in DHFR function and resistance than previously estimated.

766 To investigate the specific effects of mutations at position 199, we conducted an
767 *in silico* mutagenesis at this position using Rosetta's FlexddG method and predicted
768 changes in binding energy between PjDHFR, MTX and DHF (Fig 4D) [59]. The changes
769 in binding energy predicted by Rosetta FlexddG between PjDHFR and MTX show a slight
770 positive correlation (ρ = 0.42, p-value = 0.07) with the selection coefficients measured at
771 IC90 for each corresponding mutant, implying that mutations that destabilize interactions
772 with MTX are more likely to be highly resistant (Fig 4D). Interestingly, FlexddG predicted
773 that mutations at position 199 would generally be more destabilizing between PjDHFR
774 and MTX than between PjDHFR and DHF, which would translate to a more resistant but
775 still functional enzyme (S10 Fig). Here, most points fall above the diagonal, implying that
776 most mutations are more destabilizing between PjDHFR and MTX than between PjDHFR

777 and DHF. Most mutations that are more destabilizing correlate with measured selection
778 coefficient. In *E. coli*, biochemical properties of mutant F153S were investigated. It was
779 found to have no significant effect on the maximum rate of catalysis (k_{cat}) of FolA, but did
780 increase K_m to TMP, which is consistent with the decrease in binding affinity modeled by
781 Rosetta [68]. This evidence points towards position 199 having distal effect on binding
782 affinity, but not necessarily on catalytic rate, suggesting that mutants at this position could
783 lead to resistance without strongly compromising enzyme function, making this position
784 a good candidate for the evolution of resistance mutations.

785 **Discussion**

786 **Functional complementation and DMS are powerful tools to investigate hard-**
787 **to-study organisms**

788 The study of *Pneumocystis jirovecii* has always been difficult because of our
789 current inability to cultivate it *in vitro* or in animal models. As resistance is increasing in
790 this fungal pathogen, there is a pressing need to understand the mechanisms behind its
791 emergence. By employing functional complementation in *S. cerevisiae* and a DMS library
792 of PjDHFR, we investigated resistance to an antifolate using an exhaustive library of
793 mutants.

794 We demonstrate that under the regulation of a constitutive promoter, PjDHFR is
795 sensitive to MTX in a quantitative manner, even in the presence of wild-type genomic
796 *DFR1*. From this data, we targeted two selection pressures to conduct DMS followed by
797 a competition assay to assess the effect of mutations compared to the wild-type PjDHFR.
798 This allowed us to identify a broad range of fitness effects, both positive and negative, in
799 the presence of MTX. We also observed mutations that affect protein function and gene
800 dosage toxicity. Results from the competition assay are consistent with known resistance
801 mutations to MTX in eukaryotes, implying that some resistance mechanisms are
802 conserved across distant lineages.

803 By using the DMS library from PjDHFR to conduct competition assays in MTX at
804 two different concentrations, we were able to measure different ranges of effect on
805 resistance. Using higher selection pressure, such as IC90, allowed mutations that caused
806 higher-resistance mutations to be more enriched in regard to wild-type-like mutations and
807 showed a broader range of effect for highly resistant mutations. At a weaker concentration
808 (IC75), mutations that performed worse than wild-type were more easily identifiable than
809 at IC90, which is indicative of either mutants less resistant than the wild-type, or non-
810 functional enzymes. Finally, by conducting the experiment in the absence of MTX
811 selection, we were able to assess the absence of dominant negative effect, as well as to
812 identify the presence of gene dosage toxicity through the effects of nonsense mutations.
813 Combining the results of both datasets in the presence of MTX also helped identify
814 mutants with inconsistent effects and guided the subsequent validation work.

815 **Resistance to methotrexate occurs through different mechanisms**

816 Most mutations that confer resistance to MTX were within the substrate or cofactor
817 binding pocket, and most often, several different amino acid substitutions allowed for
818 resistance at a given position. In the case of M33, shorter chain, hydrophobic amino acids

(wild-type-like) were the substitutions that caused the lowest levels of resistance, implying that changes in local amino acid properties are key for resistance, instead of the specific amino acid present at this position. While M33 is in direct contact with MTX, F199 is not, but both displayed this pattern of “property-based” resistance. Other positions where resistance was observed, such as I10, displayed no such clear pattern. This implies that several mutational pathways are available for evolution towards resistance to MTX, and possibly other antifolates. This raises major concern about the future of resistance to antifolates in *P. jirovecii*.

Structural analysis and binding affinity modeling of the observed fitness effects allowed us to identify putatively allosteric mechanisms that cause resistance to MTX, which is supported by a previous study into the allosteric regulation of DHFR. These results hint at more complex internal protein dynamics for function and resistance within DHFRs than previously estimated [57,68,75].

Conclusion

Studying *P. jirovecii*’s DHFR and its evolution towards resistance to antifolates is of particular interest because of its insensitivity to classical antifungals. Here, we developed resources that will help study how resistance to antifolates evolves in this fungus. With the rising concerns about resistance in this pathogen, such resources will be useful to detect or predict resistance from DNA sequence alone, as well as potentially developing species-specific antifolates in the future. As other pathogenic fungi have been developing resistance to classical antifungals, it is more relevant than ever to study alternative treatment targets in fungal pathogens, such as antifolates [19].

Data availability

Raw read sequencing from competition assay is available at BioProject accession: PRJNA1013557

Demultiplexed and analyzed data with all computed scores is available in codon form in Supplementary_data_1.csv for DMSO, Supplementary_data_2.csv for IC75 and Supplementary_data_3.csv for IC90. All scripts for figures and analysis, as well as processed data, is available on GitHub at https://github.com/Landrylab/Rouleau_et_al_2023.git. Detailed files and structures for FlexddG analysis are available upon request.

Acknowledgements

We would like to thank the Landry lab for their helpful comments throughout the project. We also thank Joëlle Pelletier for her comments and insights. The project was funded by a Canadian Institutes of Health Research Foundation grant number 387697 to CRL, a FRQNT Team Grant (**FRQNT (2022-PR-298169)** and a Genome Canada and Genome Quebec grant (6569 to CRL). CRL holds the Canada Research Chair in Cellular Systems and Synthetic Biology. FDR was supported by fellowships from FRQNT, PROTEO, EvoFunPath NSERC CREATE program and the Vanier Canada Graduate Scholarship agency. SD was supported by a fellowship from FRQNT. PCD was supported by fellowships from the FRQS and from the Vanier Canada Graduate Scholarship.

References

1. Sanger F. Species Differences in Insulins. Nature. 1949;164: 529–529.

2. Futterman S. Enzymatic reduction of folic acid and dihydrofolic acid to tetrahydrofolic acid. J Biol Chem. 1957;228: 1031–1038.

3. Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM, Kroon G, et al. Divergent evolution of protein conformational dynamics in dihydrofolate reductase. Nat Struct Mol Biol. 2013;20: 1243–1249.

4. Bystroff C, Oatley SJ, Kraut J. Crystal structures of Escherichia coli dihydrofolate reductase: the NADP+ holoenzyme and the folate.NADP+ ternary complex. Substrate binding and a model for the transition state. Biochemistry. 1990;29: 3263–3277.

5. Champness JN, Achari A, Ballantine SP, Bryant PK, Delves CJ, Stammers DK. The structure of Pneumocystis carinii dihydrofolate reductase to 1.9 A resolution. Structure. 1994;2: 915–924.

6. Cody V, Pace J, Rosowsky A. Structural analysis of a holoenzyme complex of mouse dihydrofolate reductase with NADPH and a ternary complex with the potent and selective inhibitor 2,4-diamino-6-(2'-hydroxydibenz[b,f]azepin-5-yl)methylpteridine. Acta Crystallogr D Biol Crystallogr. 2008;64: 977–984.

7. Hay S, Evans RM, Levy C, Loveridge EJ, Wang X, Leys D, et al. Are the catalytic properties of enzymes from piezophilic organisms pressure adapted? Chembiochem. 2009;10: 2348–2353.

8. Krahn JM, Jackson MR, DeRose EF, Howell EE, London RE. Crystal structure of a type II dihydrofolate reductase catalytic ternary complex. Biochemistry. 2007;46: 14878–14888.

9. Lozovsky ER, Chookajorn T, Brown KM, Imwong M, Shaw PJ, Kamchonwongpaisan S, et al. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. Proc Natl Acad Sci U S A. 2009;106: 12025–12030.

10. Tarnchompoo B, Chitnumsub P, Jaruwat A, Shaw PJ, Vanichtanankul J, Poen S, et al. Hybrid Inhibitors of Malarial Dihydrofolate Reductase with Dual Binding Modes That Can Forestall Resistance. ACS Med Chem Lett. 2018;9: 1235–1240.

11. Askari BS, Krajinovic M. Dihydrofolate reductase gene variations in susceptibility to disease and treatment outcomes. Curr Genomics. 2010;11: 578–583.

12. Bilsland E, Pir P, Gutteridge A, Johns A, King RD, Oliver SG. Functional expression of parasite drug targets and their human orthologs in yeast. PLoS Negl Trop Dis. 2011;5: e1320.

899 13. Schweitzer BI, Dicker AP, Bertino JR. Dihydrofolate reductase as a therapeutic
900 target. FASEB J. 1990;4: 2441–2452.

901 14. Visentin M, Zhao R, Goldman ID. The Antifolates. Hematol Oncol Clin North Am.
902 2012;26: 629–629.

903 15. Gangjee A, Jain HD, Kurup S. Recent advances in classical and non-classical
904 antifolates as antitumor and antiopportunistic infection agents: part I. Anticancer
905 Agents Med Chem. 2007;7: 524–542.

906 16. Kemnic TR, Coleman M. Trimethoprim Sulfamethoxazole. StatPearls Publishing;
907 2022.

908 17. Brolund A, Sundqvist M, Kahlmeter G, Grape M. Molecular characterisation of
909 trimethoprim resistance in Escherichia coli and Klebsiella pneumoniae during a two
910 year intervention on trimethoprim use. PLoS One. 2010;5.
911 doi:10.1371/journal.pone.0009233

912 18. Ravikumar A, Arzumanyan GA, Obadi KA, Javanpour AA, Liu Correspondence CC,
913 Obadi MKA, et al. Scalable, Continuous Evolution of Genes at Mutation Rates above
914 Genomic Error Thresholds Resource Scalable, Continuous Evolution of Genes at
915 Mutation Rates above Genomic Error Thresholds. Cell. 2018;175: 1946–1957.e13.

916 19. DeJarnette C, Luna-Tapia A, Estredge LR, Palmer GE. Dihydrofolate Reductase Is
917 a Valid Target for Antifungal Development in the Human Pathogen Candida albicans.
918 mSphere. 2020;5. doi:10.1128/msphere.00374-20

919 20. Edman JC, Kovacs JA, Masur H, Santi DV, Elwood HJ, Sogin ML. Ribosomal RNA
920 sequence shows Pneumocystis carinii to be a member of the fungi. Nature.
921 1988;334: 519–522.

922 21. Van der Meer G, Brug SL. Infection a Pneumocystis chez l'homme et chez les
923 animaux. 1942.

924 22. Chagas C. Neue trypanosomen: vorläufige mitteilung. 1908. Available:
925 https://www.arca.fiocruz.br/handle/icict/53339

926 23. Frenkel JK. Pneumocystis jiroveci n. sp. from man: morphology, physiology, and
927 immunology in relation to pathology. Natl Cancer Inst Monogr. 1976;43: 13–30.

928 24. Hugot J-P, Demanche C, Barriel V, Dei-Cas E, Guillot J. Phylogenetic Systematics
929 and Evolution of Primate-Derived Pneumocystis Based on Mitochondrial or Nuclear
930 DNA Sequence Comparison. Syst Biol. 2003;52: 735–744.

931 25. Liu Y, Su L, Jiang SJ, Qu H. Risk factors for mortality from pneumocystis carinii
932 pneumonia (PCP) in non-HIV patients: a meta-analysis. Oncotarget. 2017;8: 59729–
933 59729.

934 26. Morris A, Norris KA. Colonization by pneumocystis jirovecii and its role in disease.
935 Clinical Microbiology Reviews. 2012. pp. 297–317. doi:10.1128/CMR.00013-12

936 27. Stern A, Green H, Paul M, Vidal L, Leibovici L. Prophylaxis for Pneumocystis
937 pneumonia (PCP) in non-HIV immunocompromised patients. Cochrane Database of
938 Systematic Reviews. 2014. doi:10.1002/14651858.CD005590.pub3

939 28. Vera C, Rueda ZV. Transmission and Colonization of Pneumocystis jirovecii. Journal
940 of Fungi 2021, Vol 7, Page 979. 2021;7: 979–979.

941 29. Nasr M, Mohammad A, Hor M, Baradeiya AM, Qasim H. Exploring the Differences in
942 Pneumocystis Pneumonia Infection Between HIV and Non-HIV Patients. Cureus.
943 2022;14: e27727.

944 30. Lee SM, Cho YK, Sung YM, Chung DH, Jeong SH, Park J-W, et al. A Case of
945 Pneumonia Caused by Pneumocystis jirovecii Resistant to Trimethoprim-
946 Sulfamethoxazole. Korean J Parasitol. 2015;53: 321–327.

947 31. Roembke F, Heinzow HS, Gosseling T, Heinecke A, Domagk D, Domschke W, et al.
948 Clinical outcome and predictors of survival in patients with pneumocystis jirovecii
949 pneumonia--results of a tertiary referral centre. Clin Respir J. 2014;8: 86–92.

950 32. Medrano FJ, Montes-Cano M, Conde M, de la Horra C, Respaldiza N, Gasch A, et
951 al. Pneumocystis jirovecii in general population. Emerg Infect Dis. 2005;11: 245–250.

952 33. Kolbrink B, Scheikholeslami-Sabzewari J, Borzikowsky C, von Samson-
953 Himmelstjerna FA, Ullmann AJ, Kunzendorf U, et al. Evolving epidemiology of
954 pneumocystis pneumonia: Findings from a longitudinal population-based study and
955 a retrospective multi-center study in Germany. Lancet Reg Health Eur. 2022;18:
956 100400.

957 34. Morris A, Wei K, Afshar K, Huang L. Epidemiology and clinical significance of
958 pneumocystis colonization. J Infect Dis. 2008;197: 10–17.

959 35. Ponce CA, Gallo M, Bustamante R, Vargas SL. Pneumocystis colonization is highly
960 prevalent in the autopsied lungs of the general population. Clin Infect Dis. 2010;50:
961 347–353.

