

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Mapping structural and environmental constraints on the mutational landscape of functional proteins using high-throughput screening and computational modeling

Permalink

<https://escholarship.org/uc/item/9qf6v85h>

Author

Thompson, Samuel M

Publication Date

2020

Peer reviewed|Thesis/dissertation

Mapping structural and environmental constraints on the mutational landscape of functional proteins using high-throughput screening and computational modeling

by
Samuel Mark Thompson

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Tanja Kortemme

FA555ADA28F1439...

Tanja Kortemme

Chair

DocuSigned by:

Geeta Narlikar

DocuSigned by:
43B...

Geeta Narlikar

David Agard

328308DD7E5546A...

David Agard

Committee Members

Mapping structural and environmental constraints on the mutational
landscape of functional proteins using high-throughput screening and
computational modeling

Samuel Mark Thompson

June 9, 2020

Copyright 2020

by

Samuel Mark Thompson

For better or for worse, my PhD has been consisted of a great deal of solitary work. This independence has been at times a challenging isolation and a disappointment, but it has also been a tremendous gift of freedom. Generous funding from the NSF and the Chuan Lyu Fellowship and the research environment created by Tanja Kortemme in our lab at UCSF have all allowed me autonomy in my project that I did not deserve. I have also been fortunate to claim time from my thesis committee members – Geeta Narlikar and David Agard – both of whom I greatly admire and who have made themselves available whenever I have asked. Beyond this, my project was also built on a great deal of previous work and many new contributions. My computational work would not have been possible without the decades of hard-work within the Rosetta community nor without specific contributions from Noah Ollikainen, Colin Smith, Andrew Leaver-Faye, and most especially from Amelie Stein and Tianjiao Zhang. The experimental components of this project have their backbone in prior work from Kim Reynolds at UT Southwestern, who has somehow managed to be an exceptional collaborator and an accessible mentor from afar. Chris Ingle, Thuy Nguyen, and Victor Salinas from UT Southwestern have been patient and generous in providing materials and helping me troubleshoot selection and sequencing experiments via the internet. Alongside Chris, Anna Sellas, Norma Neff, and Rene Sit at the Chan-Zuckerberg BioHub, and Natasha Carli and Jim McGuire at the Gladstone Institute Genomics Core facilitated or performed all of the Illumina sequencing for my selection experiments, including several re-runs of samples that I prepared. Carol Gross's perception and insight saved my project when I did not know how to go from an intriguing observation to a true scientific discovery, and Melanie Silvis and Byoung Mo Koo from her lab provided expertise and materials that were essential to the experiments that revealed what I believe will prove to be the most important findings herein. Many of the experiments to support these findings would not have been completed in time without Sunny

Zhang's diligence and careful experimentation. In preparing the manuscript and this thesis, I am reminded of how Tanja has been instrumental in shaping how I conceptualize what comprises a novel scientific result and how I communicate science in text, in images, and in speech. Of course, I would be remiss not to dedicate the final product here to my grandfather of self-same name, as he revels in seeing his name in print.

To Steven, thank you for your patience, independence, love, support, enthusiasm, and laughter. Everything has been a richer experience with you.

This thesis was written during the shelter-in-place implemented to stem the spread of the Covid-19 pandemic and was completed during a time of protest against systemic racist violence against people of color, police violence against Black Americans in particular. It would be irresponsible not to acknowledge how privileged I have been to enjoy stable housing, continued income, and good health while isolating myself at home to write. I think I will always feel that I used my thesis deadlines as an excuse not to be more involved in combating the multiple crises that we are currently facing. That is a sobering and regretful thought.

Some text and figures in this dissertation are reprints of or modified from material as it appears in "Altered expression of a quality control protease in *E. coli* reshapes the in vivo mutational landscape of a model enzyme", *eLife*, **2020**. The co-author listed in this publication directed and supervised the research that forms the basis for this dissertation.

Mapping structural and environmental constraints on the mutational landscape of functional proteins using high-throughput screening and computational modeling

Samuel Mark Thompson

Abstract

Natural protein sequences are the result of optimization on a mutational landscape with multiple competing pressures. These pressures will arise from constraints imposed by selection for a particular fold and function within a cellular context. We can categorize these pressures as structural-functional and environmental. Nevertheless, it remains a challenge to quantify mutational landscapes with thousands of mutations and dissect the contributions from multiple constraints. Similarly, we generally do not know how to encode protein design models with the structural constraints that define a complex molecular function, even simplified *in vitro* environments. Reverse-engineering the multiple structural-functional and environmental pressures that were integrated to yield the mutational landscapes that produced natural proteins would improve our understanding of the cellular milieu and our ability to engineer new protein functions.

Using *E. coli* dihydrofolate reductase (DHFR) as a model system, we developed computation and experimental methods for identifying, quantifying, and modeling structural-functional and environmental constraints on functional proteins. **Chapter 1** of this thesis is a brief introduction to the concept of mutational landscapes and to our model enzyme. **Chapter 2** of this thesis (page 11) describes a multi-state modeling framework for encoding complex functions into protein design and the application of this framework to recovering evolutionary sequence preferences in DHFR. **Chapter 3** (page 54) describes the calibration of a high-throughput selection assay for DHFR activity and the mutational landscape for a library of all possible single point mutations to

DHFR. The final chapter (**Chapter 4**, page 132) describes the quantification of broad impacts to the DHFR mutation landscape from expression of Lon protease. The results in these three chapters show the impact of structural-functional and environmental constraints on sequence preferences from mutational landscapes. These data allow us to propose methods for engineering the behavior of entire mutational landscapes by modulating environmental constraints.

Contents

1	Introduction	1
1.1	DHFR	2
2	Modeling Functional Proteins as Multi-state Ensembles for Computational Design	11
2.1	Introduction	11
2.2	Background	12
2.2.1	Computational Protein Design	12
2.2.2	Overview of Model Types	13
2.2.3	Single-state Design Models	14
2.2.4	Ensemble Design Models	18
2.2.5	Multi-state Design Models	20
2.2.6	Evaluating Design Performance	25
2.2.7	Rosetta	27
2.2.8	Challenges for Multi-state Design	27
2.3	Results	28
2.3.1	A Multi-state Ensemble Design Model	28
2.3.2	Selecting a Multi-state Design Patch in DHFR	30
2.3.3	Multi-state Design on DHFR with Ensembles	31
2.4	Discussion	35
2.4.1	Future Directions	36
2.5	Methods	37

3	Deep Mutational Scanning to Resolve Highly Active Single-Point Mutants of a Model Enzyme	54
3.1	Introduction	54
3.2	Background	55
3.2.1	Deep Mutational Scanning	55
3.3	Results	58
3.3.1	Preliminary Selection on DHFR at UT Southwestern	58
3.3.2	Optimizing the Selection Pressure	64
3.3.3	Developing Non-selective Growth Conditions	74
3.3.4	Building a DMS Library of DHFR Single Point Mutants	78
3.3.5	Construction of an In-house Turbidostat	85
3.3.6	DMS on DHFR Under Optimized Conditions	86
3.3.7	Functionally Characterized Positions in DHFR are Enriched for Disadvantageous and Null Mutations	90
3.3.8	Advantageous Mutations to DHFR	92
3.3.9	Characterizing Top Advantageous Mutations to DHFR	93
3.4	Discussion	105
3.4.1	Future Directions	106
3.5	Methods	107
4	Altered Expression of Lon Protease Reshapes the Mutational Landscape of a Model Enzyme	132
4.1	Introduction	132
4.2	Background	133

4.2.1	Protein Homeostasis in Cells	133
4.2.2	Interplay between Protein Homeostasis in Cells and Protein Function . . .	134
4.2.3	Lon Protease	135
4.3	Results	136
4.3.1	Generating a Lon-expressing Selection Strain	136
4.3.2	Repeating Selection with a Lon-expressing Strain	141
4.3.3	Population-level Mutational Analysis of Selection Data	142
4.3.4	Categorizing Positions by Mutational Impact and Position-based Analysis	144
4.3.5	Lon Impact as Δ Selection Coefficient	145
4.3.6	Structural Patterns in DHFR Selection	152
4.3.7	Comparing Sequence Preferences in Selection and Evolution	155
4.3.8	Examining the Lon Locus in <i>E. coli</i> Genome from Seminal Experiments .	158
4.4	Discussion	159
4.4.1	Discussion	161
4.5	Methods	163
5	Appendix	174

List of Figures

1.1	Hydride transfer step in the DHFR catalyzed reduction of DHF to THF	3
1.2	Crystal structure models of DHFR conformations adopted during the catalytic cycle.	4
2.1	Process flowchart for single-state, ensemble, and multi-state design models.	15
2.2	Application of the Metropolis criterion to simulated annealing.	17
2.3	A sigmoidal function for state fitness in multi-state design.	22
2.4	Parameters in the sigmoid fitness function.	23
2.5	Comparison of Workflows in Multi-state Ensemble Models to Single-state, Ensemble, and Multi-state Models.	29
2.6	Macrostates and microstates in a multi-state ensemble model.	30
2.7	Selection of residues for multi-state design in DHFR.	31
2.8	Multi-state design on DHFR using energies from enumerated amino acid substitutions.	32
2.9	Multi-state design on DHFR using first-order statistical energies from ensemble design.	34
3.1	Selection coefficients from deep mutational scanning	56
3.2	Conceptual diagram of a turbidostat to maintain constant cell culture density.	59
3.3	Counts per mutant at $t=0$ during preliminary selection.	60
3.4	Analysis of the error in preliminary selection coefficients.	61
3.5	Distribution of selection coefficients from preliminary selection experiment.	61
3.6	Heatmaps of selection coefficients from preliminary selection experiment.	62
3.7	Amino acid preferences from selection compared to preferences from DHFR orthologues.	63

3.8	Expression of FLAG-tagged DHFR and TYMS off the selection plasmid.	66
3.9	Selection strain growth rate with TYMS induction off selection plasmid.	66
3.10	Selection using an "AACGAG" mutated RBS sequence for DHFR yields selection coefficients that correlate poorly with preliminary selection.	68
3.11	Selection using an "AACGAG" mutated RBS sequence for DHFR yields selection coefficients that correlate poorly with <i>in vitro</i> kinetics.	69
3.12	Total DHFR expression from the optimized selection plasmid is qualitatively decreased relative to endogenous expression.	69
3.13	Soluble DHFR expression from the optimized selection plasmid is quantitatively decreased relative to endogenous expression.	71
3.14	Correlation between <i>in vitro</i> kinetics and growth rates for DHFR point mutants shows increasing selection pressure.	73
3.15	Selection strain growth rates in M9 minimal medium without and with the " <i>folA</i> " supplement mix.	75
3.16	Selection strain growth curves in M9 minimal medium without and with the " <i>folA</i> " supplement mix.	75
3.17	Growth rates in the turbidostat after overnight growth in M9 minimal medium supplemented with the " <i>folA</i> mix".	76
3.18	Selection strain growth curves in supplemented M9 minimal medium without and with adenosine and thymidine drop-outs to the medium.	77
3.19	Selection strain growth rates in supplemented M9 minimal medium.	77
3.20	Growth rates in the turbidostat after overnight growth in supplemented M9 minimal medium.	78

3.21 Preliminary mutant frequencies in DHFR point mutant sublibraries.	80
3.22 Preliminary background frequencies in DHFR point mutant sublibraries.	80
3.23 Comparison of mutant frequencies in sublibraries and as background.	82
3.24 Comparison of background mutant frequencies in two independent sequencing experiments.	82
3.25 Final mutant frequencies in DHFR point mutant sublibraries.	83
3.26 Final background frequencies in DHFR point mutant sublibraries.	84
3.27 Example OD measurements from a selection experiment using the in-house turbidostat.	86
3.28 Schematic for the in-house turbidostat.	87
3.29 Error and reproducibility in DHFR DMS.	88
3.30 <i>E. coli</i> DHFR deep mutational scanning uncovers many advantageous mutations.	89
3.31 Variation in selection coefficients for DHFR DMS.	90
3.32 Residues previously shown to have a functional role displayed on the DHFR structure.	91
3.33 Growth curves for top advantageous mutations.	92
3.34 Example positions with multiple advantageous mutations hypothesized to be destabilizing.	93
3.35 DHFR structure with mutational hot-spots.	94
3.36 Structural context for hotspot residues from Figure 3.35	95
3.37 Lysate activity for DHFR wild-type and point mutants on the selection plasmid.	96
3.38 <i>In vitro</i> velocities of purified DHFR wild-type and point mutants.	97
3.39 Soluble cellular abundance for DHFR wild-type and point mutants on the selection plasmid.	100

3.40	Selection coefficient compared to predictions of WT DHFR and point mutant activity from cellular abundance and <i>in vitro</i> velocity measurements.	101
3.41	Zoom in for selection coefficient compared to predictions of WT DHFR and point mutant activity from cellular abundance and <i>in vitro</i> velocity measurements. . .	102
3.42	Cellular abundance versus <i>in vitro</i> velocity for DHFR wild-type and point mutants.	103
3.43	Thermal denaturation curves monitored by CD signal at 225 nm for selected hotspot mutants.	104
4.1	Change in DHFR mutant dependent growth rates in the selection strain and a selection strain with a restored endogenous Lon promoter.	136
4.2	Change relative to WT in DHFR mutant-dependent growth rates in the selection strain and a selection strain with a restored endogenous Lon promoter.	137
4.3	Expression of Lon under a constitutive promoter in the re-engineered selection strain.	139
4.4	Change in DHFR mutant-dependent growth rates in the selection strain \pm Lon. .	140
4.5	Change relative to WT in DHFR mutant-dependent growth rates in the selection strain \pm Lon.	140
4.6	Quality of the selection under +Lon conditions.	141
4.7	Relationship between error and selection coefficient for +Lon selection.	142
4.8	Lon protease expression reshapes the mutational landscape.	143
4.9	Comparison of selection coefficients \pm Lon.	144
4.10	Δ selection coefficients show Lon impact.	147
4.11	Δ selection coefficients show Lon impact on natively buried and hydrophobic residues.	148
4.12	Correlation of Δ selection coefficients to T_m	150

4.13	Correlation of Δ selection coefficients to change in cellular abundance \pm Lon. . .	152
4.14	Structural characterization of multiple constraints on the DHFR mutational landscape.	153
4.15	Selection coefficients under the two Lon expression regimes mapped on the DHFR structure.	154
4.16	Burial of residues within each mutation response category reported as the mean number of atomic neighbors.	155
4.17	Residues in mutational response categories in the $-$ Lon selection as a function of distance from several sites in the DHFR structure.	156
4.18	Comparison of DHFR per-position sequence preferences.	157
4.19	Concept for modulating the mutational landscape with Lon to combat antibiotic resistance.	162

List of Tables

3.1	Michaelis-Menten kinetics parameters for characterized DHFR point mutants . . .	65
3.2	Predicted strength for DHFR ribosome binding sites.	70
3.3	Soluble expression of DHFR measured from lysate activity for heterologous expression from plasmid and for endogenous expression.	70
3.4	Growth rates for the ER2566 $\Delta folA/\Delta thyA$ selection strain with a panel of 12 DHFR mutants and 3 DHFR RBS sequences.	72
3.5	Single point mutant library coverage and background.	85
3.6	<i>In vitro</i> velocity for selected advantageous mutations.	98
3.7	Soluble DHFR abundance levels in molecules per cell	99
3.8	Apparent T_m values from thermal denaturation experiments.	100
4.1	Mutational response category and burial classification for DHFR positions. . . .	146
4.2	Soluble DHFR abundance levels in molecules per cell	151
5.1	<i>E. coli</i> strains	174
5.2	Plasmids	174
5.3	Primers	174
5.5	Selection coefficients $\pm Lon$	179
5.6	<i>In vitro</i> DHFR reactions	243

Acronyms

aTC anhydrotetracycline.

