Linking Protein Sequence to Structure and Function in Bacterial DHFR Enzymes

A Thesis

submitted to

Indian Institute of Science Education and Research, Pune

in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Saillesh Aravindhan Chinnaraj

Indian Institute of Science Education and Research Pune,

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

April, 2023

Supervisor: Dr. Nishad Matange

Assistant Professor, Department of Biology, Indian Institute of Science Education and Research, Pune

From May 2022 to March 2023

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE

Certificate

This is to certify that this dissertation entitled **Linking Protein Sequence to Stucture and Function in Bacterial DHFR Enzymes** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Saillesh Aravindhan Chinnaraj at Indian Institute of Science Education and Research under the supervision of Dr. Nishad Matange, Assistant Professor, Department of Biology, during the academic year 2022-2023.

Dr. Nishad Matange

Committe:

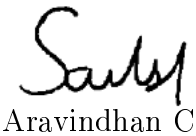
Dr. Nishad Matange

Dr. M.S. Madhusudhan

This Thesis is dedicated to my parents, my brother, and to my younger self

Declaration

I hereby declare that the matter embodied in the report entitled **Linking Protein Sequence to Structure and Function in Bacterial DHFR Enzmyes** are the results of the work carried out by me at the Department of Biology, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Nishad Matange and the same has not been submitted elsewhere for any other degree



Saillesh Aravindhan Chinnaraj

Date: 02/05/2023

Table of Contents

List of Tables

List of Figures

Abstract

Evolution has led to significant sequence variation between homologous proteins present in different organisms. Natural variation is a part of the theoretical sequence space that can fold into a specific protein structure, which in turn plays an important role in determining the functionality of the protein. Understanding how sequence variation affects protein structure and function will enable the development of effective therapeutic strategies to curb infections by bacteria. In this study, the impact of interspecific and intraspecific variation on protein structure and function in bacterial Dihydrofolate Reductase (DHFR) enzymes is investigated. Trimethoprim resistance in *Escherichia coli* was used as a phenotypic read-out to investigate the impact of single and combinatorial mutations in DHFR on protein function. The results indicate that natural sequence variation in DHFR has the potential to influence intrinsic and mutationally acquired trimethoprim resistance. The results show that the ability of intraspecific variants in *E. coli* DHFR to confer trimethoprim resistance depends on the physiochemical properties of the residues. Moreover, the results also show that a combination of mutations has drastically different phenotypic effects than their corresponding single mutations. The results display the impact of intraspecific variations on the phenotypic effects of other resistance-conferring mutations. At least some of these findings can be explained by how sequence variation alters the stability of DHFR, as well as its tolerance to mutation. This study, thus, demonstrates the potential of standing sequence variation in proteins to significantly impact the evolution of organismal traits with biomedical significance, such as antimicrobial resistance.

Acknowledgements

First and foremost, I would like to thank Dr. Nishad Matange for providing me with the resources and facilities to complete my Master's Thesis Project. His constant guidance and support in matters related to and outside the project was a constant pillar of support that enabled me to bring this project to its fruition. I would also like to thank all the members of the Bugs and Drugs lab for creating an amazing atmosphere and for their constant support and insights that allowed the smooth sailing of my project. I would like to express my gratitude to Dr. M.S. Madhusudhan whose guidance and support provided important insights into the project. I would like to thank Tejashree Kanitkar for her assistance in analyzing the data related to the project. I would like to thank Chetna Yelpure for performing the MSA for the *E. coli* strains and Arsh Chavan for providing me with the list of mutations in *E. coli* that confer resistance to TMP.

I would like to extend my appreciation to the Indian Institute of Science Education and Research, Pune for providing me with the facilities and the resources which capacitated me to complete my Thesis project. I would also like to thank the administrative staff as well as the housekeeping staff for providing me with the support that enabled the smooth functioning of my project work.

I would like to express my gratitude to my friends and family, who are my emotional pillar, for their love and support.

Contributions

Contributor's name	Contributor's role
Saillesh Chinnaraj; Dr. Nishad Matange	Conceptualization Ideas
Saillesh Chinnaraj; Dr. Nishad Matange	Methodology
Saillesh Chinnaraj	Software
Saillesh Chinnaraj	Validation
Saillesh Chinnaraj	Formal Analysis
Saillesh Chinnaraj; Dr. Nishad Matange	Investigation
Dr. Nishad Matange	Resources
Saillesh Chinnaraj; Dr. Nishad Matange	Data Curation
Saillesh Chinnaraj	Writing - original draft preparation
Dr. Nishad Matange	Writing - review and editing
Saillesh Chinnaraj; Dr. Nishad Matange	Visualization
Saillesh Chinnaraj; Dr. Nishad Matange; Rhea Vinchhi; Chinmaya Jena ; Chetna Yelpure	Supervision
Saillesh Chinnaraj; Dr. Nishad Matange	Project administration
Dr. Nishad Matange	Funding acquisition

1 Introduction

1.1 Variation in Homologous Proteins

1.1.1 Homologous Proteins - similar yet different

Homologous proteins are proteins belonging to different species that are derived from a common ancestor. These proteins often have similar structures and functions (Pearson and Sierk, 2005). However, an alternative explanation as to why two proteins are similar is that the two proteins originated from different ancestors but converged (i.e., underwent convergent evolution) due to certain functional or structural constraints. Homologous proteins can either be orthologous or paralogous. When a speciation event occurs, a particular gene in the ancestral species is replicated and is present in the two newly formed species. The copies of the ancestral gene that are present in the two newly formed species are orthologous, and the proteins produced by these genes are called orthologous proteins. For example, the α-hemoglobin molecule belonging to humans and the α-hemoglobin molecule belonging to mice are orthologous. When a gene duplication event occurs and both copies evolve beside each other in the same organism, the genes are said to be paralogous and the proteins produced by these genes are paralogous proteins (Fitch, 1970). Both orthologous and paralogous proteins generally have the same function. The key difference lies in the specificity of the protein to their respective targets. Paralogous proteins tend to bind to different targets, whether it is a ligand or a target like DNA. Orthologous proteins, on the other hand, bind to similar targets despite existing in different organisms (Mirny and Gelfand, 2002). While homologous proteins are similar, there often exists significant variation between the sequences of these proteins (Horner and Pesole, 2003).

1.1.2 The Importance of Understanding Variation in Homologous proteins

Variation is crucial for the process of evolution to occur (Hershberg, 2015). A sequence space can be defined as the space that contains all possible amino acid sequences. The sequence space contains an innumerable number of protein sequences, but only a small percentage of the theoretical sequence space is found in nature. This is due to the various structural and functional constraints involved in the folding and formation of a protein. The structural constraints that can influence the variation that is allowed in a protein include factors such as solvent accessibility, packing density, and flexibility. The functional constraints that influence the variation allowed in a protein include factors such as purifying selection and positive selection. This 'permissible' percentage represents only those sequences that can fold into a specific protein structure, which in turn plays an important role in determining the functionality of the protein (Povolotskaya and Kondrashov, 2010). Natural variation in protein sequence exists in nature and hence represents a subset of this small fraction of the sequence space that can conform into and form a protein. A schematic depicting this process

is shown in Fig 1.

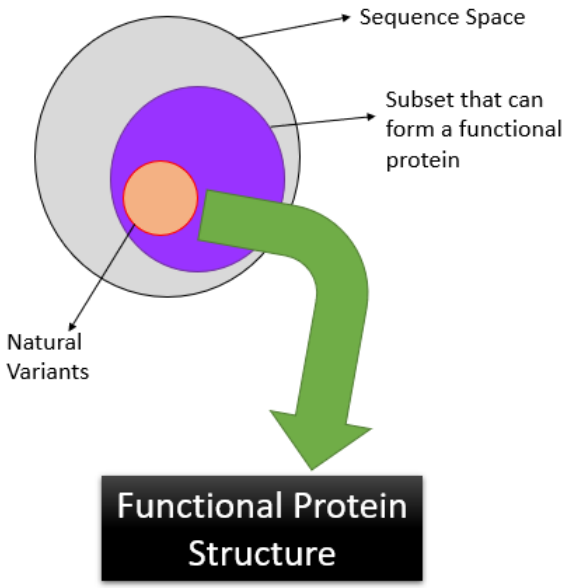


Figure 1: **A schematic representing the concept of a sequence space and the advantage of studying natural variation in proteins.**

Thus, while trying to understand the relationship between protein sequence and protein structure, it is more fruitful to study natural variation than to look at the impact of random mutations in the protein sequence on structure as in the latter, the chance that the mutation is deleterious due to structural or functional constraints being violated is much higher. This is because most mutations and variations have a neutral effect with regard to how the mutations and variations affect fitness (Kimura, 1968). Among the mutations that can affect the function of the protein, most of them have a deleterious effect on the protein (Jordon et al., 2010). Thus, studying natural variation can enable a better understanding of protein structure and function and the influence of changes in certain positions of the protein sequence on protein structure and function.

1.1.3 The Evolution of Protein Function

The variation in protein sequence and structure observed in nature depends on factors other than the structural and functional constraints of the protein. Given the right conditions, a protein can evolve new functions by accumulating variation. As shown in Fig 2, a protein can traverse the sequence space (i.e., accumulate variation or mutations) only by taking unit-evolutionary steps. In this context, unit-evolutionary steps mean that a protein can accumulate amino-acid substitutions only if the resultant protein is still functional. Thus, protein evolution involves the protein exploring the sequence space by going through func-

tional intermediates. If a certain trajectory of evolution involves the protein undegoing a certain amino-acid substitution at one of its residues such that the intermediate becomes non-functional, such an intermediate would most likely be eliminated by evolutionary forces like natural selection before another amino-acid substitution can be accommodated. (Maynard Smith, 1970; Povolotskaya and Kondrashov, 2010). However, by treading along a trajectory of functional intermediates, proteins can evolve to an extent such that they gain the ability to perform another function. This is a model by which genes (and the proteins encoded by these genes) may gain new functions (Maynard Smith, 1970).

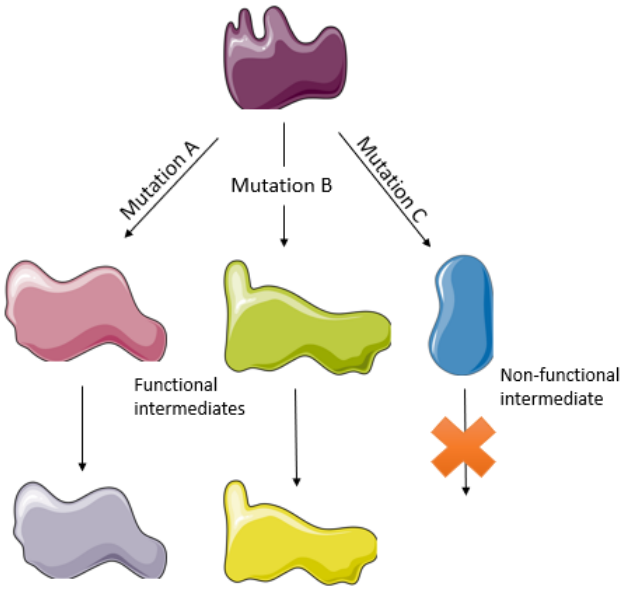


Figure 2: **A schematic depicting the possible evolutionary trajectories of a protein and the outcomes of each possibility.** When mutations result in non-functional intermediates, the evolutionary trajectory is shut down as the intermediate is likely to be eliminated by natural selection.

1.1.4 How does Variation affect Proteins?

Previous studies have reported the capability of natural variation in impacting the function of a protein. An example was a study conducted by Anwer et al. in 2014 that looked at a natural allele of the protein ELF3, which is involved in circadian clock regulation in Arabidopsis. Anwer et al. discovered that organisms with this natural allele of ELF3 were worse off in their ability to respond readjust their internal clocks in response to external cues compared to organisms with the wild-type ELF3. Moreover, the protein synthesized by the natural variant of ELF3 was less likely to localize to the nucleus compared to the protein synthesized by the wild-type ELF3 (Anwer et al., 2014). It has also been shown that in *Drosophila melanogaster* larvae, the presence of specific natural variants in the *for* gene (which codes for a cGMP-dependent protein kinase) affects the foraging behavior of the

larvae. Larvae with one natural variant of the *for* gene adapt a rover-type foraging behavior while larvae with another natural variant of the for gene adapt a sitter-type foraging behavior (Osborne et al., 1997; Dawson-Scully et al., 2007). Thus, natural variation can cause changes in protein function and can even lead to changes in the behavior of organisms. However, the changes that natural variation in genes make to the protein which in turn lead to changes in function are not well studied.

Most proteins have a set of residues that are crucial for their function. While it might be easy to assume that variation only at those residues (and nearby residues) is relevant to protein function, that is not always the case. Variation in residues located far from the active site residues is also capable of influencing function (Echave et al., 2016). The most common way by which variation in a residue can affect proteins is by influencing protein stability (Bigman and Levy, 2018). For instance, a study by Leferink et al. in 2014 showed that the mutation of a residue located far away from the active site (about 12Å) increases the active site accessibility of the enzyme copper nitrite reductase which results in a lowered affinity of the substrate (nitrite) and a lowered catalytic efficiency (Leferink et al., 2014). Thus, variation arising in residues that are not directly involved in function has the potential to affect protein structure and function.

1.1.5 The Relevance of Natural Variation in the context of disease

Natural variation has the potential to enhance our present understanding of molecular pathways and mechanisms, especially those related to human health and disease (Gasch et al., 2016). For instance, a study integrated the natural variation of *Drosophila* taken from the Drosophila Genetic Reference Panel (DGRP) (which represents polymorphisms in a natural population and comprises more than 200 inbred lines of wild fly isolates) with a *D. melanogaster* model for traumatic brain injury (TBI) and discovered that consuming a diet low in sugar after TBI protected against death in *D. melanogaster* (Katzenberger at al., 2015).

In the context of disease-causing pathogens, the functionality of a protein that performs a specific function often determines how effective certain therapeutics are in inhibiting the growth of the pathogens. As natural variation is capable of affecting protein function, understanding the impact of natural variation on protein structure and function could enable the more efficient use of therapeutics. Thus, natural variation could be used to enhance our understanding of phenomena such as antimicrobial resistance, which has become a growing threat in recent years (Reygaert et al., 2018).

1.2 Antimicrobial Resistance – A Growing Threat

1.2.1 An Overview of Antimicrobial Resistance

Antibiotics and antimicrobials are used on a daily basis to treat infection and disease. However, bacteria and microbes have gained resistance to multiple antimicrobial agents (Christaki

et al., 2019; Reygaert et al., 2018). In recent years, antimicrobial resistance (AMR) has surfaced as a pressing issue at a global scale, with multi-drug resistant bacteria emerging in all parts of the world (Christaki et al., 2019; Reygaert et al., 2018). Antimicrobial resistance refers to the ability that bacteria or microbes have or gain that allows them to avoid the mechanisms utilized by antimicrobial agents to kill or prevent their growth (Christaki et al., 2019). The emergence of AMR as a property that allows microbes to resist the effects of therapeutics has significantly boosted the impact of infectious diseases on the world population and on the world's healthcare (Reygaert et al., 2018). For instance, AMR has reduced the number of treatment options available to patients and is associated with an increase in mortality and morbidity, with infections becoming increasingly more difficult to treat and treatments requiring extended periods of time (Reygaert et al., 2018). It is predicted that AMR will cause 10 million deaths per year by 2050 and cause a massive economic impact (Pierce et al., 2020).

Resistance can be classified into two types – natural and acquired resistance. Natural resistance refers to when a microbe has a trait shared amongst the individuals of the species that confer the microbe with the ability to resist the action of antimicrobial agents (Martinez et al., 2014; Cox and Wright, 2013; Reygaert et al., 2018). Acquired resistance refers to when a microbe acquires genetic material (either through mutations or through horizontal gene transfer) that confers the microbe with the ability to resist the action of antimicrobial agents (Reygaert et al., 2018). The following section will discuss the mechanisms used by microbes to gain resistance to antimicrobial agents.

1.2.2 Mechanisms of Antimicrobial Resistance

Antibiotics and antimicrobial agents kill or inhibit the growth of microbes using several mechanisms that include inhibiting cell wall synthesis, inhibiting protein, and nucleic acid synthesis, and by inhibiting metabolic pathways that are crucial for the survival of the microbe (Reygaert et al., 2018). Microbes have, over generations of antibiotic exposure, evolved with several interesting mechanisms to resist the effects of bacteria which can be divided into 4 types:

1. **Limiting uptake of the drug:** The presence of certain structures and mechanisms in the microbial cell can prevent the uptake of a drug. An example of such a mechanism is the presence of structures called lipopolysaccharides in gram-negative bacteria, whose structure and function provide a natural barrier for the bacteria to some drugs (Blair et al., 2014). In bacteria, hydrophilic molecules can use porin channels present in the cell membrane to enter the cell (Reygaert et al., 2018) as normally they cannot cross the cell membrane due to the hydrophobicity of the membrane (Blair et al., 2014). However, mutations in the genes that code for these porin channels or a decrease in the number of porin channels could decrease drug intake and hence allow the bacteria to resist the effects of antibiotics (Kumar and Schweizer, 2005). As a mechanism for carbapenems resistance, some bacteria have reduced the number of porin channels

produced in the cell, with the production of porin channels being stopped completely at some points in some cases (Cornaglia et al., 1996). Biofilm formation is another example of bacteria limiting the uptake of antimicrobials. In the case of pathogenic bacteria, the presence and formation of biofilms provide protection against antimicrobial agents as the antimicrobial agents cannot enter the bacteria cells (Reygaert et al., 2018).

2. **Inactivation/degradation of the drug:** The drug can be inactivated or broken down by the action of certain enzymes that are produced by the microbe (Reygaert et al., 2018). The most well-known example of this mechanism is the production of β-lactamases by certain microbes, which are hydrolyzing enzymes that can inactivate the β-lactams (Reygaert et al., 2018). Antimicrobial drugs can also be inactivated by the action of enzymes that transfer chemical groups. Acetylation is a commonly used mechanism to inactivate drugs like chloramphenicol. Phosphorylation and adenylation are also well-known chemical modifications that can inactivate drugs like aminoglycosides (Blair et al., 2015; Ramirez and Tolmasky, 2010; Robicsek et al., 2005).

3. **Efflux of drugs:** Efflux pumps can be expressed constitutively or are induced at high expression levels which result in the expulsion of the drug, preventing the drug from targeting the metabolic pathways of the microbial cell (Blair et al., 2014; Villagra et al., 2012). Efflux pumps pump out substances that are toxic to the bacterial cell (including antimicrobial agents) (Blair et al., 2014; Villagra et al., 2012). An example of this is the efflux protein MacB which works along with the proteins MacA and TolC to expel macrolide drugs. Another example is the efflux pump EmrB that works along with the proteins EmrA and TolC to expel nalidixic acid in *E. coli* (Tanabe et al., 2009; Jo et al., 2017).

4. **Modification of drug target:** The target of the drug can be modified in some manner (normally through mutations in the gene that codes for the target), which enables the microbe to resist the effects of the drug. An example of such a mechanism is the modification of ribosomal subunits (through mutations or methylation of the subunits) that results in the inability of the drugs (that normally target these ribosomal subunits) to bind to their targets (Kumar et al., 2013; Roberts, 2003; Roberts, 2004; Reygaert et al., 2018). Another example is mutations in DNA gyrase and topoisomerases that change the structure of these enzymes such that antimicrobial agents (that target these enzymes) are unable to bind effectively to these targets (Hawkey, 2003; Redgrave, Sutton, and Webber et al., 2014). Sulfonamides and trimethoprim (TMP) are drugs that target important metabolic pathways related to folic acid biosynthesis, which is a necessary cofactor to produce thymidine and purines (Masters et al., 2003). Mutations in the active site of the target proteins (TMP and sulfonamides are competitive inhibitors of the enzymes DHFR and DHFS respectively, which are enzymes involved in folic acid biosynthesis) lead to structural changes in the enzyme that prevent the effective binding of the drug to the enzyme target (Huovinen, Sundstrom, Swedberg et al., 1995; Vedantam, 1998; Reygaert et al., 2018).

Among these mechanisms of antimicrobial resistance, the mechanisms that involve the modification of drug targets usually involve the loss of binding of the antimicrobial agent to the active site by modifying the structure of the active site (Matange et al., 2018). Mutations in the drug target will lead to the modification of the active site of the enzyme, which in turn generates a new function in the form of antimicrobial resistance (Matange et al., 2018). Studying the effects of mutations that modify drug targets in relation to natural variation may be more fruitful as it is known that studying these mutations has provided insights into the relationship between protein stability and function (Shafer and Schapiro, 2008; Matange et al., 2018) and as mentioned previously, protein stability is one of the most common ways through which variation affects proteins (Bigman and Levy, 2018). Trimethoprim resistance has been studied previously in detail w.r.t to stability-function relationships (Matange et al., 2018; Bershtein et al., 2012; Bershtein et al., 2015). The advantages of using the DHFR-Trimethoprim system to study natural variation in the context of protein structure and function will be described in the following sub-section.

1.2.3 Trimethoprim and Mechanisms of Trimethoprim Resistance

Trimethoprim (TMP) is an anti-folate drug that is commonly used in the treatment of urinary tract infections (Desforges et al., 1993). TMP has been used in combination with sulfonamides to inhibit the growth of common urinary tract pathogens as it was thought that this combination of drugs was synergistic in vitro (Bushby and Hitchings, 1968). Trimethoprim inhibits the enzyme dihydrofolate reductase (DHFR), which is involved in folic acid synthesis. Folic acid is a necessary co-factor for the synthesis of nucleic acids and proteins (Masters et al., 2003).

Dihydrofolate reductase (DHFR) catalyzes the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate (Fierke et al., 1987). Tetrahydrofolate is important for several metabolic pathways such as nucleic acid synthesis and methylation (Askari and Krajniovic, 2010). The reaction mechanism, transformation of the reactants to products, and the role of TMP are depicted in Fig 3. DHFR enzymes have two main sub-domains: the adenosine binding domain (to which NADPH binds) and the loop domain, which contains three loops namely the M20 loop, the F–G loop, and the G–H loop (Schnell et al., 2004). The movement of the M20 loop determines the ability of the substrate to enter the active site of DHFR (Shrimpton and Allemann, 2002).

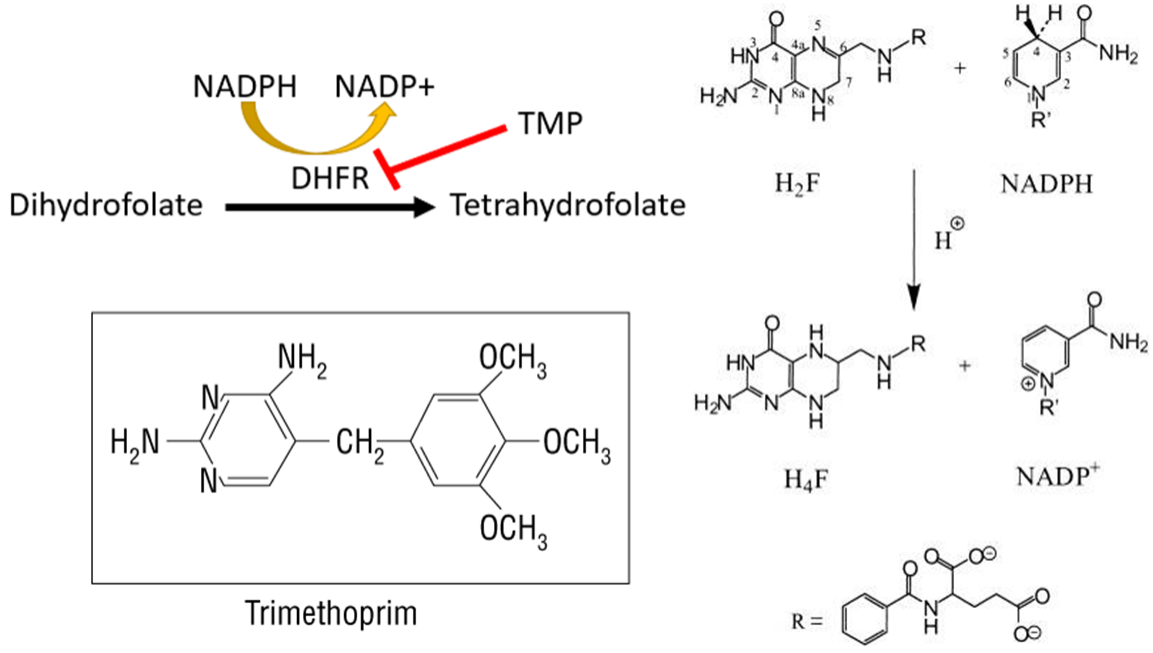


Figure 3: **The reaction mechanism of dihydrofolate reductase and the involvement of TMP in inhibiting DHFR function.** The reaction mechanism is shown in the upper left corner while the structure of TMP is shown in the lower left corner. The right side depicts the transformation of the reactants of DHFR reduction into products. H_2F = Dihydrofolic acid, H_4F = tetrahydrofolic acid. Structures of TMP and the figure from the right side were adapted from Masters et al., 2003 and Shrimpton and Allemann, 2002 respectively.