962 36. Porollo A, Meller J, Joshi Y, Jaiswal V, George Smulian A, T. Cushion M. Analysis of
963 Current Antifungal Agents and Their Targets within the Pneumocystis carinii
964 Genome. Curr Drug Targets. 2012;13: 1575–1585.

965 37. Tomblyn M, Chiller T, Einsele H, Gress R, Sepkowitz K, Storek J, et al. Guidelines
966 for preventing infectious complications among hematopoietic cell transplantation
967 recipients: a global perspective. Biol Blood Marrow Transplant. 2009;15: 1143–1238.

968 38. Leidner F, Yilmaz K, Schiffer CA. Deciphering Antifungal Drug Resistance in
969 Pneumocystis jirovecii DHFR with Molecular Dynamics and Machine Learning. Cite

970 This: J Chem Inf Model. 2021;61. doi:10.1021/acs.jcim.1c00403

971 39. Singh Y, Mirdha BR, Guleria R, Kabra SK, Mohan A, Chaudhry R, et al. Novel
972 dihydropteroate synthase gene mutation in Pneumocystis jirovecii among HIV-
973 infected patients in India: Putative association with drug resistance and mortality.
974 Journal of Global Antimicrobial Resistance. 2019;17: 236–239.

975 40. White PL, Price JS, Backx M. Therapy and Management of Pneumocystis jirovecii
976 Infection. Journal of Fungi. 2018;4: 127–127.

977 41. Zhu M, Ye N, Xu J. Clinical characteristics and prevalence of dihydropteroate
978 synthase gene mutations in Pneumocystis jirovecii-infected AIDS patients from low
979 endemic areas of China. PLoS One. 2020;15: e0238184–e0238184.

980 42. Dhaneshwar SS. Colon-specific prodrugs of 4-aminosalicylic acid for inflammatory
981 bowel disease. World J Gastroenterol. 2014;20: 3564–3571.

982 43. Kannan G, Vasantha J, Rani N, Thennarasu P, Kousalya K, Anuradha P, et al. Drug
983 usage evaluation of dapsone. Indian J Pharm Sci. 2009;71: 456–460.

984 44. Masters PA, O'Bryan TA, Zurlo J, Miller DQ, Joshi N. Trimethoprim-sulfamethoxazole
985 revisited. Archives of Internal Medicine. American Medical Association; 2003. pp.
986 402–410. doi:10.1001/archinte.163.4.402

987 45. Barclay BJ, Nagel MG, Huang T. Dihydrofolate Reductase is not the Target of
988 Trimethoprim in Saccharomyces Cerevisiae. In: Ayling JE, Nair MG, Baugh CM,
989 editors. Chemistry and Biology of Pteridines and Folates. Boston, MA: Springer US;
990 1993. pp. 551–554.

991 46. Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, et al. Designer deletion
992 strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and
993 plasmids for PCR-mediated gene disruption and other applications. Yeast. 1998;14:
994 115–132.

995 47. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier
996 DNA/PEG method. Nat Protoc. 2007;2: 31–34.

997 48. Dubé AK, Dandage R, Dibyachintan S, Dionne U, Després PC, Landry CR. Deep
998 Mutational Scanning of Protein–Protein Interactions Between Partners Expressed
999 from Their Endogenous Loci In Vivo. In: Devaux F, editor. Yeast Functional
1000 Genomics: Methods and Protocols. New York, NY: Springer US; 2022. pp. 237–259.

1001 49. Li A, Acevedo-Rocha CG, Reetz MT. Boosting the efficiency of site-saturation
1002 mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy. Appl
1003 Microbiol Biotechnol. 2018;102: 6095–6103.

1004 50. Yachie N, Petsalaki E, Mellor JC, Weile J, Jacob Y, Verby M, et al. Pooled-matrix
1005 protein interaction screens using Barcode Fusion Genetics. Mol Syst Biol. 2016;12:

1006 863–863.

1007 51. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq:
1008 paired-end assembler for illumina sequences. BMC Bioinformatics. 2012;13: 31.

1009 52. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open
1010 source tool for metagenomics. PeerJ. 2016;4: e2584.

1011 53. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-
1012 learn: Machine Learning in Python. J Mach Learn Res. 2011;12: 2825–2830.

1013 54. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al.
1014 SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods.
1015 2020;17: 261–272.

1016 55. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with
1017 python. Proceedings of the 9th Python in Science Conference. SciPy; 2010.
1018 doi:10.25080/majora-92bf1922-011

1019 56. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF
1020 ChimeraX: Structure visualization for researchers, educators, and developers.
1021 Protein Sci. 2021;30: 70–82.

1022 57. Reynolds KA, McLaughlin RN, Ranganathan R. Hot spots for allosteric regulation on
1023 protein surfaces. Cell. 2011;147: 1564–1575.

1024 58. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython:
1025 freely available Python tools for computational molecular biology and bioinformatics.
1026 Bioinformatics. 2009;25: 1422–1423.

1027 59. Barlow KA, Ó Conchúir S, Thompson S, Suresh P, Lucas JE, Heinonen M, et al. Flex
1028 ddG: Rosetta Ensemble-Based Estimation of Changes in Protein–Protein Binding
1029 Affinity upon Mutation. J Phys Chem B. 2018;122: 5389–5399.

1030 60. Charlotte Eccleston R, Manko E, Campino S, Clark TG, Furnham N. A computational
1031 method for predicting the most likely evolutionary trajectories in the step-wise
1032 accumulation of resistance mutations. bioRxiv. 2022. p. 2022.01.25.477595.
1033 doi:10.1101/2022.01.25.477595

1034 61. Schnell JR, Dyson HJ, Wright PE. Structure, dynamics, and catalytic function of
1035 dihydrofolate reductase. Annu Rev Biophys Biomol Struct. 2004;33: 119–140.

1036 62. Huang T, Barclay BJ, Kalman TI, von Borstel RC, Hastings PJ. The phenotype of a
1037 dihydrofolate reductase mutant of Saccharomyces cerevisiae. Gene. 1992;121: 167–
1038 171.

1039 63. Ma L, Jia Q, Kovacs JA. Development of a yeast assay for rapid screening of
1040 inhibitors of human-derived Pneumocystis carinii dihydrofolate reductase. Antimicrob

1041 Agents Chemother. 2002;46: 3101–3103.

1042 64. Pattishall KH, Acar J, Burchall JJ, Goldstein FW, Harvey RJ. Two distinct types of
1043 trimethoprim-resistant dihydrofolate reductase specified by R-plasmids of different
1044 compatibility groups. J Biol Chem. 1977;252: 2319–2323.

1045 65. Aranda-Díaz A, Mace K, Zuleta I, Harrigan P, El-Samad H. Robust Synthetic Circuits
1046 for Two-Dimensional Control of Gene Expression in Yeast. ACS Synth Biol. 2017;6:
1047 545–554.

1048 66. Tarassov K, Messier V, Landry CR, Radinovic S, Serna Molina MM, Shames I, et al.
1049 An in vivo map of the yeast protein interactome. Science. 2008;320: 1465–1470.

1050 67. Pelletier JN, Campbell-Valois FX, Michnick SW. Oligomerization domain-directed
1051 reassembly of active dihydrofolate reductase from rationally designed fragments.
1052 Proc Natl Acad Sci U S A. 1998;95: 12141–12146.

1053 68. Tamer YT, Gaszek IK, Abdizadeh H, Batur TA, Reynolds KA, Atilgan AR, et al. High-
1054 Order Epistasis in Catalytic Power of Dihydrofolate Reductase Gives Rise to a
1055 Rugged Fitness Landscape in the Presence of Trimethoprim Selection. Mol Biol Evol.
1056 2019;36: 1533–1550.

1057 69. Bhattacharyya S, Bershtein S, Yan J, Argun T, Gilson AI, Trauger SA, et al. Transient
1058 protein-protein interactions perturb E. coli metabolome and cause gene dosage
1059 toxicity. Elife. 2016;5. doi:10.7554/eLife.20309

1060 70. Wong LH, Sinha S, Bergeron JR, Mellor JC, Giaever G, Flaherty P, et al. Reverse
1061 Chemical Genetics: Comprehensive Fitness Profiling Reveals the Spectrum of Drug
1062 Target Interactions. PLoS Genet. 2016;12: e1006275.

1063 71. Volpato JP, Fossati E, Pelletier JN. Increasing methotrexate resistance by
1064 combination of active-site mutations in human dihydrofolate reductase. J Mol Biol.
1065 2007;373: 599–611.

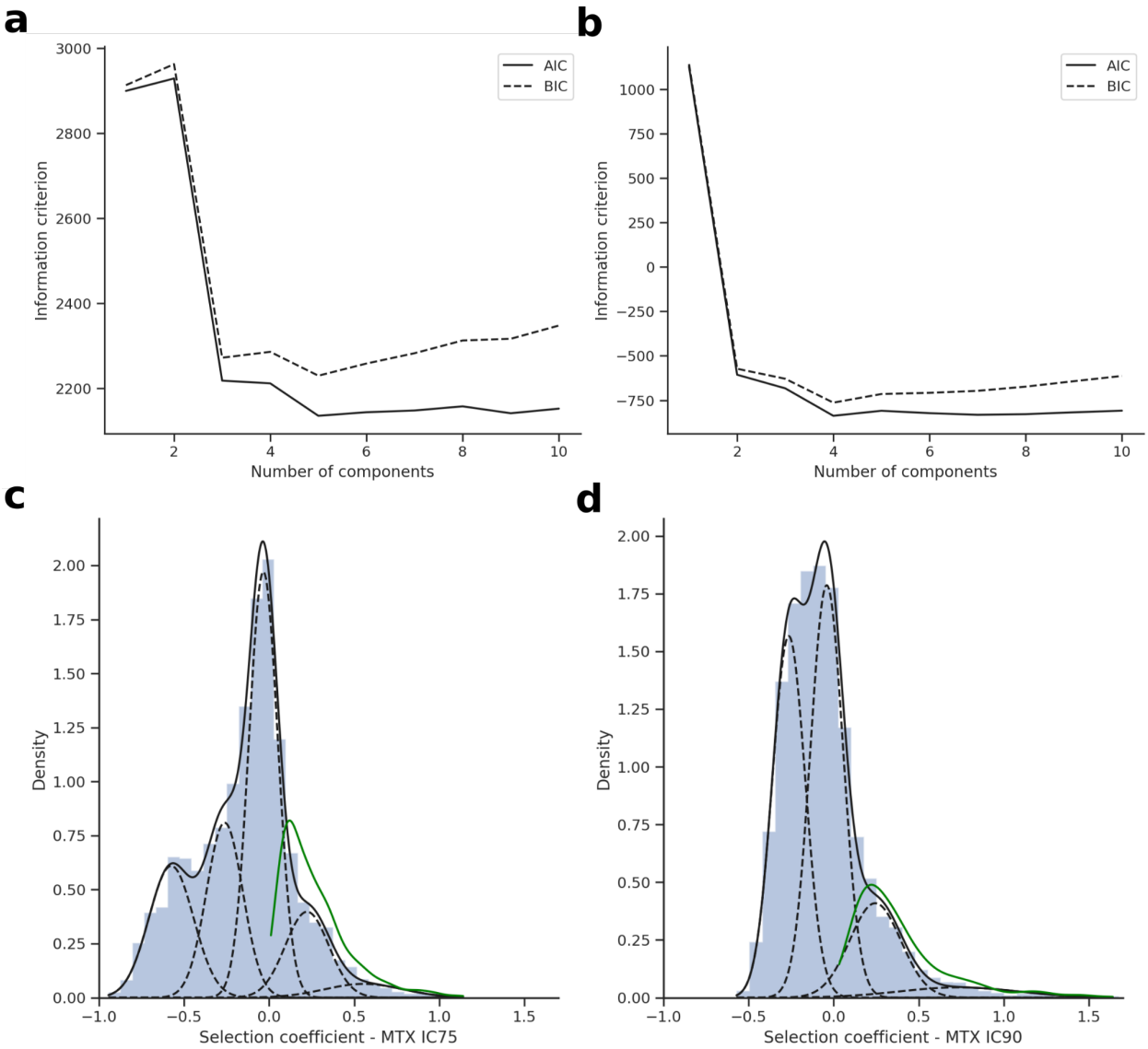
1066 72. Queener SF, Cody V, Pace J, Torkelson P, Gangjee A. Trimethoprim resistance of
1067 dihydrofolate reductase variants from clinical isolates of Pneumocystis jirovecii.
1068 Antimicrob Agents Chemother. 2013;57: 4990–4998.

1069 73. Thillet J, Absil J, Stone SR, Pictet R. Site-directed mutagenesis of mouse
1070 dihydrofolate reductase. Mutants with increased resistance to methotrexate and
1071 trimethoprim. J Biol Chem. 1988;263: 12500–12508.