COMETS Constrained Optimization of Multistate Energies by Tree Search.

DCA direct coupling analysis.

DEE dead-end elimination.

DHF dihydrofolate.

DHFR *E. coli* dihydrofolate reductase.

DMS deep mutational scanning.

DTT dithiothreitol.

EDTA ethylenediaminetetraacetic acid.

FACS fluorescence activated cell sorting.

FADS fluorescence activated droplet sorting.

HDX hydrogen-deuterium exchange mass spectroscopy.

IPTG isopropyl β -D-1-thiogalactopyranoside.

LTEE Long Term Evolution Experiment.

MES 2-ethanesulfonic acid.

MFPred Rosetta Mean Field Prediction.

MSA multiple sequence alignment.

MSE mean squared error.

NADH nicotinamide adenine dinucleotide (reduced).

NADPH nicotinamide adenine dinucleotide phosphate (reduced).

NMR nuclear magnetic resonance spectroscopy.

PCR polymerase chain reaction.

PDB Protein Data Bank.

PMSF phenylmethanesulfonyl fluoride.

QM quantum mechanics.

QMMM mixed quantum mechanics/molecular mechanics.

RBS ribosome binding site.

RECON Rosetta Restrained Convergence.

SL sublibrary.

THF tetrahydrofolate.

TYMS *E. coli* thymidylate synthase.

WT wild-type.

1 Introduction

The future of material science and biotechnology will lie in the control of complex molecular functions at ever broader scales with ever greater control. One of the most useful platforms for nanoscale engineering is protein. Within the central dogma of biology, proteins are the effector of the control sequence that is genomic DNA. Proteins' ability to perform complex functions such as catalysis has changed our atmosphere and shaped the surface of the planet. Thus, we have evidence of the quite literally world-changing potential within engineered proteins, but we do not fully know the limits of what proteins can accomplish or how to harness their potential.

Proteins are linear chains of amino acids. **Sequence space** is hyper-dimensional space of every possible amino acid over any length of protein. Within that vast combinatorial space are pockets where the protein sequence will adopt a 3D structure that is stable enough to be characterized. These regions of sequence space for well-folded proteins overlap with regions encoding proteins with complex function. Through the optimization process of evolution, natural proteins have explored these regions of functional protein sequence space. In 1932, Sewall Wright provided the conceptual metaphor of a **mutational landscape** or fitness landscape for understanding this process of optimization[1]. The fitness landscape is a topology over sequence space where organismal fitness is the elevation. Similarly, a mutational landscape is a topology over sequence space, but the elevation is growth rate, enzyme activity, or some other quantity that can be directly measured. **Mutational impacts** are the individual changes to the measured quantity for an arbitrary sequence relative to those for a reference point in sequence space (e.g. the wild-type (WT) sequence). This heuristic of topology predicts that protein sequences will mutate to ascend an upward fitness gradient such that highly functional sequences accumulate over time.

For the goal of protein engineering, it will be highly useful to understand **1)** what shapes the

mutational landscapes that optimize proteins and **2)** how we can build computational models that encode conceptually similar constraints on sequence space to optimize proteins for non-natural, designer purposes.

1.1 DHFR

We chose *E. coli* dihydrofolate reductase (DHFR) as a model enzyme system for computational and experimental mapping of mutational landscapes because DHFR has an established *in vivo* selection assay[2], has been extensively characterized biochemically[3–10], and has been structurally characterized with crystal structures that represent the major catalytic intermediates[11]. Furthermore, we anticipated that existing literature on DHFR would help us interpret mutational impacts to generate specific and generalizable hypotheses about complex protein function and then adapt existing protocols to test those hypotheses.

1.1.0.1 Biochemistry The native activity of DHFR is the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) via hydride transfer from an nicotinamide adenine dinucleotide phosphate (reduced) (NADPH) cofactor (**Figure 1.1**, page 3). The literature values for Michaelis-Menten kinetic parameters range from 3-7 s⁻¹ for k_{cat} , 0.15-3 μM for $K_{M,DHF}$ and ~0.1 μM for $K_{M,NADPH}$ [3, 4, 12, 13]. DHF is typically the limiting reagent in *in vitro* assays for DHFR activity due to the low cellular concentration of DHF relative to $K_{M,DHF}$. While DHFR turnover is 2-3 orders of magnitude lower than the diffusion limit and DHFR's kinetic parameters are comparable to the average for all measured enzymes [14, 15], the rate acceleration from DHFR is unmeasurable because the reduction of DHF by NADPH is not detected in the absence of DHFR.

The entire kinetic cycle of DHFR has been determined based on equilibrium and pre-steady

state kinetics measurements for all on- and off-rates in addition to the rate of hydride transfer[3]. The rate limiting step of the DHFR reaction is product release (12 s^{-1}), the rate of which is essentially identical to k_{cat} for the WT enzyme (7 s^{-1}), whereas the hydride transfer step is nearly 2 order of magnitude faster ($220\text{-}720 \text{ s}^{-1}$).

1.1.0.2 3D Structure Models from X-ray crystallography[11] and mixed quantum mechanics/molecular mechanics (QMMM) modeling[17] have been generated for major intermediates in the DHFR catalytic cycle (**Figure 1.2**, page 4). The conformational state of DHFR is defined by the identity of the bound ligands and the conformation of the M20 loop that folds over the active site. Models of the high-energy transfer state come from crystal structures of DHFR in complex with the inhibitor methotrexate that are believed to have transition state-like properties. Furthermore, QMMM simulations have been performed to model this transition state with the natural ligands.

Structural analyses of DHFR have revealed key aspects that are critical for enzyme function. Detailed nuclear magnetic resonance spectroscopy (NMR) experiments have linked the rate-limiting step (product release, k_{cat}) to the rate of fluctuations in the M20 loop that folds over the active site [5, 7]. The fluctuations of this loop have also been reported to be allosterically coupled with the exchange of the NADPH cofactor [18, 19]. Mutations that decrease overall enzyme velocity have acted by multiple mechanisms: disrupting hydride transfer[20], inhibiting

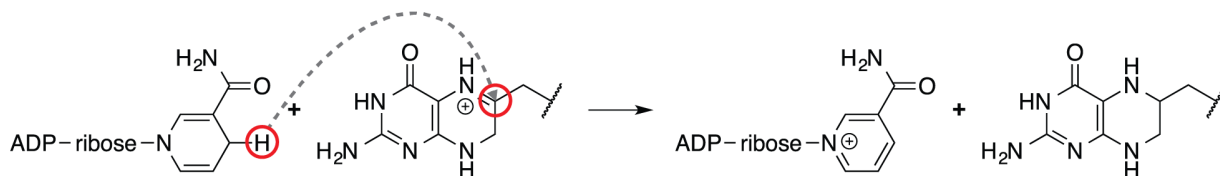


Figure 1.1: Hydride transfer step in the DHFR catalyzed reduction of DHF to THF. On the reaction side, a dashed line connects the hydride to the C6 carbon, which is electrophilic when the N5 nitrogen is protonated.

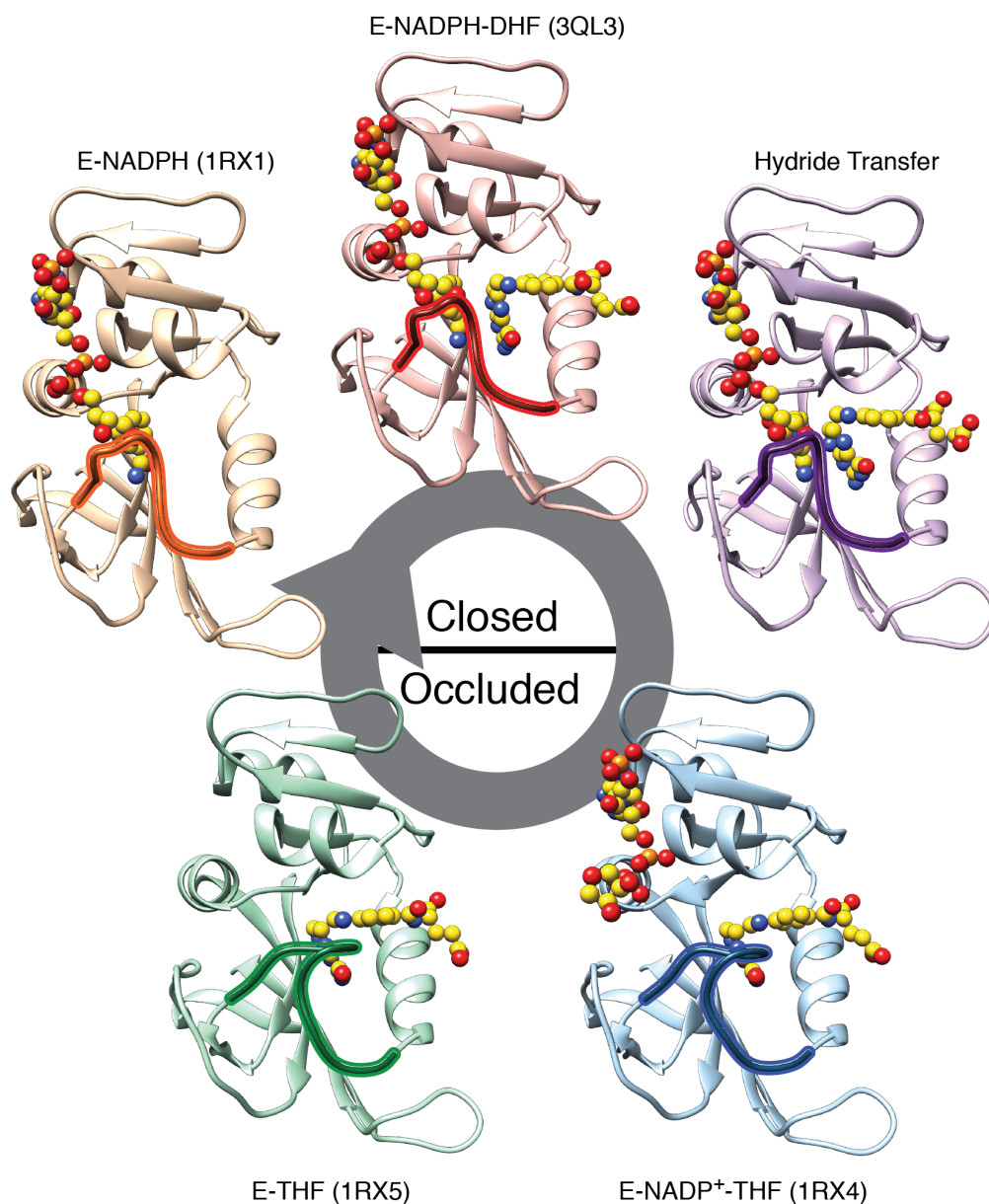


Figure 1.2: Crystal structures of DHFR (PDB IDs: 1RX1, 3QL3, 1RX4, and 1RX5) and a mixed QMMM model of the hydride transfer step[16] represent the conformational states adopted by DHFR over the catalytic cycle. The identity of each state is defined by the identity of the bound ligands (yellow spheres with heteroatom coloring) and the conformation of the M20 loop (outlined) that folds over the active site (closed or occluded). Upper models are in the closed state, and lower models are in the occluded state. All PDBs were downloaded from the PDB_REDO[17].

the protonation of nitrogen N5 and shifting the pK_a of the reaction[12], eliminating hydrogen bonds stabilizing the interactions between the M20 loop and the F-G and G-H loops[8, 9], and by altering the dynamics of the F-G and G-H loops[7, 21].

1.1.0.3 Role in Metabolism The concentration of the DHF substrate in *E. coli* cells during exponential growth is in the tens of μM , which is sufficient to saturate the WT enzyme[22]. The NADPH cofactor exists in the cytosol at a concentration of $\sim 120 \mu\text{M}$ [23], which is far above the K_d of $0.15 \mu\text{M}$ for NADPH[24]. The affinity of DHFR for nicotinamide adenine dinucleotide (reduced) (NADH) measured as $K_{M,NADH}$ is $334 \mu\text{M}$, approximately 2 orders of magnitude higher than that for NADPH[24]. This difference allows DHFR to be specific for NADPH even though the NADH concentration in the cytosol is $\sim 100 \mu\text{M}$.

The THF product of the DHFR reaction is essential for the synthesis of thymidine from uracil and other products of folate metabolism such as serine and glycine [25]. Because DHFR is essential for cell doubling, human DHFR is an important clinical target for anti cancer drugs. Because prokaryotic DHFRs diverge structurally from eukaryotic DHFRs[16], they are also an important target for antibiotics.

Despite DHFR activity being essential, DHFR is expected to only be ~ 90 [26] to ~ 400 [27] copies per cell. The measurements establishing these bounds were taken under controlled growth conditions, which is expected to have a significant impact on the expression levels. It is estimated that DHFR is buffered such that changes in expression level do not have a measurable impact on *E. coli* growth[26] for decreases in expression down to 30% and increases of up to 7-fold relative to the endogenous expression level. The activity of DHFR is tightly coupled to the activity of the downstream enzyme *E. coli* thymidylate synthase (TYMS)[25]. This is reportedly the result of system level optimization to maintain low cellular concentrations of toxic folate intermediates[25].