The DHFR protein is dynamic, in that it undergoes changes in conformation to perform its function (Sawaya and Kraut, 1997). Studies with *E. coli* DHFR have shown that the M20 loop adopts 4 conformations: open, closed, disordered, and occluded (Sawaya and Kraut, 1997). The disordered loop conformation is thought to exist due to time-averaged fluctuations between the closed and the occluded conformation (Sawaya and Kraut, 1997). The open loop conformation exists only in certain crystal structures and is stabilized by certain contacts in the crystal structure (Sawaya and Kraut, 1997). DHFR exists in the occluded loop conformation when only the substrate site of the protein is occupied. When the cofactor binds to DHFR, it leads to the protein existing in the closed loop conformation where the active site is closed and is protected from the solvent (Sawaya and Kraut, 1997; Schnell et al., 2004). The closed conformation allows the co-factor and substrate to be in close proximity such that the reduction reaction could occur (Sawaya and Kraut, 1997; Schnell et al., 2004). The closed and occluded loop conformations of DHFR differ in the identity of hydrogen bonds formed between the M20 loop and the F-G and G-H loop respectively. The hydrogen bonds that are formed in the two conformations and the conformation in which the DHFR protein exists at each step of the reaction are shown in Fig 4.

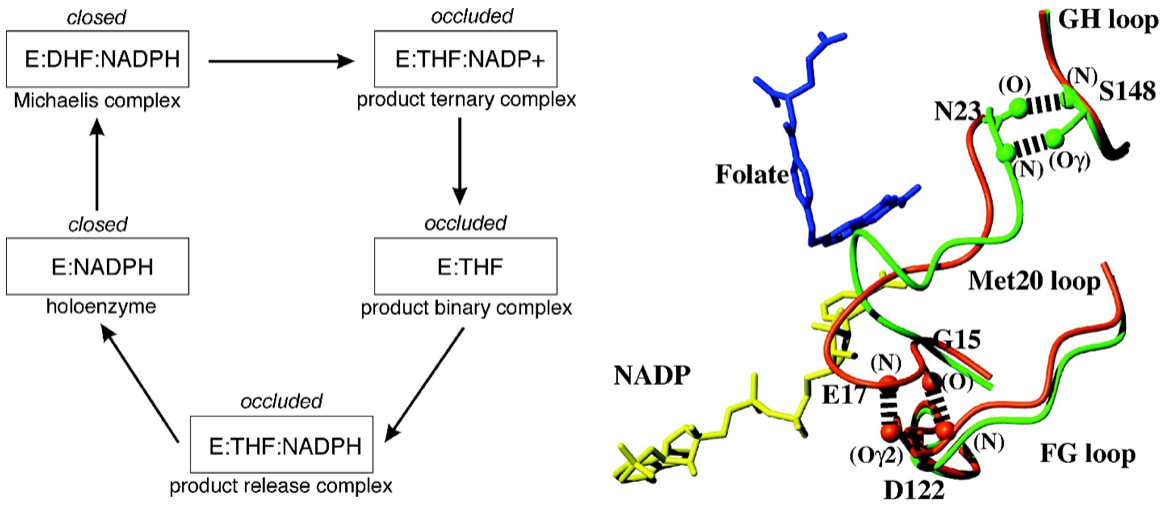


Figure 4: T**he detailed mechanism of DHFR catalysis with changes in conformation at every step**. The left image depicts the mechanism of reduction of DHFR, with the conformation of the protein mentioned for each intermediate. The right image depicts the possible conformations that DHFR can take during the reaction. The red color line depicts the closed conformation while the green color line depicts the occluded conformation. Adapted from Schnell et al., 2004

Trimethoprim structurally mimics the pteridine ring of dihydrofolic acid (which is the natural substrate of DHFR) and hence TMP functions by competing with dihydrofolic acid for binding to the active site of DHFR. This results in competitive inhibition and hence affects the reduction of dihydrofolic acid to tetrahydrofolic acid (Masters et al., 2003). Hence TMP, along with sulfonamides (which target the enzyme DHFS that functions upstream of DHFR in the folate biosynthesis pathway) are very effective when used together due to this sequential blockade of enzymes in the same metabolic pathway (Masters et al., 2003).

DHFR is an enzyme that is ubiquitously produced by all cellular organisms (Schnell et al., 2004). However, TMP binds more effectively to bacterial DHFRs compared to eukaryotic DHFRs (Matthews et al., 1985). A possible reason for this has been studied by comparing the DHFR protein in chickens and in *E. coli*. In chicken DHFR, it was discovered that the residues on the opposite side of the active site region were farther apart by about 2Å compared to the structurally equivalent residues in *E. coli* DHFR (Matthews et al., 1985). The difference in binding affinity to TMP between *E. coli* DHFR and chicken DHFR was attributed to the potential loss of a hydrogen bond between the backbone carbonyl group of Val115 and TMP in chicken DHFR (Matthews et al., 1985).

1.3 Aim of the Project

This project focuses on the study of sequence variation in homologous proteins and how it impacts the structure and hence function of these proteins in the context of the *E. coli* dihydrofolate reductase (DHFR) system. DHFR is a small globular protein and as a result, DHFRs from several bacteria have been characterized structurally and functionally. DHFR is competitively inhibited by the antibiotic trimethoprim, a routinely used antibiotic for treating bacterial infections of the urinary tract, making DHFR an appealing system for phenotypic studies (Cao et al, 2018, Matange et al, 2018). First, the impact of interspecific sequence variation in DHFR between different bacteria, intraspecific sequence variation between different strains of *E. coli*, and mutations (single and in combination) in DHFR that are selected by trimethoprim exposure have been collated. Second, structural analysis using crystal structures of DHFR from different bacteria has been performed to predict how effective trimethoprim would be against different sequence variants of DHFR. Finally, the consequences of sequence variation in DHFR have been tested in the context of trimethoprim resistance by mutagenesis and phenotypic characterization in *E. coli*.

2 Materials and Methods

2.1 Materials

2.1.1 Consumables and Equipment

Glassware and beakers were obtained from Borosil Glass Works Ltd. (India) and Schott Duran (Germany). Plasticware was obtained from Tarsons Products Pvt. Ltd. (India) and HiMedia. Sterile 96-well plates were obtained from HiMedia. Consumables like Parafilm, pipette tips, and aluminum fold were obtained from Tarsons Products Pvt. Ltd, Corning Life Sciences, and FoilPlus, respectively. 1.5 mL Eppendorf tubes were obtained from Eppendorf. Pipettes were obtained from Gilson.

2.1.2 Instruments Used

Instrument	Usage	System Details
GeneBox	Agarose Gel imaging, Chemiluminescence imaging of Western Blot	GeneSys
Incubator (with shaking facility)	Incubation of bacterial cells at 37°C	New Brunswick Innova 42
Laminar Air Flow Cabinet	Microbiology work (inoculation, 96 well plate preparation for broth dilution, transformation, and competent cell preparation)	MicroFilt India
Refrigerators (-80°C, -20°C, 4°C)	Storage of glycerol stocks and competent cells, plasmid, primers, antibiotics, and other chemicals	Panasonic
96 Well Plate Reader	Measurement of Optical Density of cells in 96 well plates for broth dilution	Ensight by Perkin Elmer
Thermal Cycler	PCR reactions	Eppendorf Master cycler x50s
Mini-centrifuge	Plasmid isolation and spinning down cultures	Minispin by Eppendorf
Thermal Cycler	PCR reactions	Eppendorf vapo.protect
Biosafety cabinet	Microbiology work (inoculation, 96 well plate preparation for broth dilution, transformation, and competent cell preparation)	Krew instruments
Centrifuge	Centrifugation of 50mL falcon tubes for competent cell preparation	Eppendorf Centrifuge 5810 R
Thermomixer	Heat shock of cells for transformation	Eppendorf Thermomixer C
Spectrophotometer	Verifying purity of plasmid DNA	Thermo scientific NANODROP 2000c

Table 1: **List of instruments used in this study**.

2.1.3 Chemicals

HiMedia	Sigma Lifesciences	MPBio	Others	Present in Labarotory stock
LB powder	CH_3CH_2OH	Tris Chloride	H_2O_2 (Millipore)	Whatman filter paper grade 3
LA powder	$CaCl_2$	Tris (Free Base)	Luminol substrate (Millipore)	Ammonium Persulfate
Ethidium Bromide	DMSO	Glycine	Glycerol (Invitrogen)	tetramethylethylene-diamine (TEMED)
NaCl	Isopropanol		SDS (Merck)	Tween 20
NaOH	EDTA			Bromophenol Blue
CH_3COOH				β-mercaptoethanol
				Acrylamide
				Bisacrylamide
				PVDF membrane (Imolbilon)

Table 2: **Chemicals used in this study.**

2.1.4 Antibiotics Used

Antibiotic	Source	Catalogue No	Solvent
Ampicillin	MP Bio	Cat#194526	Autoclaved Milli-Q
Trimethoprim	Sigma Lifesciences	Cat#T7883	DMSO

Table 3: **Antibiotics used in this study.**

2.1.5 DNA Ladders/markers used

Name of Ladder/Marker	Range	Source
Takara 1kb DNA ladder	1-10kb	DSS Takara Biosciences
Invitrogen 1kb+ DNA ladder	100bp-15kb	Invitrogen
Agilent 1kb DNA ladder	250bp - 10kb	Agilent
Gel Loading Dye Purple	nil	New England Biolabs

Table 4: **DNA ladders and markers used in this study.**

2.1.6 Protein Markers used

Name of Ladder/Marker	Range	Source
Precision Plus Protein Dual Color Standard	10-250 kDa	Bio-Rad
Precision Plus Protein Unstained Protein Standard	10-250 kDa	Bio-Rad

Table 5: **Protein markers used in this study.**

2.1.7 Enzymes

All enzymes and buffers were obtained from New England Biolabs. Buffers were selected such that all enzymes in the reaction would have the maximum possible activity.

Name of Enzyme	Catalogue No
*Ava*I	R0152S
*Bgl*II	R0144S
*Bsu*36I	R0524S
*Dpn*I	R0176S
*Hind*III	R0104S
*Nae*I	R0190S
*Nco*I	R0193S
*Sal*I	R0138S

Table 6: **Restriction enzymes used in this study**

RNase A (Cat#10109142001) was obtained from Sigma Lifesciences. The primary antibody (Polyclonal anti-DHFR) and the secondary antibody (Anti-rabbit HRP) were prepared in-house. PRIMESTAR MAX PCR (Cat#R045B) mix was obtained from Takara Biosciences.

2.1.8 Primers and Plasmids

All plasmids used in this study, which include pPRO-His-*folA*, pPRO-His-*folA* W30R, pPRO-His-*folA* P21L, pPRO-His-*folA* F153A, and pPRO-His-*folA* I94L were obtained from Dr. Nishad Matange.

Primer Name	Primer Sequence (5' - 3')	Primer Application
folA_W47R_fwd_SalI	CGCCATACCCGCGAATCAATCG GTCGACCGTTGCCAGGACGCAAA	Forward Primer to induce the mutation W47R in DHFR
folA_P105A_fwd_HindIII	ACAGTTCTTGGCTAAAGCGC AAAAGCTTTATCTGACGCATATCG	Forward Primer to induce the mutation P105A in DHFR
folA_L28Q_fwd_NaeI	GGAACCTGCCGGCCGA TCAAGCCTGGTTTAAACGC	Forward Primer to induce the mutation L28Q in DHFR
folA_L28F_fwd_NaeI	GGAACCTGCCGGCCGA TTTTGCCTGGTTTAAACGC	Forward Primer to induce the mutation L28F in DHFR
folA_L28K_fwd_NaeI	GGAACCTGCCGGCCGA TAAAGCCTGGTTTAAACGC	Forward Primer to induce the mutation L28K in DHFR
folA_Y151D_fwd_BglII	CTCTCACAGCGATTGCTTTG AGATCTTGGAGCGGCGGGGCC	Forward Primer to induce the mutation Y151D in DHFR
folA_Y151L_fwd_BglII	TCTCACAGCTTATGCTTTG AGATCTTGGAGCGGCGGGGCC	Forward Primer to induce the mutation Y151L in DHFR
folA_Y151F_fwd_BglII	CTCACAGCTTTTGCTTTG AGATCTTGGAGCGGCGGGGCC	Forward Primer to induce the mutation Y151F in DHFR
folA_W47R_rev_*Sal*I	CTGGCAACGGTCGACCGATT GATTCGCGGGTATGGCGGCCCATA	Reverse Primer to induce the mutation W47R in DHFR
folA_P105A_rev_HindIII	GCGTCAGATAAAGCTTTTGCG CTTTAGCCAAGAACTGTTCATAAAC	Reverse Primer to induce the mutation P105A in DHFR
folA_L28Q_rev_NaeI	TAAACCAGGCTTGATCGG CCGGCAGGTTCCACGGCATGGC	Reverse Primer to induce the mutation L28Q in DHFR
folA_L28F_rev_NaeI	TAAACCAGGCAAAATCGG CCGGCAGGTTCCACGGCATGGC	Reverse Primer to induce the mutation L28F in DHFR
folA_L28K_rev_NaeI	TAAACCAGGCTTTATCGG CCGGCAGGTTCCACGGCATGGC	Reverse Primer to induce the mutation L28K in DHFR
folA_Y151D_rev_BglII	CGCCGCTCCAAGATCTCAA AGCAATCGCTGTGAGAGTTCTGCG	Reverse Primer to induce the mutation Y151D in DHFR
folA_Y151L_rev_BglII	CGCCGCTCCAAGATCTCAA AGCATAAGCTGTGAGAGTTCTGCG	Reverse Primer to induce the mutation Y151L in DHFR
folA_Y151F_rev_BglII	GCCGCTCCAAGATCTCAAA GCAAAAGCTGTGAGAGTTCTGCG	Reverse Primer to induce the mutation Y151F in DHFR
Seq_rev	GATTTAATCTGTATCAGG	Reverse Sequencing primer for *folA* gene in the pPRO-His-*folA* plasmid

Table 7: **Primers used in this study and the purpose of use of each primer.**

2.1.9 Composition of Buffers and Solutions

Alkaline lysis Solution 1

Contains 25 mM Tris-Cl (pH 8.0) and 10 mM EDTA (pH 8.0).

Alkaline lysis Solution 2

Contains 1% weight by volume SDS along with 0.2 N NaOH.

Alkaline lysis Solution 3

Contains 11.5mL glacial acetic acid and 60mL of 5M potassium acetate dissolved in 100mL Milli-Q.

CaCl2 (0.1M)

Contains 1.1g of CaCl2 dissolved in 100mL of autoclaved Milli-Q water.

Laemmli Buffer (4X)

Contains 4mL glycerol (100%), 0.8g SDS, 2mL of 1M Tris-HCl (pH 6.8), 0.1g bromophenol blue, and 50µL of ß-mercaptoethanol dissolved in 1mL Milli-Q.

Resolving Buffer

Contains 1.5M Tris-HCl (pH 8.8).

Stacking Buffer

Contains 1.5M Tris-HCl (pH 6.8).

TAE (50X)

Contains 57.1mL of glacial acetic acid, 242g of tris-base, and 100 mL of 0.5M EDTA (pH 8) dissolved in 1L purified water.

TE (1X)

Contains Tris-Cl of desired pH with a concentration of 10mM and EDTA (pH 8.0) with a concentration of 1mM.

TBST

Contains 0.9% NaCl, 5mL of 2M tris-HCl (pH 7.5), and 0.1% Tween-20 dissolved in 1L of purified water.

TGM

Contains 14g of Glycine, 3g of Tris-base, and 10% ethanol dissolved in 1L purified water.

2.2 Methods

2.2.1 Preparation of Media and Culture Conditions

Preparation of Luria-Bertani Broth (LB) LB was prepared by adding 2.5g of Luria Bertani Broth, Miller (HiMedia ref: GM1245-500G) in 100mL purified water in a 250mL glass flask. The mixture was autoclaved at 15 psi pressure for 20 mins at 121°C.

Preparation of Luria-Bertani Agar (LA) LA was prepared by adding 8g of Luria Bertani Agar, Miller (HiMedia ref: GM151-500G) in 200mL purified water in a 500mL glass flask. The mixture was autoclaved at 15 psi pressure for 20 mins at 121°C.

Strain and culture conditions *E. coli* K-12 MG1655 was used as the wild-type strain for all experiments. *E. coli* DH10B strain was used for the manipulation of plasmid DNA. LB was used to culture the bacteria and LA was used to plate and streak bacteria. The cultures were grown at 37°C with shaking at 180 rpm. The plasmids pPRO-His-*folA*, pPRO-His-*folA*-W30R, pPRO-His-*folA*-P21L, pPRO-His-*folA*-F153A, and pPRO-His-*folA*-I94L were obtained from Dr. Nishad Matange and were used for site-directed mutagenesis. Chemicals were purchased from HiMedia, MP Chemicals, and Sigma Lifesciences. Restriction enzymes were purchased from New England Biolabs. Solvents used in the experiments were obtained from the Bio store in IISER Pune. The Polyclonal rabbit anti-DHFR antibody and the Goat anti-rabbit Horseradish peroxidase (HRP) antibody were obtained from Dr. Nishad Matange. Glycerol stocks were prepared by adding 0.5mL of 50% glycerol to 0.5mL of saturated culture and stored at -80 $^{\circ}$ C.

2.2.2 Sequence and Structural Analysis

Sequence and Structural Alignments All protein sequence alignments were performed using the ClustalW algorithm using the software Mega-X (Kumar et al., 2018). Sequences of *E. coli* and *S. enterica* strains were obtained from the culture collection of the NCTC 3000 project (NCTC 3000 Project. (n.d.).). Variations were identified keeping *E. coli* MG1655 and *S. enterica* (GenBank ID: SUG98162.1) as references. Multiple Sequence Alignment (MSA) of sequences belonging to different *E. coli* strains was performed by Chetna Yelpure. For MSA with *S. enterica* strains, sequences with either a part of the sequence missing (compared to the other sequences) or with extra residues were removed from the final alignment. Structural alignments were performed in PyMol (PyMOL 2.0) using the align command. Visualization of protein structures and measurement of distances between residues in proteins were performed in PyMol.

Protein stability predictions The $\Delta\Delta G$ values for specific mutations were predicted using the prediction tools I-Mutant 2.0 (Capriotti et al., 2005) and Site-Directed Mutator (SDM) (Topham et al., 1997). In I-Mutant 2.0, all predictions were performed with temperatures at 37°C and pH at 7. The predicted values were plotted in the form of a heatmap using R. The amino acids were arranged in the order of increasing hydrophilicity. The order of hydrophilic residues was determined from the interface scale (Wimley and White, 1996).

Plotting, T-tests, correlation analysis, and other analysis All plotting, T-tests, and correlation analysis were performed using the R programming language.

2.2.3 Site-Directed Mutagenesis

The protocol for SDM was adapted from Shenoy and Visweswariah, 2003. In brief, the plasmid containing the *folA* gene was mutagenized in a PCR using mutagenic primers. The PCR mix was then transformed into *E. coli* DH10B competent cells after which colonies were picked and the presence of the mutation was confirmed using restriction digestion. Those plasmid isolates that had the desired mutations were further sent for sequencing for confirmation of the presence of the mutation and to ensure no miscellaneous mutations have occurred. The protocol has been briefly described in Fig 5.

Polymerase Chain Reaction (PCR) A control and a test PCR reaction were set up for every mutant. The control reaction was set up by adding 0.5µL of template plasmid DNA, 1µL of 10µM forward mutagenic primer, 1µL of reverse mutagenic primer, and 17.5µL of autoclaved Milli-Q. The test reaction was set up by adding 0.5µL of template plasmid DNA, 1µL of 10µM forward primer, 1µL of reverse primer, 10µL of PRIMESTAR MAX PCR mix (xxxx), and 7.5µL of autoclaved Milli-Q. All PCR reactions were set up with the same conditions, which are as follows: initial denaturation at 95°C for 2 mins - 35 cycles of 95°C for 30 seconds, 45°C for 15 seconds, 72°C for 2.5 mins - final extension at 72°C for 5 mins. A part (2µL) of the control and test PCR mixture was run in an agarose gel electrophoresis experiment to confirm that the PCR had taken place and sufficient amplified DNA was present. To the control and test solutions, 0.5µL of *Dpn*I was added and the solutions were incubated overnight at 37°C to digest hemimethylated and methylated DNA.

Preparation of competent cells To inoculate the bacterial strain, 10µL of the glycerol stock of the bacterial strain (either *E. coli* MG1655 or *E. coli* DH10B) was added into 3mL LB. The culture was incubated overnight at 37°C with shaking at 180 rpm. 1% of the overnight culture was inoculated into 100 LB and was incubated for 1.5 hours at 37°C with shaking at 180 rpm. The culture was then transferred into two 50mL falcon tubes and placed on ice. The tubes were centrifuged for 10 mins at 3000 RCF at 4°C. The supernatant was discarded and the pellet was resuspended in 20mL chilled 0.1M $CaCl_2$. The tubes were then incubated on ice for 15 mins, then centrifuged again for 10 mins at 3000 RCF at 4°C. The supernatant was discarded and the pellet was washed twice with 10mL chilled 0.1M $CaCl_2$. The supernatant was then discarded and the pellet was resuspended into 1mL of 0.1M $CaCl_2$

dissolved in 15% glycerol. The resultant suspension was aliquoted into 1.5mL tubes such that each tube holds 50µL of the suspension. The cells were then frozen and stored at -80°C.

Transformation of plasmid DNA into competent cells Competent cells and plasmid DNA were thawed on ice. To transform the competent cells, 9µL of SDM (PCR) products were added to a 1.5mL tube containing competent cells (50µL) and the mixture was incubated on ice for 15 mins. The mixture was then subjected to a heat shock by keeping the tube containing the mixture at 42°C for 90 seconds. The tube was then placed back on ice. To allow the cells to recover, 1mL of LB was added to the mixture and the mixture was incubated at 37°C with shaking at 180 rpm for 1.5 hours. The tube was centrifuged for 1 min at 13000 rpm to pellet down the cells. The supernatant was discarded and the pellet was resuspended in the LB remaining in the tube. The remaining mixture was plated on LA plates supplemented with 100µg/mL ampicillin. The plate is then incubated at 37°C overnight.

Isolation of Plasmid DNA A single bacterial colony after transformation was picked up and inoculated into 3mL LB and was incubated overnight at 37°C with shaking at 180 rpm. The overnight cultures were transferred into 1.5mL tubes and the culture was centrifuged for 1 min at 13000 rpm to pellet down the cells. The cells were resuspended into 300µL of alkaline lysis solution 1 and mixed until the solution had a uniform suspension. To this, 300µL of alkaline lysis solution 2 was added and mixed gently. After 1 min, 300µL of alkaline lysis solution 3 was added and mixed gently. 300µL of chloroform was added and centrifuged for 3 min at 13000 rpm. The aqueous phase was transferred to a fresh 1.5mL tube and 650µL of isopropanol was added. The solution was mixed well and centrifuged for 3 min at 13000 rpm to pellet down plasmid DNA and RNA. The supernatant was discarded and the pellet was washed using 0.5mL chilled 70% ethanol. The solution was centrifuged for 1 min at 13000 rpm and the pellet was dried and resuspended in 30µL TE buffer supplemented with 5µL of 20mg/mL RNase A.

Isolation of Plasmid DNA for sequencing A single bacterial colony after transformation was picked up and inoculated into 3mL LB and was incubated overnight at 37°C with shaking at 180 rpm. The overnight cultures were transferred into 1.5mL tubes and the culture was centrifuged for 1 min at 13000 rpm to pellet down the cells. The cells were resuspended into 300µL of alkaline lysis solution 1 and mixed until the solution had a uniform suspension. To this, 300µL of alkaline lysis solution 2 was added and mixed gently. After 1 min, 300µL of alkaline lysis solution 3 was added and mixed gently. The mixture was centrifuged and the aqueous supernatant was transferred to a fresh 1.5 mL tube. To this, 2µL of 20mg/mL RNase A was added to each sample and was incubated at 37°C for 3 hours. After the incubation, 300µL of chloroform was added and centrifuged for 3 min at 13000 rpm. The aqueous phase was transferred to a fresh 1.5mL tube and 650µL of isopropanol was added. The solution was mixed well and centrifuged for 3 min at 13000 rpm to pellet down plasmid DNA and RNA. The supernatant was discarded and the pellet was washed

using 0.5mL chilled 70% ethanol. The solution was centrifuged for 1 min at 13000 rpm and the pellet was dried and resuspended in 30µL TE buffer.