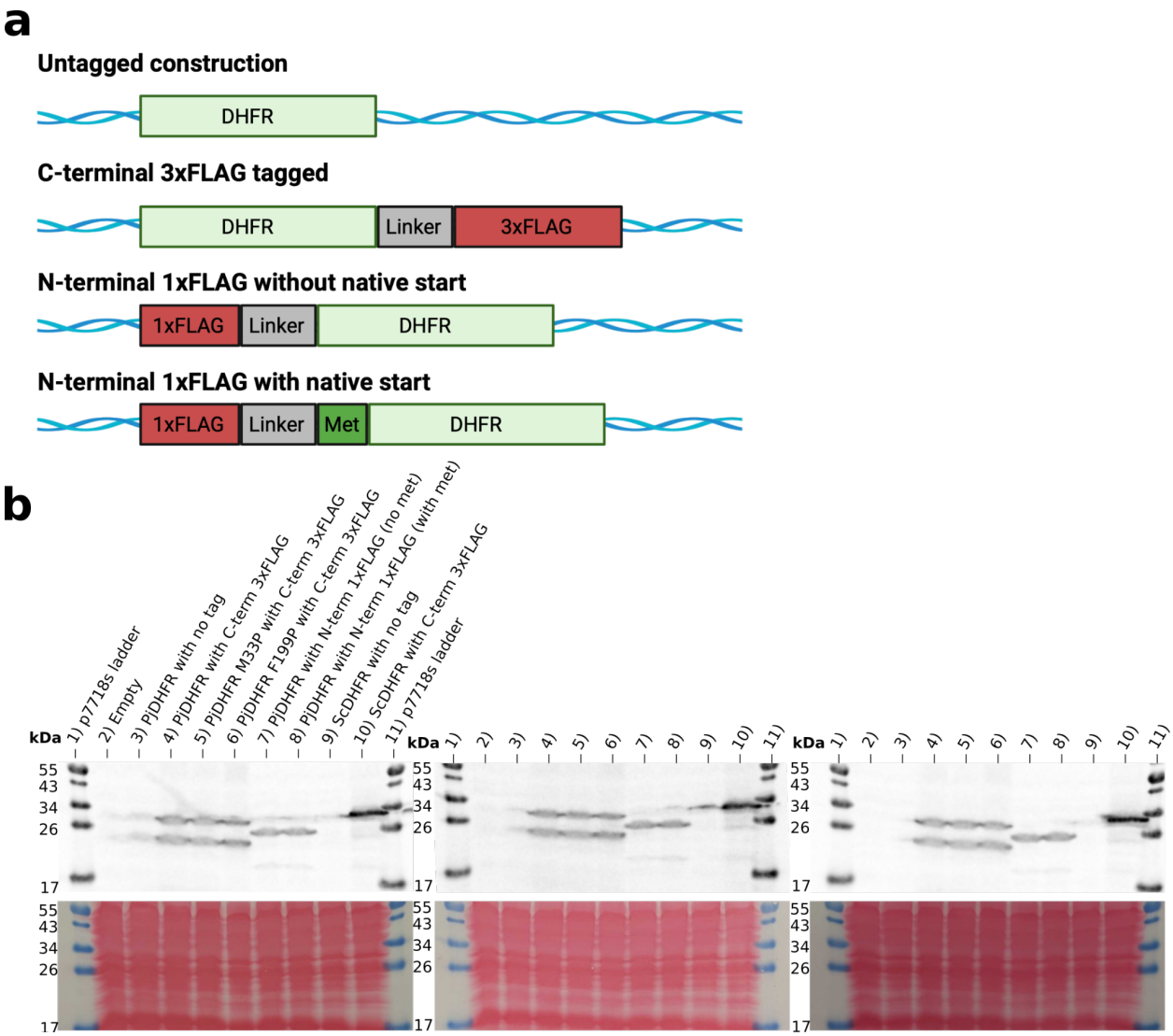
1072 74. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat
1073 Rev Genet. 2007;8: 610–618.

1074 75. McCormick JW, Russo MA, Thompson S, Blevins A, Reynolds KA. Structurally
1075 distributed surface sites tune allosteric regulation. Elife. 2021;10.
1076 doi:10.7554/eLife.68346

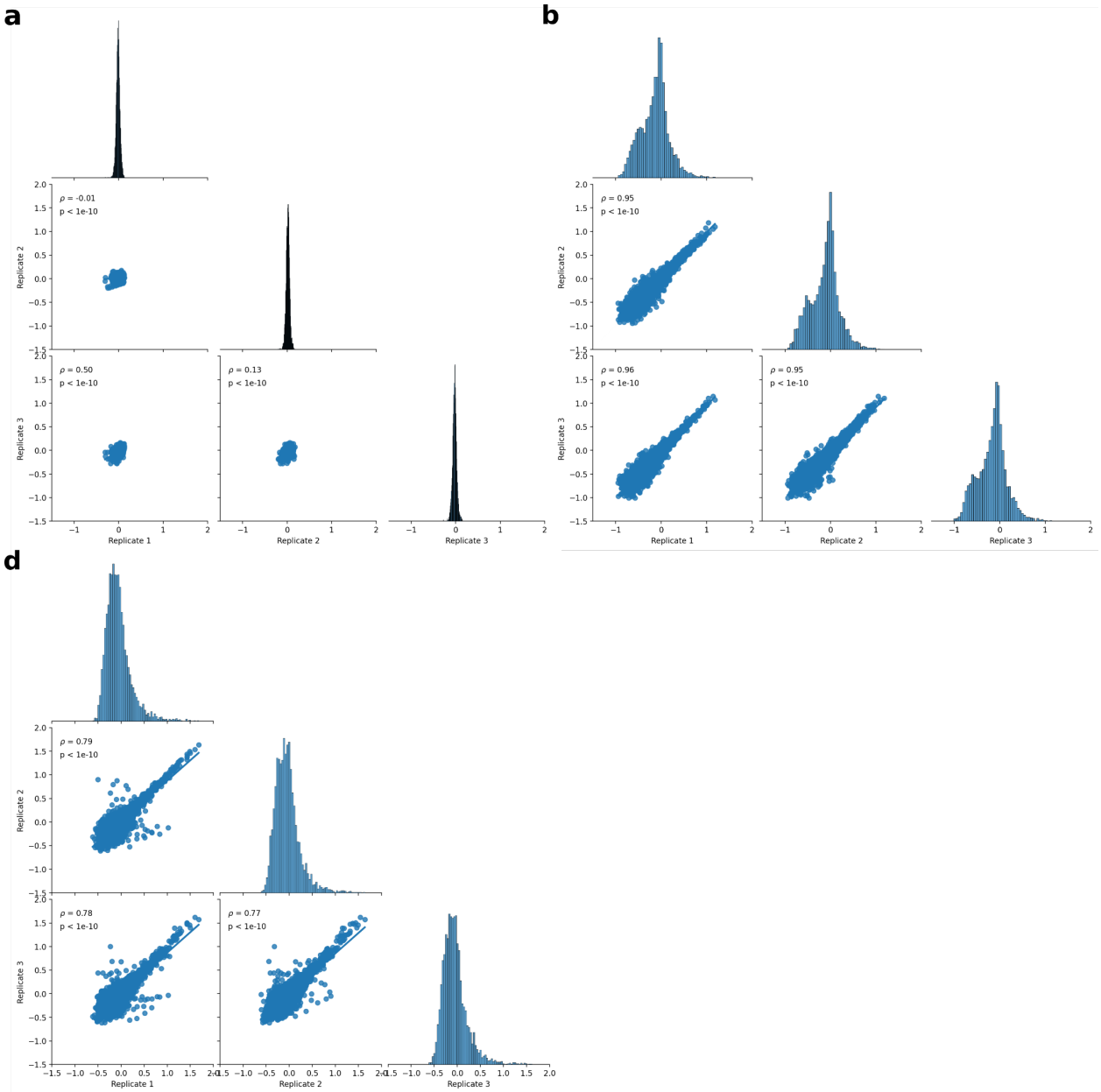
1077



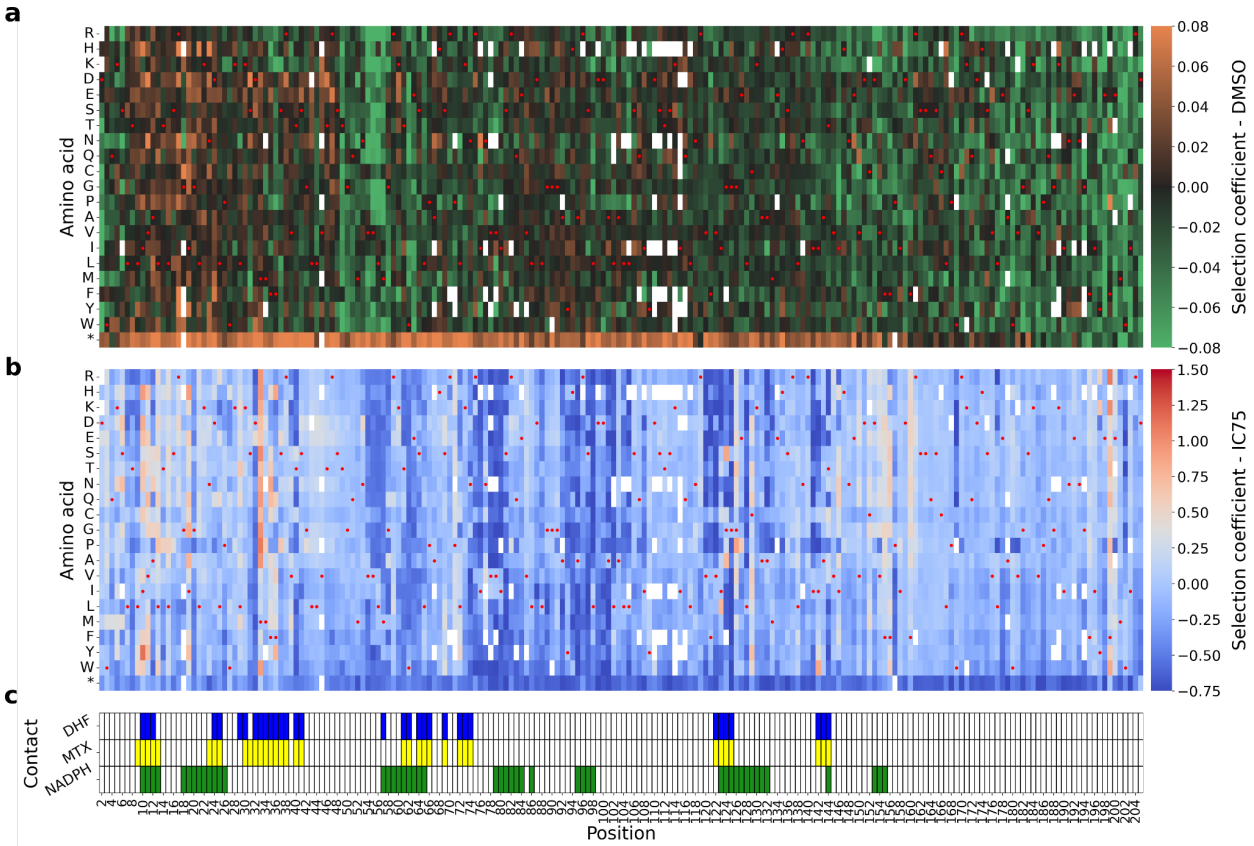
S1 Fig.: Gaussian mixture model optimization for MTX IC75 and IC90 conditions. Optimization of information criterion for **a)** IC75and **b)** IC90. Best Gaussian mixture model (dashed lines represent underlying Gaussians) to recapitulate the overlying distribution (black line/histogram) for **c)** IC75 and **d)** IC90. Density curve of significant mutants for Benjamini-Hochberg (green) correction is visible.



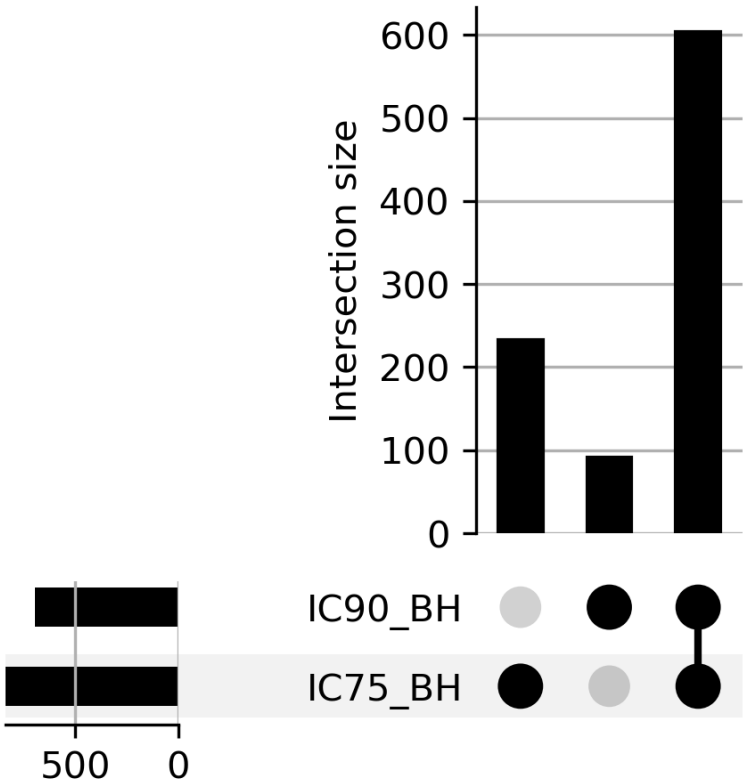
S2 Fig.: PjDHFR protein expression analyses in *Saccharomyces cerevisiae* strain IGA130. a) Construction schematics of tagged proteins expressed in yeast that can further be detected in protein cell extracts by western blots. All proteins tagged at their C-termini are tagged using 3xFLAG-tag, and all proteins tagged at their N-terminal are tagged using 1xFLAG-tag. N-terminal-tagged proteins were constructed with or without their native start codon. Constructions are not to scale. Schematic was made with BioRender. **b)** Western blots from biological replicates for all constructions. In C-terminal-tagged constructions with PjDHFR (wells 4-5-6), two bands can be seen, corresponding to the full length (top band) and the truncated protein from an alternative start codon (possibly M33 or M34, lower band). This lower band is absent from the wells with the N-terminal constructions (wells 7-8), supporting the hypothesis of an alternative stop codon versus a post-translational cleavage.