1.1.0.4 *In vivo* Assay Critical for our experiments, an *in vivo* assay that reports on DHFR activity for a panel of mutants had already been developed by our collaborator, Kim Reynolds[2], based off a *folA* (DHFR) and *thyA* (TYMS) double knock out of ER2566 strain *E. coli*[28]. The TYMS deletion arises spontaneously after deletion of DHFR from the *E. coli* genome[29], so in the selection assay, both genes are supplemented on a pACYC-Duet plasmid. The selection pressure in the assay comes from DNA replication and folate intermediate toxicity. The *in vivo* assay had been shown to report on DHFR activity[2], and selection could be optimized by tuning the relative expression levels of DHFR and TYMS.

References

- [1] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the sixth international congress of genetics*, pages 356–366, 1932.
- [2] K. A. Reynolds, R. N. McLaughlin, and R. Ranganathan. Hot spots for allosteric regulation on protein surfaces. *Cell*, 147(7):1564–75, 2011.
- [3] Carol A. Fierke, Kenneth A. Johnson, and Stephen J. Benkovic. Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from escherichia coli. *Biochemistry*, 26(13):4085–4092, 1987.
- [4] C. A. Fierke and S. J. Benkovic. Probing the functional role of threonine-113 of escherichia coli dihydrofolate reductase for its effect on turnover efficiency, catalysis, and binding. *Biochemistry*, 28(2):478–86, 1989.
- [5] C. J. Falzone, P. E. Wright, and S. J. Benkovic. Dynamics of a flexible loop in dihydrofolate reductase from escherichia coli and its implication for catalysis. *Biochemistry*, 33(2):439–42, 1994.
- [6] Z. Huang, C. R. Wagner, and S. J. Benkovic. Nonadditivity of mutational effects at the folate binding site of escherichia coli dihydrofolate reductase. *Biochemistry*, 33(38):11576–85, 1994.
- [7] C. E. Cameron and S. J. Benkovic. Evidence for a functional role of the dynamics of glycine-121 of escherichia coli dihydrofolate reductase obtained from kinetic analysis of a site-directed mutant. *Biochemistry*, 36(50):15792–800, 1997.
- [8] G. P. Miller and S. J. Benkovic. Strength of an interloop hydrogen bond determines

- the kinetic pathway in catalysis by escherichia coli dihydrofolate reductase. *Biochemistry*, 37(18):6336–42, 1998.
- [9] G. P. Miller, D. C. Wahnou, and S. J. Benkovic. Interloop contacts modulate ligand cycling during catalysis by escherichia coli dihydrofolate reductase. *Biochemistry*, 40(4):867–75, 2001.
- [10] V. Stojkovic, L. L. Perissinotti, D. Willmer, S. J. Benkovic, and A. Kohen. Effects of the donor-acceptor distance and dynamics on hydride tunneling in the dihydrofolate reductase catalyzed reaction. *J Am Chem Soc*, 134(3):1738–45, 2012.
- [11] M. R. Sawaya and J. Kraut. Loop and subdomain movements in the mechanism of escherichia coli dihydrofolate reductase: crystallographic evidence. *Biochemistry*, 36(3):586–603, 1997.
- [12] J. R. Appleman, E. E. Howell, J. Kraut, and R. L. Blakley. Role of aspartate 27 of dihydrofolate reductase from escherichia coli in interconversion of active and inactive enzyme conformers and binding of nadph. *J Biol Chem*, 265(10):5579–84, 1990.
- [13] Y. T. Tamer, I. K. Gaszek, H. Abdizadeh, T. A. Batur, K. A. Reynolds, A. R. Atilgan, C. Atilgan, and E. Toprak. High-order epistasis in catalytic power of dihydrofolate reductase gives rise to a rugged fitness landscape in the presence of trimethoprim selection. *Mol Biol Evol*, 36(7):1533–1550, 2019.
- [14] A. Bar-Even, R. Milo, E. Noor, and D. S. Tawfik. The moderately efficient enzyme: Futile encounters and enzyme floppiness. *Biochemistry*, 54(32):4969–77, 2015.
- [15] A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21):4402–10, 2011.

- [16] C. T. Liu, P. Hanoian, J. B. French, T. H. Pringle, S. Hammes-Schiffer, and S. J. Benkovic. Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proc Natl Acad Sci U S A*, 110(25):10159–64, 2013.
- [17] R. P. Joosten, F. Long, G. N. Murshudov, and A. Perrakis. The pdb redo server for macromolecular structure model optimization. *IUCrJ*, 1(Pt 4):213–20, 2014.
- [18] D. Oyen, R. B. Fenwick, P. C. Aoto, R. L. Stanfield, I. A. Wilson, H. J. Dyson, and P. E. Wright. Defining the structural basis for allosteric product release from e. coli dihydrofolate reductase using nmr relaxation dispersion. *J Am Chem Soc*, 139(32):11233–11240, 2017.
- [19] D. Oyen, R. B. Fenwick, R. L. Stanfield, H. J. Dyson, and P. E. Wright. Cofactor-mediated conformational dynamics promote product release from escherichia coli dihydrofolate reductase via an allosteric pathway. *J Am Chem Soc*, 137(29):9459–68, 2015.
- [20] K. F. Wong, T. Selzer, S. J. Benkovic, and S. Hammes-Schiffer. Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase. *Proc Natl Acad Sci U S A*, 102(19):6807–12, 2005.
- [21] R. V. Mauldin, P. J. Sapienza, C. M. Petit, and A. L. Lee. Structure and dynamics of the g121v dihydrofolate reductase mutant: lessons from a transition-state inhibitor complex. *PLoS One*, 7(3):e33252, 2012.
- [22] Y. K. Kwon, W. Lu, E. Melamud, N. Khanam, A. Bogner, and J. D. Rabinowitz. A domino effect in antifolate drug action in escherichia coli. *Nat Chem Biol*, 4(10):602–8, 2008.
- [23] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nat Chem Biol*, 5(8):593–9, 2009.

- [24] D. P. Baccanari, D. Stone, and L. Kuyper. Effect of a single amino acid substitution on escherichia coli dihydrofolate reductase catalysis and ligand binding. *J Biol Chem*, 256(4):1738–47, 1981.
- [25] A. F. Schober, A. D. Mathis, C. Ingle, J. O. Park, L. Chen, J. D. Rabinowitz, I. Junier, O. Rivoire, and K. A. Reynolds. A two-enzyme adaptive unit within bacterial folate metabolism. *Cell Rep*, 27(11):3359–3370.e7, 2019.
- [26] S. Bershtein, W. Mu, A. W. Serohijos, J. Zhou, and E. I. Shakhnovich. Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. *Mol Cell*, 49(1):133–44, 2013.
- [27] A. Schmidt, K. Kochanowski, S. Vedelaar, E. Ahrne, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, and M. Heinemann. The quantitative and condition-dependent escherichia coli proteome. *Nat Biotechnol*, 34(1):104–10, 2016.
- [28] P. T. Rajagopalan and S. J. Benkovic. Preorganization and protein dynamics in enzyme catalysis. *Chem Rec*, 2(1):24–36, 2002.
- [29] E. E. Howell, P. G. Foster, and L. M. Foster. Construction of a dihydrofolate reductase-deficient mutant of escherichia coli by gene replacement. *J Bacteriol*, 170(7):3040–5, 1988.

2 Modeling Functional Proteins as Multi-state Ensembles for Computational Design

2.1 Introduction

A central tenet of protein biophysics is that the amino acid sequence gives rise to a 3-D structure. This was tested by Anfinsen in seminal experiments that showed that activity could be restored for denatured ribonuclease A by gradually removing the denaturant[1]. From this result, it was inferred that protein sequence was sufficient to encode the 3D structure of a protein. The inference of sequence-structure relationships has broadly held true for the past 60 years. It has also lead to the notion that an optimized sequence can be predicted for protein structures, even structures that do not exist in nature.

The field of protein design engages in the study of how to generate designer protein sequences for specific protein folds, what are the limits of designable proteins, and how can we engineer proteins to perform useful work. In this field, *de novo* protein design is a particularly challenging area that has seen great successes over the past 25 years[2–7]. Notable milestones include the design of protein folds without natural homologues[8], the design of modular helical protein architecture through parametric design[9, 10], the design of buried hydrogen-bond networks[11], the design of multi-pass transmembrane proteins[12], the modification of dynamic motions in proteins[13], guided library design[14, 15], the design of protein biosensors[16], and the design of large protein assemblies and nanocages[17–19]. A continuing challenge in this area is advancing protein design beyond static structural topologies to proteins with multiple conformations, dynamic conformational changes, and complex functions.

In this chapter, we developed an approach for modeling proteins as a multi-state system of

ensembles. We demonstrate this system using a natural enzyme as a model system, and we analyze the computational efficiency and performance of our multi-state ensemble method using a sequence tolerance benchmark on our model system.

2.2 Background

Protein engineering is a broad field, and the approaches taken in a project are largely determined by the goal. For many applications, modifying an existing protein through rational design or a library selection is the only practicable approach. Dually, the goal at the outset of a protein engineering project may be curtailed by what is possible with existing protein scaffolds. Therefore, the ability to design a *de novo* proteins with designer-specific properties would open new opportunities for biomedical engineering[20–26], chemical synthesis[27–30], and nanoscale structural engineering[6, 17].

2.2.1 Computational Protein Design

Applications of computational protein design methods include the modification of existing proteins and the *de novo* design of new protein structures and topologies. In either case, the key input to current state-of-the-art computational protein design methods is the knowledge of the desired protein structure. The computational design method then samples combinatorial sequence space and predicts a sequence that is optimized for the desired structure based on a biophysical scorefunction.

As with predictive models in general, this approach consists of three major components: representation, sampling, and scoring. **Representation** refers to what information and what level of complexity is used in depicting the subject of the model. Depending on the scale of the model, a protein may be modeled as a field of electron density, a cloud of points representing

each nucleus, a string of beads representing each residue, or a single point. Representation plays a central role in sampling and scoring by dictating the available degrees of freedom, the functional form for modeling forces, and the applicability of computational algorithms. **Sampling** refers to both the fineness and the breadth of change that is examined by the model. In sampling sequences, all amino acids may be sampled at all positions, or the model may be restricted to one or a few amino acids at most positions. In sampling structural conformations, the model may sample movements in Cartesian space or sample torsional rotations. Similarly, specific regions may be allowed finer sampling of conformational space (e.g. functionally important residues) and other regions (e.g. the protein backbone) may be assigned a fixed conformation. **Scoring** refers to how the model evaluates the optimality of the current solution. For protein design, a scorefunction is used as an indicator for sequences that will successfully fold into the desired conformation. Terms in the scorefunction may be drawn from statistical data or equations in the form of first principles. Scorefunctions are generally parameterized to recapitulate experimental data from biophysical experiments and quantum mechanics (QM) modeling of molecular systems. Ultimately, the desired functional properties of the real world protein must be encoded into the model on the basis of the input structural representation, sampling algorithm, and scorefunction.

2.2.2 Single-state, Ensemble, Multi-state Design Models

One promising approach for encoding complex function into computational protein models is to change the representation to make models more realistic and capture more properties of known proteins. Because we know that natural proteins undergo conformational changes that are required for function, computational models might be improved by using representations that incorporate multiple conformations. The re-design of natural proteins can also serve as a useful

test for computational methodologies, both in simulations and in wet-lab experiments.

2.2.2.1 Design Model Categories The current state-of-the-art computational design is a bespoke process, with each protocol tailored to the project. Thus, there have been many varied and inventive approaches to adding information into computational design models. For the purpose of discussing the representation of protein conformations in simulation, we will group computational design protocols into three general categories: single-state, ensemble, and multi-state design (**Figure 2.1**, page 15). Each of these approaches comes with important differences in how they approach representation, sampling, and scoring.

2.2.3 Single-state Design Models

Single-state design is arguably the current state-of-the-art. It has been used in the design of a wide range of *de novo* protein topologies[8, 19], *de novo* catalytic proteins[29], nanomolar affinity small molecule binding domains[31], and higher order protein assemblies[17, 18, 32]. Relative to models with more complex representation, single-state design is computationally efficient. Replicate simulations can also be run independently in a highly parallelized fashion. Generally, many outputs from replicate simulations are run through additional down-stream filtering criteria to select designs for final analysis and experimental characterization.

2.2.3.1 Representation Molecules in single-state design are typically represented as full-atom models. Every nucleus in the protein is represented as a point. Each atom is assigned one of a set of atom types (e.g. aliphatic carbon, aromatic nitrogen, etc.) that defines its atomic properties, including partial charge, hydrogen-bonding capacity, Van der Waals radius, Lennard-

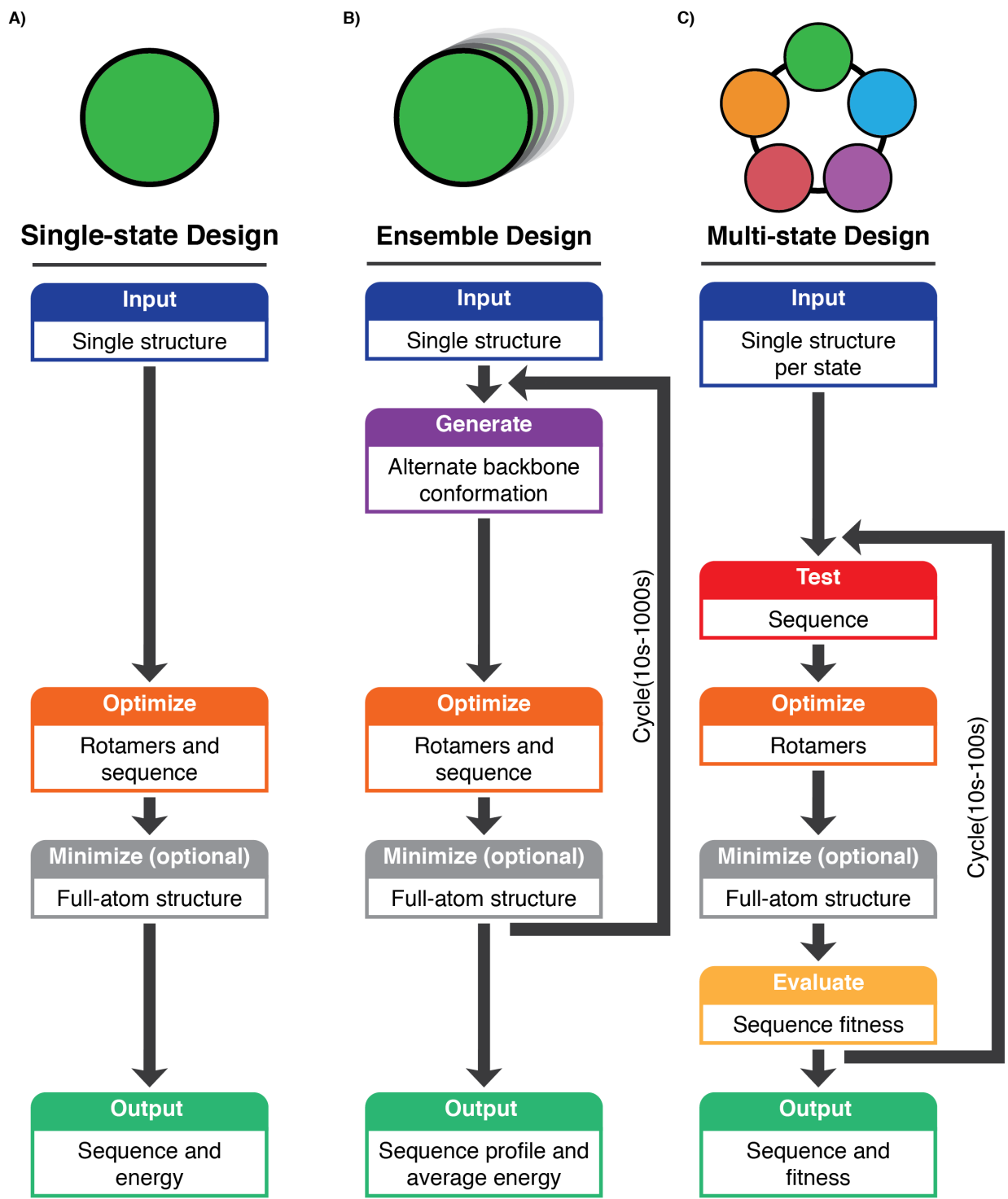


Figure 2.1: Conceptual diagram and process flowchart for **A)** single-state, **B)** ensemble, and **C)** multi-state design models. Circles represent the backbone conformations used in the model. Ensembles sampling the same energy well are shown as stacked circles of uniform color. Models sampling different energy wells are shown as spaced circles that are colored uniquely. Below each diagram is a flowchart of generic steps used in each approach.