Agarose Gel Electrophoresis 1% agarose solution was prepared using 1x TAE buffer as solvent. The solution was boiled and was allowed to cool down and solidify after supplementing the solution with 1mg/mL of ethidium bromide. Gel electrophoresis was performed by supplying a voltage of 150V. A gel loading buffer was added to the samples and the samples were subjected to gel electrophoresis along with a DNA ladder until a sufficient resolution was achieved.

Restriction Digestion The reaction mixture was created by adding 2µL of restriction enzyme buffer, 0.5µL of each restriction enzyme, and 3µL of plasmid DNA, and autoclaved Milli-Q was added to make up the total volume of the solution to be 20µL. This solution was incubated at 37°C for 3 hours. The samples were subjected to agarose gel electrophoresis and the gel was imaged to confirm the presence of the corresponding restriction enzyme within the plasmid DNA.

Preparation for sequencing Two of the plasmid DNA samples with a confirmed restriction site for every mutation were transformed into *E. coli* DH10B competent cells. To transform the competent cells with plasmid DNA that were confirmed to contain the required mutations (through restriction digestion), 1µL of the plasmid DNA was used to transform 25µL of DH10B competent cells. After transformation, 2 colonies for every sample DNA were picked and cultured overnight in LB at 37°C with shaking at 180 rpm. Plasmid DNA was extracted from these cultures using either of two ways:

1. By using the HiMedia HiPurA Plasmid DNA Miniprep Purification kit.
2. By following an alternate plasmid isolation protocol (refer to the isolation of plasmid DNA for sequencing).

The purified plasmid DNA was sent for Sanger sequencing to Barcode Biosciences. The sequence obtained after sequencing was aligned with the pPRO-His-*folA* sequence to confirm that the expected mutations have been generated. The data for all mutations generated in this study along with the template plasmid DNA used, mutagenic primers used, and the restriction site introduced for every mutation is summarized in Appendix 1.

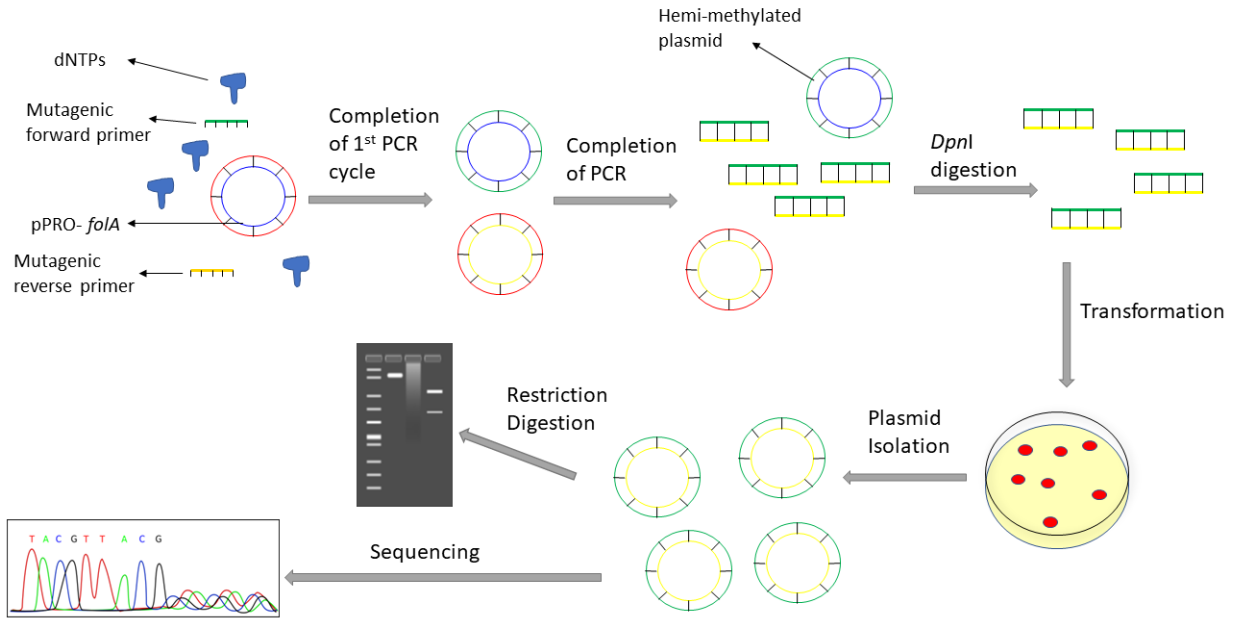


Figure 5: **Schematic depicting the various steps involved in site-directed mutagenesis.**

2.2.4 Broth Dilution to calculate IC50

Preparation of 96-well plate The preparation of the 96-well plate is depicted in Fig 6. In a 96-well plate, the peripheral wells were filled with autoclaved Milli-Q. LB was supplemented with ampicillin such that the final concentration of ampicillin comes to 100µg/mL. The LB supplemented with ampicillin was used for all subsequent steps. The wells were filled by adding 135µL of LB, except the wells belonging to the 2nd column and the 2nd, 4th, and 6th rows of the plate (B2, D2, and F2 respectively). To the aforementioned wells, either 255µL of LB (maximum TMP concentration = 500µg/mL) followed by 15µL of 10mg/mL of TMP or 240µL of LB (maximum TMP concentration = 2mg/mL) followed by 30µL of 20mg/mL TMP was added such that the total volume is 270µL. The wells were mixed by using a pipette to dissolve TMP in LB. From the wells labeled as B2, D2, or F2, 135µL of the solution was picked up from this well and transferred to the adjacent well such that the adjacent well now had a volume of 270µL. This step (serial dilution) was repeated (each well was mixed using a pipette after every transfer) to all the wells excluding those belonging to the 11th column of the plate (which serve as drug-free controls). The serial dilutions were performed such that each serial dilution was performed over 2 rows of wells.

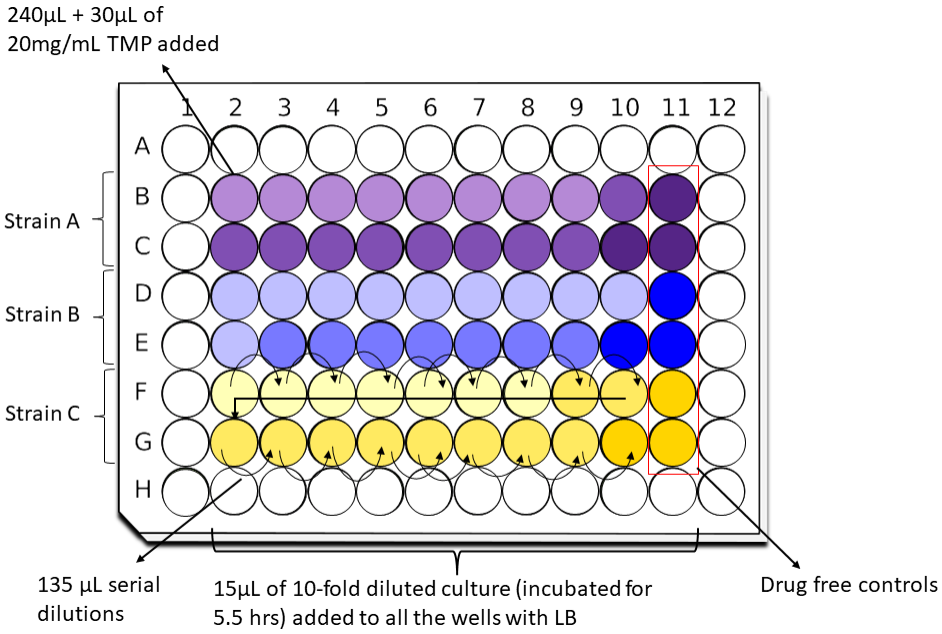


Figure 6: **Schematic depicting preparation of the 96-well plate for broth dilution.**

Growth and measurement of IC50 of bacterial strain with respect to trimethoprim To inoculate the bacterial strain, 5-10µL of the glycerol stock of a bacteria strain was added into 1mL LB and 37°C with shaking at 180 rpm for 5.5 hours (until saturation). A 10-fold dilution of the saturated culture was prepared by adding 100µL of the saturated culture to 900µL of LB supplemented with 100µg/mL ampicillin in a 1.5 mL tube. To each of the wells (excluding the outer wells filled with Milli-Q), 15µL of the diluted culture was added. The 96-well plate was then covered with aluminum foil and was put in a ziplock bag after which the plate was incubated for 15 hours at 37°C. After 15 hours, the plate was unwrapped and the OD600 of each well was measured.

Calculation of IC50 values The OD600 values of the 2 wells (drug-free controls) were averaged and were used to normalize the OD600 values for the remaining wells. The OD600 values of wells containing TMP were plotted against the log values of the antibiotic concentrations. A non-linear regression (4 parametric inhibition response curves) was fitted to the data to calculate IC50 values. The regression curve used was adapted from GraphPad version 9. The plotting and fitting of the data to the curve was performed using R.

2.2.5 Western Blot Analysis

Preparation of cell lysates To inoculate the bacterial strain, 10µL of glycerol stock of a bacterial strain was added into 1mL LB supplemented with 100µg/mL ampicillin and was grown overnight at 37°C with shaking at 180 rpm. 1% of the overnight culture was inoculated into 3mL LB supplemented with 100µg/mL ampicillin and was incubated at 37°C

with shaking at 180 rpm for 3 hours. This culture was then centrifuged at 13000 rpm for 1 min to pellet down the cells. The supernatant was discarded and 25µL of 4X Laemmli buffer and 75µL of Milli-Q filtered water were added. The mixture was then boiled at 95°C for 5 mins after which the mixture was allowed to cool down. The lysate was stored at -20°C and the lysate was boiled at 95°C for 5 mins every time the lysate was thawed after freezing.

SDS-PAGE, Transfer to PVDF membranes, and processing of the blot The gel was prepared such that the gel contained a 15% resolving gel with a 5% stacking gel at the top. The samples were loaded into the gel such that 5µL of the lysate was added into each well along with the Precision Plus Dual Color Protein Standard protein ladder. The gel was run in TGS buffer at a constant voltage of 150V until the dye ran out of the gel. The transfer cassette was placed with the negative side facing downward. The following materials were added in a specific manner as follows: A sponge was placed followed by a Whatman grade 3 filter paper. The polyacrylamide gel was placed on top of the filter paper. The PVDF membrane was activated using methanol and was then placed on top of the gel. A Whatman grade 3 filter paper was added on top followed by another sponge. The transfer cassette was placed into the transfer apparatus in the presence of ice-cold TGM. The transfer occurred at a constant current of 200mA for 2 hours. After the transfer, the membrane was blocked using a solution of TBST supplemented with 5% BSA. The membrane was then washed with TBST. The membrane was incubated overnight at 4°C in the presence of 10mL TBST + 0.5% BSA supplemented with 100ng/mL of anti-DHFR polyclonal primary antibody. The membrane was washed 3 times with TBST. The membrane was incubated in the presence of 10mL TBST + 0.5% BSA supplemented with the secondary antibody (diluted to a ratio of 1:10000). The membrane was washed again thrice using TBST. The blot was developed by using H_2O_2 and the chemiluminescence substrate in a 1:1 ratio and adding it to the membrane. The membrane was incubated with H_2O_2 and the chemiluminescence substrate for a minute and then was visualized in a gel documentation system with readings taken every 30 seconds for 5 mins.

3 Results and Discussion

3.1 Consequences of Interspecific Variation on Protein Structure and Function

In this project, interspecific and intraspecific sequence variations were analyzed to understand how it impacts protein structure and function.

To study interspecific variation between different organisms containing the DHFR enzyme, the structures of DHFR belonging to 15 different species were obtained from the PDB database. A multiple sequence alignment (MSA) was performed using the protein sequences of the DHFR enzymes belonging to the different species and the residues previously reported to be associated with TMP resistance in *E. coli* were compared. This MSA was curated using the structures of the DHFR enzymes. In order to curate the alignment, each of the structures of the sequences in the MSA was structurally aligned to the structure of *E. coli* DHFR bound to trimethoprim (TMP) (PDB ID:7MYM) and the residues associated with TMP resistance in *E. coli* were compared to their equivalent residues in the DHFR proteins belonging to the other species. The results of the above analysis are summarized in Table 8. A plot showing the frequencies of each residue at the positions equivalent to those residues that have been previously reported to be associated with TMP resistance was constructed and is shown in Fig 7.

PDB code	Organism	Residue										
		W30RGCY	A26STV	D27E	F153ALSV	I94L	L28R	P21LQ	R98P	Y151D	I5F	M20I
7MYM	*E. coli*	W	A	D	F	I	L	P	R	Y	I	M
2W9H	*Staphylococcus aureus*	H (31)	N (27)	D (28)	F (152)	F (93)	L (29)	P (22)	T (97)	H (150)	L (6)	L (21)
1DG5	*Mycobacterium tuberculosis*	H (30)	E (26)	D (27)	L (153)	I (94)	Q (28)	P (21)	Q (98)	Y (151)	I (5)	I (20)
3FL9	*Bacillus anthracis*	Y (31)	S (27)	E (28)	Y (155)	F (96)	L (29)	P (22)	Q (100)	Y (153)	M (6)	L (21)
2HM9	*Lacticaseibacillus casei*	Y (29)	D (25)	D (26)	Y (155)	A (97)	L (27)	P (20)	Q (101)	H (153)	L (4)	L (19)
3IA4	*Moritella profunda*	L (31)	A (27)	E (28)	F (155)	I (96)	L (29)	P (22)	T (100)	Y (153)	I (6)	M (21)
7MYL	*Klebsiella pneumoniae*	L (31)	G (27)	E (28)	Y (151)	S (97)	Q (29)	P (22)	E (101)	Y (149)	M (6)	I (21)
4M7V	*Enterococcus faecalis*	F (38)	N (34)	D (35)	Y (164)	G (105)	L (36)	P (29)	V (109)	H (162)	I (13)	L (28)
2W3W	*Mycobacterium avium*	R (34)	E (30)	D (31)	F (161)	I (102)	L (32)	P (25)	Q (106)	Y (159)	V (9)	I (24)
7K6C	*Mycobacterium abscessus*	R (41)	E (37)	D (38)	F (163)	I (107)	Q (39)	P (32)	E (111)	Y (161)	I (16)	I (31)
7RZO	*Stenotrophomonas maltophilia*	H (39)	D (35)	D (36)	F (165)	I (106)	F (37)	P (30)	E (110)	F (163)	I (14)	M (29)
1VDR	*Haloferax volcanii*	Q (32)	A (28)	D (29)	L (152)	I (100)	K (30)	P (22)	A (104)	F (150)	V (6)	L (21)
1ZDR	*Bacillus stearothermophilus*	Y (30)	A (26)	D (27)	F (155)	I (96)	L (28)	P (21)	E (100)	H (153)	I (5)	L (20)
4KD7	*Homo sapiens*	Y (33)	N (29)	E (30)	F (179)	V (115)	F (31)	P (23)	S (119)	Y (177)	I (7)	L (22)
3QLW	*Candida albicans*	Y (35)	K (31)	E (32)	Y (186)	I (112)	I (33)	P (26)	E (116)	Y (184)	V (10)	M (25)

Table 8: **Variation at positions associated with TMP resistance in *E. coli*.** The mutations mentioned at the top of every column in the residues section of the table are mutation associated with TMP resistance.

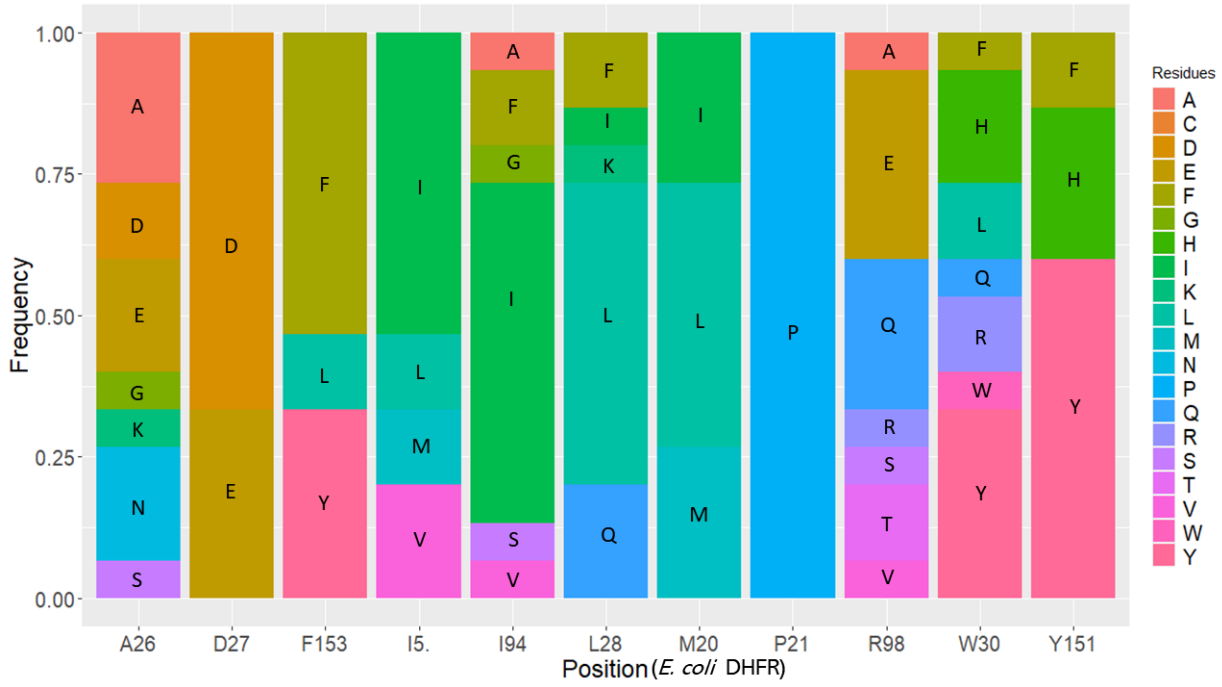
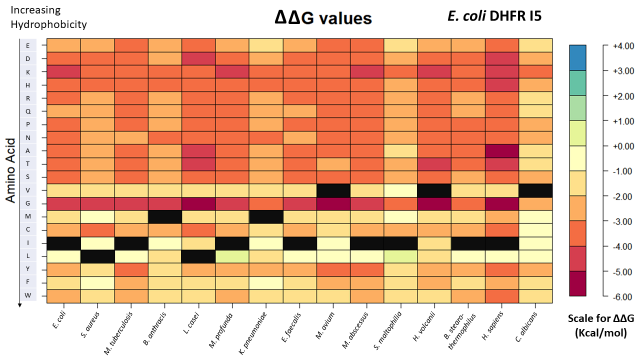


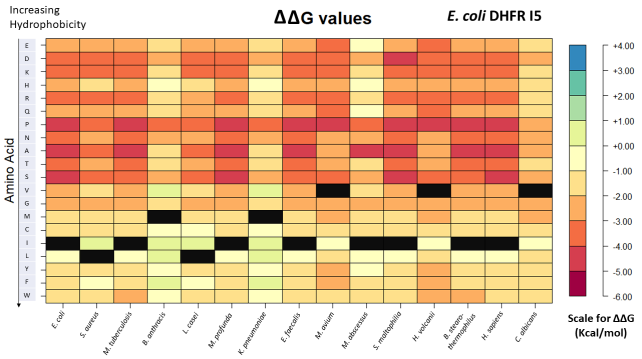
Figure 7: **Frequency Bar plot depicting variation at positions associated with TMP resistance in *E. coli* and their frequencies.**

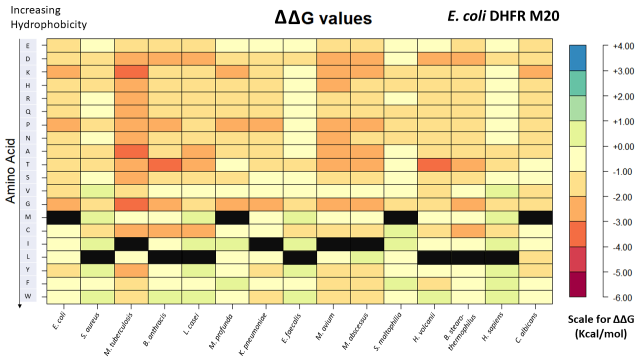
As seen in Fig 7 and Table 8, it seems that the degree of variation observed among DHFR proteins at those positions previously associated with TMP resistance is different at different positions. Positions such as P21 (exhibits no variation), D27 (2 types of residues, D or E), and Y151 (3 types of residues, Y, F, or H) (w.r.t *E. coli* DHFR) exhibit less variation while other residues like A26 (7 types of residues) and R98 (7 types of residues) exhibit a large degree of variation. This suggests that interspecific variation at a given position cannot be correlated with the ability of the residue at that position to influence protein function (in this case, resistance to TMP). This is expected as each of these residues has different physiochemical properties and there is a multitude of factors (such as whether the residue binds to the substrate, what interactions a residue has with the substrate, or whether the residue can affect protein stability and hence function) that can affect the potential of a residue to affect protein function. Hence interspecific variation alone probably cannot provide information about the ability of the residue at that position to influence protein function. An interesting observation is that most of the residues previously associated with TMP resistance are hydrophobic. While the list of resistance-conferring mutations may not be exhaustive, 7 of the 11 positions previously associated with TMP resistance are hydrophobic, as can be seen in Table 8. This could possibly hint at a correlation between the hydrophobicity of a residue and its ability to influence protein function (in this case, resistance to TMP). Further experimentation and analysis would be required to confirm such a correlation. Regardless, Table 8 highlights the potential contribution that the hydrophobicity of a residue may have toward the ability of the residue to influence protein function.

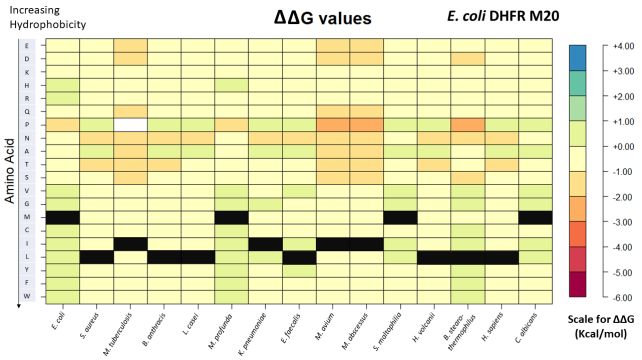
Variation at specific residues in a protein can influence protein structure which may lead to the perturbation of protein function and previous work has shown that some resistance-conferring mutations can compromise protein stability in DHFR (Matange et al., 2018). To understand how substituting amino acids found in the variants at the sites associated with TMP resistance in DHFR affect the stability of DHFR, $\Delta\Delta$G values were predicted by performing in-silico mutagenesis using two tools, I-Mutant 2.0 and Site-Directed Mutator (SDM). In brief, at the sites associated with TMP resistance in *E. coli* DHFR, the corresponding amino acid was mutated to all the remaining 19 commonly occurring amino acids. The above analysis was performed using two different tools in order to prevent bias that may occur due to the method that the tools use in order to predict $\Delta\Delta$G values. The plots depicting the $\Delta\Delta$G values for all such sites in all the DHFR structures taken from the PDB database are shown in Fig 8.

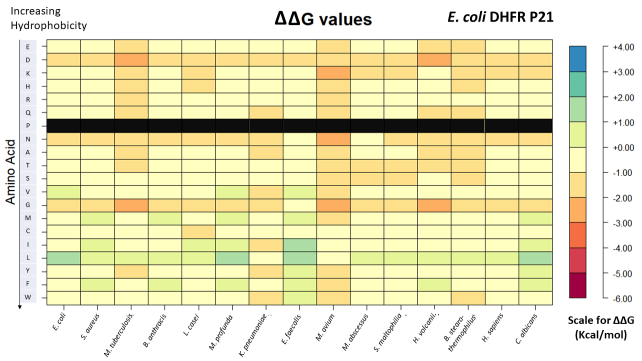
The effects of documented mutations that lead to TMP resistance in *E. coli* DHFR on stability were assessed using the above analysis. The vast majority of mutations were predicted by I-Mutant and SDM to negatively impact protein stability across all the DHFR proteins belonging to 15 different species for which this analysis was performed. The $\Delta\Delta$G predictions for I5 and I94 show that most mutations at these positions are destabilizing. This could hint at the importance of the residue that is originally present in each of the DHFR structures analyzed for protein structure or function. The $\Delta\Delta$G predictions for position A26 (w.r.t *E. coli* DHFR) show that a significant proportion of mutations seem to have a stabilizing/neutral effect on stability for both tools across all the homologous proteins tested. However, the pattern of mutations (i.e., which mutations are stabilizing and which mutations are destabilizing) are different across I-Mutant 2.0 and SDM. In the case of the mutations D27E, P21Q, and M20I, the protein stability predictions by the two tools yielded different results (I-Mutant 2.0 predicts that most mutations negatively decrease protein stability while SDM predicts that most mutations increase protein stability) so no conclusions can be drawn from the data of these mutations. The mutation P21L is the only exception where I-Mutant 2.0 and SDM predict that this mutation increases overall protein stability. Since the predictions were performed on the structure of *E. coli* DHFR bound to TMP (PDB ID:7MYM), the predictions suggest that mutations previously reported to be associated with TMP resistance decrease the stability of DHFR bound to TMP which could imply that the DHFR: TMP protein complex is destabilized. This in turn implies the possibility that the mutations associated with TMP resistance have the effect of decreasing the affinity of DHFR to TMP, resulting in resistance to TMP. Mutations that don't drastically decrease the overall stability of DHFR but can destabilize the DHFR: TMP complex can thus be considered to be mutations with good prospects to confer TMP resistance. These results show the correlation between mutations that impact protein stability and mutations that have the potential to confer resistance, displaying the importance of protein stability in antibiotic resistance.