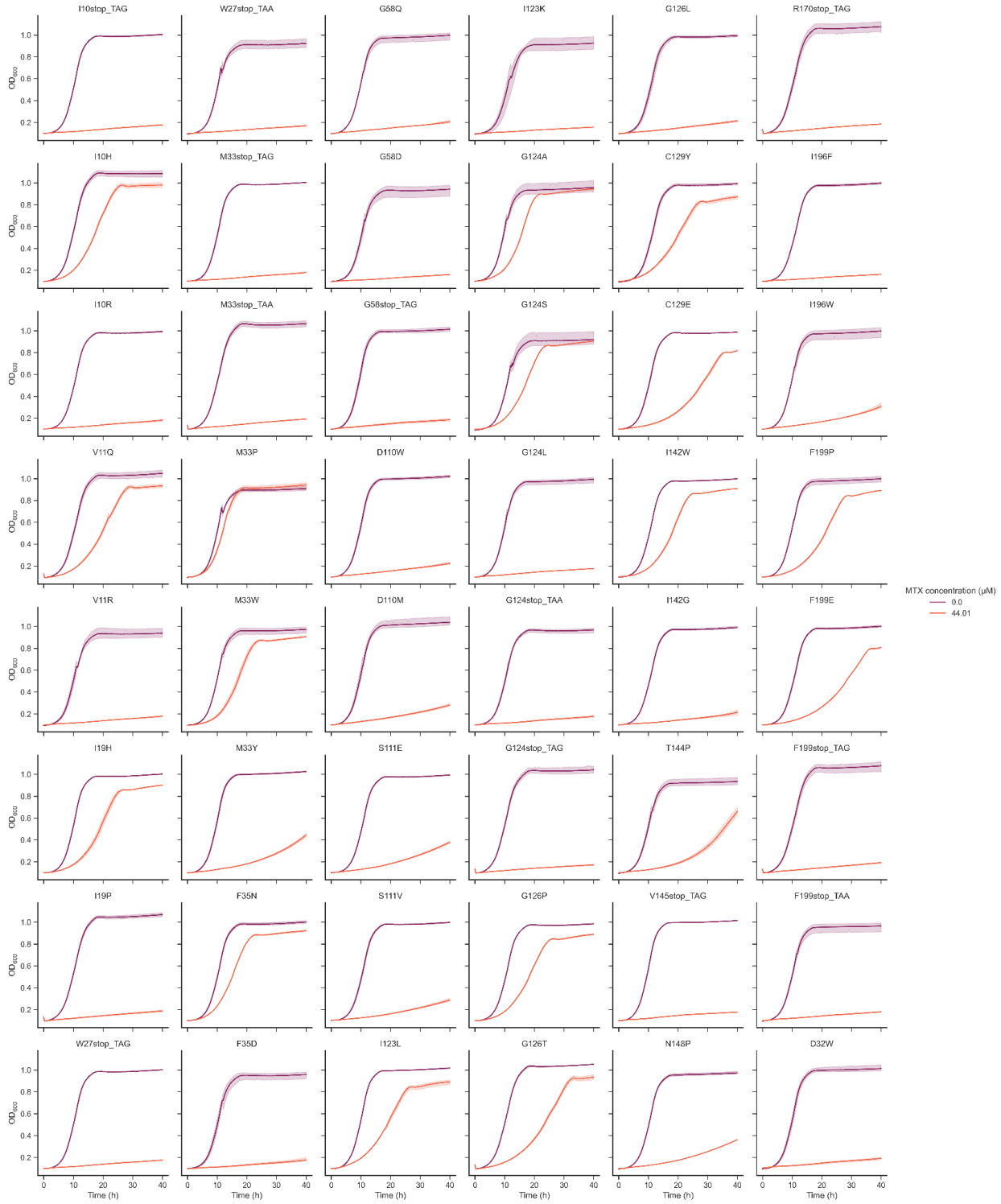
S3 Fig.: Correlation among biological replicates of competition assays in control and MTX conditions. a) Distributions of selection coefficients for amino acids in the different replicates in DMSO condition. Axes are scaled to allow comparison of the different conditions. Since the distribution of selection coefficients in DMSO is much narrower than in MTX, data appears as a group centered on 0. Density plots show the distribution of selection coefficients for the replicate on the X axis. **b)** Distributions of selection coefficients for amino acids in the different replicates in MTX IC75 condition. **c)** Distributions of selection coefficients for amino acids in the different replicates in MTX IC90 condition. Statistical tests are Spearman's rank correlation.



S4 Fig.: Effects of amino acid substitutions on the PjDHFR in the abscence of methotrexate and at IC75 a) Selection coefficient of a given amino acid (y-axis) at a given position (x-axis) without MTX. Red points show the wild-type sequence of PjDHFR. **b)** Selection coefficient of a given amino acid (y-axis) at a given position (x-axis) at a MTX concentration corresponding to 75% of growth inhibition (IC75). **c)** Positions of contacting residues along PjDHFR modeled on structural alignments between PjDHFR and orthologous *P. carinii* DHFR (PDB: 3CD2 (MTX and NADPH) and 4CD2 (DHF)). Contact was established as amino acids with an α carbon located less than 8 Å from MTX, DHF or NADPH.

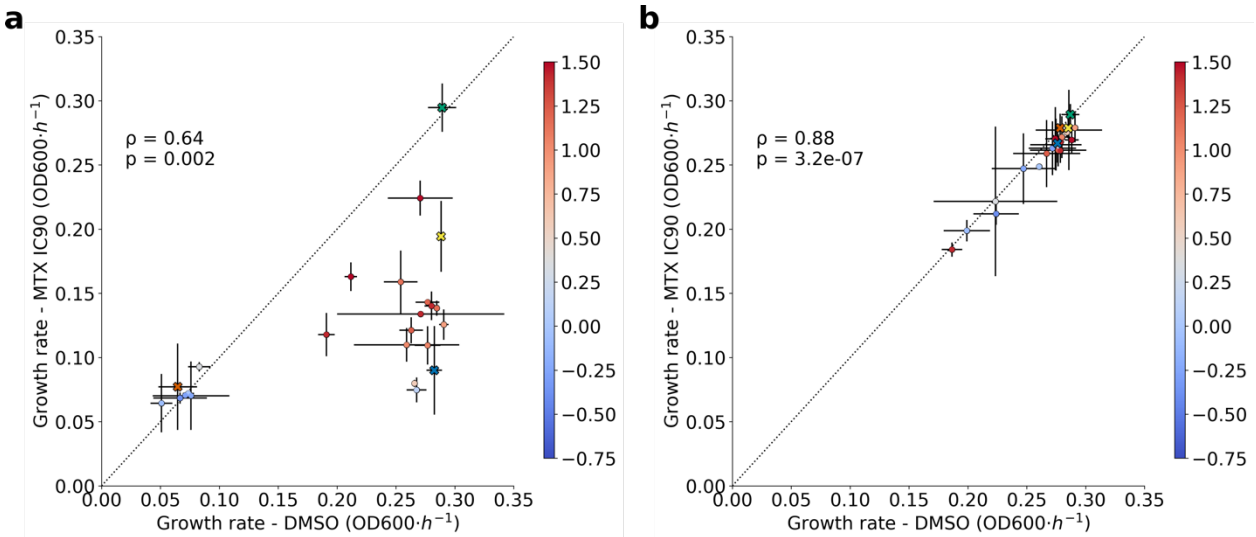


S5 Fig.: Intersects between statistically significant mutants in IC75 and IC90.
Upset plot showing the intersections of the Benjamini Hochberg-FDR corrected groups (control for the false discovery rate of significantly resistant mutants at 5% confidence). The X-axis bars represent the size of each group and the Y axis represent intersection size for each group. For maximum confidence, mutations considered significant in both conditions should be considered.

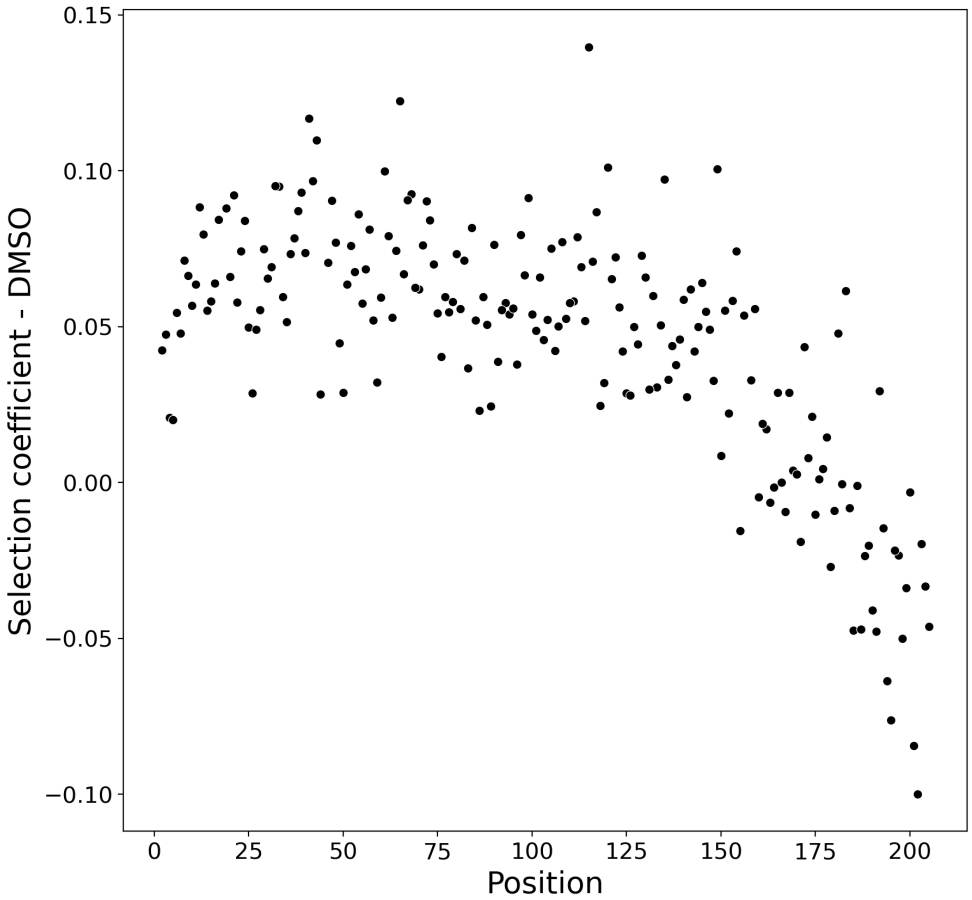


S6 Fig.: Individual growth curves for validation of mutations in DMSO and IC90. Mutants are identified on top of their individual growth curves. For stop codons, we constructed both TAG (from DMS using NNK degenerate primers) and TAA as extra controls to ensure that TAG stop codons did not allow read through. All curves were done in triplicate, with error intervals between replicates appearing around the curve in a lighter color. Strains were grown in the same media as the competition assay for 40 hours, and validation growth rates were measured from these curves. Small discrepancies intervals in some of the curves in DMSO can be attributed to the formation of bubbles within the wells of the plates. These intervals were not

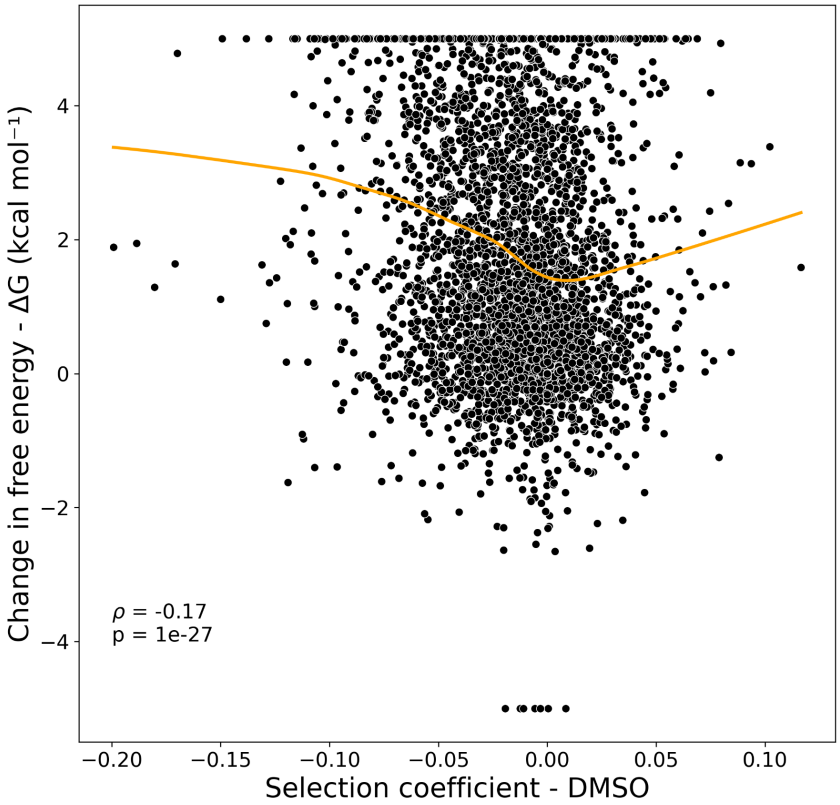
1127 considered when measuring growth rate.
1128



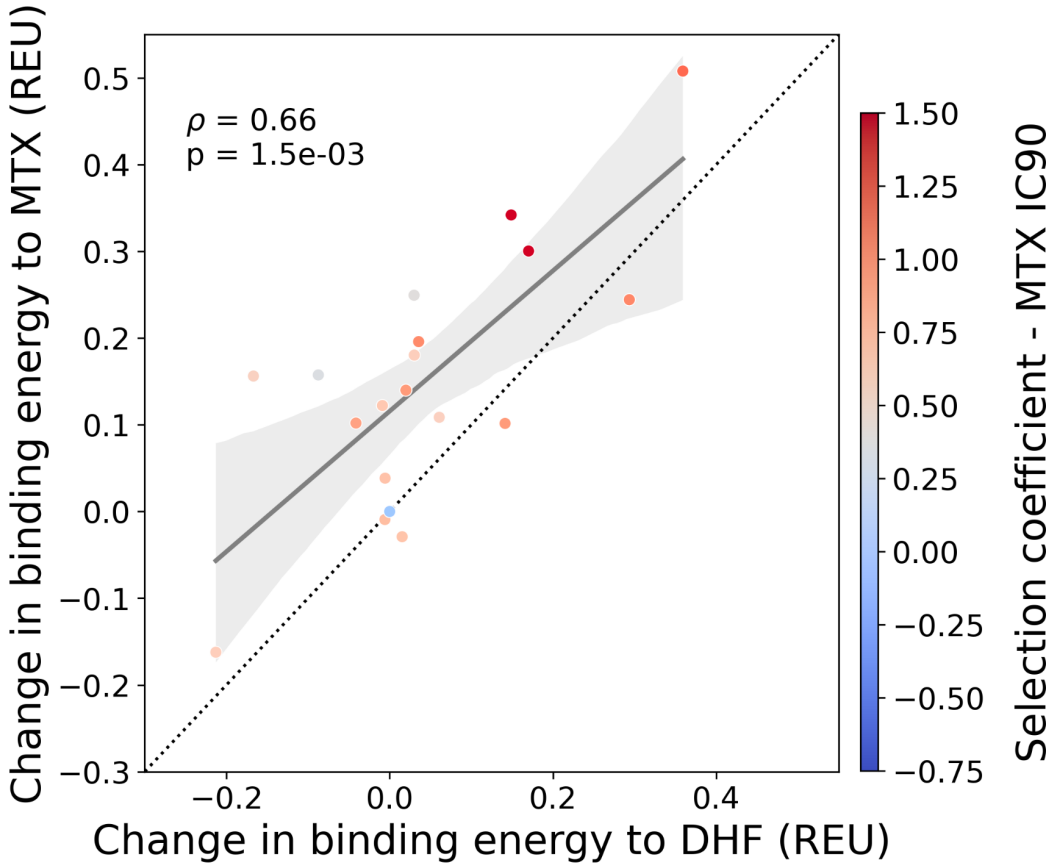
S7 Fig.: Comparison of growth rates from validations growth curves in strain FDR0001 for some manually reconstructed mutants. a) Growth rates in DMSO are plotted on the X axis, as a measure of complementation, and growth rates in the presence of MTX IC90 are plotted on the Y axis, as a measure of resistance. These growth curves were conducted in the absence of β-estradiol. Most sensitive mutants cannot complement the deletion of *DFR1* in strain FDR0001 (*dfr1Δ*). Error bars represent growth rate values measured across the four biological replicates in both conditions. Data points marked by crosses represent the Empty plasmid (orange), the wild-type PjDHFR (blue), ScDHFR (yellow) and mDHFR (green). Other round data points are colored by their selection coefficients at IC90. **b)** Same as panel a), but in the presence of β-estradiol, which leads to expression of MTX insensitive DfrB1. Statistical tests are Spearman's rank correlation.