Jones potential well depth, and desolvation potential. Electrons and electron density are not explicitly represented in full-atom representation as they are in quantum mechanical simulations.

Protein backbones remain in a fixed conformation throughout single-state design. This is a feature that differentiates single-state design from other methodologies (**Figure 2.1**, page 15). The only degrees of freedom are in the side-chains, which are allowed to rotate around their chi angles. The discrete conformation of a side-chain is referred to as a rotamer. Rotamer libraries have been generated that sample the most probable rotamers based on statistics from the Protein Data Bank (PDB)[33, 34], and some design algorithms use continuous probability functions for continuous rotamer sampling[35, 36].

2.2.3.2 Sampling Sampling in single-state design is performed by changing the rotamers (side-chain conformations) that are on the backbone. This allows for sampling the space of rotamer configurations, where a configuration is a set of rotamer conformations over the entire protein. Sampling can be performed with rotamers for a fixed amino acid sequence or with a set of rotamers that include rotamers for multiple amino acids to simultaneously sample sequence space and rotamer configuration space.

Monte Carlo simulated annealing is one method that optimizes the rotamers to output low energy conformations[8, 37]. In each step of the simulated annealing algorithm, a randomly selected rotamer for a randomly selected position is trialed and evaluated by the Metropolis Criterion (**Figure 2.2**, page 17). Thereby, a change to the model is accepted if it lowers the model energy, but moves that increases the model energy are accepted with a probability that is dependent on the change in energy (ΔE) multiplied by a temperature factor (kT). During the course of the simulation, the temperature factor is modulated such that the simulation is able to sample unfavorable moves at high temperatures and settle in a low energy well as the

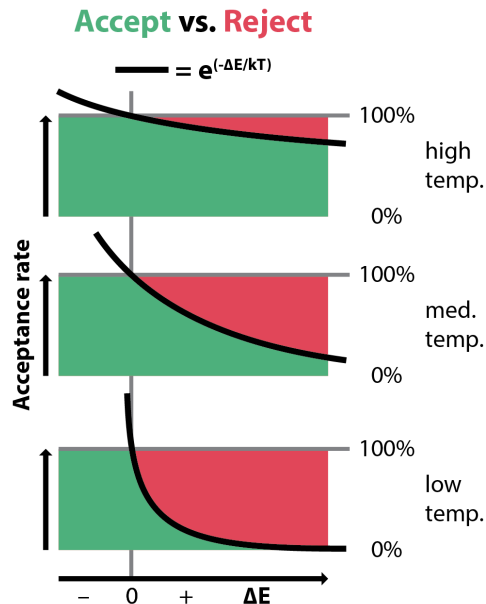


Figure 2.2: Application of the Metropolis criterion to simulated annealing. The probability (equation, top) of accepting (green) or rejecting (red) a Monte Carlo move (e.g. rotamer substitution) is shown as a function of the change in biophysical energy after the move (ΔE) at three different temperatures (labels, right).

temperature cools. One caveat to simulated annealing is that it is not ensured to converge on the lowest energy rotamer configuration, and the output is stochastic. Convergence on a solution (e.g. rotamer configuration, sequence, or score) can only be tested by multiple repetitions of the simulation. The probability of the simulation converging on a small set of similar solutions is a function of the size of rotamer configuration space, thus at the number of design positions is limited. The maximum number of designable positions in current implementations of simulated annealing will be affected by factors such as protein topology and the number of rotamers per positions, but less than 100 positions is a reasonable limit for many applications.

In contrast, dead-end elimination (DEE) is a provable method that outputs the global minimum energy rotamer configuration[38, 39]. While the importance of the global minimum energy rotamer configuration depends on the appropriateness of the model, identifying the global minimum avoids local minima that can potentially trap a simulated annealing trajectory. One further caveat to DEE is that the simulation frequently does not converge on a solution

and no output is produced. Strategies that implement stochastic steps in sampling side-chains allow for the identification of non-optimal but low energy configurations[40], but the probability of convergence decreases as a function of the combinatorial size of rotamer configuration space, so a reasonable number of design positions in DEE is 10-20 positions with the same caveats as for simulated annealing.

2.2.3.3 Scoring A scorefunction for computational design or molecular modeling must have predictive power for what sequences and rotamer configurations will stabilize a given fold. Scorefunctions can be statistical potentials drawn from large datasets such as multiple sequence alignments [41, 42] and PDB statistics [36, 43–46]. Scorefunctions can also be explicitly parameterized to only include terms based on physical first principles [39, 47, 48]. Many scorefunctions are hybrids of the two, containing terms for both statistical physical potentials [49–51].

2.2.3.4 Optimization Single-state design is frequently followed by a minimization protocol that allows the backbone and side-chain torsion angles to adjust to the final design sequence. Many single-state design protocols now iterate between design and minimization steps such as the FastDesign protocol in the Rosetta protein design software suite. This level of backbone movement is a deviation from our definition of single-state design, but it is generally lower magnitude than that in design simulations that explicitly sample backbone movement.

2.2.4 Ensemble Design Models

In contrast to single-state models, ensemble models are not constrained to a fixed backbone. To distinguish ensemble models from multi-state models, we define an ensemble as conformations

that sample the same well on the energy landscape or functional conformation and that do not change the secondary structure assignment of residues.

2.2.4.1 Representation Ensemble models represent a protein with backbone flexibility. A single backbone conformation with a set of ϕ , ψ , and ω torsion angles is referred to as a conformer. Ensembles of conformers are generated by sampling the local conformers around a single input structure or by using multiple conformers from ensemble structural refinement methods. Sampling algorithms for generating backbone ensembles include Rosetta Backrub[52–54], KIC[55–57] PertMin[58, 59], or molecular dynamics trajectories[47, 50].

2.2.4.2 Sampling For each of the backbone conformers in the ensemble model, the rotamer configuration is optimized using the same algorithms as are used in single-state design. When the sequence space being sampled is very large the only output may be predictions of amino acid preferences [60, 61]. When the ensemble predictions are being compared to specific stability measurements for point mutants, the model is used to enumerate over that limited sequence space [59, 62].

2.2.4.3 Scoring Conformers and Ensembles After performing design on all members of the ensemble, aggregate properties can be extracted. If the sequence for each conformer is optimized independently, then statistics can be collected to model the amino acid preferences at each position[60, 63]. For design, individual conformers with their optimized sequences and rotamer configurations can be examined more carefully and experimentally characterized[16, 64]. In cases involving an enumerated sequence space, the energy values over the ensemble can be averaged together for each sequence. Boltzmann-weighted averaging (**Equation 2.1**, page 20) is frequently used in this step even though the sampling of

conformational space in the ensemble is not sufficient to represent even the low energy microstates in the partition function. Rather, Boltzmann averaging effectively filters out the contribution to the average for a sequence from conformers that are significantly higher in energy than the median energy with a given sequence over the conformer ensemble.

$$E_{state} = \frac{\sum_i^n E_i \cdot e^{(-E_i/kT)}}{\sum_i^n e^{(-E_i/kT)}} \text{ for each } i \text{ conformer in ensemble of size } n \quad (2.1)$$

2.2.4.4 Optimization As with single-state design, the members of the ensemble may be minimized after rotamer sampling. This step has been demonstrated to decrease error in predictions of experimental data[59, 62].

2.2.4.5 Hybrid Models Alternative methods have been developed that sample combined Monte Carlo moves of backbone conformations with sequence and rotamer substitutions[57, 65]. These methods may have a role in future multi-state design methodologies, particularly for small sequence spaces where ensemble generation and minimization steps may be inefficient relative to the output.

2.2.5 Multi-state Design Models

We define multi-state models as having a framework to represent a macromolecular system with a group of functionally distinct conformations or states. We note that some definitions of multi-state models require the inclusion of negative states (off-target interactions or unfolded states). Under these definitions, a model that only includes positive states is a multi-constraint model. We chose to use the term multi-state here because we examine models that have the mathematical capacity to include negative states. The inclusion of negative states is not explicitly addressed here.

2.2.5.1 Representation In a multi-state model, a protein is represented as a group of states. While ensemble models represent a protein with conformers from the same well on the energy landscape, states in a multi-state model are conformers from different wells. States are frequently conformations that can be characterized individually by structural methodologies such as x-ray crystallography and NMR spectroscopy. The structural difference between states may constitute a fold change, a change in small molecule binding, a protein-protein complex interaction, or loop motions that sample multiple separate low energy conformations.

2.2.5.2 Sampling Rotamer Configuration and Scoring Individual States Rotamer optimization and scoring forms an internal loop in the multi-state design algorithm (**Figure 2.1C**, page 15). Optimization of rotamer configuration is performed using the same algorithms used in single-state design, except that the sequence is not optimized simultaneous to the rotamer configuration. As with ensemble design applied to enumerated sequence spaces, sequences on each individual state are evaluated by their score with an optimized rotamer configuration. Because each change to the sequence is followed by rotamer optimization for that sequence on all states, full atom minimization is not practical in many cases.

2.2.5.3 Sampling Sequences The external loop in multi-state design is an algorithm for sampling changes to sequences. Many algorithms have been employed, including random Monte Carlo sampling [66] and genetic algorithms[67–69] that both sample randomly and merge successful random moves. Sampling of sequence space is typically slow. For runtimes of 1 week or less, the probability that repeated simulations will converge on similar optimal sequences decreases as a function of the number of design positions and the number of states, and the maximum for systems with only 2 states is ~25-30 designable positions.

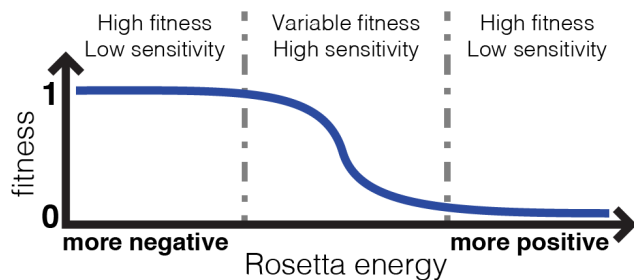


Figure 2.3: Fitness for each state is modeled by a sigmoidal function. Fitness (blue line) is a function of the biophysical energy of the state. The sigmoid has three regimes (labels, top) based on the overall fitness of the state and how responsive the fitness is to changes in the biophysical energy as a result of mutation.

2.2.5.4 Objective Function for Design Each sequence that is sampled must be evaluated, and design moves in sequence space must be accepted or rejected. Monte Carlo sampling utilizes the Metropolis criterion and genetic algorithms propagate a population by carrying a high scoring fraction of sequences from the current round of sampling to the next. In both cases, the scores must be integrated from all states into an objective function or fitness function that is optimized over the sampled sequence space. Many simulations use the straightforward mean energy of all states, but this method can yield poorly optimized outputs if contributions from one state dominate the other states. Furthermore, this approach does not take into account the observation that the states in a functional protein (e.g. an enzyme bound to the transition state of the chemical reaction step or an enzyme bound to the reaction products) will need to be differentially stabilized.

Warszawski et al. developed an objective function for multi-state computational design that used a sigmoidal fitness model. Scores for each state (E_{state}) were converted to a fitness score representing the predicted probability that the trial sequence stabilizes the state (**Equation 2.2**, page 23). Because the sigmoid functional form for each state have a responsive regime surrounded by two regimes where fitness is not responsive to changes in energy, compromise is possible between low-fitness and high-fitness states even if all states can significantly improve in energy

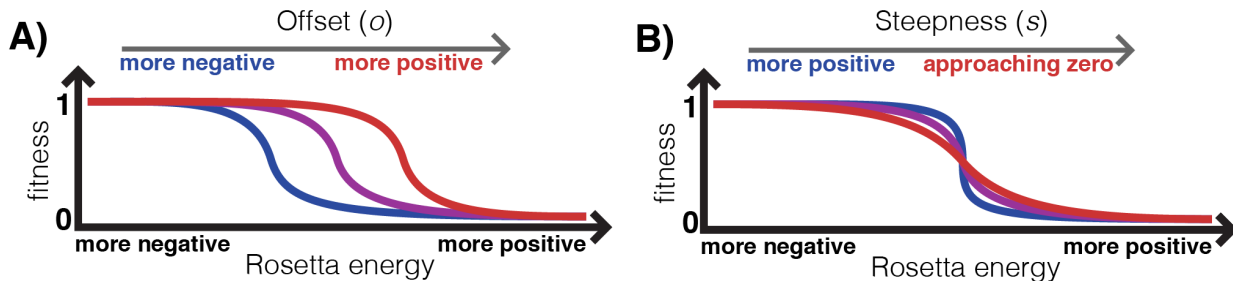


Figure 2.4: Two user-set parameters modulate the sigmoid fitness function. **A)** The offset parameter (o) sets the center of the fitness curve on the energy axis. **B)** The steepness parameter (s) sets the rate of change along the fitness curve.