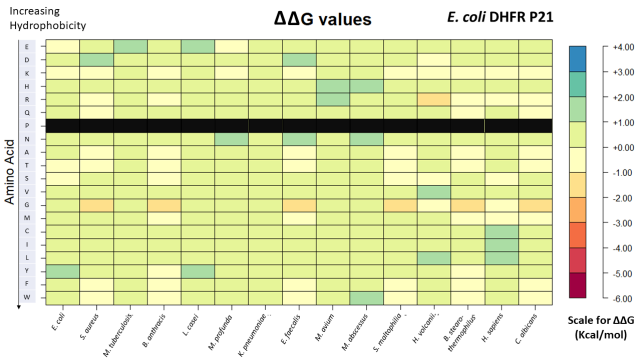

Increasing Hydrophobicity
ΔΔG values
E. coli DHFR I5
Amino Acid
E. coli
S. aureus
M. tuberculosis
B. anthracis
L. casei
M. profundo
K. pneumoniae
E. faecalis
M. avium
M. abscessus
S. multophilia
H. volcanii
B. stearo-thermophilus
H. sapiens
C. albicans
Scale for ΔΔG (Kcal/mol)

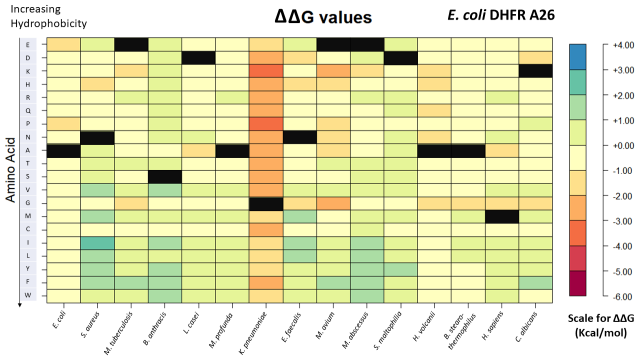

Increasing Hydrophobicity
ΔΔG values
E. coli DHFR I5
Amino Acid
E. coli
S. aureus
M. tuberculosis
B. anthracis
L. casei
M. profundo
K. pneumoniae
E. faecalis
M. avium
M. abscessus
S. multophilia
H. volcanii
B. stearo-thermophilus
H. sapiens
C. albicans
Scale for ΔΔG (Kcal/mol)

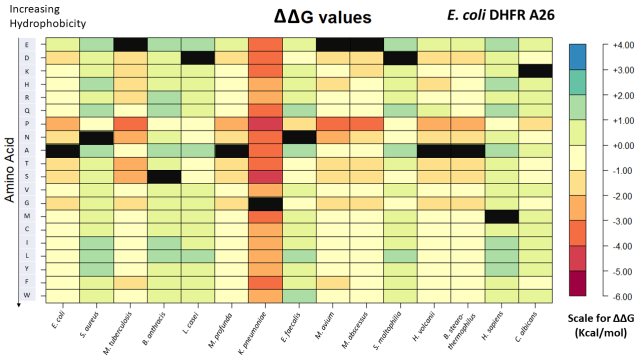

Increasing Hydrophobicity
ΔΔG values
E. coli DHFR M20
Amino Acid
E. coli
S. aureus
M. tuberculosis
B. anthracis
L. casei
M. profundo
K. pneumoniae
E. faecalis
M. avium
M. abscessus
S. multophilia
H. volcanii
B. stearo-thermophilus
H. sapiens
C. albicans
Scale for ΔΔG (Kcal/mol)

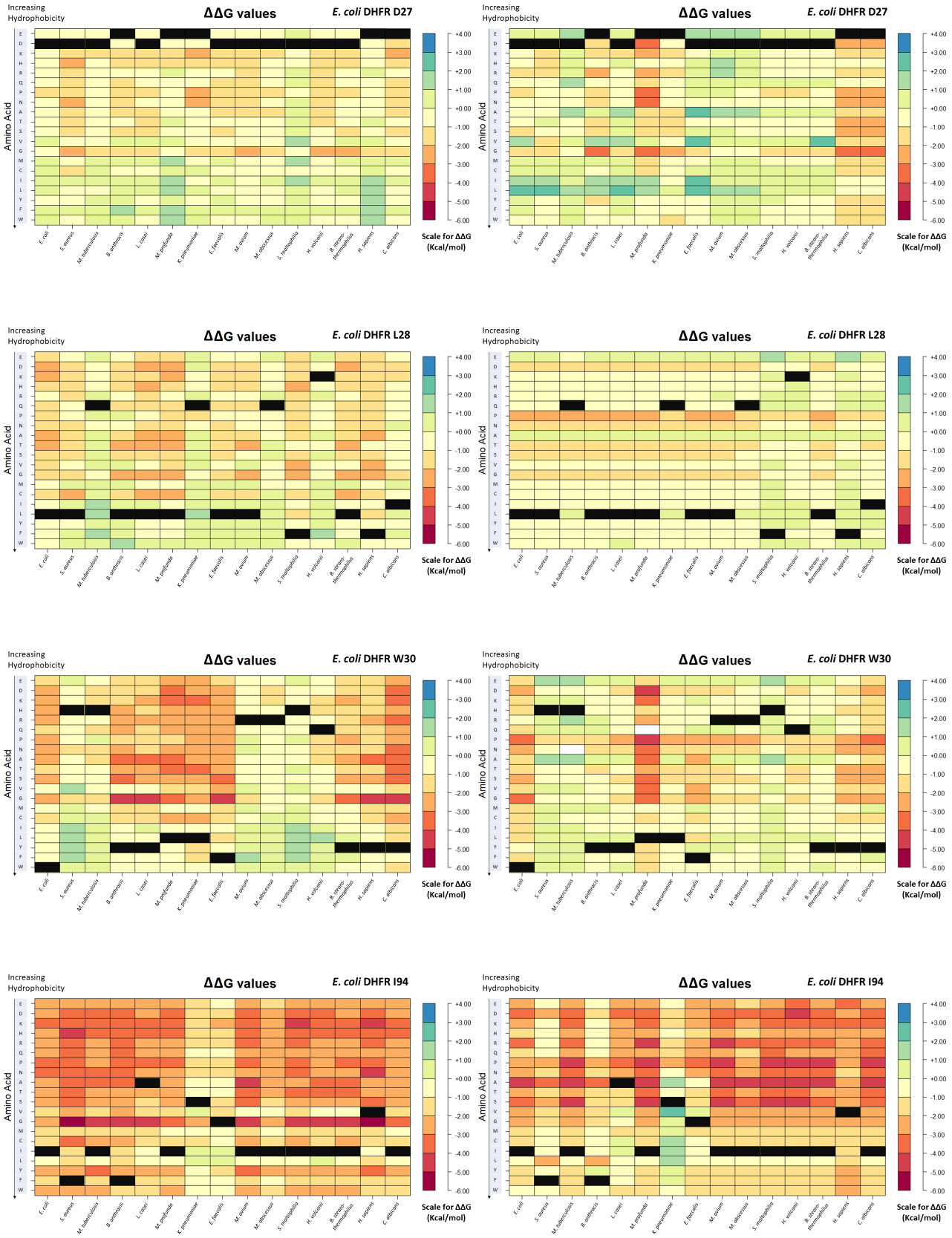

Increasing Hydrophobicity
ΔΔG values
E. coli DHFR M20
Amino Acid
E. coli
S. aureus
M. tuberculosis
B. anthracis
L. casei
M. profundo
K. pneumoniae
E. faecalis
M. avium
M. abscessus
S. multophilia
H. volcanii
B. stearo-thermophilus
H. sapiens
C. albicans
Scale for ΔΔG (Kcal/mol)


Increasing Hydrophobicity
ΔΔG values
E. coli DHFR P21
Amino Acid
E. coli
S. aureus
M. tuberculosis
B. anthracis
L. casei
M. profundo
K. pneumoniae
E. faecalis
M. avium
M. abscessus
S. multophilia
H. volcanii
B. stearo-thermophilus
H. sapiens
C. albicans
Scale for ΔΔG (Kcal/mol)


Increasing Hydrophobicity
ΔΔG values
E. coli DHFR P21
Amino Acid
E. coli
S. aureus
M. tuberculosis
B. anthracis
L. casei
M. profundo
K. pneumoniae
E. faecalis
M. avium
M. abscessus
S. multophilia
H. volcanii
B. stearo-thermophilus
H. sapiens
C. albicans
Scale for ΔΔG (Kcal/mol)


Increasing Hydrophobicity
ΔΔG values
E. coli DHFR A26
Amino Acid
E. coli
S. aureus
M. tuberculosis
B. anthracis
L. casei
M. profundo
K. pneumoniae
E. faecalis
M. avium
M. abscessus
S. multophilia
H. volcanii
B. stearo-thermophilus
H. sapiens
C. albicans
Scale for ΔΔG (Kcal/mol)


Increasing Hydrophobicity
ΔΔG values
E. coli DHFR A26
Amino Acid
E. coli
S. aureus
M. tuberculosis
B. anthracis
L. casei
M. profundo
K. pneumoniae
E. faecalis
M. avium
M. abscessus
S. multophilia
H. volcanii
B. stearo-thermophilus
H. sapiens
C. albicans
Scale for ΔΔG (Kcal/mol)


Increasing Hydrophobicity
ΔΔG values
E. coli DHFR D27
Amino Acid
E. coli
S. aureus
M. tuberculosis
B. anthracis
L. casei
M. profundo
K. pneumoniae
E. faecalis
M. avium
M. abscessus
S. multophilia
H. volcanii
B. stearo-thermophilus
H. sapiens
C. albicans
Scale for ΔΔG (Kcal/mol)
+4.00
+3.00
+2.00
+1.00
+0.00
-1.00
-2.00
-3.00
-4.00
-5.00
-6.00
Increasing Hydrophobicity
ΔΔG values
E. coli DHFR D27
Amino Acid
Scale for ΔΔG (Kcal/mol)
Increasing Hydrophobicity
ΔΔG values
E. coli DHFR L28
Amino Acid
Scale for ΔΔG (Kcal/mol)
Increasing Hydrophobicity
ΔΔG values
E. coli DHFR L28
Amino Acid
Scale for ΔΔG (Kcal/mol)
Increasing Hydrophobicity
ΔΔG values
E. coli DHFR W30
Amino Acid
Scale for ΔΔG (Kcal/mol)
Increasing Hydrophobicity
ΔΔG values
E. coli DHFR W30
Amino Acid
Scale for ΔΔG (Kcal/mol)
Increasing Hydrophobicity
ΔΔG values
E. coli DHFR I94
Amino Acid
Scale for ΔΔG (Kcal/mol)
Increasing Hydrophobicity
ΔΔG values
E. coli DHFR I94
Amino Acid
Scale for ΔΔG (Kcal/mol)

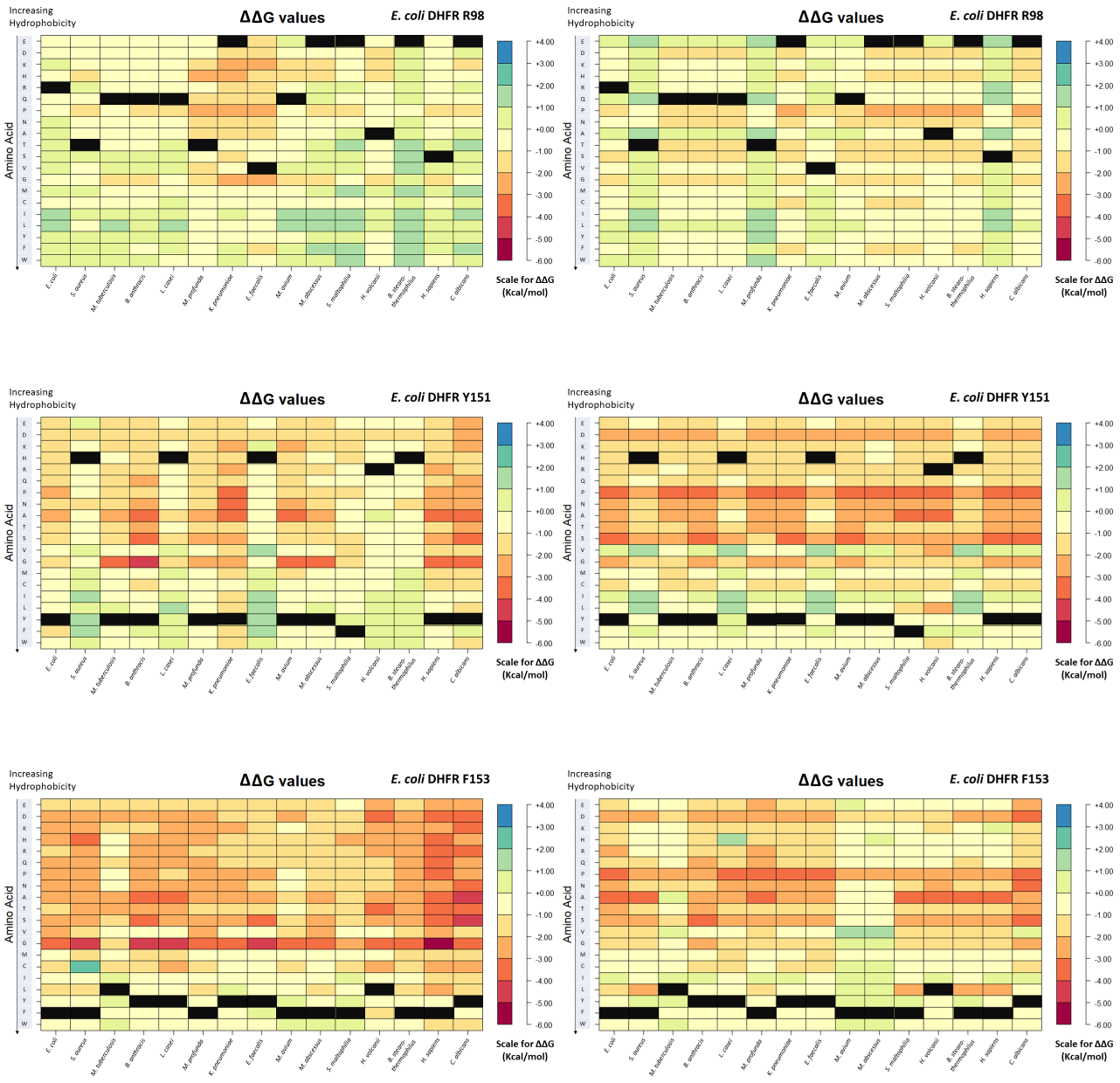


Figure 8: **ΔΔG values obtained when sites corresponding to each of the positions associated with TMP resistance in *E. coli* DHFR in the structures of DHFR of various organisms were mutated to the remaining commonly observed 19 amino acids**. The ΔΔG values were predicted using I-Mutant 2.0 (left) and Site-Directed Mutator (right). The scale depicts the range of ΔΔG values that each color represents. The black-colored bars represent the residue originally present for structures of DHFR at each of the positions corresponding to those positions associated with TMP resistance in *E. coli* DHFR.

The ΔΔG values predicted from the two tools were compared to check whether the predictions were consistent between the two tools used. The data for the percentage of matching

predictions between the two tools for those positions associated with TMP resistance in *E. coli* DHFR is shown in Table 9. In this context, matching predictions mean that the effect of a particular mutation (i.e. whether the mutation increases or decreases protein stability) was predicted by I-Mutant 2.0 and SDM to be the same.

Position (w.r.t *E. coli* DHFR)	Number of predictions matching between I-Mutant 2.0 and SDM	Percentage of matching predictions at the given position (%)
I5	273	95.79
M20	210	73.68
P21	83	29.12
A26	191	67.02
D27	172	60.35
L28	163	57.19
W30	215	75.44
I94	271	95.09
R98	136	47.72
Y151	246	86.32
F153	253	88.77

Table 9: Number (and percentage) of matching predictions between I-Mutant 2.0 and SDM at the positions associated with TMP resistance in *E. coli* DHFR.

Only those positions for which the percentage of matching predictions was above a specific cutoff (above 60%) were used for further analysis. This was done in order to ensure that the predictions are not an artifact of the method used by the two tools to predict $\Delta\Delta$G values. Based on this cutoff, the $\Delta\Delta$G predictions for the following positions (w.r.t *E. coli* DHFR) were further analyzed: I5, M20, A26, D27, W30, I94, Y151, and F153.

If a particular position in a protein structure is naturally more variable, it could imply that amino acids other than the one naturally present can be tolerated at that position. This would mean that there are no strict requirements for amino acids at a position with more variability and that more mutations have a neutral or stabilizing effect on the overall protein at this position compared to a more conserved position. If variation at a particular position can be linked to the amount of stabilizing/neutral mutations (compared to all possible mutations) across homologs, then variation could be used as an indicator of how mutations in general affect protein stability and hence fitness at a particular position. In order to understand the relationship between the variability of a given position and the cost of mutating the natural residue w.r.t protein stability, the data in Fig 8 were analyzed further. In order to observe if a similar pattern can be obtained from the data in Fig 8, the variability of a position was compared to the number of mutations that are stabilizing/neutral from both tools using the data from Fig 8. This was done only for those positions whose predictions

were consistent among the two tools. The results are summarized in Table 10 and Fig 9. Fig 9 depicts the plot between the variability of positions to the number of stabilizing mutations. Variability was denoted by the number of amino acids that were found at positions of the DHFR enzymes belonging to different species corresponding to those positions associated with TMP resistance in *E. coli* DHFR.

Position	Variability (number of amino acids found in the position)	Number of stabilizing/neutral mutations (I-Mutant 2.0) (out of 285 possible mutations)	Number of stabilizing/neutral mutations (SDM)(out of 285 possible mutations)
I5	5	3	11
M20	3	27	58
A26	7	111	114
D27	2	83	112
W30	7	50	76
I94	6	6	14
Y151	3	43	36
F153	3	12	34

Table 10: **Data comparing the variability of a position (number of amino acids found in the position) to the number of stabilizing/neutral mutations predicted by the two tools.**

In order to check for the possible link between the variability observed at a particular position and the number of stabilizing/neutral mutations observed, the correlation between the variability of a position and the number of stabilizing/neutral mutations was measured separately for I-Mutant 2.0 and SDM. The correlation was performed using Spearman's rank correlation test. The value of ρ (Spearman's rank coefficient) was calculated to be 0.0736 for both tools. A scatterplot depicting the relationship between the variability of a position and the number of stabilizing/neutral mutations found at the same position is shown in Fig 9. This implies that based on the data, there is no association between variability at a position and the number of stabilizing/neutral mutations observed at the same position. So, it is very less likely that variability at a particular position is linked to the amount of stabilizing mutations observed at the same position and hence variability is a poor indicator of how mutations affect protein stability and hence fitness at a particular position. The data from Table 10 and the poor correlation between the variability of a position and the number of stabilizing/neutral mutations suggest that not all amino acids can be accommodated at a given position. The data in Fig 9 and Table 10 thus suggests that the variation that is observed in the data (Fig 7 and Table 8) is non-random and that only amino acids with certain physiochemical characteristics are preferred at each position.

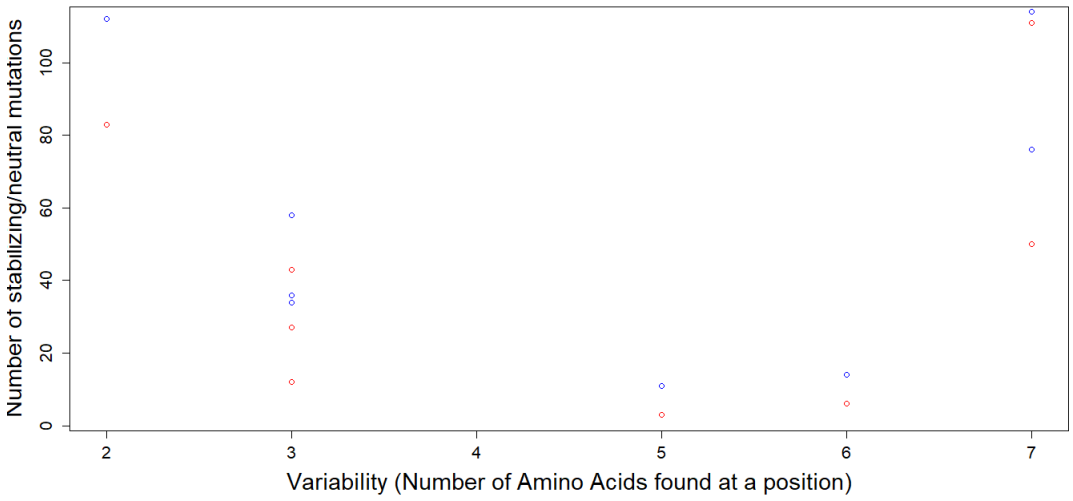


Figure 9: **A scatterplot depicting the relationship between variability of a position and the number of stabilizing/neutral mutations found at the same position.** The red points are data obtained from I-Mutant 2.0 while the blue points are data obtained from SDM.

An interesting observation at the positions equivalent to Y151 and W30 in *E. coli* DHFR is that mutations to hydrophobic residues tend to have a stabilizing/neutral effect while mutations to hydrophilic residues tend to have a destabilizing effect on the protein. In both these positions, when the residue originally present in the protein at that position is hydrophobic. Most other mutations (especially mutations to hydrophilic residues) have a destabilizing effect on the protein. However, when the residue originally present in the protein at that position is hydrophilic, then a significant number of mutations (especially mutations to hydrophobic residues) have a stabilizing/neutral effect on the protein. The pattern for the preference of hydrophobic residues at these positions is somewhat conserved across both I-Mutant 2.0 and SDM. These results imply that hydrophobic residues are more preferred w.r.t protein stability at the positions corresponding to Y151 and W40 in *E. coli* DHFR. There are, however, several DHFR proteins belonging to different organisms in which the residue originally present at the position equivalent to Y151 and W30 is hydrophilic. For those DHFR structures where the residue originally present at positions corresponding to Y151 and W30 in *E. coli* DHFR was hydrophilic, a possible explanation is that these residues are associated with some function of that particular DHFR structure. The preference for hydrophobic residues at such positions highlights the importance of hydrophobicity in certain positions in stabilizing the protein.

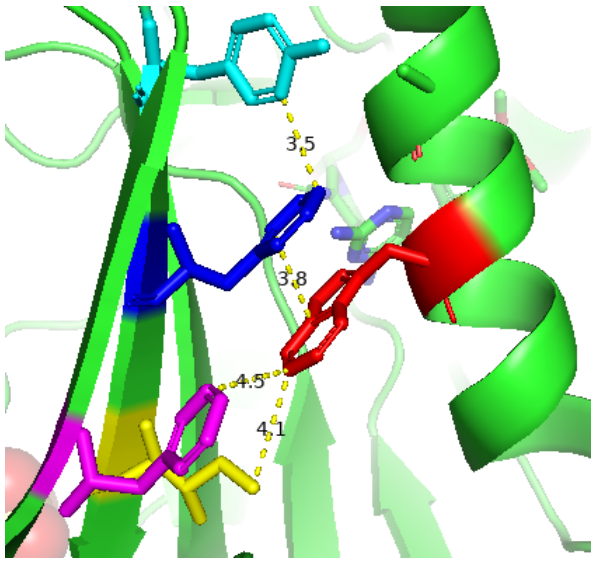


Figure 10: **Structure of *E. coli* DHFR bound to TMP (PDB ID:7MYM) with particular focus on positions W30 (red) and Y151 (cyan).** Left: The residue in red depicts W30, the residue in blue depicts F153, the residue in yellow depicts I155, the residue in cyan depicts Y151, and the residue in magenta depicts F137. The molecular distance between the residues (measured using PyMol) are shown in yellow.