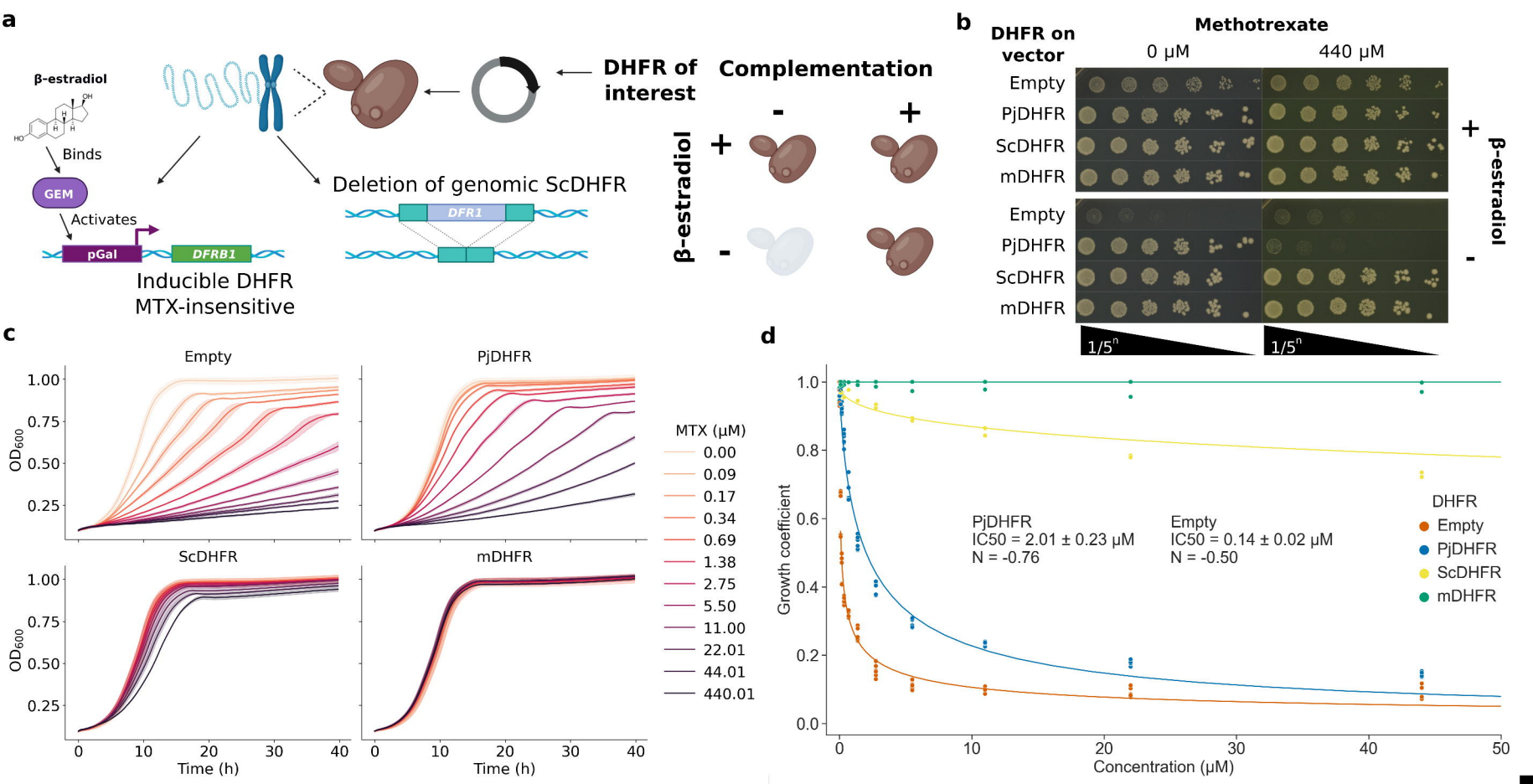
S8 Fig.: Selection coefficients of stop codons in DMSO across positions.
Each point represents the selection coefficient of a stop codon at a given position (the median selection coefficient across triplicates.

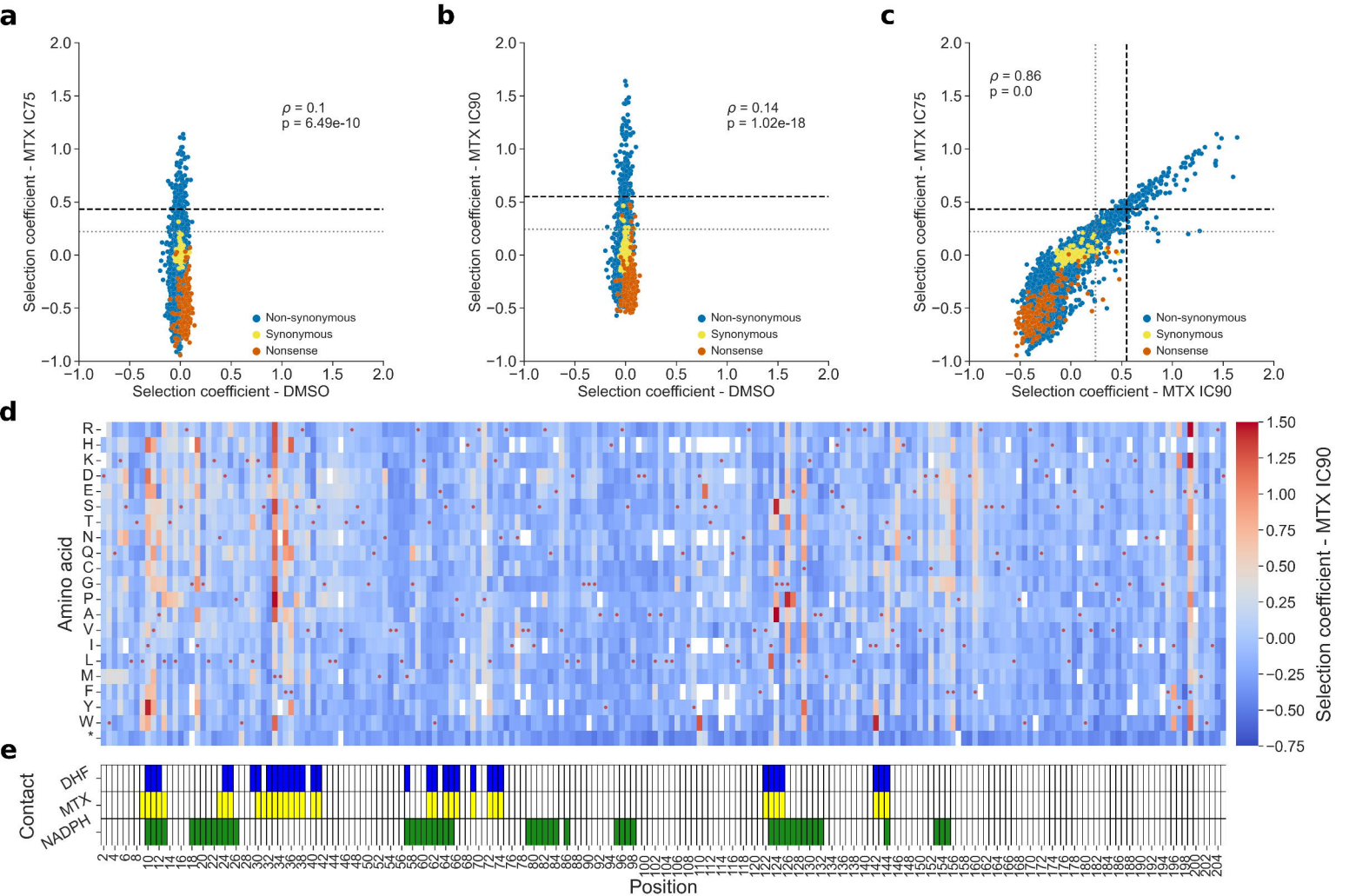


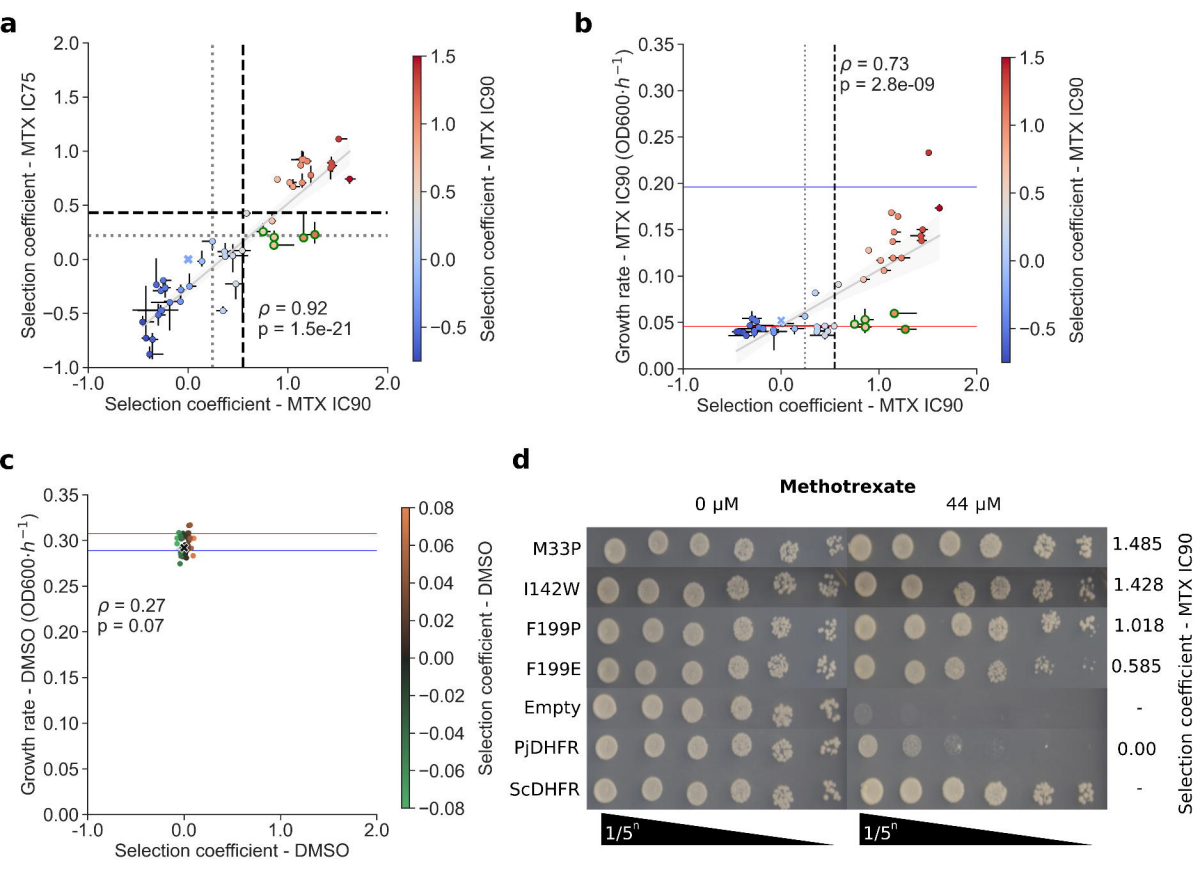
S9 Fig.: Selection coefficients in DMSO and change in free energy of the protein (ddG) do not show the expected trend. Using FoldX, changes in protein free energy were computed for all mutants forming a complex with DHF and compared to their selection coefficients in DMSO. Extreme values (-5<x<5 change in free energy) were scaled down to 5/-5, as extreme values of ddG measured by FoldX can often be attributed to clashes. Orange line represents the Lowess smoothed curve fitted using statsmodels default settings. Statistical tests are Spearman's rank correlation.