(**Figure 2.3**, page 22). A user-defined offset parameter (o) determines the midpoint, and a steepness parameter (s) modulates the magnitude of the fitness differential with respect to energy (**Figure 2.4**, page 23). Effectively, the offset is an energy normalization factor and the steepness term is an inverse temperature factor. Ultimately, the total fitness for the multistate system is calculated as the joint probability of fitness over all states (**Equation 2.4**, page 23).

$$fitness_{state} = f = \frac{1}{1 + e^{s(E_{state} - o)}} \quad (2.2)$$

$$fitness_{total} = F = \prod_j^m f_j = \prod_j^m \frac{1}{1 + e^{s(E_{j,state} - o)}} \quad \text{for each } j \text{ state} \quad (2.3)$$

2.2.5.5 Alternate Approaches The Constrained Optimization of Multistate Energies by Tree Search (COMETS) algorithm was developed to identify provably optimal sequences for a multistate simulation using a tree search algorithm [70]. As with other provable optimization methods like dead-end elimination, COMETS slows relative to stochastic methods as the number of design positions increases. Alternative approaches have focused on increasing the performance of multi-state design with algorithms that trade exactness for efficiency.

The CLEVER algorithm uses a cluster expansion method that fits arbitrary functions to

recapitulate the energies from single state design for a training set of sequences[37, 71]. The resulting scorefunction predictions can then be used to rapidly sample sequence space without requiring time intensive rotamer optimization. This algorithm was used to design heterodimeric coiled coils that bind with high specificity as well as affinity [72] where the average error in scorefunction energy prediction is 1-2 units. These units are approximated as kcal/mol based on correlation with experimental measurements of $\Delta\Delta G_{mutation}$, so the error in the scorefunction predictions from CLEVER is still within a regime that is usable with computational design. For other protein topologies, the CLEVER cluster expansion results in errors of 5-30 energy units.

Rosetta Mean Field Prediction (MFPred) uses a mean-field approach to estimate sequence preferences[61]. The mean field approach does not perform rotamer optimization, but it performs iterative calculations of probability-weighted averaging of all two-body energies, where rotamer probabilities are calculated from the energies in the previous round. Prediction performance with MFPred was dependent on pre-minimization of the input structures, which is expected to give strong bias to the starting sequence. Nevertheless, predictions of sequence preferences from MFPred had higher sequence entropy than predictions from single-state design (**Equation 2.4**, page 24).

$$entropy = -\sum_{ala}^{tyr} p_i \text{Log}_{20} p_i \text{ for each } p_i \text{ probability of each } i \text{ amino acid} \quad (2.4)$$

Finally, the Rosetta Restrained Convergence (RECON) algorithm allows the sequence on each state to optimize independently and then drives states to converge, unlike multi-state approaches that enforce convergence throughout the optimization[73, 74]. During the RECON simulation, positions that converge to a single amino acid on all states are fixed, and a convergence bonus is ramped in the scorefunction over the rounds. In the final round, uncovered positions are

designed with a greedy algorithm as follows. An unconverged position is selected randomly and all sequence substitutions are sampled at that site. The best amino acid is chosen, and the position selection continues until all positions are assigned an amino acid. When RECON was applied to a small benchmark set of proteins, its performance improved over genetic algorithms for design cases of more than 30 positions, but RECON generally output predictions with very little variation over repetitions.

2.2.6 Evaluating Design Performance

There are multiple metrics for evaluating the performance of a computational design model. Native sequence recovery tests the ability of a model to recapitulate the native sequence for an input structure (e.g. PDB structure). Sequence profile similarity tests the ability of the model to predict the first-order amino acid frequencies in a large dataset such as a multiple sequence alignment (MSA) from natural sequences or amino acid preferences from a selection experiment. It can therefore be used to quantify how a model predicts the degeneracy within a family of related sequences by examining if a model accurately distinguishes conserved and functionally important residues from residues that can more freely mutate. In contrast to profile similarity, covariation metrics such as mutual information[75, 76], direct coupling analysis (DCA)[77, 78], and statistical coupling analysis (SCA)[79, 80] attempt to quantify the interactions between residues from the second-order joint probabilities in the MSA.

2.2.6.1 Caveats for Design Metrics These metrics are in general rapid and easy to implement for purely computational benchmarks, but they are also imperfect comparisons to design simulations. Design models do not accurately represent the dynamic nature of protein structures, model the free energies (ΔH and ΔS) of folded proteins, or represent complex cellular and evolutionary

environments in which the proteins were optimized. These issues with model appropriateness only exacerbate the general challenge of identifying a single natural sequence from a large degenerate sequence space that encodes proteins with nearly identical structure. For example, common performance for native sequence recovery is in the range of 30-45% for state of the art methods, and even these scores are confounded by bias to the starting sequence in the initial model.

2.2.6.2 Comparison to Experimental Measurements The best test of a design is the experimental characterization of the designed protein, but this process is far slower and more costly than a computational benchmark, especially if the molecular structure is determined. For proteins with a measureable activity, methods for high-throughput screening such as deep mutational scanning[25, 81–84] and MITOMI[85, 86] open avenues for rapidly evaluating design outcomes. Additionally, methods of DNA library synthesis and screening have been developed to examine the fold stability of thousands of small miniproteins (~50 aa) in a single experiment[87, 88].

Other experimental datasets can also be useful in evaluating a model. It is expected that a design algorithm that can successfully predict sequences that function and/or fold into a desired structure will also have predictive power for the biophysical impact of point mutations. When paired with protein structures, large databases of $\Delta\Delta G_{folding}$ and $\Delta\Delta G_{binding}$ measurements can be used to evaluate changes to representation, scoring, and sampling in a model. Estimates of ~0.01-0.1 kcal/mol for the non-additive error per residue have been suggested for a design model to have an total expected error of ~1 kcal/mol[89], so improvement in $\Delta\Delta G_{folding}$ and $\Delta\Delta G_{binding}$ predictions is expected to result in improvements in the generation of functional sequences. As protein design expands into the design of protein dynamics, metrics such as the prediction of side-chain order parameters from NMR experiments[52, 90, 91] and room-temperature X-ray crystallography[92–95] are expected to be increasingly important.

2.2.7 Rosetta

The work in this chapter is performed in the Rosetta protein design suite[3, 51]. Rosetta is a state-of-the-art software package for macromolecular design. It has been used to design *de novo* protein folds[8, 9, 19], buried hydrogen-bond networks[10], orthogonal binding partners[11], small molecule binders[16, 26, 31], metalloproteins[96], catalytic proteins[30, 97–99], multipass membrane proteins[12], and higher-order protein assemblies[32]. Multiple multi-state design methods have been implemented in Rosetta, including MPI_MSD[68], RECON[73], MFPred[61], and Rosetta MSF[69].

2.2.7.1 The Rosetta Scorefunction The Rosetta scorefunction is a hybrid scorefunction of physical and statistical terms. It contains a Coulombic potential for electrostatics, a 6-12 potential for attractive and repulsive Van der Waals interactions, and a hydrogen-bonding potential. It contains statistical potentials for side-chain rotamers and backbone torsion angles. It is not polarizable, and it does not currently contain pi-pi or cation-pi interaction terms. The solvation term is a Lazaridis-Karplus implicit solvation model. The current scorefunction is REF2015 and it has been benchmarked against a series of performance and experimental tests[51].

2.2.8 Challenges for Multi-state Design

Despite the many developments for multi-state design protocols, a consistent set of challenges remains around computational efficiency because the optimization of rotamer configuration and amino acid sequence are separated into two processes. Multi-state design protocols require significantly more computational resources than independent single-state design simulations run over the same number of states. Multi-state models also restrict the number of design positions to

fewer than are needed to generate a well-folded miniprotein. Therefore the practical application of multi-state design methods has been limited to the design of residues on the interface of helical bundles [72, 100, 101] and 22-23 loop residues in the loops of an antibody[23, 24], and no method has been applied to a design project that evaluates thousands of varied designs in high-throughput. It would be a notable advance if the successful design of a protein fold with >100 residues was computationally tractable with a multi-state method. An even greater advance would be to apply the model to the design of a protein with complex function.

2.3 Results

To advance computational protein design methodologies for functional proteins, we sought to develop a model that could encode multiple structural constraints for a functional protein, make predictions about the extent of tolerated sequence space, with sufficient computational efficiency to design a 5-20 kDa monomeric protein.

2.3.1 A Multi-state Ensemble Design Model

We chose to combine the two models types that separately accomplish each of these goals: multi-state models and flexible-backbone ensemble models (**Figure 2.5BC**, page 29). A protein is represented in the multi-state ensemble mode by macrostates of functionally required conformations (e.g. crystal structures of catalytic intermediates or of multiple binding conformations)(**Figure 2.5D**, page 29). Macrostates are represented as ensembles of flexible-backbone microstates as in ensemble design (**Chapter 2.2.4**, page 18). Energies for sequences (E_i) are calculated as averages over the ensemble of microstates. Fitnesses for each macrostate (f) are calculated from the macrostate energies using the sigmoidal model[66] that has been modified to add an exponential weight for each state (k). Finally, the total energy of

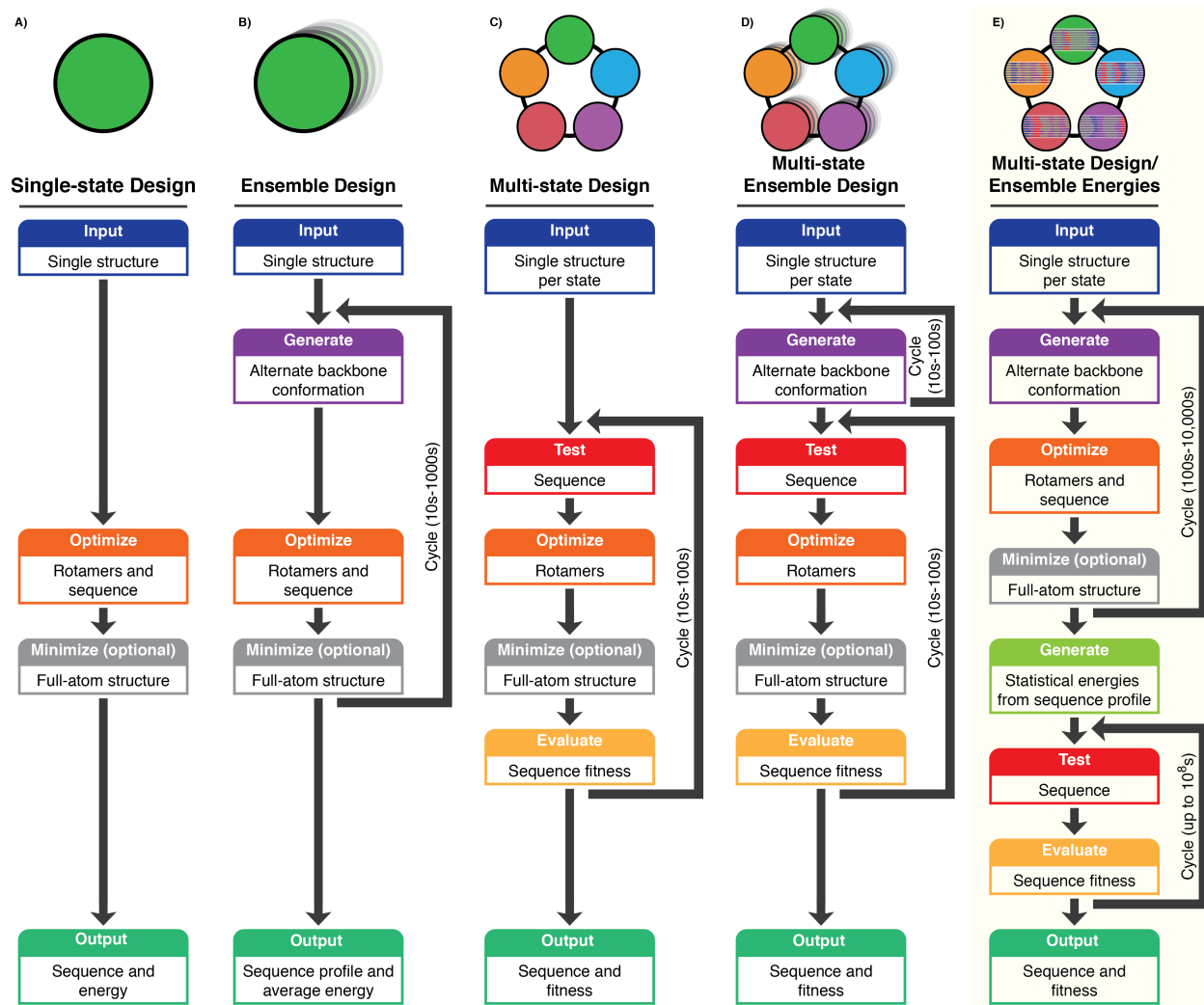


Figure 2.5: Comparison of Workflows in Multi-state Ensemble Models to Single-state, Ensemble, and Multi-state Models. Conceptual diagram and process flowchart comparing **A)** single-state, **B)** ensemble, and **C)** multi-state design models to proposed workflows for **D)** multi-state ensemble design models using enumerated energies **E)** and multi-state design with statistical energies from ensembles. Circles represent the backbone conformations used in the model. Ensembles sampling the same energy well are shown as stacked circles of uniform color. Statistical energies are represented by the MSAs from sequence tolerance. Models sampling different energy wells are shown as spaced circles that are colored uniquely. Below each diagram is a flowchart of generic steps used in each approach.