A possible explanation as to why hydrophobic residues are preferred at the positions corresponding to Y151 and W30 in *E. coli* DHFR can be seen in Fig 10. W30 (in *E. coli* DHFR) is known to be a part of a hydrophobic tetrad along with the residues F153, F137, and I155. The interactions that W30 has with the aforementioned residues may act as a hydrophobic clamp that stabilizes the protein (Matange et al., 2018). The hydrophobicity of the residue at position W30 is crucial for maintaining these interactions, explaining the preference for hydrophobic residues at this position. The residue Y151 also seems to engage in a potential hydrophobic interaction with F153 (as shown in Fig 10), which provides a potential explanation as to why hydrophobic residues are preferred at that position.

The $\Delta\Delta$G predictions between the two tools are largely different for the positions L28, P21, and R98. At position L28, a similar preference for hydrophobic residues (as was the case for Y151 and W30) can be seen according to I-Mutant 2.0 $\Delta\Delta$G predictions. However, this pattern is not conserved in the $\Delta\Delta$G predictions of SDM. L28 in *E. coli* DHFR is part of the dihydrofolate (DHF) binding site and is also known to form hydrophobic interactions with the folate inhibitor TMP (Krucinska et al., 2022; Abdizadeh et al., 2017).

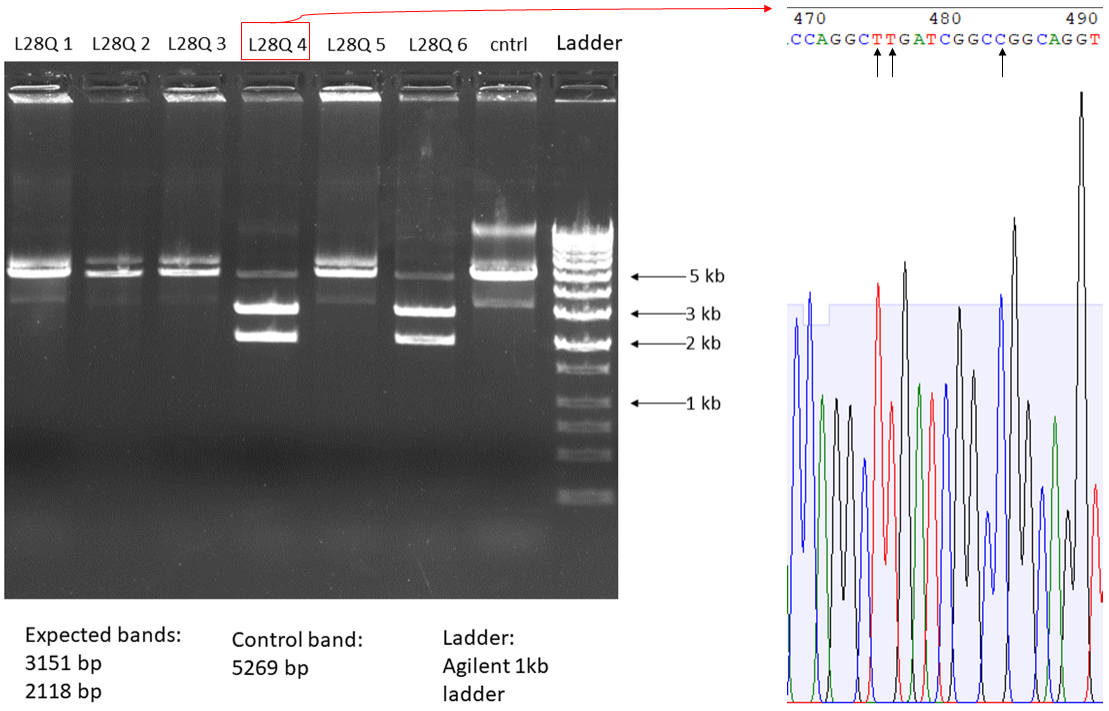


Figure 11: **Confirmation of the mutation L28Q through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzyme *Nae*I. The right side depicts the sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The black arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. The red arrow depicts the plasmid isolate that was selected and sent for sequencing. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer to Materials and Methods). cntrl = plasmid (pPRO-His-*folA*) used as the negative control.

To verify if such a preference for hydrophobic residues exists at the position corresponding to L28 in *E. coli* DHFR and to understand how natural variation in specific sites of the protein affects the functionality of the protein and to experimentally characterize such variants, the variations observed in the corresponding sites of DHFR belonging to the other species (as seen in Fig 7 and Table 8) were brought into the *E. coli* DHFR. This was achieved using a plasmid (pPROB) containing the *E. coli folA* gene (which synthesizes the DHFR enzyme) and mutating the *folA* gene in the plasmid using site-directed mutagenesis. The mutated plasmid was then transformed in the *E. coli* MG1655 strain. The IC50 values of the MG1655 strain containing the plasmid with the mutated *folA* gene were calculated for TMP using a 2-fold serial broth dilution. The above was performed for the position corresponding to L28 in *E. coli* DHFR. The data showing the confirmation of the expected mutations using restriction digestion followed by sequencing for the mutations L29Q, L28F, and L28K are shown in Fig 11, 12, and 13 respectively. The IC50 values are summarized in Fig 14.

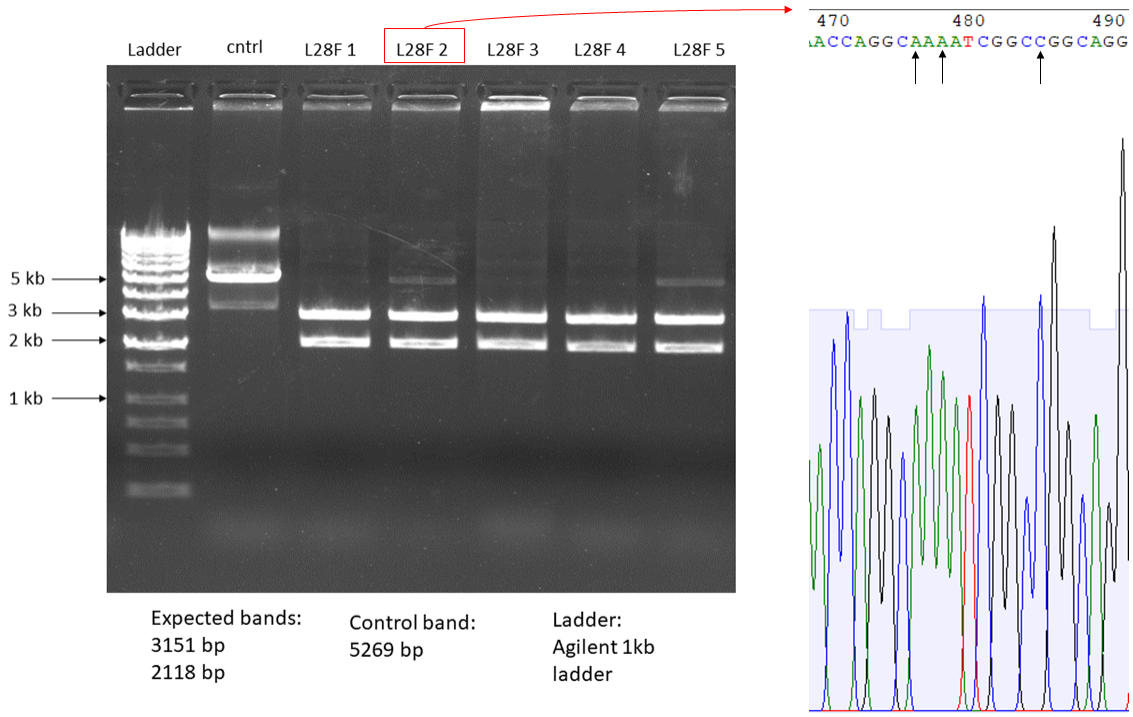


Figure 12: **Confirmation of the mutation L28F through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzyme *Nae*I. The right side depicts the sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The black arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. The red arrow depicts the plasmid isolate that was selected and sent for sequencing. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer to Materials and Methods). cntrl = plasmid (pPRO-His-*folA*) used as the negative control.

The IC50 values of TMP show that MG1655 pPRO-His-*folA* L28K and MG1655 pPRO-His-*folA* L28Q confer TMP resistance, with the IC50 values of the DHFR mutants being at least 6 times greater than that of wild type (MG1655 pPRO-His-*folA*). The MG1655 pPRO-His-*folA* L28F strain did not have a significantly higher IC50 value compared to the MG1655 pPRO-His-*folA* strain, implying that this mutation does not confer resistance. Furthermore, in order to confirm mutant protein expression in the cells containing the mutagenized plasmid, immunoblotting was performed. The results are summarized in Fig 15.

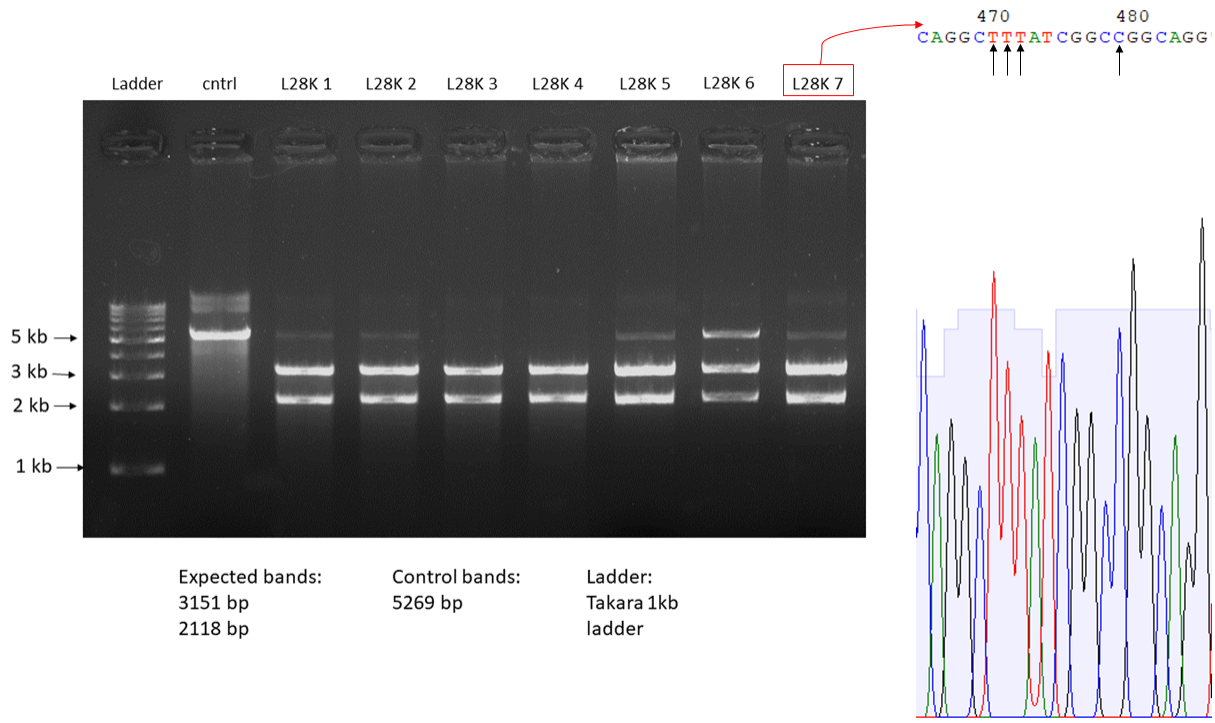


Figure 13: **Confirmation of the mutation L28K through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzyme *Nae*I. The right side depicts the sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The black arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. The red arrow depicts the plasmid isolate that was selected and sent for sequencing. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer to Materials and Methods). cntrl = plasmid (pPRO-His-*folA*) used as the negative control.

Why do the mutations L28K and L28Q confer TMP resistance, while the mutation L28F does not? As mentioned previously, this could be due to an indirect effect wherein mutations to hydrophilic residues lead to decreased protein stability (when DHFR is bound to TMP). However, a more likely explanation is as follows.

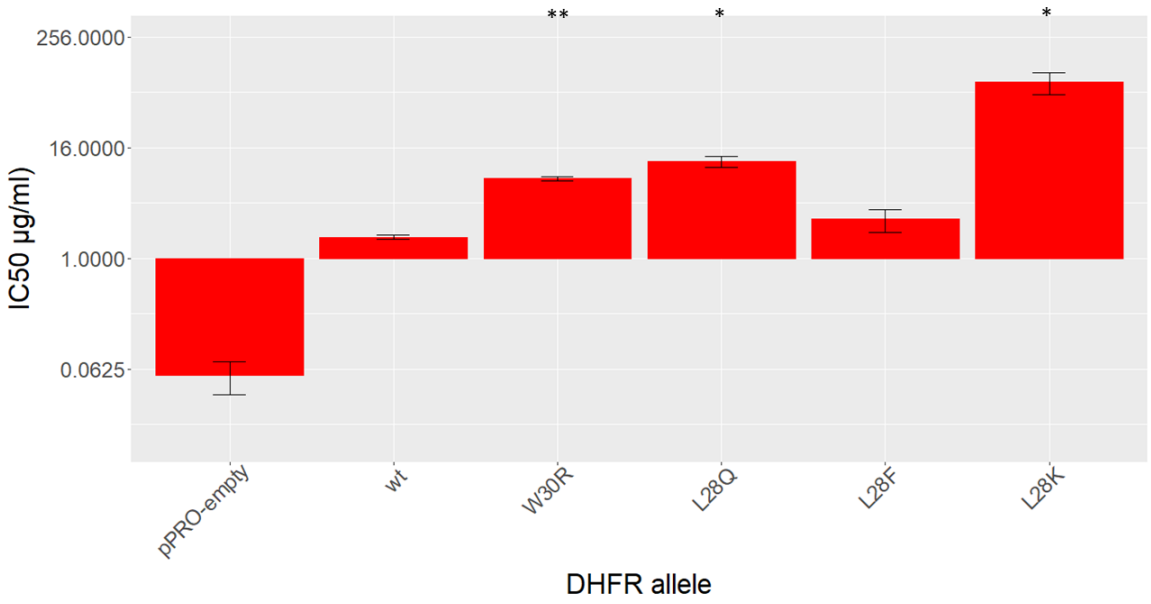


Figure 14: **IC50 values of *E. coli* MG1655 with plasmids containing *folA* with the mutations W30R, L28Q, L28F, and L28K.** * implies p-value <0.05, ** implies p-value <0.005. p-values were determined using the Welsh two-sample t-test. The t-tests were performed by comparing the mutant DHFR to wt DHFR. The IC50 values plotted are the mean of 3 replicates. The error bars represent the standard deviation derived from the 3 replicates. wt = wild-type = MG1655 pPRO-His-*folA*. The y-axis has been broken and the IC50 values from 15µg/ml to 60µg/ml are not represented in the graph (Xu et al., 2021). The y-axis has been transformed into a log2 scale.

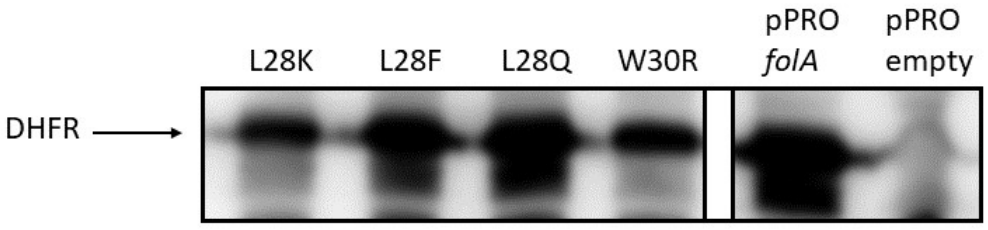


Figure 15: **Confirmation of protein expression using Western Blotting.** Immunoblotting was performed on cell lysates of MG1655 pPRO-empty plasmid, MG1655 pPRO-His-*folA*, and the mutants W30R, L28Q, L28F, L28K, and W47R using anti-DHFR as the primary antibody and Goat-anti-rabbit as the secondary antibody. The ladder used is the Bio-Rad Precision Plus Protein Dual Color Standard. The molecular weight of the His-tagged DHFR protein is ~25 kDa.

It has been previously shown that the L28R mutation in *E. coli* confers TMP resistance

by a unique mechanism wherein the newly introduced arginine residue forms extra hydrogen bonds with the aminobenzoyl glutamate tail of dihydrofolate, leading to the stabilization of the substrate within the enzyme binding site and the procurement of TMP resistance by the L28R DHFR protein (Abdizadeh et al., 2017). The L28Q and L28K mutations in DHFR could confer resistance to TMP by a similar mechanism wherein the glutamine or lysine residue forms extra hydrogen bonds with the substrate, preventing efficient binding of TMP to the DHFR mutant. The presence of a phenylalanine residue at the 28th position of DHFR may not result in these hydrogen bonds forming, explaining the absence of an increase in IC50 values in the MG1655 pPRO-His-*folA* L28F strain.

The IC50 values suggest that substituting native residues with certain properties (the presence of an amine group in this case) may have a predictable impact on the sensitivity of DHFR to TMP. The data also suggest that natural variation is capable of altering intrinsic resistance to antibiotics. DHFR proteins with amino acids with an amine group at the position corresponding to L28 in *E. coli* DHFR may intrinsically be more resistant to TMP compared to *E. coli* DHFR. For instance, the DHFR protein belonging to *Mycobacterium tuberculosis* has a glutamine group (refer to Table 8) at the position corresponding to L28 in *E. coli* DHFR. The IC50 value of DHFR belonging to *Mycobacterium tuberculosis* (*Mtb*) against TMP is 16μM (Raju et al., 2015) which is higher than the IC50 value of *E. coli* DHFR against TMP (approx. 5.85μM). However, the higher IC50 values exhibited by *Mtb* DHFR are likely to be a result of multiple differences in protein sequence (and hence structure and function) and not just this difference observed at the position equivalent to L28 in *E. coli* DHFR. The data (from Fig 14) hence exhibits the ability of natural variation to affect protein function (in this case, intrinsic resistance to antibiotics).

3.2 The Impact of Single and Combinatorial Mutations on Protein Structure and Function

The position W30 in *E. coli* DHFR has been studied with respect to protein stability and function previously. It has been shown that the mutations W30R, W30C, and W30G confer TMP resistance in *E. coli* DHFR and lead to misfolding and aggregation of *E. coli* DHFR. Moreover, it has also been shown that mutations at W30 of *E. coli* DHFR destabilize the protein due to loss of hydrophobic interactions and certain mutations at W30 of *E. coli* DHFR have a fitness cost associated with them (Matange et al., 2018). It has been observed previously that when *E. coli* MG1655 is subjected to adaptive laboratory evolution under sub-inhibitory concentrations of TMP, the bacteria gain the mutation W30R (in some of the evolution trajectories) in the DHFR protein which confers resistance to TMP. It has also been observed that when *E. coli* MG1655 *Δlon* is subjected to adaptive laboratory evolution under sub-inhibitory concentrations of TMP, the mutation Y151D co-occurs with the mutation W30R in some of the evolutionary trajectories (Nishad Matange, unpublished).

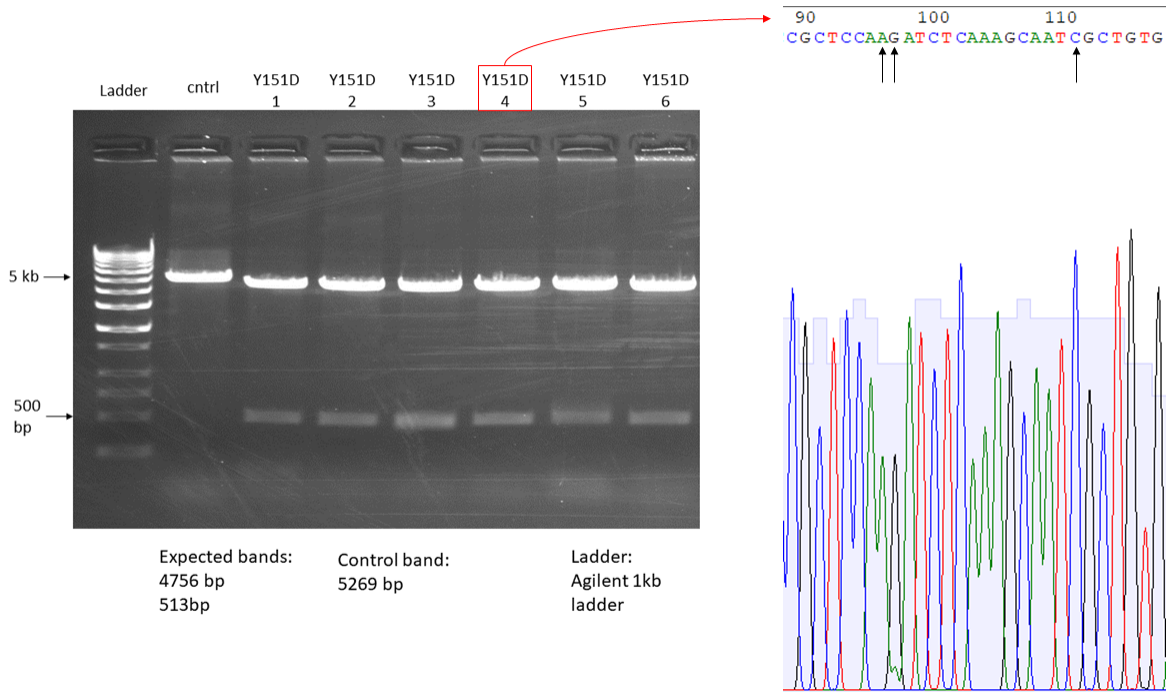


Figure 16: **Confirmation of the mutation Y151D through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzymes *Bgl*II and *Nco*I. The right side depicts sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer Materials and Methods). cntrl = plasmid (pPRO-His-*folA*) used as the negative control.

In a separate study, it has been shown that when *E. coli* was allowed to evolve in the presence of TMP, it was observed that the mutation Y151D only occurred when the mutation W30R was present (Oz et al., 2014). In either case, the mutation Y151D, which presumably can affect TMP resistance, always co-occurs with the mutation W30R in *E. coli* DHFR.

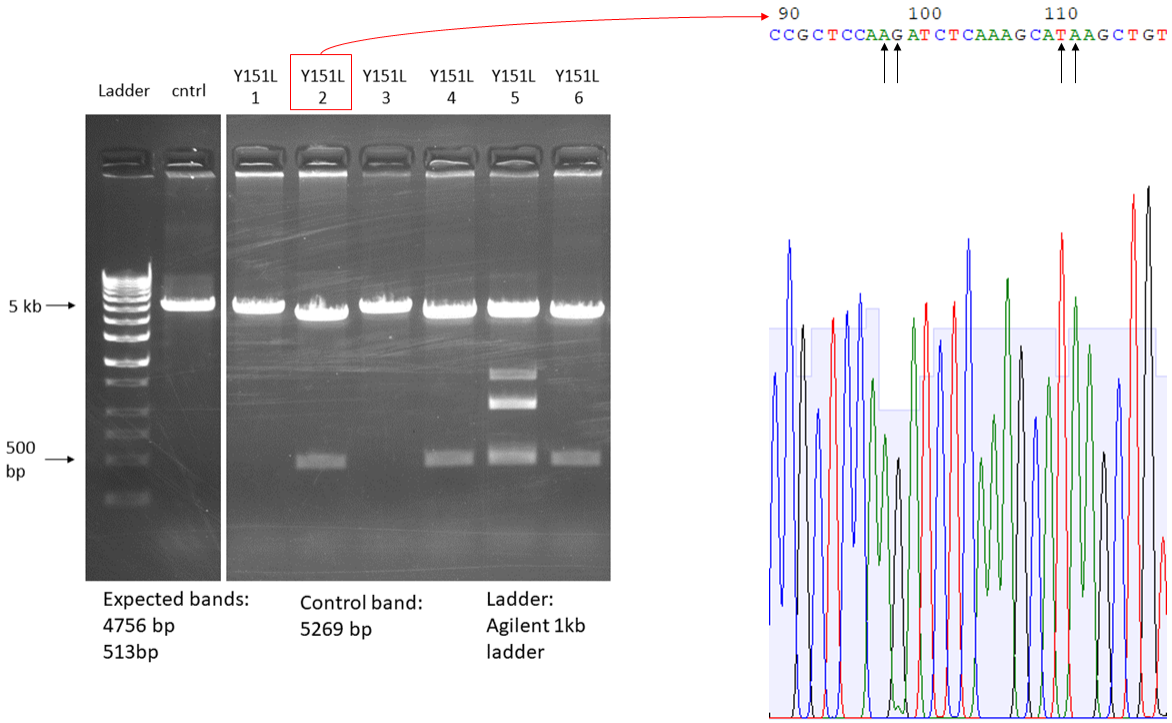


Figure 17: **Confirmation of the mutation Y151L through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzymes *Bgl*II and *Nco*I. The right side depicts sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer Materials and Methods). cntrl = plasmid (pPRO-His-*folA*) used as the negative control.