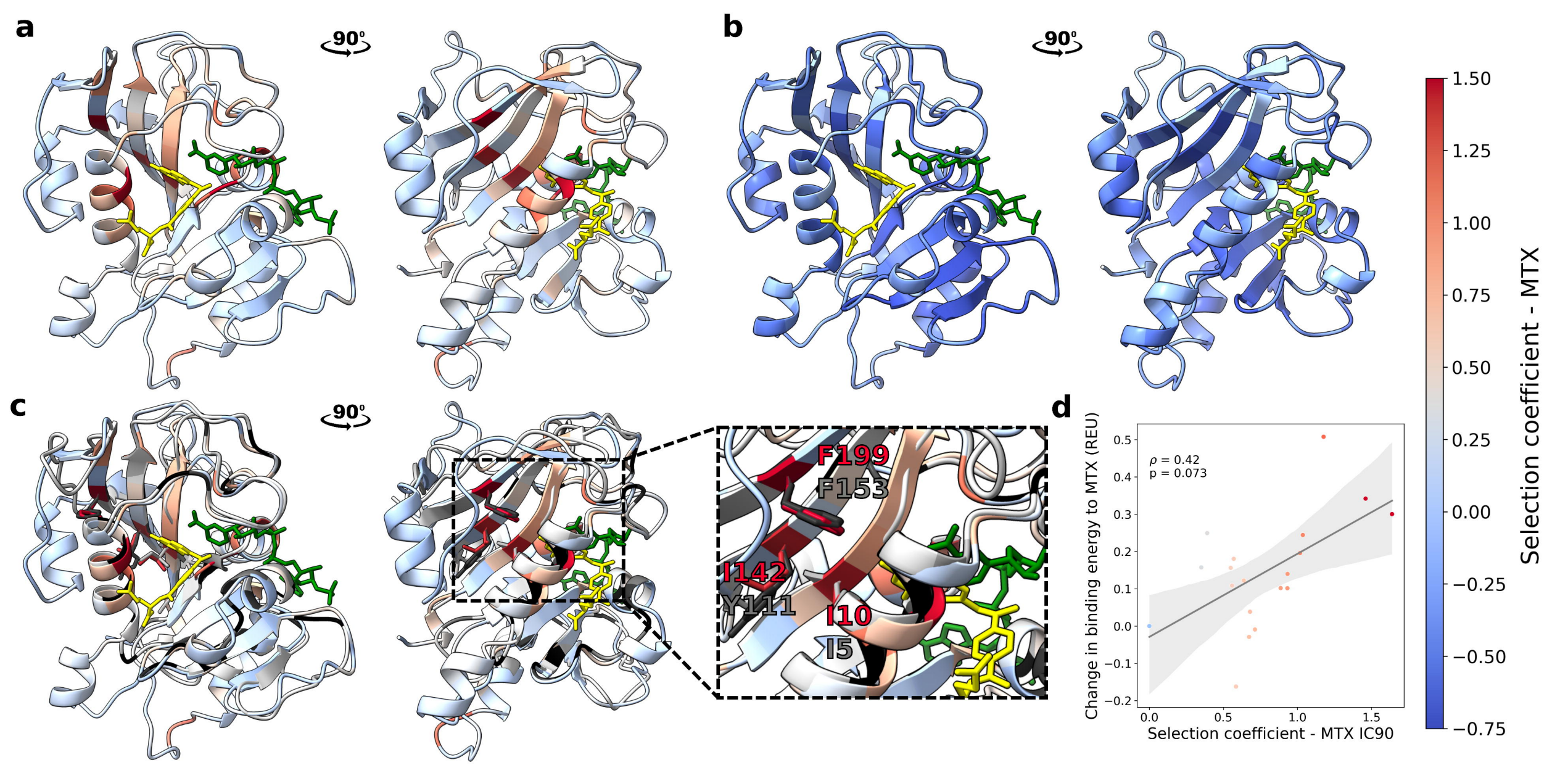


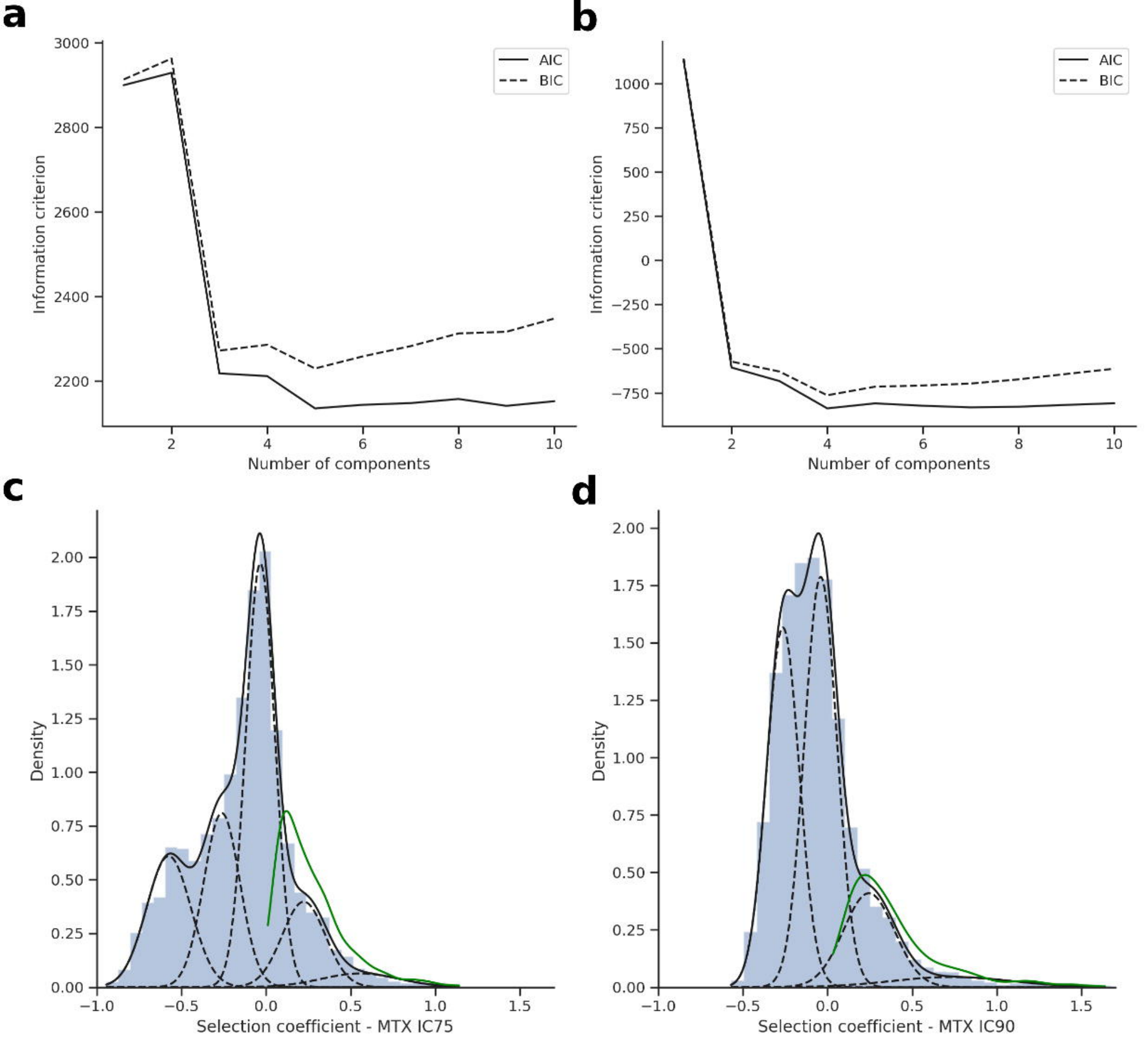
S10 Fig.: Comparison of changes in binding energy measured between PjDHFR F199 mutations with MTX and with DHF. On the X axis, we plot the change in binding energy between F199 mutants and DHF, and on the Y axis, between F199 mutants and MTX. The diagonal shows equal changes in binding energy to both molecules. Units are REU (Rosetta Energy Units). Statistical tests are Spearman's rank correlation.

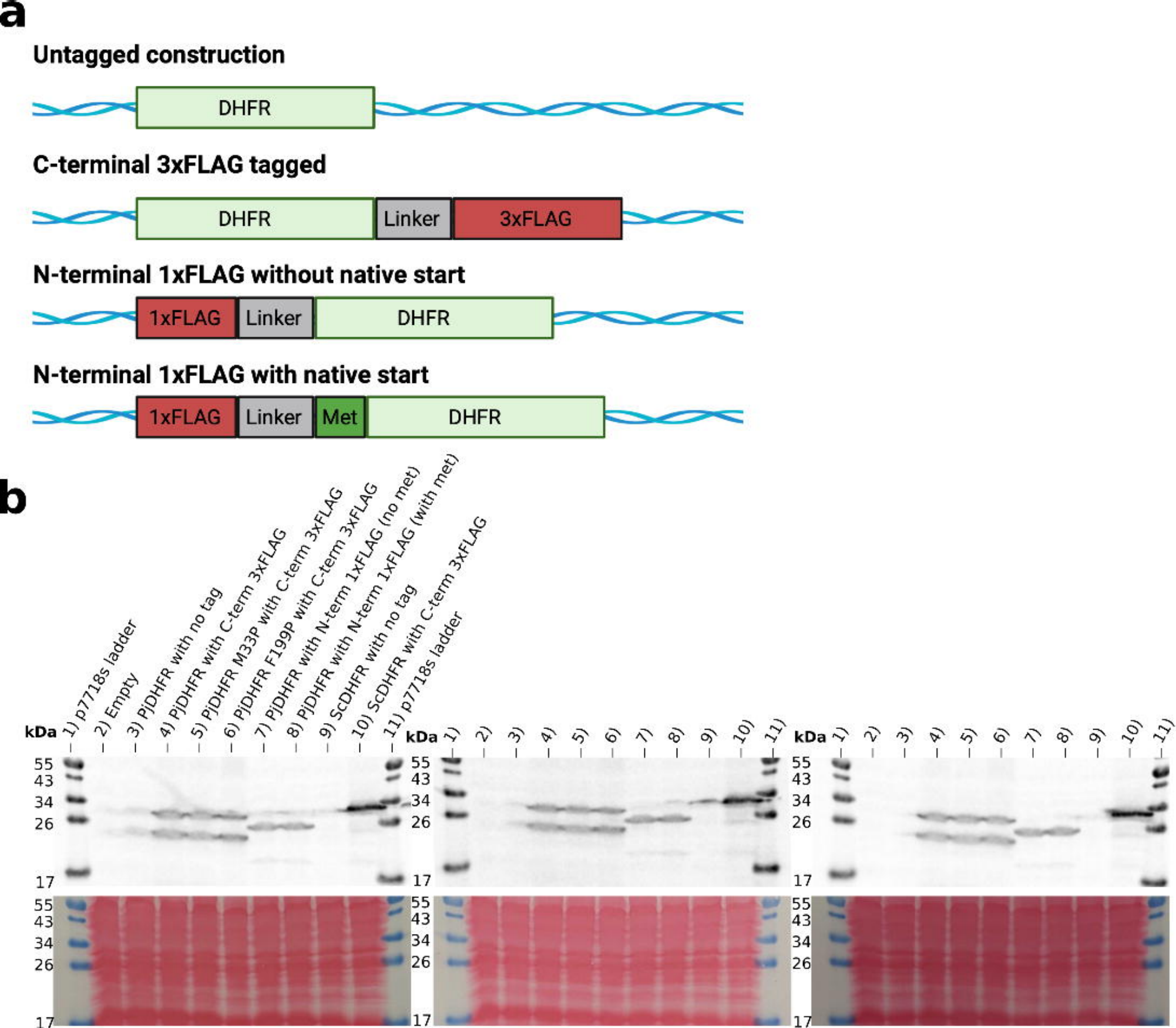

a
β-estradiol
Binds
GEM
Activates
pGal
DFRB1
Inducible DHFR
MTX-insensitive
Deletion of genomic ScDHFR
DFR1
DHFR of interest
Complementation
β-estradiol
+
-
b
DHFR on vector
Methotrexate
0 μM
440 μM
Empty
PjDHFR
ScDHFR
mDHFR
1/5^n
c
Empty
PjDHFR
ScDHFR
mDHFR
OD600
Time (h)
MTX (μM)
0.00
0.09
0.17
0.34
0.69
1.38
2.75
5.50
11.00
22.01
44.01
440.01
d
Growth coefficient
Concentration (μM)
PjDHFR
IC50 = 2.01 ± 0.23 μM
N = -0.76
Empty
IC50 = 0.14 ± 0.02 μM
N = -0.50
DHFR
Empty
PjDHFR
ScDHFR
mDHFR

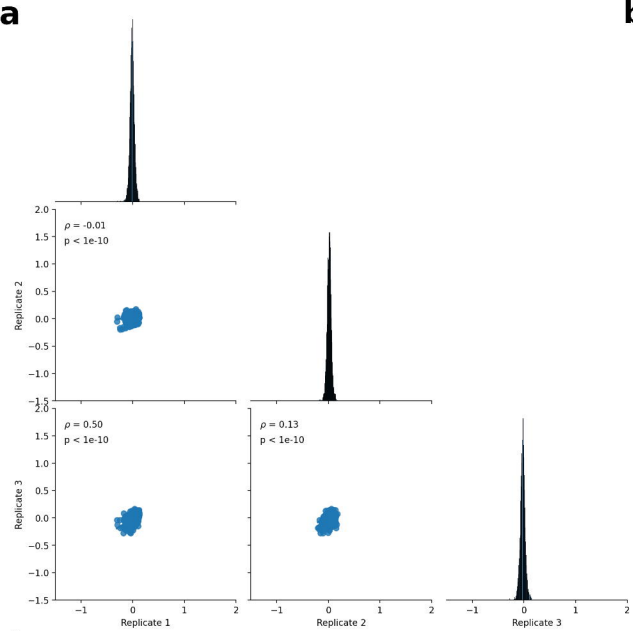

a
ρ = 0.1
p = 6.49e-10
Selection coefficient - MTX IC75
Selection coefficient - DMSO
Non-synonymous
Synonymous
Nonsense
b
ρ = 0.14
p = 1.02e-18
Selection coefficient - MTX IC90
Selection coefficient - DMSO
Non-synonymous
Synonymous
Nonsense
c
ρ = 0.86
p = 0.0
Selection coefficient - MTX IC75
Selection coefficient - MTX IC90
Non-synonymous
Synonymous
Nonsense
d
Amino acid
R H K D E S T N Q C G P A V I L M F Y W *
Selection coefficient - MTX IC90
e
Contact
DHF
MTX
NADPH
Position

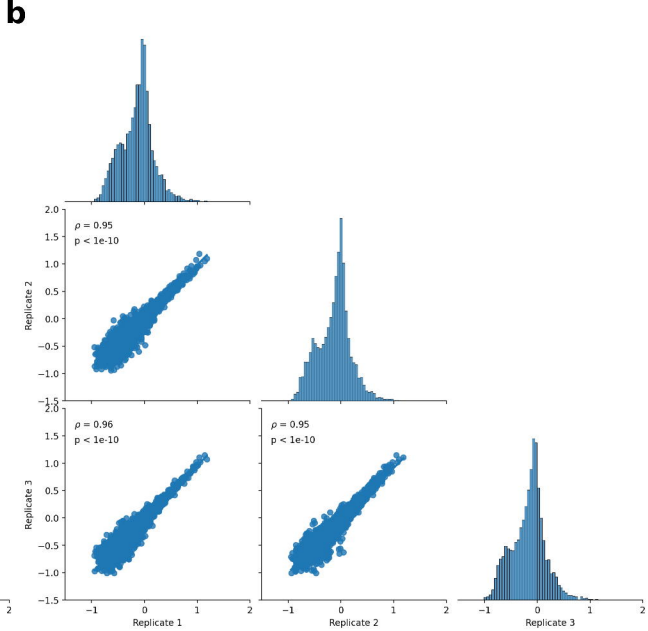

a
Selection coefficient - MTX IC75
Selection coefficient - MTX IC90
ρ = 0.92
p = 1.5e-21
Selection coefficient - MTX IC90
b
Growth rate - MTX IC90 (OD600·h⁻¹)
ρ = 0.73
p = 2.8e-09
Selection coefficient - MTX IC90
Selection coefficient - MTX IC90
c
Growth rate - DMSO (OD600·h⁻¹)
ρ = 0.27
p = 0.07
Selection coefficient - DMSO
Selection coefficient - DMSO
d
Methotrexate
0 μM
44 μM
M33P
I142W
F199P
F199E
Empty
PjDHFR
ScDHFR
1.485
1.428
1.018
0.585
-
0.00
-
Selection coefficient - MTX IC90
1/5ⁿ
1/5ⁿ