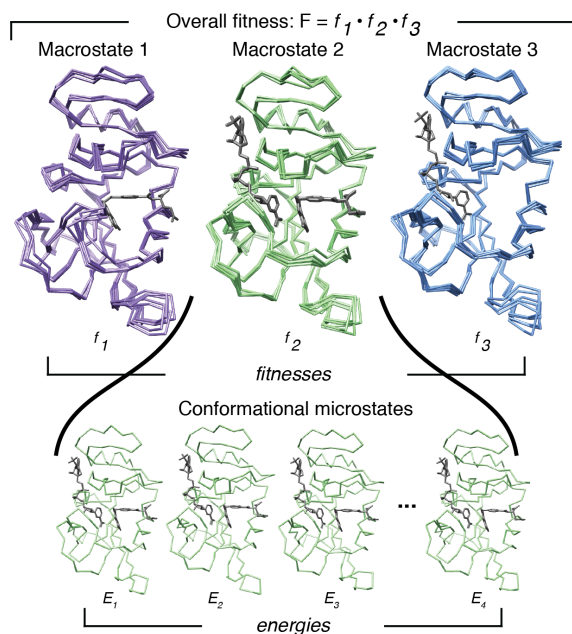


Figure 2.6: Macrostates and microstates in a multi-state ensemble model. (Top row) Macrostates of a protein with complex function (top row) are depicted as ensembles of flexible-backbone microstates using DHFR as an example (PDB IDs: 1RX5, 3QL3, 1RX1). Energies extracted from the microstates (bottom row) and fitnesses are calculated as described in **Chapter 2.3.1** (page 28).

the system (F) is calculated at the product of the macrostate fitnesses, as in the original sigmoidal model (**Equation 2.5**, page 30). Because the fitnesses (f and F) are probability-like terms, the log of these terms is an energy-like term. Therefore, exponential weights in the fitness are scalar multipliers for each macrostate energy.

$$fitness_{total} = F = \prod_j^m f_j^{k_j} = \prod_j^m \left(\frac{1}{1 + e^{s(E_{j,state} - o)}} \right)^{k_j} \text{ for each } j \text{ state} \quad (2.5)$$

2.3.2 Selecting a Multi-state Design Patch in DHFR

To focus our simulations on the residues that change their chemical environment during the DHFR catalytic cycle, we calculated the maximum difference in ϕ and ψ angles (**Figure 2.7**, page 31). Other metrics such as contact order change gave similar but less clear results. In contrast, the

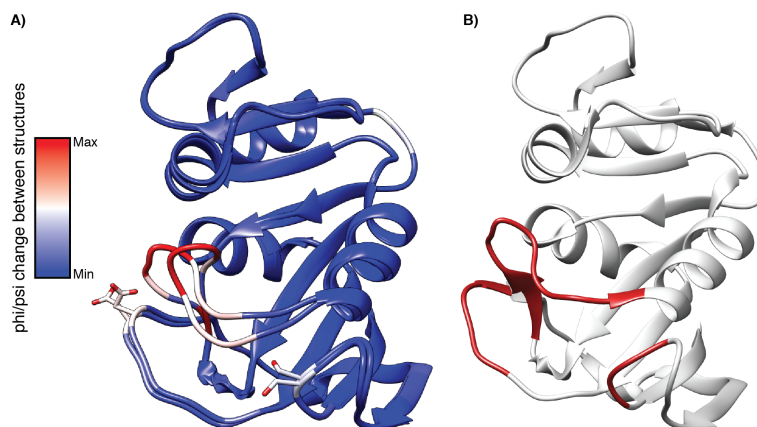


Figure 2.7: Selection of residues for multi-state design in DHFR. **A)** Maximum cumulative change in phi and psi torsion angles for all PDB structures in **Figure 1.2** (page 4). Torsion angle differences are represented on the closed (PDB ID: 3QL3) and occluded (PDB ID: 1RX5) conformations of DHFR according to the heatmap. **B)** Positions selected for multi-state design (13-24,119-123,147-149) according to maximum change in the psi and psi angles between DHFR conformations from PDB structures are colored in red on the closed conformation.

change in ϕ and ψ angles clearly highlighted residues in the highly mobile loops. We selected 20 residues in three contiguous sections (13-24, 119-123, 147-149) for multi-state design.

2.3.3 Multi-state Design on DHFR with Ensembles

To test if the multi-state ensemble model improved predictions of functional sequences in a computational design model, we tested if the model could improve predictions of sequence tolerance for DHFR relative to predictions from ensemble design on each of the individual DHFR macrostates.

2.3.3.1 Multi-state Ensemble Design on DHFR with Enumerated Sequence Energies

Energies (E_i) were calculated for all single point mutations and used to calculate amino acid probabilities at each position (see **Methods, Chapter 2.5.0.1**, page 37). For the multi-state ensemble model, the global steepness parameter (s) and the macrostate weight parameters were optimized to best recapitulate the amino acid frequencies in the MSA of bacterial DHFR sequences from OpenSeq[103]. In comparing the computational simulations to the MSA of bacterial DHFR

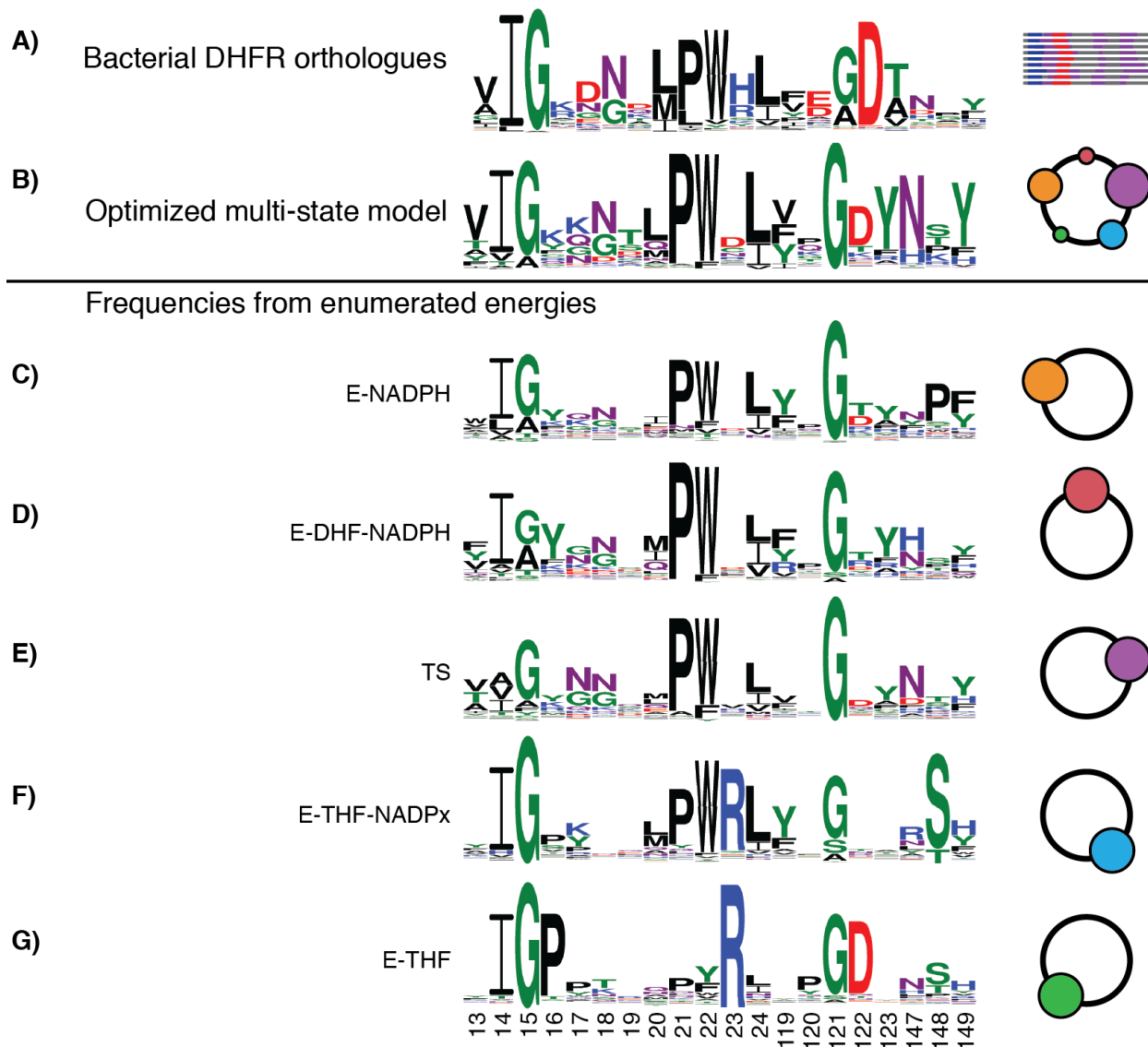


Figure 2.8: Multi-state design on DHFR using energies from enumerated amino acid substitutions. Sequence profiles are shown for **A)** a multiple sequence alignment of bacterial DHFR orthologues from OpenSeq[102], **B)** multi-state design with optimized weights, and **C-G)** the individual macrostates from the DHFR structure (**Figure 1.2**, page 4) as labeled (left). For profiles from simulations, a diagram (right) indicates the state weights correspond to the colors in **Figure 1.2**: E-NADPH (orange), E-DHF-NADPH (red), transition state (purple), E-THF-NADPx (blue), E-THF (green).

orthologues, we observe that the simulations generally recapitulate highly conserved and highly variable positions, particularly for positions 13-24. In contrast, the predictions from the individual macrostates are either more conserved or more variable at each position compared to the MSA. Compared to the predictions from individual macrostates, the multi-state ensemble model shows improved recovery of the amino acid frequencies from the MSA better than the single macrostate models. In the optimized weights for the multi-state ensemble model, we see that the transition state is the highest weighted state followed by E-NADPH (closed conformation) and E-THF-NADP_x states (occluded conformation), see **Figure 1.2** (page 4). Taken together, these results demonstrate that the inclusion of multiple functional conformations of DHFR into a design model can improve predictions of sequence tolerance.

2.3.3.2 Multi-state Design on DHFR with Statistical Energies from Ensembles The simulations from enumeration are bogged down by computational times of approximately 85 cpu hours per mutation with 200 microstates for each of 5 macrostates. Because each sequence is individually threaded onto the model, these energies do not include information about combinatorial mutations and multi-state ensemble simulations are limited by the same slow sampling of sequence space that limits most multi-state design algorithms (**Figure 2.5D**, page 29). We therefore sought to develop a computational approach that gives similar performance to multi-state ensemble design but is more computationally efficient. We developed an approach for performing multi-state design using statistical energies from ensemble design – multi-state design with ensemble energies – that approximates the Rosetta scorefunction over flexible backbone ensembles for each state (**Figure 2.5E**, page 29). This multi-state design with ensemble energies method relies on simulations that are very similar to those for multi-state ensemble design excepts that all positions are designed simultaneously on each microstate rather than threading individual sequences onto

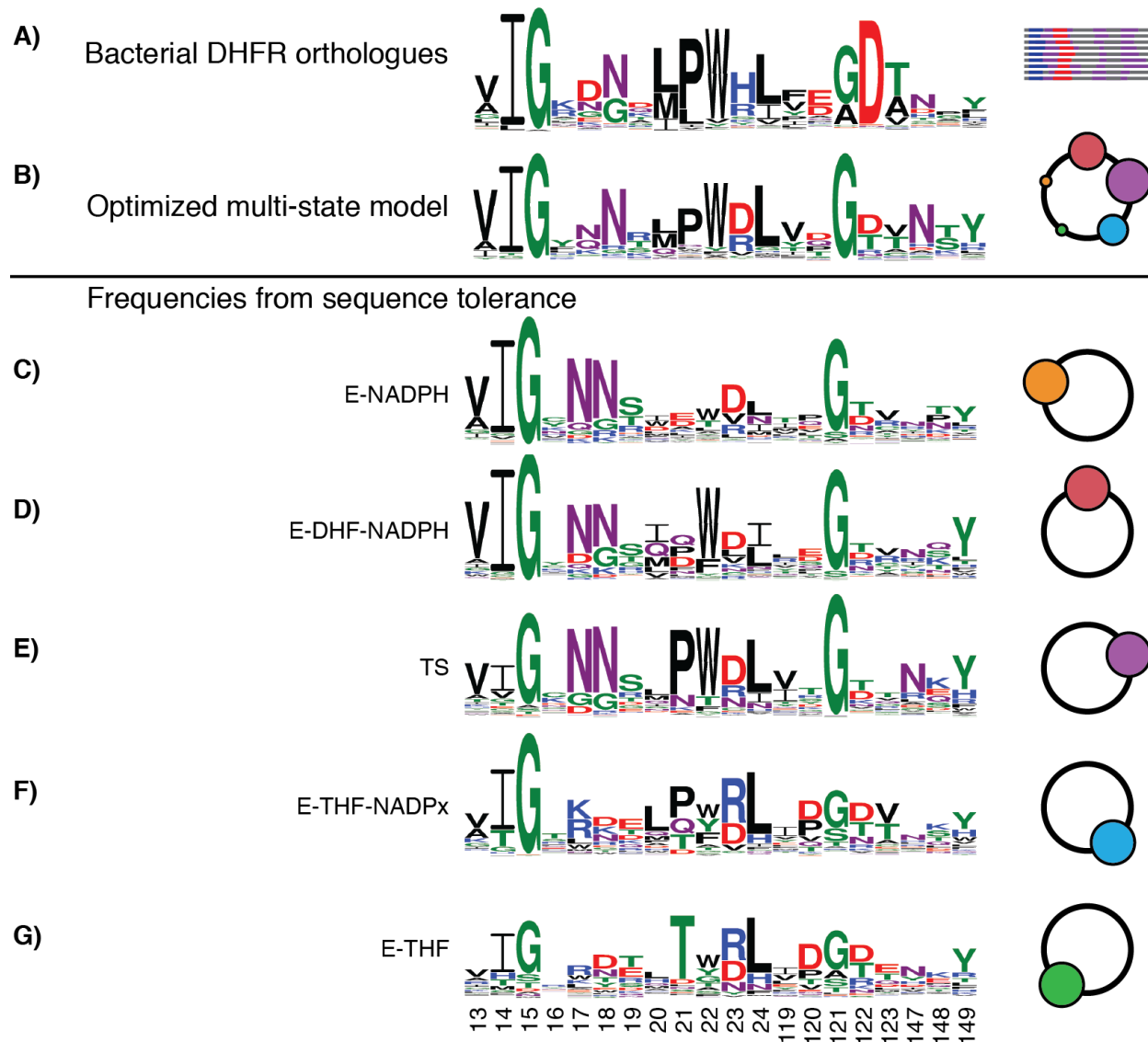


Figure 2.9: Multi-state design on DHFR using first-order statistical energies from ensemble design. Sequence profiles are shown for **A)** a multiple sequence alignment of bacterial DHFR orthologues from OpenSeq[102], **B)** multi-state design with optimized weights, and **C-G)** the individual macrostates from the DHFR structure (**Figure 1.2**, page 4) as labeled (left). For profiles from simulations, a diagram (right) indicates the state weights colors correspond to the colors in **Figure 1.2**: E-NADPH (orange), E-DHF-NADPH (red), transition state (purple), E-THF-NADPx (blue), E-THF (green).

the model (see **Methods, Chapter 2.5.0.2**, page 38).