To understand how the mutation Y151D impacts TMP resistance when it occurs on its own and when it co-occurs with *E. coli* DHFR, site-directed mutagenesis was performed on a plasmid (pPROB) containing the *E. coli folA* gene (which synthesizes the DHFR enzyme). The mutated plasmid was then transformed in the *E. coli* MG1655 strain. The mutations Y151D, Y151L, and Y151F as well as the double mutants W30R-Y151D and W30R-Y151F were introduced into *E. coli* DHFR. This was done because the residues Asp, Leu, and Phe represent amino acids with different physio-chemical characteristics, and understanding how mutating Y151 to these residues affects TMP resistance and protein stability can enable a better mechanistic understanding of how mutations in Y151 affect TMP resistance and hence protein function. The double mutants (W30R along with Y151D, Y151L, or Y151F) were generated by performing site-directed mutagenesis using primers that introduce the Y151D, Y151L, or Y151F mutation into pPRO-His-*folA* that already contains the mutation W30R. The data showing the confirmation of the expected mutations using restriction digestion followed by sequencing for the mutations Y151D, Y151L, Y151F, and the double mutants

W30R-Y151D and W30R-Y151F are shown in Fig 16-20 respectively. The IC50 values for the strains containing the mutagenized plasmid are summarized in Fig 21.

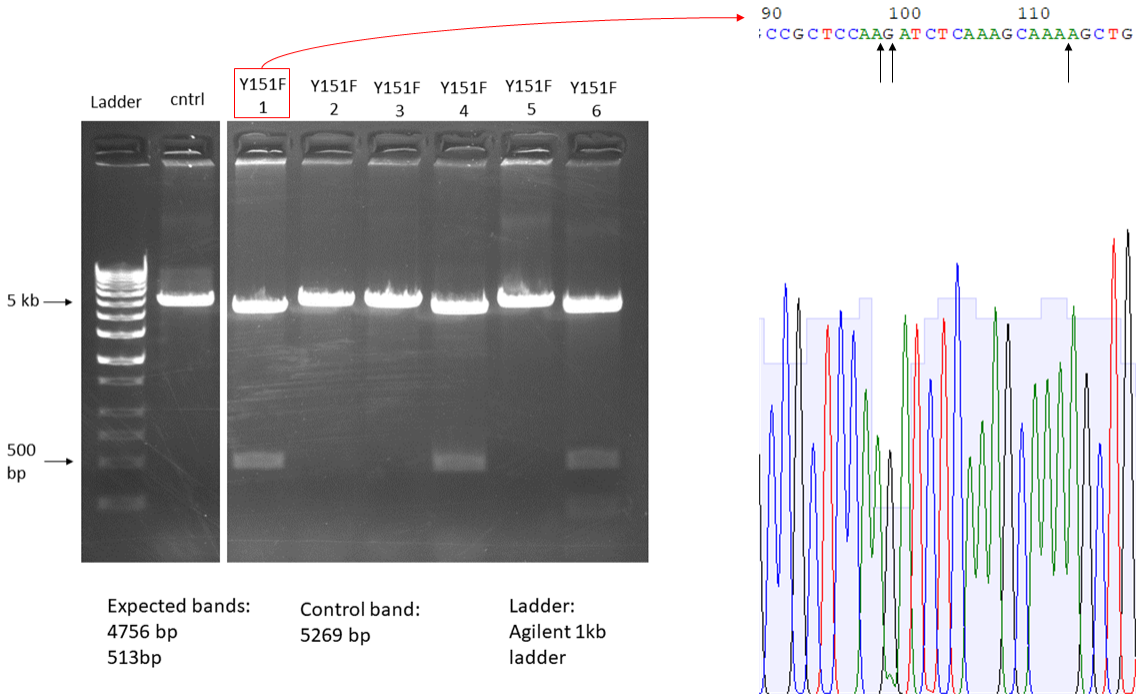


Figure 18: **Confirmation of the mutation Y151F through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzymes *Bgl*II and *Nco*I. The right side depicts sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer Materials and Methods). cntrl = plasmid (pPRO-His-*folA*) used as the negative control.

The IC50 values of the *E. coli* MG1655 derivatives carrying the plasmid containing the Y151D, Y151L, and Y151F mutations in the *folA* gene show an interesting pattern (Fig 21). The mean IC50 values of the mutants Y151L and Y151F are statistically significantly different from wild-type *E. coli* DHFR. However, there isn't a huge difference in the actual IC50 values, suggesting that the phenotypic effect of the Y151L and Y151F variants are not different from that of wild-type. The IC50 value of the mutant Y151D is significantly less than that of wild-type DHFR (2 sample t-test p-value ~ 0.0004) and is similar to the IC50 value of MG1655 pPRO-empty (2 sample t-test p-value = 0.8386) (i.e., MG1655 containing the plasmid without the *folA* gene inserted). These data suggest that the mutation Y151D in *E. coli* DHFR sensitizes the protein to TMP, making the bacteria more susceptible to TMP. Since only the mutation Y151D in *E. coli* DHFR leads to such a phenotype while the mutations Y151L and Y151F do not, it is possible that loss of hydrophobicity at Y151 of

E. coli DHFR leads to a phenotype wherein the bacteria become more susceptible to TMP. Loss of hydrophobicity at Y151 of *E. coli* DHFR could perturb hydrophobic interactions between Y151 and other nearby residues (for instance, F153 as shown in Fig 10 of Part 1 of Results and Discussion) which in turn could increase the affinity of DHFR to TMP, leading to the bacteria becoming more susceptible to TMP.

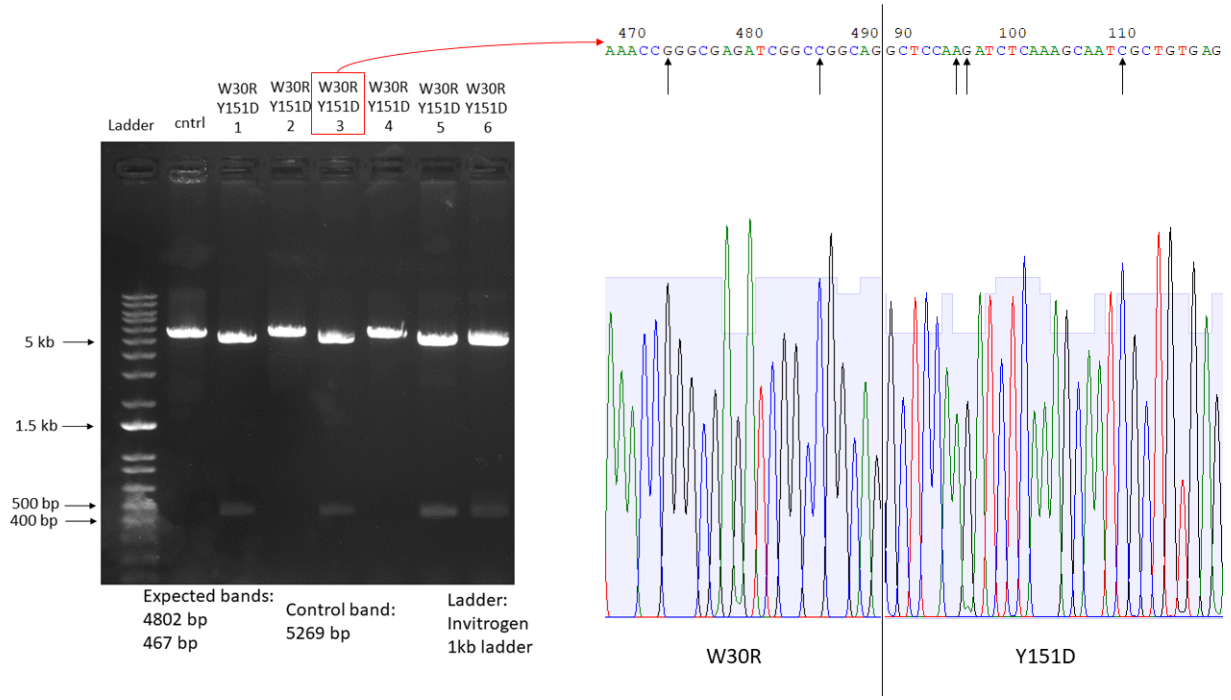


Figure 19: **Confirmation of the double mutation W30R-Y151D through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzymes *Bgl*II and *Bsu*36I. The right side depicts sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The two sections shown are the sections in the *folA* gene where the expected mutations have been generated. The arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer Materials and Methods). cntrl = plasmid (pPRO-His-*folA* W30R) used as the negative control.

Immunoblotting can be used to confirm whether *E. coli* DHFR with the mutation Y151D is being expressed in the cells containing the mutagenized plasmid. The results of these experiments are still pending.

To understand the effects of the mutations in position Y151 of *E. coli* DHFR co-occurring with the mutation W30R, the double mutants W30R-Y151D and W30R-Y151F were generated in *E. coli* DHFR. Site-directed mutagenesis was performed on a plasmid (pPROB) containing the *E. coli folA* gene (which synthesizes the DHFR enzyme) with the W30R mu-

tation and the doubly mutated plasmid was then transformed in the *E. coli* MG1655 strain. The IC50 values for the strains containing the mutagenized plasmid are summarized in Fig 21.

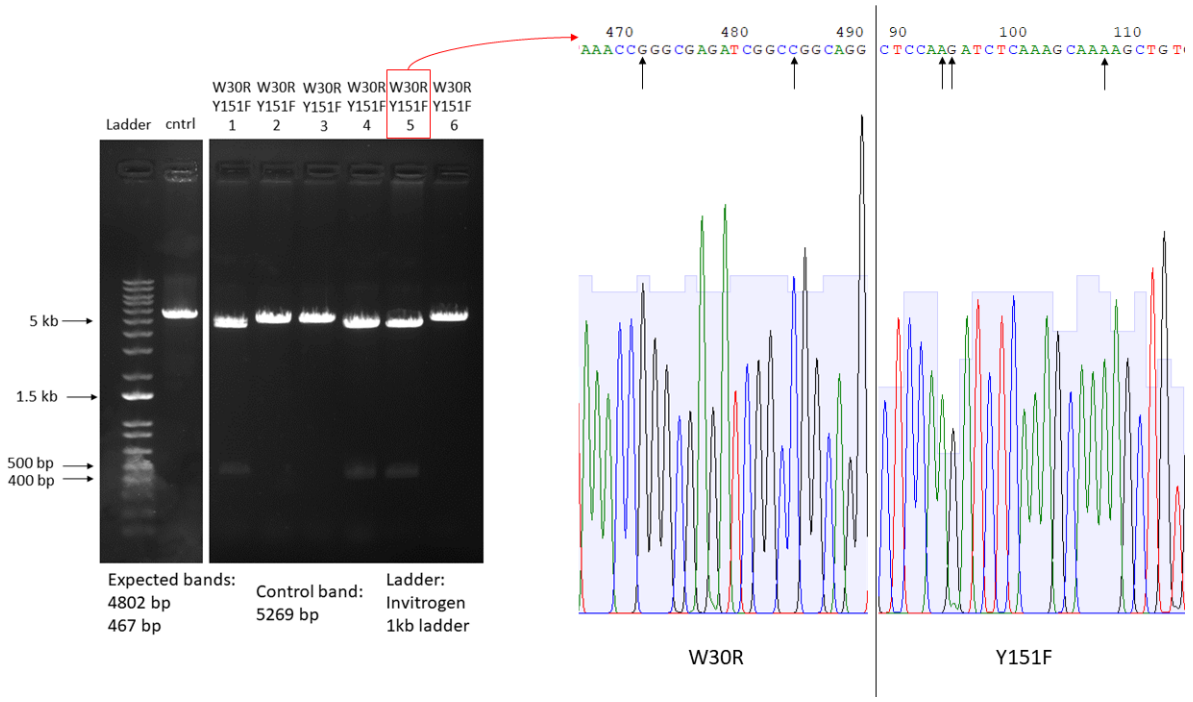


Figure 20: **Confirmation of the double mutation W30R-Y151F through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzymes *Bgl*II and *Bsu*36I. The right side depicts sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The two sections shown are the sections in the *folA* gene where the expected mutations have been generated. The arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer Materials and Methods). cntrl = plasmid (pPRO-His-*folA* W30R) used as the negative control.

The mean IC50 values of the mutant W30-Y151F are statistically significantly different from wild-type *E. coli* DHFR. However, there isn't a huge difference in the actual IC50 values, suggesting that the phenotypic effect of the W30R-Y151F double mutant is not different from that of the wild-type. The phenotypic effect (IC50 value) observed for the W30R-Y151F double mutant is hence an additive effect of the individual phenotypic effects of the W30R mutant and the Y151F mutant. The IC50 value of the W30R-Y151D double mutant in a *Δlon* background was significantly higher than that of wild-type DHFR.

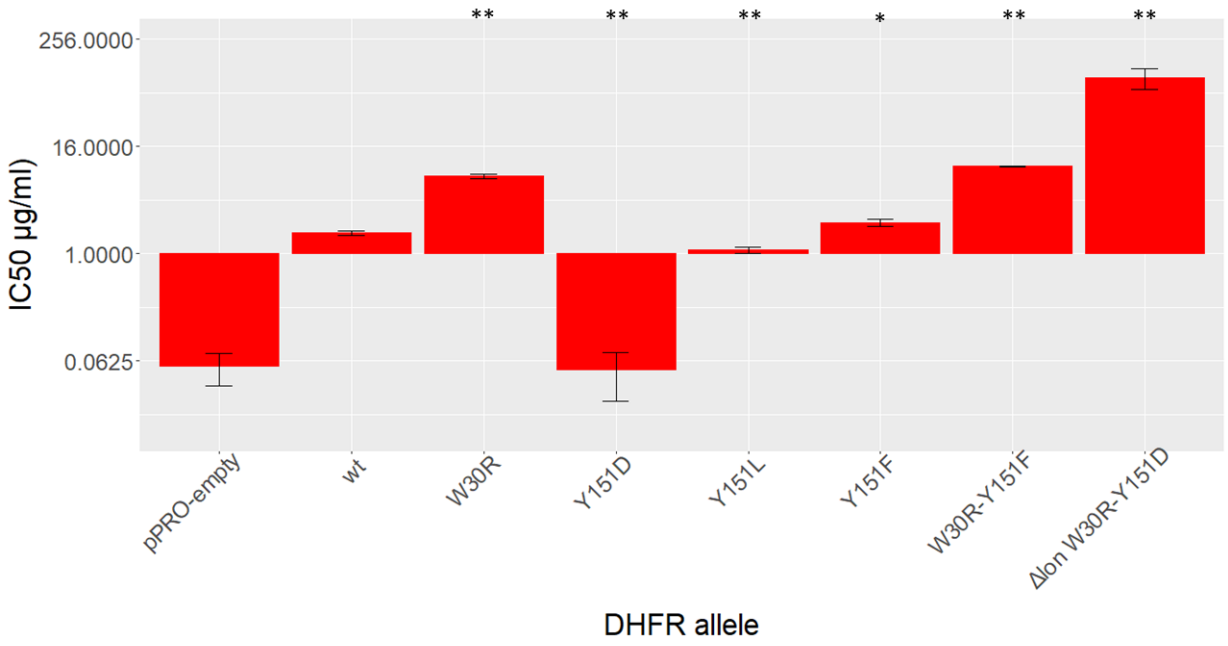


Figure 21: **IC50 values of *E. coli* MG1655 with plasmids containing *folA* with the mutations W30R, Y151D, Y151L, Y151F, and the double mutant W30R-Y151F and W30R-Y151D (in a *Δlon* background).** * implies p-value <0.05, ** implies p-value <0.005. p-values were determined using the Welsh two-sample t-test. The t-tests were performed by comparing the mutant DHFR to wt DHFR. The IC50 values plotted are the mean of 3 replicates. The error bars represent the standard deviation derived from the 3 replicates. wt = wild-type = MG1655 pPRO-His-*folA*. The y-axis has been broken and the IC50 values from 15µg/ml to 60µg/ml are not represented in the graph (Xu et al., 2021). The y-axis has been transformed into a log2 scale.

The strain MG1655 pPRO-His-*folA* W30R-Y151D behaved in a peculiar manner. The dose-response curves showed a pattern similar to a linear decrease in OD values as TMP concentration was increased instead of the characteristic sigmoidal shape that inhibitory dose-response curves tend to have (Fig 22). Due to this, the IC50 values for the W30R-Y151D mutant in a wild-type background could not be obtained. The evolutionary trajectories obtained from the adaptive laboratory evolution experiments (Nishad Matange, unpublished) show that the DHFR double mutant W30R-Y151D only occurred in MG1655 *Δlon* strains and did not occur in the MG1655 wild-type strain. This could hint at the reason why IC50 values for MG1655 pPRO-His-*folA* W30R-Y151D could not be obtained. It could be that the DHFR protein with the mutations W30R and Y151D cannot be sustained in the MG1655 wild-type strain and can only be sustained in the MG1655 *Δlon* strain. The gene *lon* codes for the Lon protease, which is a serine protease that degrades abnormal or mutated proteins and plays a crucial role in the maintenance of mitochondrial homeostasis (Pinti et al., 2016). A possible explanation as to why the W30R-Y151D double mutant is allowed

only in the MG1655 Δlon strain is that the W30R-Y151D double mutant DHFR protein is abnormal (either unstable or folds in an abnormal manner such that the behavior of the protein is abnormal) and is degraded by the Lon protease. Hence, the mutant protein can only be accommodated in the absence of the Lon protease (in the Δlon strain) such that a phenotypic effect w.r.t TMP resistance is displayed.

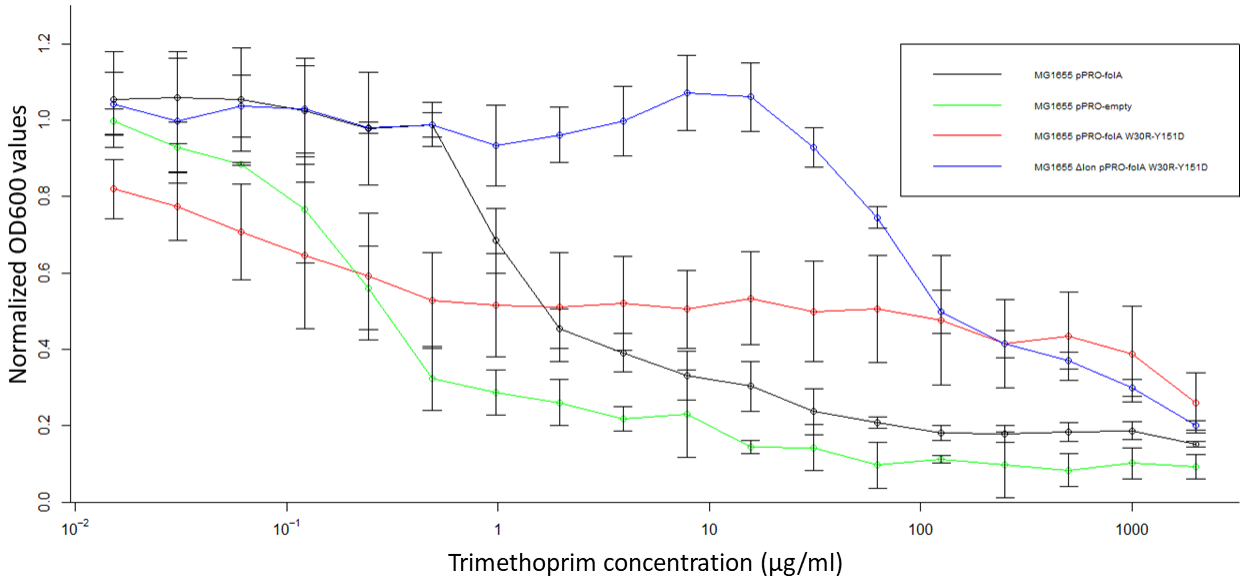


Figure 22: **Dose-response curves for *E. coli* MG1655 with plasmids containing *folA* with the mutations W30R, Y151D, the double mutant W30R-Y151F, and W30R-Y151D (in a Δlon background).**

In order to verify whether *E. coli* DHFR with the double mutation W30R-Y151D is being expressed in the cells containing the mutagenized plasmid and whether the W30R-Y151D mutant shows lower protein levels, immunoblotting experiments were planned. These experiments are still ongoing.

3.3 Consequences of Intraspecific Variation on Protein Function

To study intraspecific variation and the impact of such variation on protein structure and function, a multiple sequence alignment (MSA) was created using DHFR protein sequences belonging to various strains of the same species. The DHFR sequences were obtained from the NCTC 3000 project. The NCTC 3000 is a collection that contains whole genome sequence information of several type and reference bacteria strains. The strains represent various species of specific diseases and are obtained from various geographical conditions ((NCTC 3000 Project. (n.d.).). This analysis was performed for *E. coli* as well as *S. enterica* strains. Fig 23 and 24 represent plots showing the variation observed in the sequence alignments performed using strains of *E. coli* and *S. enterica*. The positions 47 (W or R) and 105 (P

or A) were variable in *E. coli* DHFR while the positions 23 (N or S), 61 (V or I), 66 (P or S), 88 (A or V), 89 (P or S), 127 (D or N), and 159 (R or a deletion) were variable in *S. enterica* DHFR.

The sequence alignment using DHFR sequences that belong to various *E. coli* strains (Fig 23 and Table 11) shows that there is a high degree of conservation between the various DHFR proteins. However, there exists some variation between the proteins and this variation occurs at 2 sites which are the W47 and the P105 residues. The alignment shows that some proteins contain an arginine (R) residue instead of tryptophan at the 47th residue, while other proteins contain an alanine (A) residue instead of a proline at the 105th position of *E. coli* DHFR. However, the sequence alignment using sequences belonging to various *S. enterica* strains (Fig 24 and Table 12) reveals that relative to the *E. coli* MSA, many more positions are variable. Most of the positions have a low frequency of the variant occurring. Positions 23, 88, 89, 127, and 159 have only one sequence (out of the 66 sequences analyzed) for which the variants exist. For positions 61 and 66, the frequency of variation is higher (and hence more significant and less likely to be due to errors during sequencing). A pairwise sequence alignment of the DHFR protein sequence (performed using the EMBOSS Needle alignment) (Rice et al., 2000) between the type strain *S. enterica* (*S. enterica* subsp. enterica serotype Typhimurium LT2) and the DHFR sequence (taken from the PDB file 7MYM) reveal that the DHFR protein sequence belonging to the 2 species are exactly alike (i.e., sequence identity = 100%). Despite the similarity between the DHFR sequences of the two strains, it can be observed from Fig 23 and 24 that the number of positions at which the sequence is variable as well as the pattern of positions at which the sequence is variable is vastly different for the two species. These data suggest that the two species have their own unique pattern of variation (in the case of the DHFR protein) despite the similarity between the DHFR sequences between the two strains. The reason for the presence of the unique pattern of variation (despite the similarity between the DHFR sequences of the two species) could be due to the two species evolved independently of each other under different environmental conditions.

To understand the effects that the intraspecific variations have on protein stability, $\Delta\Delta$G values were predicted using two online tools, I-Mutant 2.0 and SDM. The results of the $\Delta\Delta$G prediction are summarized in Tables 13 and 14.

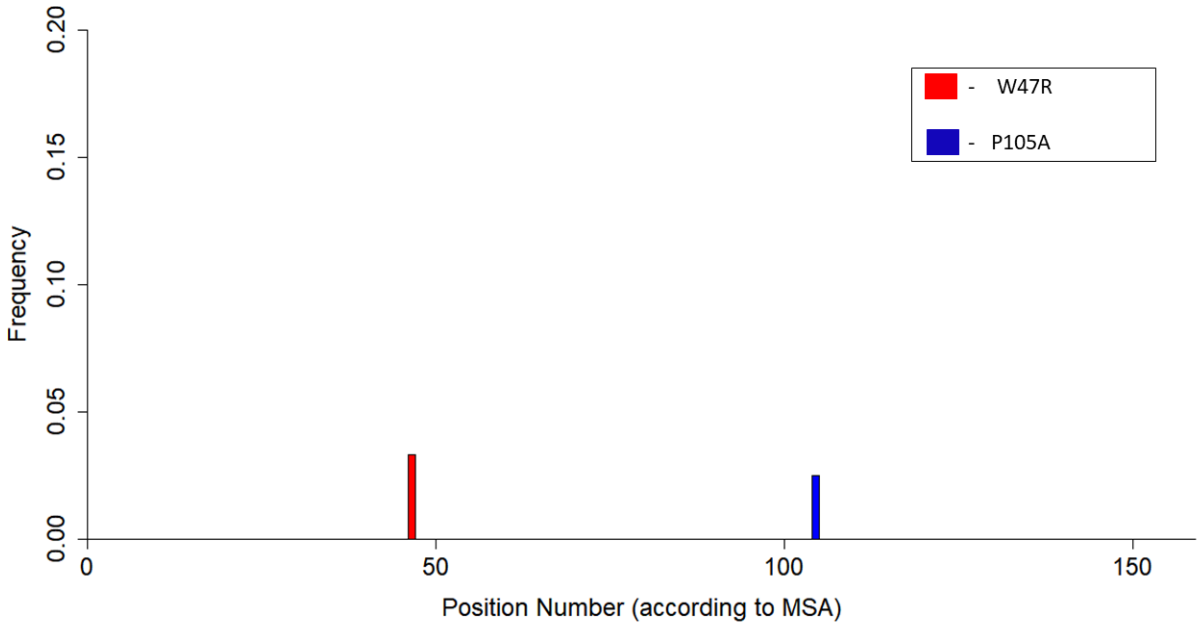


Figure 23: **Frequency of variation from the multiple sequence alignment created using DHFR sequences from *E. coli* strains.** The frequency (number of sequences with a residue different from the majority of other sequences) was plotted against the position number of the residues according to the multiple sequence alignment. The legend shows the variants for which the frequency is depicted in the figure.