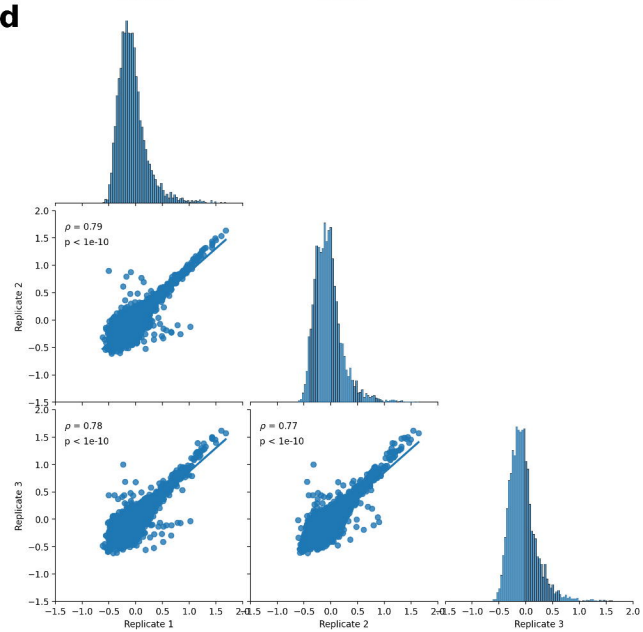

a
90°
b
90°
c
90°
F199
F153
I142
Y111
I10
I5
d
ρ = 0.42
p = 0.073
Change in binding energy to MTX (REU)
Selection coefficient - MTX IC90
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
0.0
0.5
1.0
1.5
Selection coefficient - MTX
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

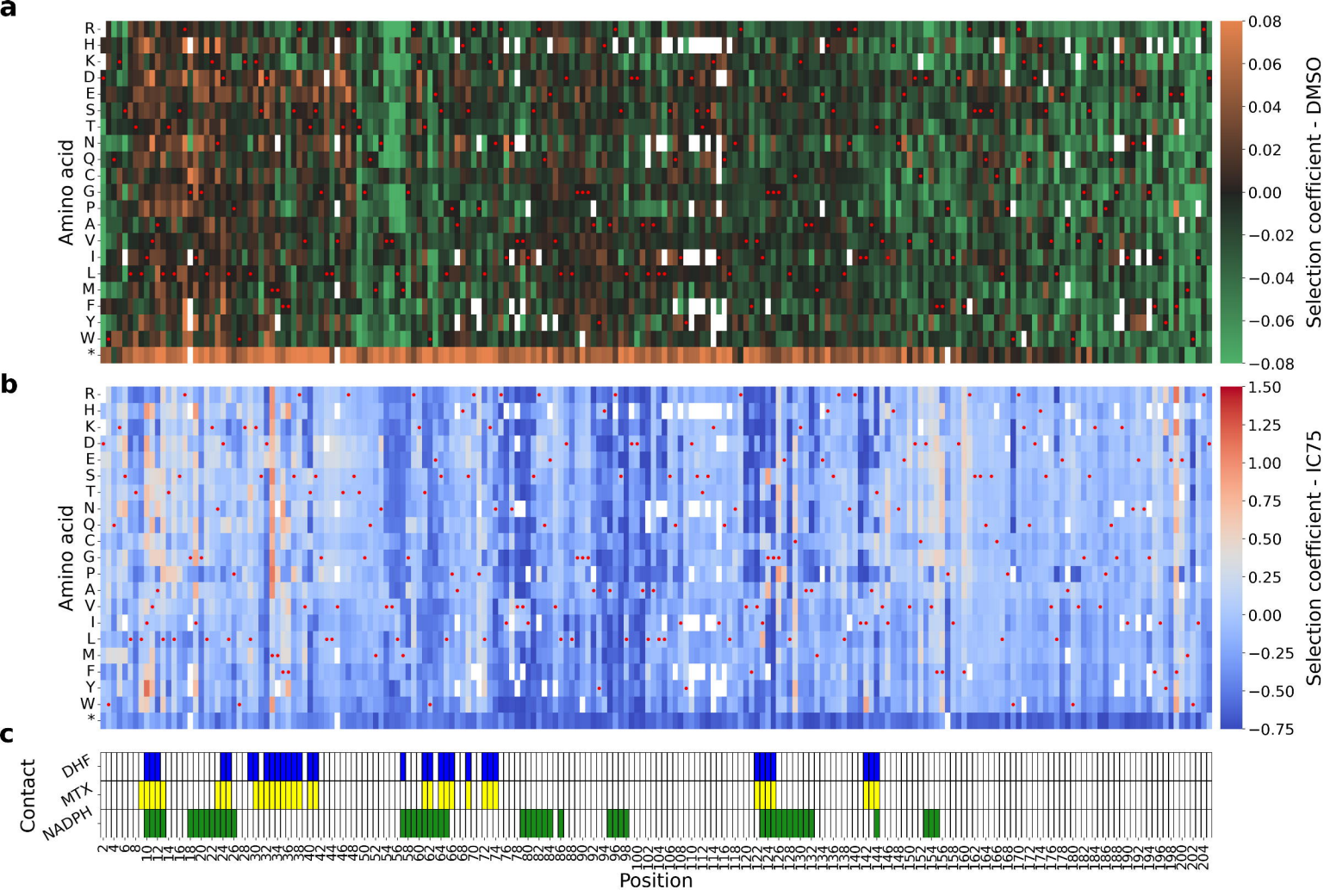

a
b
c
d
AIC
BIC
Information criterion
Number of components
Density
Selection coefficient - MTX IC75
Selection coefficient - MTX IC90

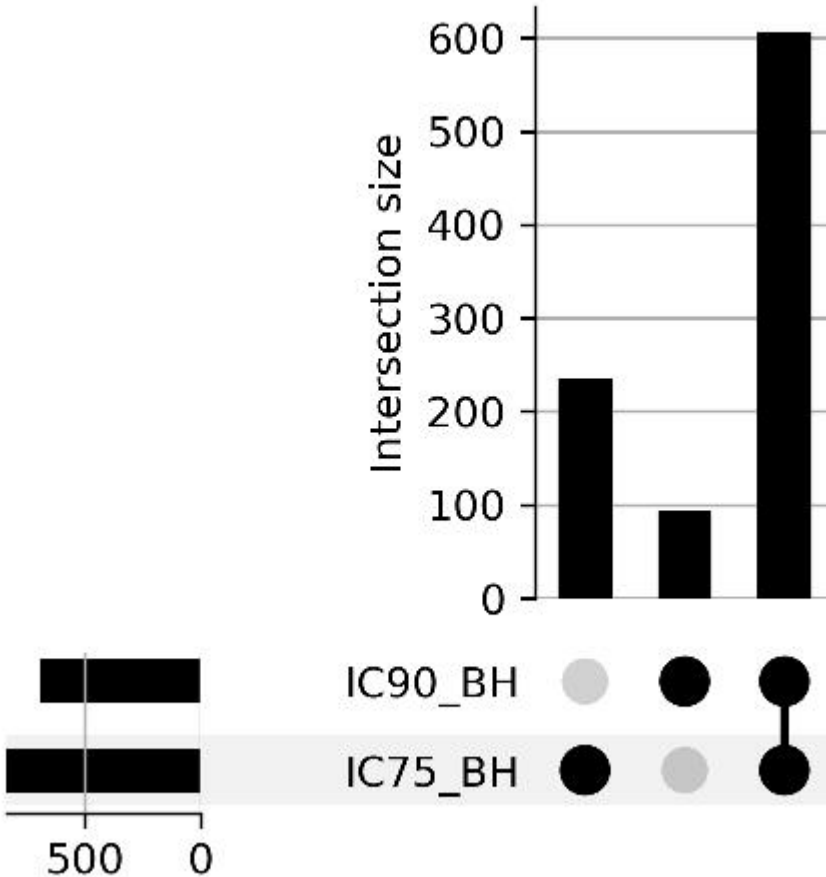

a
Untagged construction
DHFR
C-terminal 3xFLAG tagged
DHFR
Linker
3xFLAG
N-terminal 1xFLAG without native start
1xFLAG
Linker
DHFR
N-terminal 1xFLAG with native start
1xFLAG
Linker
Met
DHFR
b
1) p7718s ladder
2) Empty
3) PjDHFR with no tag
4) PjDHFR with C-term 3xFLAG
5) PjDHFR M33P with C-term 3xFLAG
6) PjDHFR F199P with C-term 3xFLAG
7) PjDHFR with N-term 1xFLAG (no met)
8) PjDHFR with N-term 1xFLAG (with met)
9) ScDHFR with no tag
10) ScDHFR with C-term 3xFLAG
11) p7718s ladder
kDa
55
43
34
26
17

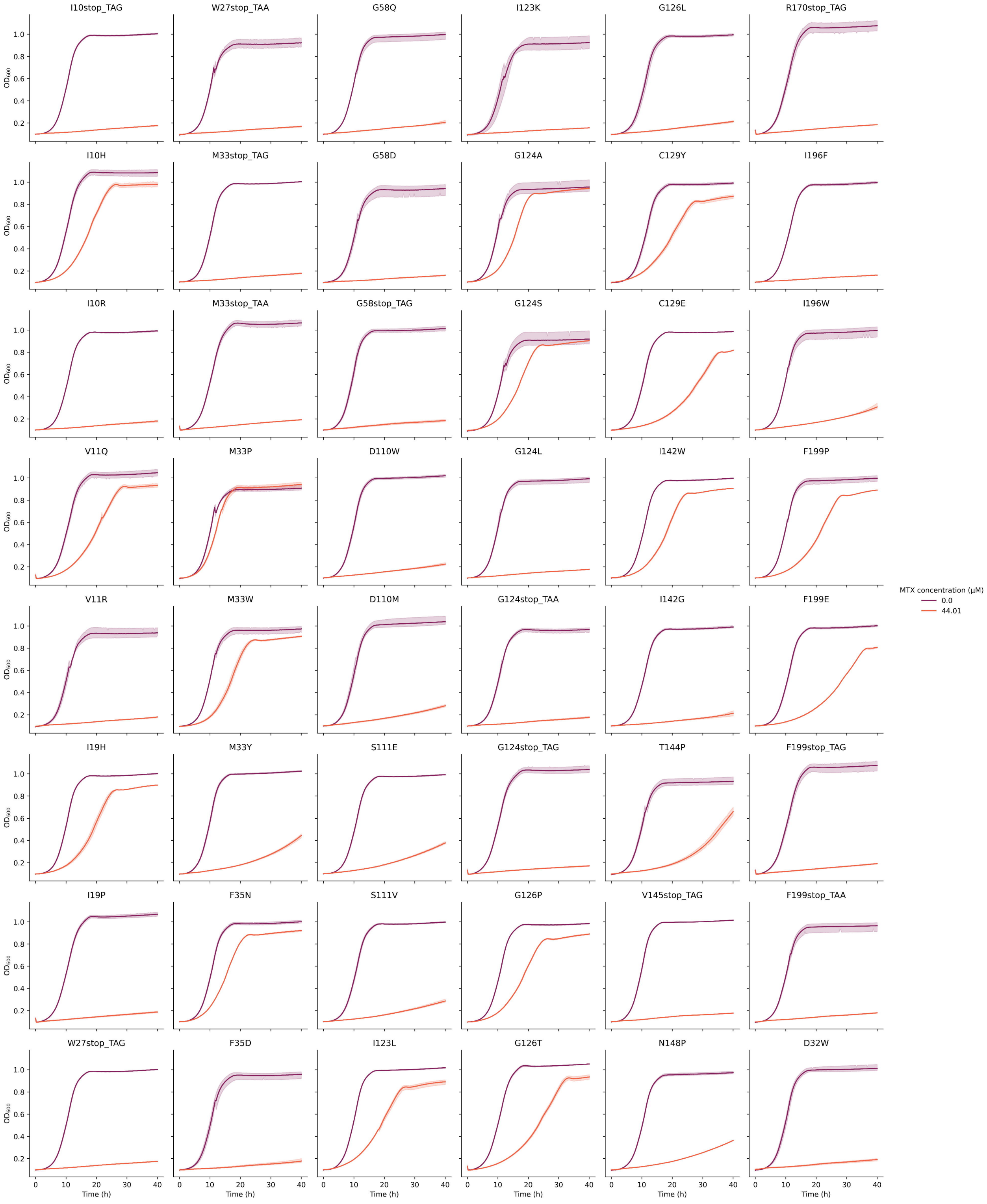

a
ρ = -0.01
p < 1e-10
ρ = 0.50
p < 1e-10
ρ = 0.13
p < 1e-10
Replicate 1
Replicate 2
Replicate 3

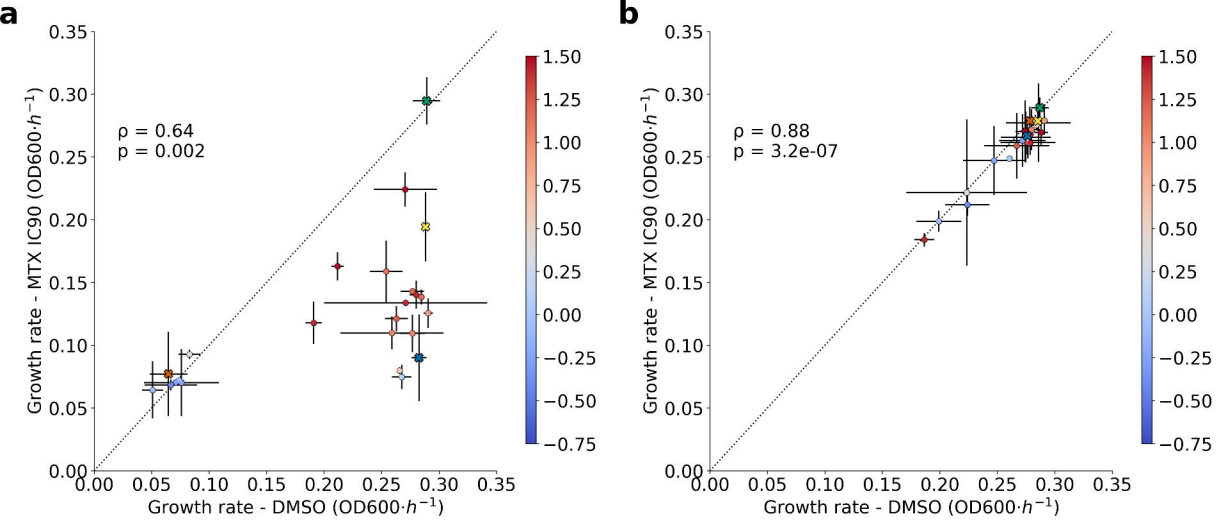

b
ρ = 0.95
p < 1e-10
ρ = 0.96
p < 1e-10
ρ = 0.95
p < 1e-10
Replicate 1
Replicate 2
Replicate 3