Simulations with multi-state design required approximately 6 cpu hours per mutation with 200 microstates for each of 5 macrostates, a 14-fold speed up over our multi-state ensemble method. Furthermore, the value of 6 cpu hours is an over-estimate because more sequence information can be extracted from the same simulations. In examining performance, we see comparable sequence tolerance predictions for energies derived from sequence enumeration and from ensemble design (**Figure 2.9**, page 34). The optimized weights show a consistent pattern with those from the simulation using enumerated energies (**Chapter 2.3.3.1**, page 31). As with the enumerated energies, the transition state is given the highest weight while one closed conformation and one occluded conformation are also highly weighted. Overall, the performance of the model for prediction of sequence tolerance from statistical energies is comparable to those from enumerated energies, but the resulting profile is more variable. These preliminary results demonstrate that using statistical energies from ensemble design is sufficient for multi-state design.

2.4 Discussion

Despite approximately 2 decades of work [100], there is still a need for computational design protocols that **1)** can include multiple structural constraints, **2)** sample alternate backbone conformations within a macrostate, and **3)** are computationally efficient. Multistate design with ensemble ensemble energies has could potentially address this need because the computational cost is in the generation of the energy tables. Once the statistical energies are generated from the ensembles, the search through sequence space is dependent on a look-up table. This approach is very similar in concept to the efficient simulated annealing algorithm. This pre-generation of the energy tables for multi-state design is also performed with the CLEVER algorithm[71, 72], but inaccuracies in the approximated energies have limited the application of

this method[104]. In multi-state design with ensemble energies, the computational time per sequence on DHFR would drop to ~ 1 cpu min per sequence if only 1×10^6 trial sequences are evaluated, representing a 5,000-fold increase in computational efficiency over our initial multi-state ensemble model. Rapidly sampling 1×10^8 or more sequences in this design protocol is feasible and would further improve the overall efficiency of the methodology.

2.4.1 Future Directions

The multi-state design with ensemble energies method could potentially scale favorably for the design of proteins with >100 residues because of its expected computational efficiency, although multi-state and multi-state ensemble design may be useful in simulations to refining sequences that are already close to functional regions of sequence space. In the simulations from **Chapter 2.3.3.2** (page 33), we used first order statistical energies, but it is possible to extract second-order statistical energies using direct coupling analysis (DCA) calculations[42], although larger ensembles will likely be necessary to ensure reproducibility. Further testing will determine if using enumeration to refine sequences or generating second-order statistical potentials is more efficient approach. Regardless, the multi-state design with ensemble energies methodology would benefit from additional computational testing with more natural protein systems, and the gold standard for testing the model will be the generation of functional protein sequences that have near-WT levels of activity. In addition to DHFR, enzymes such as chorismate mutase, TEM-1 beta-lactamase, and various proteases and nucleases could be useful test cases. The performance of the method for predicting degenerate regions of functional sequence space can be tested by sampling sequences that have a range of sequence identities compared to the sequence of the starting structural model. At the point of writing, we do not have a method for predicting the functionally required macrostates for a given function, but the computational redesign of naturally

existing proteins with complex function (e.g. catalysis) may help elucidate the constraints from function on sequence space. Ultimately, a theory that allows the *a priori* prediction of the required macrostates would allow the design of protein that performs an arbitrary function that does not exist in nature

2.5 Methods

All simulations were performed using Rosetta 3.12. Example scripts, input files, and command lines are available on https://github.com/keleayon/2020_DHFR_MSD.git.

2.5.0.1 Calculating Amino Acid Energies by Enumeration Simulations Enumerated energies for mutations were generated over flexible backbone ensembles. For each input PDB structure from the 5 states in **Figure 1.2** (page 4), an ensemble of 200 backbone conformers was generated using the Backrub application[53] with 10,000 - 15,900 trial moves evaluated by the Metropolis Criterion with $kT = 0.6$. For each backbone in each ensemble, a simulated annealing simulation was run using the FixBB application where one residue was mutated to a specific amino acid and sidechains were allowed to repack in a 10 Å shell around the designed residue, where the shell is defined as the distance between β carbons. FixBB simulations with the 5 ensembles were repeated over every possible single point mutation. For each macrostate, the microstate energies were averaged with Boltzmann weighted averaging (**Equation 2.1**, page 20) using a temperature of $kT = 2$ REU. Using the mutational energies for each macrostate, the weights for each state were optimized as described in **Chapter 2.5.0.3** (page 39). The final weights for each state were E-NADPH: 0.579, E-DHF-NADPH: 0.126, TS: 1.0, E-DHF-NADP_x: 0.406, E-THF: 0.108. With these weights and energies, fitnesses were calculated for the multi-state system (**Equation 2.5**, page 30) for each mutation. A representative energy was calculated

(**Equation 2.6**, page 38) and amino acid frequencies were calculated based on a Boltzmann-weighted ensemble (**Equation 2.7**, page 38).

$$E_{j,representative} = Ln\left(\frac{1}{F_j} - 1\right) \text{ for each } j \text{ mutation} \quad (2.6)$$

$$\text{Boltzmann frequency}_j = \frac{e^{s(E_{j,representative}-O)}}{\sum_i^m e^{s(E_{j,representative}-O)}} = \frac{\frac{1}{F} - 1}{\sum_j^m \frac{1}{F} - 1} \text{ for each } j \text{ mutation} \quad (2.7)$$

2.5.0.2 Calculating Amino Acid Energies from Sequence Tolerance Simulations Statistical energies for mutations were generated over flexible backbone ensembles. For each input PDB structure from the 5 states in **Figure 1.2** (page 4), an ensemble of 200 backbone conformers was generated using the Backrub application with 10,000 - 15,900 trial moves evaluated by the Metropolis Criterion with $kT = 0.6$. For each backbone in each ensemble, all positions were designed with the FixBB simulated annealing algorithm. Because DHFR is too large to design in a single simulation, design was performed using three overlapping sets of residues that ranged from 80-90 positions each. In each round of these simulations, these designed positions were allowed to sample all amino acids. In each round, positions that were not designed were allowed to repack. The output structure and sequences from one round was the input for the next round.

For each macrostate, the 200 output sequences were collected and the amino acid counts were tabulated at each position. A value of 1 was added to each count to prevent any amino acids from having a frequency of 0.0. Amino acid frequencies were calculated from the amino acid counts. Energies for each amino acid were calculated as the natural log (Ln) of the amino acid frequency. Using the mutational energies for each macrostate, the weights for each state were optimized as described in **Chapter 2.5.0.3** (page 39). The final weights for each state were E-NADPH: 0.0,

E-DHF-NADPH: 0.687, TS: 1.0, E-DHF-NADP_x: 0.456, E-THF: 0.0. With these weights and energies, amino acid frequencies were calculated as described in **Chapter 2.5.0.1** (page 37).

2.5.0.3 Optimizing Parameters The method for optimizing parameters was developed by Tianjiao Zhang. With mutational energies for each macrostate as inputs, amino acid frequencies were calculated from fitnesses starting with random steepness and weight values. The values for these parameters were optimized using a Cuckoo-search with random Levy flights[105] to minimize the difference in frequencies in the computational predictions and in a MSA of DHFR orthologues.

$$\text{Objective} = 1 - \sum_{i,j} |\text{frequency}_{ij,MSA} - \text{frequency}_{ij,prediction}| \quad (2.8)$$

for each i amino acid at each j position

References

- [1] C. B. Anfinsen, E. Haber, M. Sela, and Jr. White, F. H. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A*, 47(9):1309–14, 1961.
- [2] B. I. Dahiyat and S. L. Mayo. De novo protein design: fully automated sequence selection. *Science*, 278(5335):82–7, 1997.
- [3] B. Kuhlman. Designing protein structures and complexes with the molecular modeling program rosetta. *J Biol Chem*, 294(50):19436–19443, 2019.
- [4] P. Gainza, H. M. Nisonoff, and B. R. Donald. Algorithms for protein design. *Curr Opin Struct Biol*, 39:16–26, 2016.
- [5] B. Kuhlman and P. Bradley. Advances in protein structure prediction and design. *Nat Rev Mol Cell Biol*, 20(11):681–697, 2019.
- [6] P. S. Huang, S. E. Boyken, and D. Baker. The coming of age of de novo protein design. *Nature*, 537(7620):320–7, 2016.
- [7] K. W. Kaufmann, G. H. Lemmon, S. L. Deluca, J. H. Sheehan, and J. Meiler. Practically useful: what the rosetta protein modeling suite can do for you. *Biochemistry*, 49(14):2987–98, 2010.
- [8] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302(5649):1364–8, 2003.
- [9] P. S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio,

- T. Gonen, B. Luisi, and D. Baker. High thermodynamic stability of parametrically designed helical bundles. *Science*, 346(6208):481–485, 2014.
- [10] Z. Chen, S. E. Boyken, M. Jia, F. Busch, D. Flores-Solis, M. J. Bick, P. Lu, Z. L. VanAernum, A. Sahasrabudhe, R. A. Langan, S. Bermeo, T. J. Brunette, V. K. Mulligan, L. P. Carter, F. DiMaio, N. G. Sgourakis, V. H. Wysocki, and D. Baker. Programmable design of orthogonal protein heterodimers. *Nature*, 565(7737):106–111, 2019.
- [11] J. B. Maguire, S. E. Boyken, D. Baker, and B. Kuhlman. Rapid sampling of hydrogen bond networks for computational protein design. *J Chem Theory Comput*, 14(5):2751–2760, 2018.
- [12] P. Lu, D. Min, F. DiMaio, K. Y. Wei, M. D. Vahey, S. E. Boyken, Z. Chen, J. A. Fallas, G. Ueda, W. Sheffler, V. K. Mulligan, W. Xu, J. U. Bowie, and D. Baker. Accurate computational design of multipass transmembrane proteins. *Science*, 359(6379):1042–1046, 2018.
- [13] J. A. Davey, A. M. Damry, N. K. Goto, and R. A. Chica. Rational design of proteins that exchange on functional timescales. *Nat Chem Biol*, 13(12):1280–1285, 2017.
- [14] C. A. Voigt, S. L. Mayo, F. H. Arnold, and Z. G. Wang. Computationally focusing the directed evolution of proteins. *J Cell Biochem Suppl*, Suppl 37:58–63, 2001.
- [15] R. A. Chica, M. M. Moore, B. D. Allen, and S. L. Mayo. Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc Natl Acad Sci U S A*, 107(47):20257–62, 2010.
- [16] A. A. Glasgow, Y. M. Huang, D. J. Mandell, M. Thompson, R. Ritterson, A. L. Loshbaugh, J. Pellegrino, C. Krivacic, R. A. Pache, K. A. Barlow, N. Ollikainen, D. Jeon, M. J. S. Kelly,

- J. S. Fraser, and T. Kortemme. Computational design of a modular protein sense-response system. *Science*, 366(6468):1024–1028, 2019.
- [17] Z. Chen, M. C. Johnson, J. Chen, M. J. Bick, S. E. Boyken, B. Lin, J. J. De Yoreo, J. M. Kollman, D. Baker, and F. DiMaio. Self-assembling 2d arrays with de novo protein building blocks. *J Am Chem Soc*, 141(22):8891–8895, 2019.
- [18] H. Shen, J. A. Fallas, E. Lynch, W. Sheffler, B. Parry, N. Jannetty, J. Decarreau, M. Wagenbach, J. J. Vicente, J. Chen, L. Wang, Q. Dowling, G. Oberdorfer, L. Stewart, L. Wordeman, J. De Yoreo, C. Jacobs-Wagner, J. Kollman, and D. Baker. De novo design of self-assembling helical protein filaments. *Science*, 362(6415):705–709, 2018.
- [19] P. S. Huang, K. Feldmeier, F. Parmeggiani, D. A. F. Velasco, B. Höcker, and D. Baker. De novo design of a four-fold symmetric tim-barrel protein with atomic-level accuracy. *Nat Chem Biol*, 12(1):29–34, 2016.
- [20] K. L. Spiller and G. Vunjak-Novakovic. Clinical translation of controlled protein delivery systems for tissue engineering. *Drug Deliv Transl Res*, 5(2):101–15, 2015.
- [21] D. A. Wang. Delivery of engineered therapeutics and medication. *Pharm Res*, 28(6):1237–40, 2011.
- [22] S. Kobsa and W. M. Saltzman. Bioengineering approaches to controlled protein delivery. *Pediatr Res*, 63(5):513–9, 2008.
- [23] A. M. Sevy, S. Panda, Jr. Crowe, J. E., J. Meiler, and Y. Vorobeychik. Integrating linear optimization with structural modeling to increase hiv neutralization breadth. *PLoS Comput Biol*, 14(2):e1005999, 2018.

- [24] A. M. Sevy, N. C. Wu, I. M. Gilchuk, E. H. Parrish, S. Burger, D. Yousif, M. B. M. Nagel, K. L. Schey, I. A. Wilson, Jr. Crowe, J. E., and J. Meiler. Multistate design of influenza antibodies improves affinity and breadth against seasonal viruses. *Proc Natl Acad Sci U S A*, 116(5):1597–1602, 2019.
- [25] T. A. Whitehead, A. Chevalier, Y. Song, C. Dreyfus, S. J. Fleishman, C. De Mattos, C. A. Myers, H. Kamisetty, P. Blair, I. A. Wilson, and D. Baker. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol*, 30(6):543–8, 2012.
- [26] M. J. Bick, P. J. Greisen, K. J. Morey, M. S. Antunes, D. La, B. Sankaran, L. Reymond, K. Johnsson, J. I. Medford, and D. Baker. Computational design of environmental sensors for the potent opioid fentanyl. *Elife*, 6, 2017.
- [27] Y. Li and P. C. Cirino. Recent advances in engineering proteins for biocatalysis. *Biotechnol Bioeng*, 111(7):1273–87, 2014.
- [28] S. J. Bertolani, D. A. Carlin, and J. B. Siegel. Computational introduction of catalytic activity into proteins. *Methods Mol Biol*, 1414:213–31, 2016.
- [29] W. S. Mak and J. B. Siegel. Computational enzyme design: transitioning from catalytic proteins to enzymes. *Curr Opin Struct Biol*, 27:87–94, 2014.
- [30] J. B. Siegel, A. L. Smith, S. Poust, A. J. Wargacki, A. Bar-Even, C. Louw, B. W. Shen, C. B. Eiben, H. M. Tran, E. Noor, J. L. Gallaher, J. Bale, Y. Yoshikuni, M. H. Gelb, J. D. Keasling, B. L. Stoddard, M. E. Lidstrom, and D. Baker. Computational protein design enables a novel one-carbon assimilation pathway. *Proc Natl Acad Sci U S A*, 112(12):3704–9, 2015.