The $\Delta\Delta$G predictions for the *E. coli* variants were different among the two tools used. I-Mutant 2.0 predicts that the intraspecific variations have a destabilizing effect on the protein while SDM predicts that the intraspecific variants have largely a stabilizing/neutral effect on the protein. The $\Delta\Delta$G predictions (by I-Mutant 2.0) for *S. enterica* variants show that most of the variants have a destabilizing effect on the protein. $\Delta\Delta$G prediction using SDM was not possible for *S. enterica* variants as a crystal structure of *S. enterica* DHFR is not available on the PDB database. This result is unexpected as these variations exist in nature and destabilizing variants are usually preferred as they generally cannot be selected over a more stable variant. A possible explanation is that these variants may provide an advantage over the wild-type sequence in specific environmental conditions.

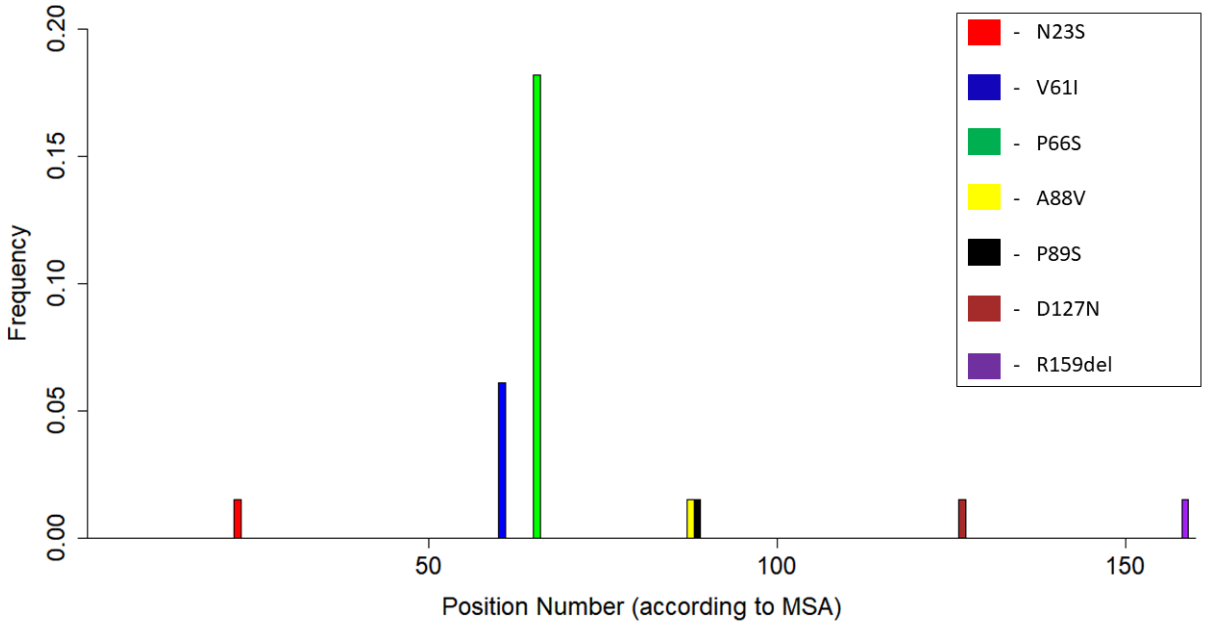


Figure 24: **Frequency of variation from the multiple sequence alignment created using DHFR sequences from *S. enterica* strains.** The frequency (number of sequences with a residue different from the majority of other sequences) was plotted against the position number of the residues according to the multiple sequence alignment. The legend shows the variants for which the frequency is depicted in the figure. Del = deletion

Variant	Number of sequences in which this variant occurs	Frequency of variation
W47R	4	0.033
P105A	3	0.025

Table 11: **Table depicting the number of sequences in which a particular variation occurs as well as the frequency of variation from the multiple sequence alignment created using DHFR sequences belonging to various *E. coli* strains.**

Variant	Number of sequences in which this variant occurs	Frequency of variation
N23S	1	0.015
V61I	4	0.061
P66S	12	0.182
A88V	1	0.015
P89S	1	0.015
D127N	1	0.015
R159del	1	0.015

Table 12: **Table depicting the number of sequences in which a particular variation occurs as well as the frequency of variation from the multiple sequence alignment created using DHFR sequences belonging to various *S. enterica* strains.** del = deletion

Variant	ΔΔG predictions (I-Mutant 2.0)	ΔΔG predictions (SDM)
W47R	-1.89	-0.06
P105A	-1.16	0.67

Table 13: **ΔΔG predictions for the intraspecific variation observed in the multiple sequence alignment created using DHFR sequences belonging to various *E. coli* strains.**

Variant	ΔΔG predictions (I-Mutant 2.0)
N23S	0.34
V61I	-0.94
P66S	-1.72
A88V	-1.46
P89S	-1.21
D127N	-0.7
R195del	nil

Table 14: ΔΔG predictions for the intraspecific variation observed in the multiple sequence alignment created using DHFR sequences belonging to various *S. enterica* strains.

To understand the effects that the intraspecific variations have on protein stability, ΔΔG values of the variants were predicted using two online tools, I-Mutant 2.0 and SDM. The results of the ΔΔG prediction are summarized in Tables 13 and 14.

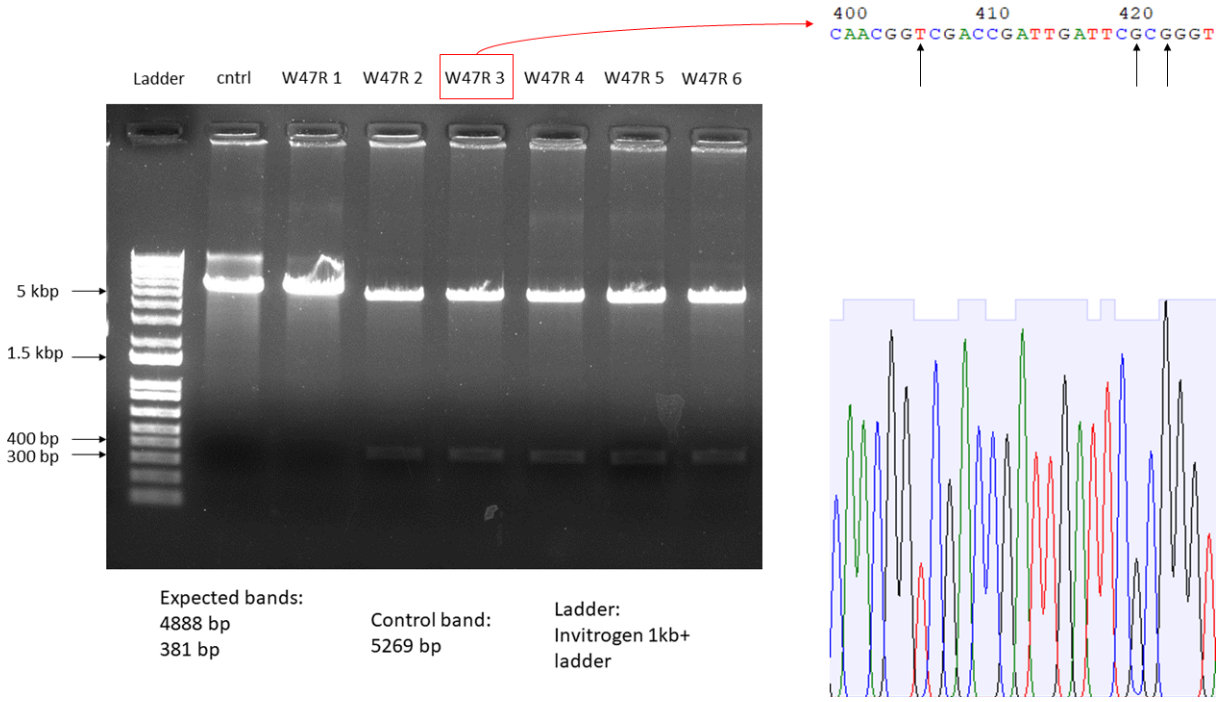


Figure 25: **Confirmation of the mutation W47R through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzymes *Hind*III and *Sal*I. The right side depicts sequencing result of one of the plasmids showing the expected DNA fragments after restriction digestion. The arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer Materials and Methods). cntrl = plasmid (pPRO-His-*folA*) used as the negative control.

The ΔΔG predictions for the *E. coli* variants were different among the two tools used. I-Mutant 2.0 predicts that the intraspecific variations have a destabilizing effect on the protein while SDM predicts that the intraspecific variants have largely a stabilizing/neutral effect on the protein. The ΔΔG predictions (by I-Mutant 2.0) for *S. enterica* variants show that most of the variants have a destabilizing effect on the protein. ΔΔG predictions using SDM were not possible for *S. enterica* variants as a crystal structure of *S. enterica* DHFR is not available on the PDB database. This result is unexpected as these variations exist in nature and destabilizing variants are usually preferred as they generally cannot be selected over a more stable variant. A possible explanation is that these variants may provide an advantage over the wild-type sequence in specific environmental conditions.

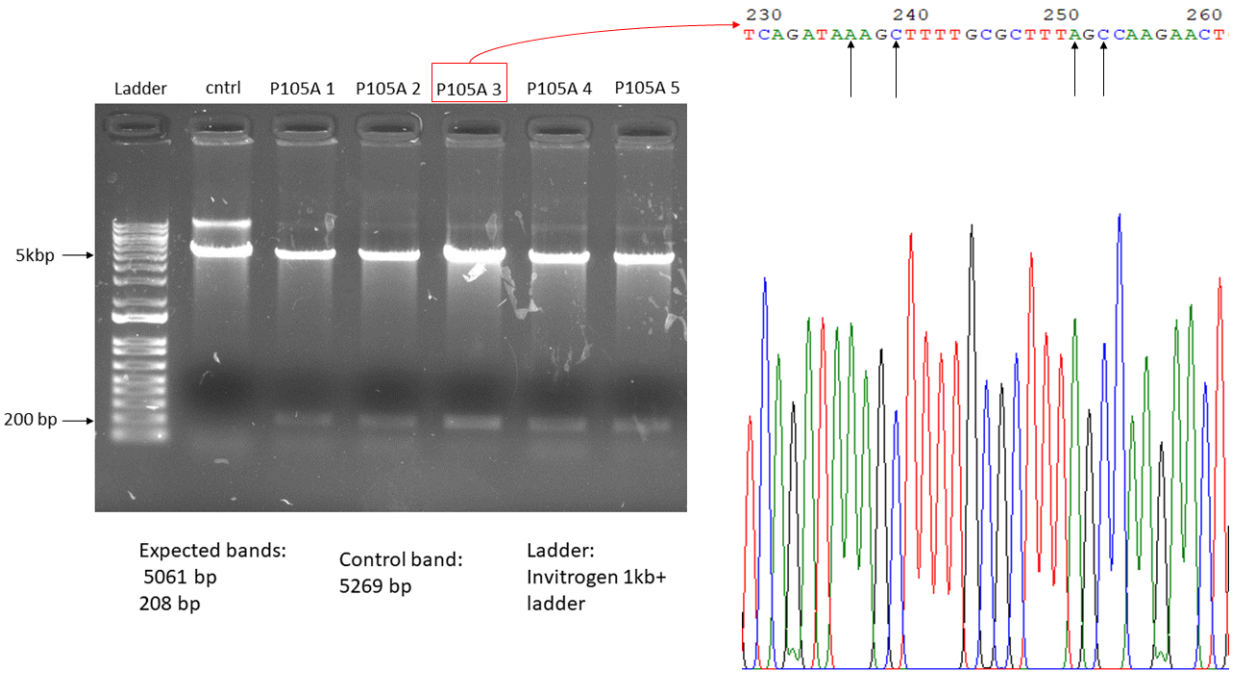


Figure 26: **Confirmation of the mutation P105A through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzyme *Hind*III. The right side depicts sequencing result of one of the plasmids showing the expected DNA fragments after restriction digestion. The arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer Materials and Methods). cntrl = plasmid (pPRO-His-*folA*) used as the negative control.

In order to ascertain the effects of an arginine residue at the 47th position and an alanine residue at the 105th position of *E. coli* DHFR, the W47 and the P105 residues of *E. coli* were mutated into an arginine and an alanine residue respectively. These mutations were brought into *E. coli* DHFR using the pPROB plasmid containing the *folA* gene and performing site-directed mutagenesis on the plasmid to generate mutants of interest (W47R and P105A) and the IC50 values of the MG1655 strain containing the above mutations were measured using 2-fold broth dilution. The data showing the confirmation of the expected mutations using restriction digestion followed by sequencing for the mutations W47R and P105A are shown in Fig 25 and 26 respectively. The IC50 values are summarized in Fig 27.

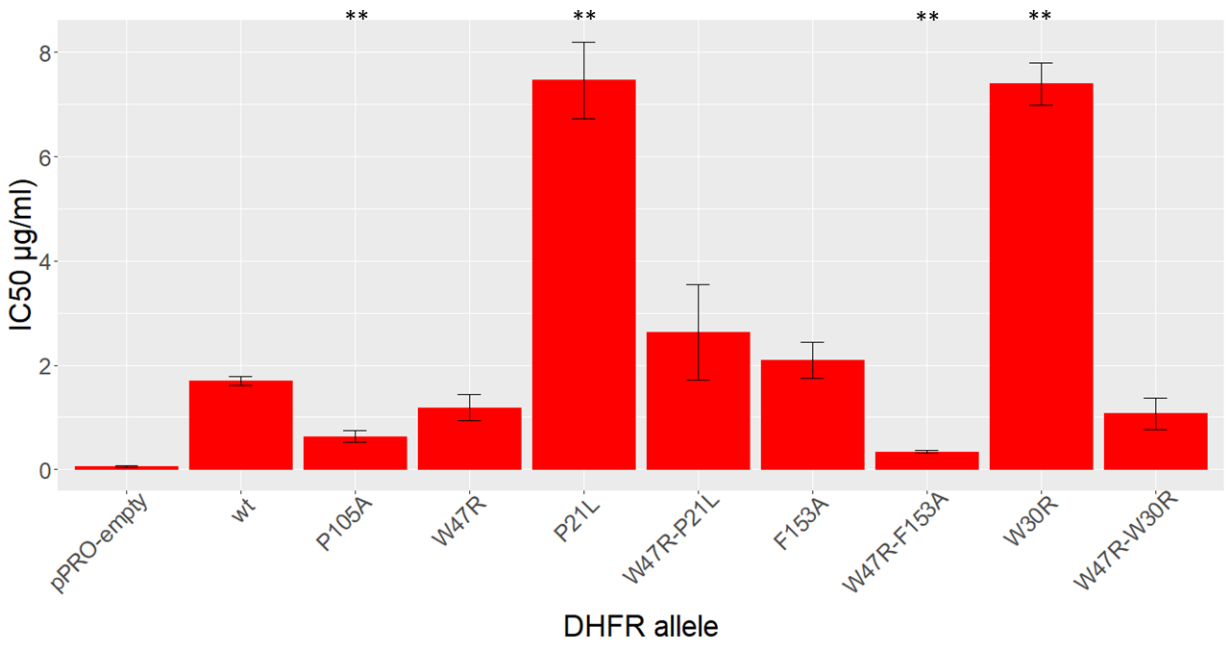


Figure 27: **IC50 values of *E. coli* MG1655 with plasmids containing *folA* with the mutations W47R, P21L, F513A, W30R, and the double mutants W47R-P21L, W47R-F153A, and W47R-W30R.** * implies p-value <0.05, ** implies p-value <0.005. p-values were determined using the Welsh two-sample t-test. The t-tests were performed by comparing the strains containing the mutant DHFR to the strains containing the wt DHFR. The IC50 values plotted are the mean of 3 replicates. The error bars represent the standard deviation derived from the 3 replicates. wt = wild-type = MG1655 pPRO-His-*folA*. The y-axis has been broken and the IC50 values from 15µg/ml to 60µg/ml are not represented in the graph.

The mean IC50 value of the mutant P105A is statistically significantly different from wild-type *E. coli* DHFR (two sample t-test p-value ~ 0.0003) (Fig 26). However, there isn't a huge difference in the actual mean IC50 values, suggesting that the phenotypic effect of the W30R-Y151F double mutant is not different from that of wild-type. The data shows that the intraspecific variants W47R and P105A do not have significantly different intrinsic resistance to TMP in *E. coli*. Hence the data shows that intraspecific variants (in this scenario) have no effect on protein function (in this case, resistance to TMP).

The W47R variant has similar levels of intrinsic resistance compared to the wild-type. However, the W47R variant could have epistatic effects on the protein. It is possible that the presence of the W47R variant modifies the structural landscape of DHFR in a manner such that the effect of a subsequent mutation in the protein is altered. The W47R variant could alter the effects of other mutations such that the net effect of having a mutation in *E. coli* DHFR is vastly different from having the same mutation in *E. coli* DHFR along with

the W47R variant. To verify if this was the case, the W47R variant was introduced into *E. coli* DHFR along with other previously known resistance-conferring mutations and the IC50 values of the MG1655 strain containing the mutation of interest (the mutation was introduced into pPRO-His-*folA* which was then transformed into *E. coli* MG1655) as well as the IC50 values of the MG1655 strain containing the mutation of interest along with the W47R variant (the mutation of interest as well as W47R was introduced into pPRO-His-*folA* which was then transformed into *E. coli* MG1655). The double mutants (W47R along with a mutation of interest) were generated by performing site-directed mutagenesis using primers that introduce the W47R mutation into pPRO-His-*folA* that already contains the mutation of interest.

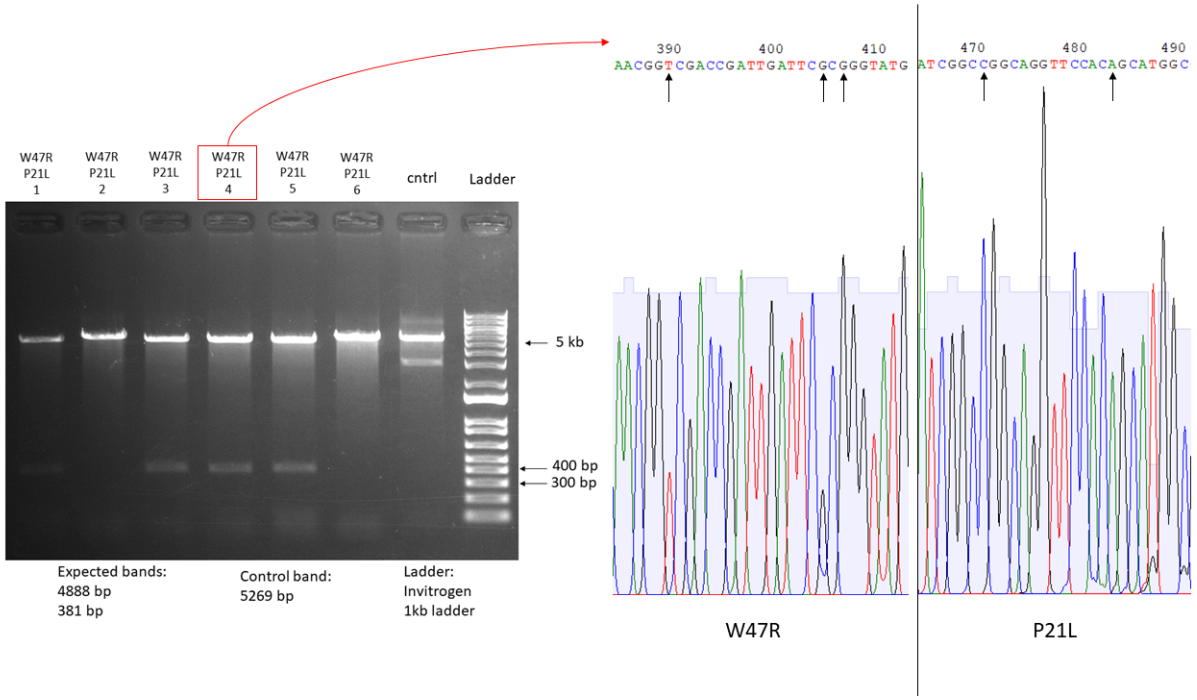


Figure 28: **Confirmation of the double mutation W47R-P21L through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzymes *Hind*III and *Sal*I. The right side depicts sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The two sections shown are the sections in the *folA* gene where the expected mutations have been generated. The black arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. The red arrow depicts the plasmid isolate that was selected and sent for sequencing. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer Materials and Methods). cntrl = plasmid (pPRO-His-*folA* P21L) used as negative control.

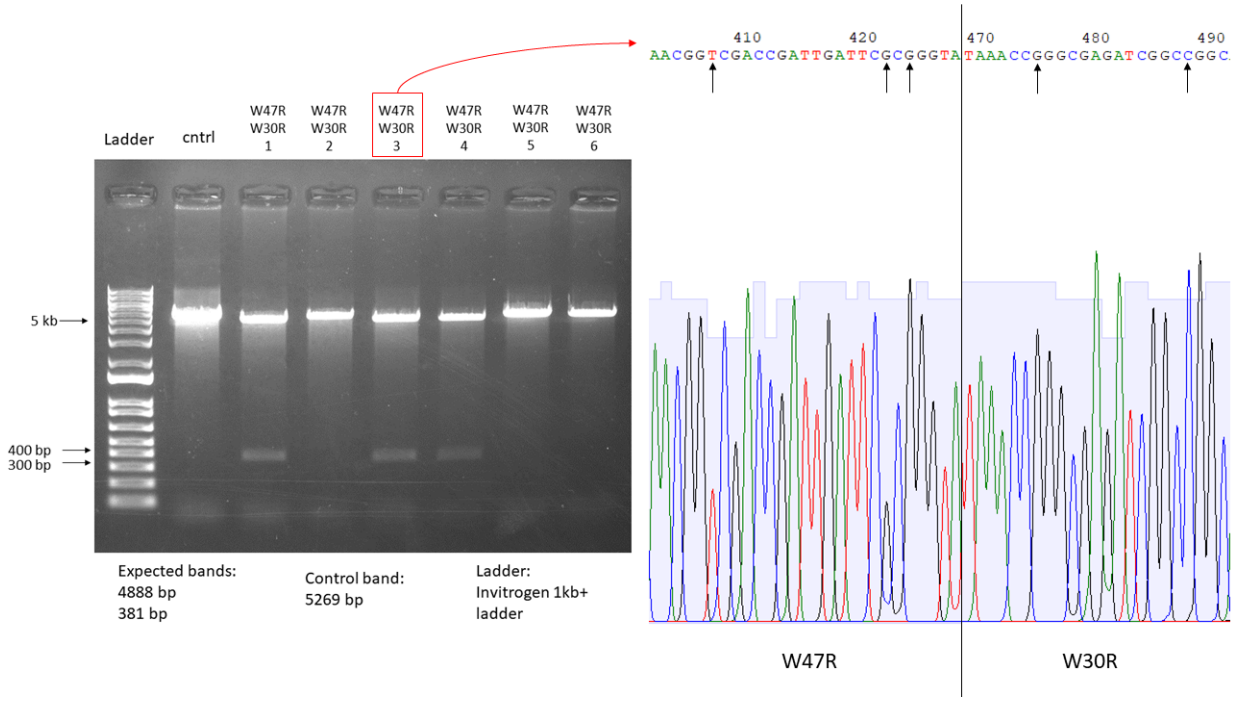


Figure 29: **Confirmation of the double mutation W47R-W30R through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzymes *Hind*III and *Sal*I. The right side depicts sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The two sections shown are the sections in the *folA* gene where the expected mutations have been generated. The black arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. The red arrow depicts the plasmid isolate that was selected and sent for sequencing. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer Materials and Methods). cntrl = plasmid (pPRO-His-*folA* W30R) used as negative control.