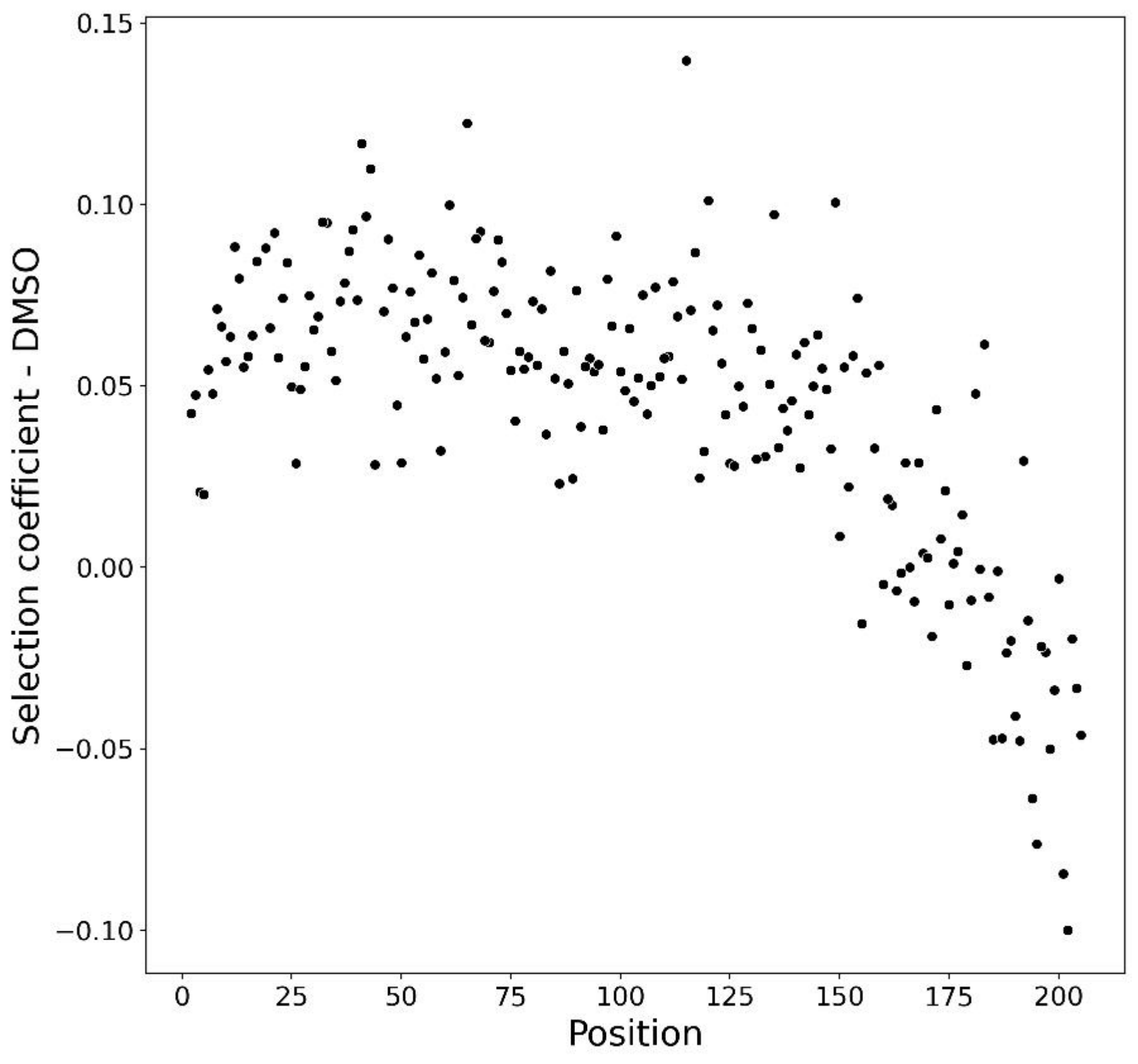

d
ρ = 0.79
p < 1e-10
ρ = 0.78
p < 1e-10
ρ = 0.77
p < 1e-10
Replicate 1
Replicate 2
Replicate 3

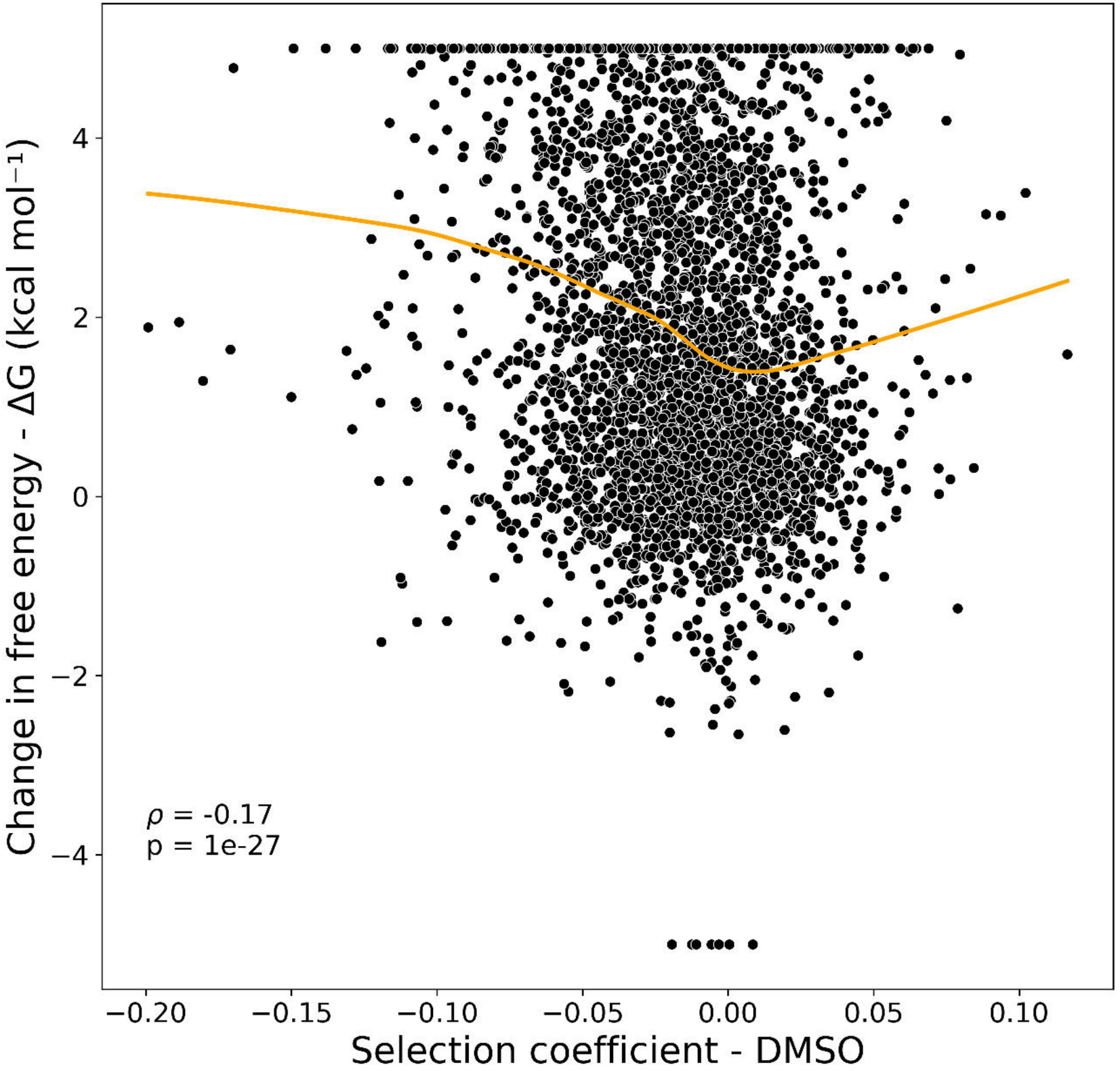

a
Amino acid
R
H
K
D
E
S
T
N
Q
C
G
P
A
V
I
L
M
F
Y
W
*
Selection coefficient - DMSO
0.08
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.08
b
Amino acid
Selection coefficient - IC75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
c
Contact
DHF
MTX
NADPH
Position

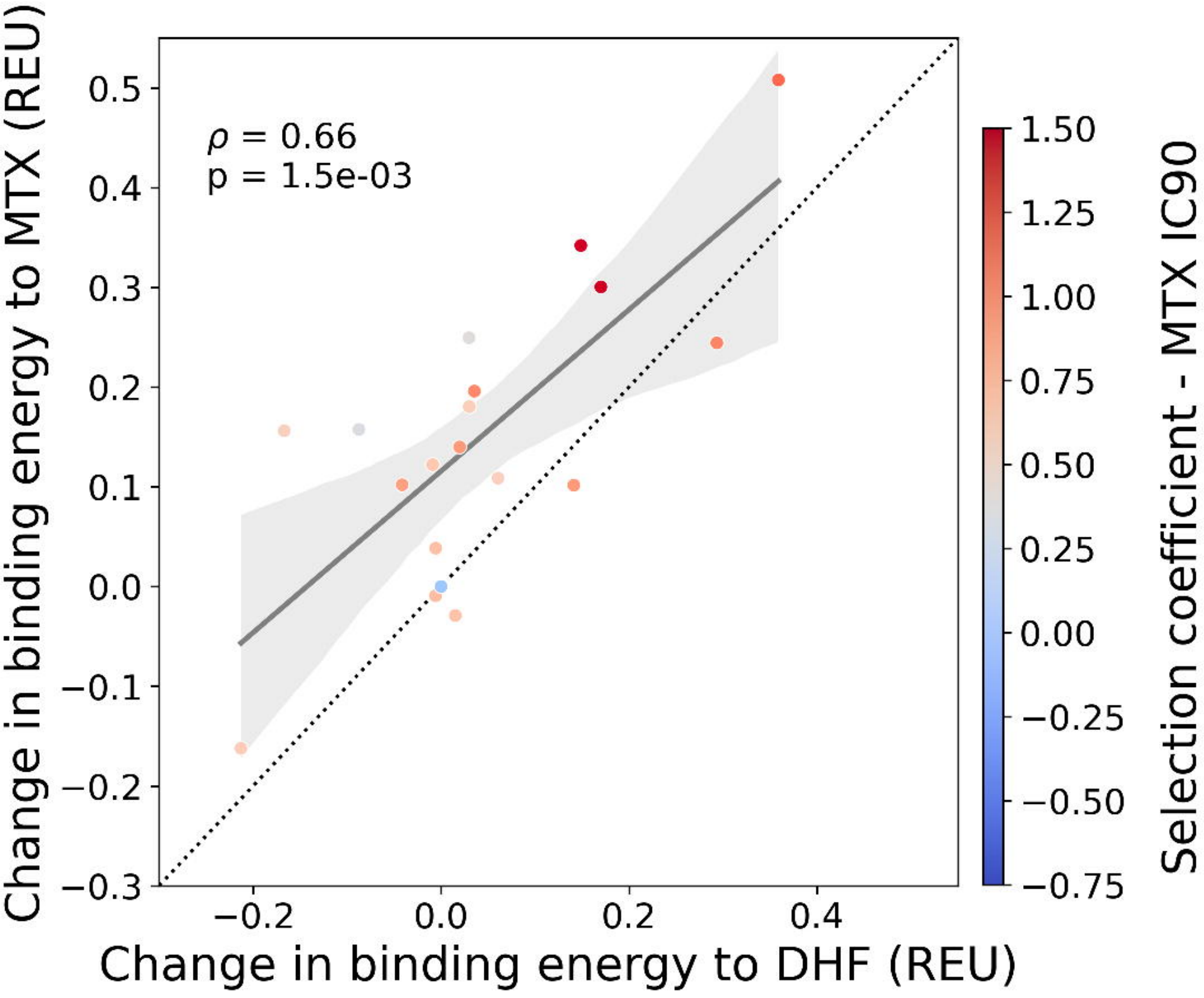

Intersection size
600
500
400
300
200
100
0
IC90_BH
IC75_BH
500
0


I10stop_TAG
W27stop_TAA
G58Q
I123K
G126L
R170stop_TAG
I10H
M33stop_TAG
G58D
G124A
C129Y
I196F
I10R
M33stop_TAA
G58stop_TAG
G124S
C129E
I196W
V11Q
M33P
D110W
G124L
I142W
F199P
V11R
M33W
D110M
G124stop_TAA
I142G
F199E
I19H
M33Y
S111E
G124stop_TAG
T144P
F199stop_TAG
I19P
F35N
S111V
G126P
V145stop_TAG
F199stop_TAA
W27stop_TAG
F35D
I123L
G126T
N148P
D32W
OD600
Time (h)
MTX concentration (μM)
0.0
44.01


a
ρ = 0.64
p = 0.002
Growth rate - MTX IC90 (OD600·h⁻¹)
Growth rate - DMSO (OD600·h⁻¹)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
b
ρ = 0.88
p = 3.2e-07
Growth rate - MTX IC90 (OD600·h⁻¹)
Growth rate - DMSO (OD600·h⁻¹)


0.15
0.10
0.05
0.00
−0.05
−0.10
Selection coefficient - DMSO
0
25
50
75
100
125
150
175
200
Position


Change in free energy - ΔG (kcal mol⁻¹)
4
2
0
−2
−4
ρ = -0.17
p = 1e-27
−0.20
−0.15
−0.10
−0.05
0.00
0.05
0.10
Selection coefficient - DMSO


ρ = 0.66
p = 1.5e-03
Change in binding energy to MTX (REU)
Change in binding energy to DHF (REU)
Selection coefficient - MTX IC90
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.2
0.0
0.2
0.4
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75