To determine potential mutations that could be affected in an epistatic manner by the W47R variant, the structure of *E. coli* DHFR was analyzed to search for positions previously associated with TMP resistance in *E. coli* DHFR located in the vicinity of W47. If a position lies near W47, then there is a possibility that W47 (or the R47 variant) directly interacts with the residue (or the mutated residue that confers resistance) in the position involved in resistance, affecting how mutations at that position affect TMP resistance. However, no positions associated with TMP resistance were located near W47. The closest position to W47R that is associated with TMP resistance is I94 which is located around 8Å away from W47. Since no position associated with TMP resistance is located in the vicinity of W47, positions associated with TMP resistance were chosen such that they were located at different parts of the linear amino acid chain of the protein. The positions chosen were: P21, W30, I94, and F153. The mutations P21L, W30R, I94L, and F153A have been previously associated with TMP resistance in *E. coli* DHFR (Matange et al., 2018).

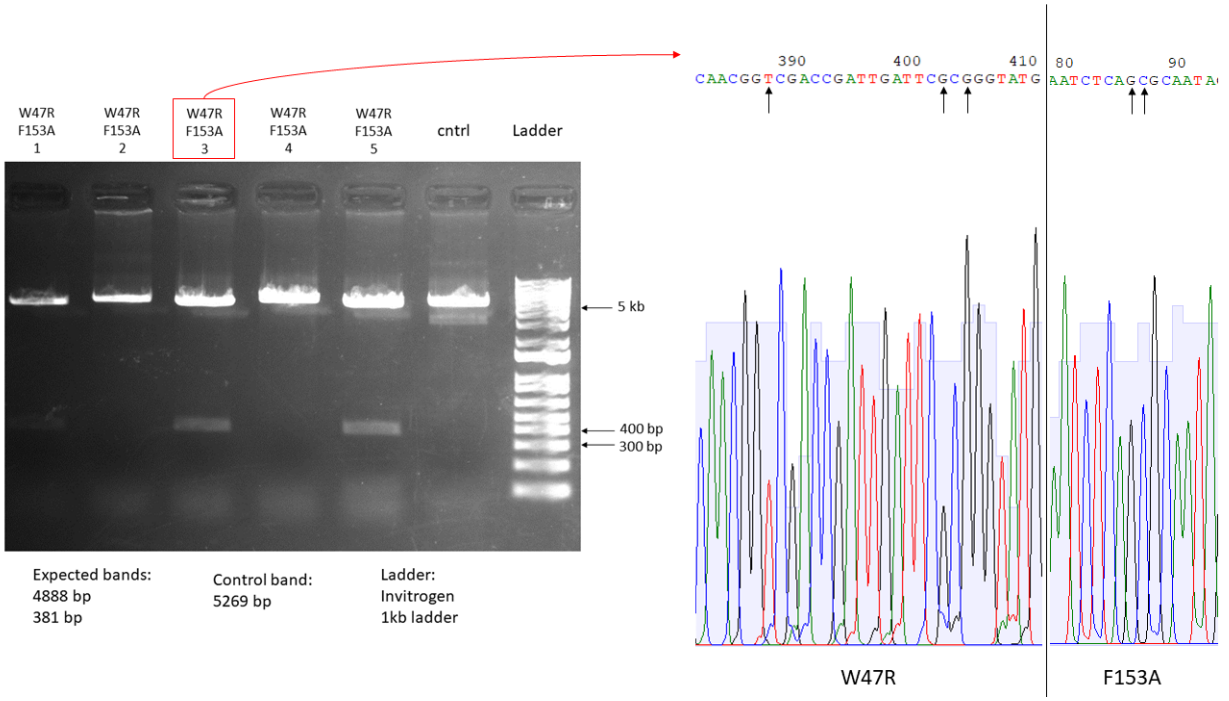


Figure 30: **Confirmation of the double mutation W47R-F153A through Restriction Digestion and Sequencing.** The left side depicts agarose (1%) gel images of mutagenized plasmid DNA after restriction digestion using the enzymes *Hind*III and *Sal*I. The right side depicts sequencing results of one of the plasmids showing the expected DNA fragments after restriction digestion. The two sections shown are the sections in the *folA* gene where the expected mutations have been generated. The black arrows point towards positions where the expected mutation has been generated in the *folA* gene in the plasmid. The red arrow depicts the plasmid isolate that was selected and sent for sequencing. Sequencing was performed using the pBAD33 reverse sequencing primer (for primer sequence refer Materials and Methods). cntrl = plasmid (pPRO-His-*folA* F153A) used as negative control.

To understand whether the W47R variant can alter the phenotypic effect of resistance-conferring mutations, double mutations were generated in *E. coli* DHFR. The double mutants W47R-P21L, W47R-W30R, W47R-I94L, and W47R-F153A were generated in *E. coli* DHFR. Site-directed mutagenesis was performed on a plasmid (pPROB) containing the *E. coli folA* gene (which synthesizes the DHFR enzyme) with a mutation of interest (P21L, W30R, I94L, and F153A) and the doubly mutated plasmid was then transformed in the *E. coli* MG1655 strain. The data showing the confirmation of the expected mutations using restriction digestion followed by sequencing for the double mutations W47R-P21L, W47R-W30R, and W47R-F153A are shown in Fig 28-30 respectively. The site-directed mutagenesis is ongoing for the W47R-I94L double mutant The IC50 values for the strains containing the mutagenized plasmids (except for the mutation W47R-I94L) are summarized in Fig 27.

The IC50 values of the *E. coli* MG1655 pPRO-His-*folA* double mutants (i.e., W47R-

P21L, W47R-W30R, and W47R-F153A) are significantly lower than that of the respective *E. coli* MG1655 pPRO-His-*folA* single mutants (i.e., P21L, W30R, and F153A respectively). The 2 sample t-test p-values for the 3 pairs of mutants are as follows: P21L vs W47R-P21L = 0.002, F153A vs W47R-F153A = 0.012, and W30R vs W47R-W30R = 5×10^{6}.

In each case (single vs double mutant), the IC50 values of the double mutants are at least 3 times lower than that of the single mutant (Fig 27). Overall, the data shows that there is a significant decrease in IC50 values when the resistance-conferring mutation is present along with the W47R variant (i.e., in a W47R background) than when the resistance-conferring mutation is present in wild-type *E. coli* DHFR. The data suggests that the W47R variant alters the effects of resistance-conferring mutations, in this case by causing a decrease in the IC50 values when the resistance-conferring mutation is present in a W47R background compared to when the resistance-conferring mutation is present in a wild-type background. The W47R variant is hence capable of affecting the extent of antibiotic resistance conferred by a mutation. Overall, the data suggest that intraspecific variation in a protein is capable of altering the phenotypic effects of other mutations (mutations in the same protein) on the same protein.

What is the mechanism behind the epistatic effect that the W47R variant has on resistance-conferring mutations? W47R is predicted by I-Mutant 2.0 and SDM to have a destabilizing effect on *E. coli* DHFR (albeit the $\Delta\Delta$G predictions by SDM predict that the W47R variant has a mildly destabilizing or neutral effect). However, this destabilizing effect may not be prominent enough to alter the phenotypic effect (i.e., TMP resistance) of DHFR. Since most resistance-conferring mutations have a destabilizing effect on the protein (as shown by $\Delta\Delta$G predictions by I-Mutant 2.0 and SDM, refer to Section 1 of the Results and Discussion Section), it is possible that having a destabilizing resistance-conferring mutation along with the W47R variant (which is also predicted to be destabilizing) causes a synergistic effect that further destabilizes the protein. This in turn could lower DHFR protein levels in the bacterial cell, resulting in the phenotype observed in the above data.

Immunoblotting can be performed on the lysates of the MG1655 strains containing the double mutants in order to verify whether protein levels expression is affected in any manner in the double mutants. These experiments are still ongoing.

4 Conclusions

The investigation of intraspecific variation in DHFR has revealed that the variation observed in nature is non-random and that certain positions in a protein have preferences with regard to the amino acid that occupies that position. The analysis of the position L28 in *E. coli* DHFR has revealed that amino acids with an amine group (such as Arg, Lys, and Gln) can confer TMP resistance at that position in *E. coli* DHFR. This result can be explained by the ability of the amine group to form extra hydrogen bonds with the substrate (dihydrofolate), leading to the stabilization of the substrate in the active site of DHFR and preventing the effective binding of TMP to DHFR. Hence, interspecific variation is capable of affecting protein function, in this case through a change in the active site that allowed for the formation of additional hydrogen bonds between the substrate and the protein.

The investigation and analysis of the position W30 and Y151 in *E. coli* DHFR reveal that the mutation Y151D increases the sensitivity of *E. coli* to TMP, which could be associated with the loss of hydrophobicity that position due to the mutation. The double mutant W30R-Y151D (which emerges in evolution experiments) is capable of conferring TMP resistance only in a Δlon background which can be due to the protein with the double mutation being abnormal and can only be accommodated in the absence of the Lon protease. Hence, the effect of a single mutation and a combination of mutations can have vastly different effects on the protein and these effects are often contingent on the genetic background of the organism.

The pattern of intraspecific variation in the DHFR proteins between two closely related species (*E. coli* and *S. enterica*) is different, showing that each species has its own unique pattern of intraspecific variation. The analysis of the intraspecific variants W47R and P105A has shown that these variants don't have a significant effect on TMP resistance. However, the presence of the variant W47R affects the extent of resistance provided by resistance-conferring mutations. This effect could be due to the effect of having multiple destabilizing mutations in the protein could lower DHFR protein levels in the bacterial cell, resulting in the phenotype observed. Hence, intraspecific variation in a protein may have indirect effects (i.e., can modify the effects of other mutations) on protein function.

5 Future Directions

In order to confirm that the ΔΔG predictions by the two tools are accurate and are not due to the possible biases of the two tools, the ΔΔG values for all the positions associated with TMP resistance could be predicted using a third tool, such as PackPred. The ΔΔG values predicted by the three tools then can be analyzed together, thus reducing the possibility that any patterns observed in the data are due to the inherent biases of the tools.

The interspecific variants L28K and L28Q confer resistance to TMP while the variant L28F does not. In order to understand if protein stability is affected in any way and whether this would have an impact on the ability of the mutation to confer resistance, protein stability assays can be carried out on all the variants.

To verify whether the presence of the amine group is crucial in the ability of L28 mutant in *E. coli* DHFR to confer resistance, site-directed mutagenesis could be performed to bring in other mutations (such as the mutating L28 to different hydrophilic and hydrophobic residues) to verify if the presence of the amine group is the only important factor that needs to be considered for the mutations to confer resistance to TMP.

If the MG1655 *Δlon* W30R-Y151D variant performs better than the MG1655 W30R variant due to the destabilizing effect of the double mutant, then this hypothesis can be verified by performing protein stability assays on both strains. By comparing mutant protein levels in the two strains, it could be verified whether it is the effect of the double mutant on protein stability that alters the phenotypic effect of the protein.

To study intraspecific variation in *S. enterica* DHFR and understand the relationship between variation and protein structure, a structure of *S. enterica* DHFR can be obtained using AlphaFold. This predicted structure can then be used to obtain ΔΔG predictions using I-Mutant 2.0 and SDM which will facilitate a better understanding of how protein sequence can affect protein structure. The variants observed can further be brought into *S. enterica* DHFR using site-directed mutagenesis and studied to understand the effect intraspecific variation will have on protein function.

The *E. coli* DHFR intraspecific variant P105A does not seem to have an impact on IC50 values (and hence antibiotic resistance) on its own. Other resistance-conferring mutations can be brought into *E. coli* DHFR along with the variant P105A (similar to what was done w.r.t W47R) to understand whether this variant is also capable of affecting the ability of resistance-conferring mutations to confer resistance to antibiotics.

References

[1] Abdizadeh, H, Tamer, YT, Acar, O, Toprak, E, Atilgan, AR, and Atilgan, C (2017). Increased substrate affinity in the Escherichia coli L28R dihydrofolate reductase mutant causes trimethoprim resistance. Physical Chemistry Chemical Physics 19, 11416–11428.

[2] Anwer, MU, Boikoglou, E, Herrero, E, Hallstein, M, Davis, AM, Velikkakam James, G, Nagy, F, and Davis, SJ (2014). Natural variation reveals that intracellular distribution of Elf3 protein is associated with function in the circadian clock. ELife 3.

[3] Bershtein, S, Choi, J-M, Bhattacharyya, S, Budnik, B, and Shakhnovich, E (2015). Systems-Level Response to Point Mutations in a Core Metabolic Enzyme Modulates Genotype-Phenotype Relationship. Cell Reports 11, 645–656.

[4] Bershtein, S, Mu, W, and Shakhnovich, EI (2012). Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. Proceedings of the National Academy of Sciences 109, 4857–4862.

[5] Bigman, LS, and Levy, Y (2018). Stability effects of protein mutations: The role of long-range contacts. The Journal of Physical Chemistry B122, 11450–11459.

[6] Blair, JM, Richmond, GE, Piddock, LJ (2014) Multidrug efflux pumps in Gram-negative bacteria and their role in antibiotic resistance. Future Microbiol 9, 1165–1177.

[7] Blair, JM, Webber, MA, Baylay, AJ (2015). Molecular mechanisms of antibiotic resistance. Nat Rev Microbiol 13, 42–51.

[8] BUSHBY, SRM, and HITCHINGS, GH (1968). TRIMETHOPRIM, A SULPHONAMIDE POTENTIATOR. British Journal of Pharmacology and Chemotherapy 33, 72–90.

[9] C Reygaert, W (2018). An overview of the antimicrobial resistance mechanisms of bacteria. AIMS Microbiology 4, 482–501. Cao, H, Gao, M, Zhou, H, and Skolnick, J (2018). The crystal structure of a tetrahydrofolate-bound dihydrofolate reductase reveals the origin of slow product release. Communications Biology 1.

[10] Capriotti, E, Fariselli, P, and Casadio, R (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Research 33, W306–W310.

[11] Christaki, E, Marcou, M, and Tofarides, A (2019). Antimicrobial resistance in bacteria: Mechanisms, evolution, and persistence. Journal of Molecular Evolution 88, 26–40.

[12] Cornaglia, G, Mazzariol, A, Fontana, R, and Satta, G (1996). Diffusion of carbapenems through the outer membrane of Enterobacteriaceae and correlation of their activities with their periplasmic concentrations. Microbial Drug Resistance 2, 273–276.

[13] Cox, G, and Wright, GD (2013). Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions. International Journal of Medical Microbiology 303, 287–292.

[14] Culture Collections. Culture Collections. Available at: https://www.culturecollections.org.uk/collections/nctc-3000-project.aspx.

[15] Dawson-Scully, K, Armstrong, GA, Kent, C, Robertson, RM, and Sokolowski, MB (2007). Natural variation in the thermotolerance of neural function and behavior due to a cgmp-dependent protein kinase. PLoS ONE 2.

[16] Desforges, JF, Stamm, WE, and Hooton, TM (1993). Management of Urinary Tract Infections in Adults. New England Journal of Medicine 329, 1328–1334.

[17] Echave, J, Spielman, SJ, and Wilke, O. (2016). Causes of evolutionary rate variation among protein sites. Nature Reviews Genetics 17, 109–121.

[18] Fierke, CA, Johnson, KA, and Benkovic, SJ (1987). Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from Escherichia coli. Biochemistry 26, 4085–4092.

[19] Fitch, WM (1970). Distinguishing homologous from analogous proteins. Systematic Zoology 19, 99.

[20] Gasch, AP, Payseur, BA, and Pool, JE (2016). The power of natural variation for Model Organism Biology. Trends in Genetics 32, 147–154.

[21] Hawkey, PM, (2003). Mechanisms of quinolone action and microbial response. J Antimicrob Chemoth 1, 28–35.

[22] Hershberg, R (2015). Mutation—the engine of evolution: Studying mutation and its role in the evolution of bacteria: Figure 1. Cold Spring Harbor Perspectives in Biology 7.

[23] Horner, D, and Pesole, G (2003). The estimation of relative site variability among aligned homologous protein sequences. Bioinformatics 19, 600–606.

[24] Huovinen, P, Sundström, L. Swedberg, G, and Sköld, O (1995). Trimethoprim and sulfonamide resistance. Antimicrobial Agents and Chemotherapy 39, 279–289.

[25] Jo, I, Hong, S, Lee, M, Song, S, Kim, J-S, Mitra, AK, Hyun, J, Lee, K, and Ha, N-C (2017). Stoichiometry and mechanistic implications of the MacAB-TolC tripartite efflux pump. Biochemical and Biophysical Research Communications 494, 668–673.

[26] Jordan, DM, Ramensky, VE, and Sunyaev, SR (2010). Human allelic variation: Perspective from protein function, structure, and evolution. Current Opinion in Structural Biology 20, 342–350.

[27] Katzenberger, RJ, Chtarbanova, S, Rimkus, SA, Fischer, JA, Kaur, G, Seppala, JM, Swanson, LC, Zajac, JE, Ganetzky, B, and Wassarman, DA (2015). Death following traumatic brain injury in drosophila is associated with intestinal barrier dysfunction. ELife 4.

[28] KIMURA, MOTOO (1968). Evolutionary rate at the molecular level. Nature 217, 624–626.

[29] Krucinska, J, Lombardo, MN, Erlandsen, H, Estrada, A, Si, D, Viswanathan, K, and Wright, D L (2022). Structure-guided functional studies of plasmid-encoded dihydrofolate reductases reveal a common mechanism of trimethoprim resistance in gram-negative pathogens. Communications Biology 5.

[30] Kumar, A, and Schweiser, H (2005). Bacterial resistance to antibiotics: Active efflux and reduced uptake. Advanced Drug Delivery Reviews 57, 1486–1513.

[31] Kumar, S, Mukherjee, MM, Varela, MF (2013). Modulation of bacterial multidrug resistance efflux pumps of the major facilitator superfamily. Int J Bacteriol.

[32] Kumar, S, Stecher, G, Li, M, Knyaz, C, and Tamura, K (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution 35, 1547–1549.

[33] Leferink, NG, Antonyuk, SV, Houwman, JA, Scrutton, NS, Eady, RR, and Hasnain, SS (2014). Impact of residues remote from the catalytic centre on enzyme catalysis of copper nitrite reductase. Nature Communications 5.

[34] Martinez, JL (2014). General principles of antibiotic resistance in bacteria. Drug Discovery Today: Technologies 11, 33–39.

[35] Masters, PA, O'Bryan, TA, Zurlo, J, Miller, DQ, and Joshi, N (2003). Trimethoprim-Sulfamethoxazole Revisited. Archives of Internal Medicine 163, 402.

[36] Matange, N, Bodkhe, S, Patel, M, and Shah, P (2018). Trade-offs with stability modulate innate and mutationally acquired drug resistance in bacterial dihydrofolate reductase enzymes. Biochemical Journal 475, 2107–2125.

[37] Matthews, DA, Bolin, JT, Burridge, JM, Filman, DJ, Volz, KW, and Kraut, J (1985). Dihydrofolate reductase. The stereochemistry of inhibitor selectivity. Journal of Biological Chemistry 260, 392–399.

[38] MAYNARD SMITH, JOHN (1970). Natural selection and the concept of a protein space. Nature 225, 563–564.

[39] Mirny, LA, and Gelfand, MS (2002). Using orthologous and paralogous proteins to identify specificity-determining residues in bacterial transcription factors. Journal of Molecular Biology 321, 7–20.

[40] Osborne, A, Robichon, A, Burgess, E, Butland, S, Shaw, RA, Coulthard, A, Pereira, HS, Greenspan, RJ, and Sokolowski, MB (1997). Natural behavior polymorphism due to a cgmp-dependent protein kinase of drosophila. Science 277, 834–836.

[41] Oz, T, Guvenek, A, Yildiz, S, Karaboga, E, Tamer, YT, Mumcuyan, N, Ozan, VB, Senturk, GH, Cokol, M, Yeh, P, and Toprak, E (2014). Strength of selection pressure is an important parameter contributing to the complexity of antibiotic resistance evolution. Molecular Biology and Evolution. 31, 2387–2401.

[42] Pearson, WR, and Sierk, ML (2005). The limits of protein sequence comparison? Current Opinion in Structural Biology 15, 254–260.

[43] Pierce, J, Apisarnthanarak, A, Schellack, N, Cornistein, W, Maani, AA, Adnan, S, and Stevens, MP (2020). Global Antimicrobial Stewardship with a focus on low- and middle-income countries: A position statement for the International Society for Infectious Diseases. International Journal of Infectious Diseases 96, 621–629.

[44] Pinti, M, Gibellini, L, Nasi, M, De Biasi, S, Bortolotti, CA, Iannone, A, and Cossarizza, A (2016). Emerging role of Lon protease as a master regulator of mitochondrial functions. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1857, 1300–1306.

[45] Povolotskaya, IS, and Kondrashov, FA (2010). Sequence space and the ongoing expansion of the protein universe. Nature 465, 922–926.

[46] Prism - GraphPad. Prism - GraphPad. Available at: https://www.graphpad.com/scientific-software/prism/

[47] Raju, A, Kulkarni, S, Ray, MK, Rajan, MGR, and Degani, MS (2015). E84G mutation in dihydrofolate reductase from drug resistant strains of Mycobacterium tuberculosis (Mumbai, India) leads to increased interaction with Trimethoprim. International Journal of Mycobacteriology 4, 97–103.

[48] Ramirez, M.S, and Tolmasky, M.E (2010). Aminoglycoside modifying enzymes. Drug Resistance Updates 13, 151–171.

[49] Redgrave, L.S, Sutton, SB, Webber, MA, and Piddock, LJV (2014). Fluoroquinolone resistance: Mechanisms, impact on bacteria, and role in evolutionary success. Trends in Microbiology 22, 438–445.

[50] Rice, P, Longden, I, and Bleasby, A (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics. 16(6):276-277.

[51] Roberts, MC, (2003). Tetracycline therapy: update. Clin Infect Dis 36, 462–467.

[52] Roberts, MC, (2004). Resistance to macrolide, lincosamide, streptogramin, ketolide, and oxazolidinone antibiotics. Mol Biotechnol 28, 47–62.

[53] Robicsek, A, Strahilevitz, J. Jacoby, GA, Macielag, M, Abbanat, D, Hye Park, C, Bush, K, and Hooper, DC (2005). Fluoroquinolone-modifying enzyme: A new adaptation of a common aminoglycoside acetyltransferase. Nature Medicine 12, 83–88.

[54] S. Askari, B, and Krajinovic, M (2010). Dihydrofolate Reductase Gene Variations in Susceptibility to Disease and Treatment Outcomes. Current Genomics 11, 578–583.

[55] Sawaya, MR, and Kraut, J (1997). Loop and Subdomain Movements in the Mechanism of Escherichia coli Dihydrofolate Reductase: Crystallographic Evidence. Biochemistry 36, 586–603.

[56] Schnell, JR, Dyson, HJ, and Wright, PE (2004). Structure, Dynamics, and Catalytic Function of Dihydrofolate Reductase. Annual Review of Biophysics and Biomolecular Structure 33, 119–140.

[57] Shafer, RW, and Schapiro, JM (2008). HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. AIDS Rev. 10, 67–84.

[58] Shenoy, AR, and Visweswariah, SS (2003). Site-directed mutagenesis using a single mutagenic oligonucleotide and DpnI digestion of template DNA. Analytical Biochemistry 319, 335–336.

[59] Shrimpton, P, and Allemann, RK (2002). Role of water in the catalytic cycle of *E. coli* dihydrofolate reductase. Protein Science 11, 1442–1451.

[60] Tanabe, M, Szakonyi, G, Brown, KA, Henderson, PJF, Nield, J, and Byrne, B (2009). The multidrug resistance efflux complex, EmrAB from Escherichia coli forms a dimer in vitro. Biochemical and Biophysical Research Communications 380, 338–342.

[61] The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. Topham CM, Srinivasan N, Blundell TL (1997). Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. Protein Eng 10, 7–21.

[62] Vedantam, G, Guay, GG, Austria, NE, Doktor, SZ, and Nichols, BP (1998). Characterization of Mutations Contributing to Sulfathiazole Resistance in Escherichia coli. Antimicrobial Agents and Chemotherapy 42, 88–93.

[63] Villagra, NA, Fuentes, JA, Jofre, MR, Hidalgo, AA, Garcia, P, and Mora, GC (2012). The carbon source influences the efflux pump-mediated antimicrobial resistance in clinically important Gram-negative bacteria. Journal of Antimicrobial Chemotherapy 67, 921–927.

[64] Wimley, WC, & White, S H (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nature Structural & Molecular Biology, 3(10), 842–84.

[65] Xu, S, Chen, M, Feng, T, Zhan, L, Zhou, L, and Yu, G (2021). Use ggbreak to Effectively Utilize Plotting Space to Deal With Large Datasets and Outliers. Frontiers in Genetics 12.

Appendix

1. List of mutations generated in pPRO-His-folA using site-directed mutagenesis and the template and mutagenic primers used as well as the restriction site introduced.