- [31] C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard, and D. Baker. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*, 501(7466):212–216, 2013.
- [32] Y. Hsia, J. B. Bale, S. Gonen, D. Shi, W. Sheffler, K. K. Fong, U. Nattermann, C. Xu, P. S. Huang, R. Ravichandran, S. Yi, T. N. Davis, T. Gonen, N. P. King, and D. Baker. Design of a hyperstable 60-subunit protein dodecahedron. [corrected]. *Nature*, 535(7610):136–9, 2016.
- [33] Jr. Dunbrack, R. L. and M. Karplus. Backbone-dependent rotamer library for proteins. application to side-chain prediction. *J Mol Biol*, 230(2):543–74, 1993.
- [34] Jr. Dunbrack, R. L. Rotamer libraries in the 21st century. *Curr Opin Struct Biol*, 12(4):431–40, 2002.
- [35] P. Gainza, K. E. Roberts, and B. R. Donald. Protein design using continuous rotamers. *PLoS Comput Biol*, 8(1):e1002335, 2012.
- [36] M. V. Shapovalov and Jr. Dunbrack, R. L. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*, 19(6):844–58, 2011.
- [37] C. Negron and A. E. Keating. Multistate protein design using clever and classy. *Methods Enzymol*, 523:171–90, 2013.
- [38] J. Desmet, M. De Maeyer, B. Hazes, and I. Lasters. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, 356(6369):539–42, 1992.
- [39] B. I. Dahiyat and S. L. Mayo. Protein design automation. *Protein Sci*, 5(5):895–903, 1996.

- [40] J. Desmet, J. Spriet, and I. Lasters. Fast and accurate side-chain topology and energy refinement (faster) as a new method for protein structure optimization. *Proteins*, 48(1):31–43, 2002.
- [41] K. A. Reynolds, W. P. Russ, M. Socolich, and R. Ranganathan. Evolution-based design of proteins. *Methods Enzymol*, 523:213–35, 2013.
- [42] P. Tian, J. M. Louis, J. L. Baber, A. Aniana, and R. B. Best. Co-evolutionary fitness landscapes for sequence design. *Angew Chem Int Ed Engl*, 57(20):5674–5678, 2018.
- [43] P. Xiong, M. Wang, X. Zhou, T. Zhang, J. Zhang, Q. Chen, and H. Liu. Protein design with a comprehensive statistical energy function and boosted by experimental selection for foldability. *Nat Commun*, 5:5330, 2014.
- [44] P. Xiong, X. Hu, B. Huang, J. Zhang, Q. Chen, and H. Liu. Increasing the efficiency and accuracy of the abacus protein sequence design method. *Bioinformatics*, 36(1):136–144, 2020.
- [45] F. Zheng and G. Grigoryan. Sequence statistics of tertiary structural motifs reflect protein stability. *PLoS One*, 12(5):e0178272, 2017.
- [46] V. Frappier, J. M. Jenson, J. Zhou, G. Grigoryan, and A. E. Keating. Tertiary structural motif sequence statistics enable facile prediction and design of peptides that bind anti-apoptotic bfl-1 and mcl-1. *Structure*, 27(4):606–617.e5, 2019.
- [47] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J Comput Chem*, 25(9):1157–74, 2004.
- [48] M. J. Hsieh and R. Luo. Physical scoring function based on amber force field and poisson-boltzmann implicit solvent for protein structure prediction. *Proteins*, 56(3):475–86, 2004.

- [49] T. Lazaridis and M. Karplus. Effective energy function for proteins in solution. *Proteins*, 35(2):133–52, 1999.
- [50] B. R. Brooks, 3rd Brooks, C. L., Jr. Mackerell, A. D., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. Charrmm: the biomolecular simulation program. *J Comput Chem*, 30(10):1545–614, 2009.
- [51] R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O’Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, Jr. Dunbrack, R. L., R. Das, D. Baker, B. Kuhlman, T. Kortemme, and J. J. Gray. The rosetta all-atom energy function for macromolecular modeling and design. *J Chem Theory Comput*, 13(6):3031–3048, 2017.
- [52] G. D. Friedland, N. A. Lakomek, C. Griesinger, J. Meiler, and T. Kortemme. A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. *PLoS Comput Biol*, 5(5):e1000393, 2009.
- [53] C. A. Smith and T. Kortemme. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J Mol Biol*, 380(4):742–56, 2008.
- [54] F. Zheng, J. Zhang, and G. Grigoryan. Tertiary structural propensities reveal fundamental sequence/structure relationships. *Structure*, 23(5):961–971, 2015.
- [55] D. J. Mandell, E. A. Coutsiias, and T. Kortemme. Sub-angstrom accuracy in protein loop

- reconstruction by robotics-inspired conformational sampling. *Nat Methods*, 6(8):551–2, 2009.
- [56] A. Stein and T. Kortemme. Improvements to robotics-inspired conformational sampling in rosetta. *PLoS One*, 8(5):e63090, 2013.
- [57] A. L. Loshbaugh and T. Kortemme. Comparison of rosetta flexible-backbone computational protein design methods on binding interactions. *Proteins*, 88(1):206–226, 2020.
- [58] J. A. Davey and R. A. Chica. Multistate computational protein design with backbone ensembles. *Methods Mol Biol*, 1529:161–179, 2017.
- [59] J. A. Davey, A. M. Damry, C. K. Euler, N. K. Goto, and R. A. Chica. Prediction of stable globular proteins using negative design with non-native backbone ensembles. *Structure*, 23(11):2011–21, 2015.
- [60] C. A. Smith and T. Kortemme. Predicting the tolerated sequences for proteins and protein interfaces using rosettabackrub flexible backbone design. *PLoS One*, 6(7):e20451, 2011.
- [61] A. B. Rubenstein, M. A. Pethe, and S. D. Khare. Mfpred: Rapid and accurate prediction of protein-peptide recognition multispecificity using self-consistent mean field theory. *PLoS Comput Biol*, 13(6):e1005614, 2017.
- [62] K. A. Barlow, Ó Conchúir S, S. Thompson, P. Suresh, J. E. Lucas, M. Heinonen, and T. Kortemme. Flex ddg: Rosetta ensemble-based estimation of changes in protein-protein binding affinity upon mutation. *J Phys Chem B*, 122(21):5389–5399, 2018.
- [63] N. Ollikainen and T. Kortemme. Computational protein design quantifies structural constraints on amino acid covariation. *PLoS Comput Biol*, 9(11):e1003313, 2013.

- [64] N. F. Polizzi, Y. Wu, T. Lemmin, A. M. Maxwell, S. Q. Zhang, J. Rawson, D. N. Beratan, M. J. Therien, and W. F. DeGrado. De novo design of a hyperstable non-natural protein-ligand complex with sub-Å accuracy. *Nat Chem*, 9(12):1157–1164, 2017.
- [65] N. Ollikainen, R. M. de Jong, and T. Kortemme. Coupling protein side-chain and backbone flexibility improves the re-design of protein-ligand specificity. *PLoS Comput Biol*, 11(9):e1004335, 2015.
- [66] S. Warszawski, R. Netzer, D. S. Tawfik, and S. J. Fleishman. A "fuzzy"-logic language for encoding multiple physical traits in biomolecules. *J Mol Biol*, 426(24):4125–4138, 2014.
- [67] M. Babor and T. Kortemme. Multi-constraint computational design suggests that native sequences of germline antibody h3 loops are nearly optimal for conformational flexibility. *Proteins*, 75(4):846–58, 2009.
- [68] A. Leaver-Fay, R. Jacak, P. B. Stranges, and B. Kuhlman. A generic program for multistate protein design. *PLoS One*, 6(7):e20937, 2011.
- [69] P. Löffler, S. Schmitz, E. Hupfeld, R. Sterner, and R. Merkl. Rosetta:msf: a modular framework for multi-state computational protein design. *PLoS Comput Biol*, 13(6):e1005600, 2017.
- [70] M. A. Hallen and B. R. Donald. comets (constrained optimization of multistate energies by tree search): A provable and efficient protein design algorithm to optimize binding affinity and specificity with respect to sequence. *J Comput Biol*, 23(5):311–21, 2016.
- [71] G. Grigoryan, F. Zhou, S. R. Lustig, G. Ceder, D. Morgan, and A. E. Keating. Ultra-fast evaluation of protein energies directly from sequence. *PLoS Comput Biol*, 2(6):e63, 2006.

- [72] G. Grigoryan, A. W. Reinke, and A. E. Keating. Design of protein-interaction specificity gives selective bzip-binding peptides. *Nature*, 458(7240):859–64, 2009.
- [73] A. M. Sevy, T. M. Jacobs, Jr. Crowe, J. E., and J. Meiler. Design of protein multi-specificity using an independent sequence search reduces the barrier to low energy sequences. *PLoS Comput Biol*, 11(7):e1004300, 2015.
- [74] M. F. Sauer, A. M. Sevy, Jr. Crowe, J. E., and J. Meiler. Multi-state design of flexible proteins predicts sequences optimal for conformational change. *PLoS Comput Biol*, 16(2):e1007339, 2020.
- [75] S. D. Dunn, L. M. Wahl, and G. B. Gloor. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics*, 24(3):333–40, 2008.
- [76] R. J. Dickson, L. M. Wahl, A. D. Fernandes, and G. B. Gloor. Identifying and seeing beyond multiple sequence alignment errors using intra-molecular protein covariation. *PLoS One*, 5(6):e11082, 2010.
- [77] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci U S A*, 108(49):E1293–301, 2011.
- [78] F. Morcos, T. Hwa, J. N. Onuchic, and M. Weigt. Direct coupling analysis for protein contact prediction. *Methods Mol Biol*, 1137:55–70, 2014.
- [79] S. W. Lockless and R. Ranganathan. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, 286(5438):295–9, 1999.

- [80] V. H. Salinas and R. Ranganathan. Coevolution-based inference of amino acid interactions underlying protein function. *Elife*, 7, 2018.
- [81] D. M. Fowler and S. Fields. Deep mutational scanning: a new style of protein science. *Nat Methods*, 11(8):801–7, 2014.
- [82] P. A. Romero, T. M. Tran, and A. R. Abate. Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc Natl Acad Sci U S A*, 112(23):7159–64, 2015.
- [83] P. Bandaru, N. H. Shah, M. Bhattacharyya, J. P. Barton, Y. Kondo, J. C. Cofsky, C. L. Gee, A. K. Chakraborty, T. Kortemme, R. Ranganathan, and J. Kuriyan. Deconstruction of the ras switching cycle through saturation mutagenesis. *Elife*, 6, 2017.
- [84] E. Firnberg, J. W. Labonte, J. J. Gray, and M. Ostermeier. A comprehensive, high-resolution map of a gene’s fitness landscape. *Mol Biol Evol*, 31(6):1581–92, 2014.
- [85] C. S. Nelson, C. K. Fuller, P. M. Fordyce, A. L. Greninger, H. Li, and J. L. DeRisi. Microfluidic affinity and chip-seq analyses converge on a conserved foxp2-binding motif in chimp and human, which enables the detection of evolutionarily novel targets. *Nucleic Acids Res*, 41(12):5991–6004, 2013.
- [86] A. K. Aditham, T. C. Shimko, and P. M. Fordyce. Bet-seq: Binding energy topographies revealed by microfluidics and high-throughput sequencing. *Methods Cell Biol*, 148:229–250, 2018.
- [87] A. Chevalier, D. A. Silva, G. J. Rocklin, D. R. Hicks, R. Vergara, P. Murapa, S. M. Bernard, L. Zhang, K. H. Lam, G. Yao, C. D. Bahl, S. I. Miyashita, I. Goreshnik, J. T. Fuller, M. T. Koday, C. M. Jenkins, T. Colvin, L. Carter, A. Bohn, C. M. Bryan, D. A. Fernández-Velasco,

- L. Stewart, M. Dong, X. Huang, R. Jin, I. A. Wilson, D. H. Fuller, and D. Baker. Massively parallel de novo protein design for targeted therapeutics. *Nature*, 550(7674):74–79, 2017.
- [88] G. J. Rocklin, T. M. Chidyausiku, I. Goreshnik, A. Ford, S. Houlston, A. Lemak, L. Carter, R. Ravichandran, V. K. Mulligan, A. Chevalier, C. H. Arrowsmith, and D. Baker. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science*, 357(6347):168–175, 2017.
- [89] K. A. Dill. Additivity principles in biochemistry. *J Biol Chem*, 272(2):701–4, 1997.
- [90] Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *Journal of the American Chemical Society*, 104(17):4546–4559, 1982.
- [91] Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. analysis of experimental results. *Journal of the American Chemical Society*, 104(17):4559–4570, 1982.
- [92] J. S. Fraser, H. van den Bedem, A. J. Samelson, P. T. Lang, J. M. Holton, N. Echols, and T. Alber. Accessing protein conformational ensembles using room-temperature x-ray crystallography. *Proc Natl Acad Sci U S A*, 108(39):16247–52, 2011.
- [93] D. A. Keedy, H. van den Bedem, D. A. Sivak, G. A. Petsko, D. Ringe, M. A. Wilson, and J. S. Fraser. Crystal cryocooling distorts conformational heterogeneity in a model michaelis complex of dhfr. *Structure*, 22(6):899–910, 2014.
- [94] R. B. Fenwick, H. van den Bedem, J. S. Fraser, and P. E. Wright. Integrated description of protein dynamics from room-temperature x-ray crystallography and nmr. *Proc Natl Acad Sci U S A*, 111(4):E445–54, 2014.

- [95] D. A. Keedy, L. R. Kenner, M. Warkentin, R. A. Woldeyes, J. B. Hopkins, M. C. Thompson, A. S. Brewster, A. H. Van Benschoten, E. L. Baxter, M. Uervirojnangkoorn, S. E. McPhillips, J. Song, R. Alonso-Mori, J. M. Holton, W. I. Weis, A. T. Brunger, S. M. Soltis, H. Lemke, A. Gonzalez, N. K. Sauter, A. E. Cohen, H. van den Bedem, R. E. Thorne, and J. S. Fraser. Mapping the conformational landscape of a dynamic enzyme by multitemperature and xfel crystallography. *Elife*, 4, 2015.
- [96] J. H. Mills, S. D. Khare, J. M. Bolduc, F. Forouhar, V. K. Mulligan, S. Lew, J. Seetharaman, L. Tong, B. L. Stoddard, and D. Baker. Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. *J Am Chem Soc*, 135(36):13393–9, 2013.
- [97] L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, 3rd Barbas, C. F., D. Hilvert, K. N. Houk, B. L. Stoddard, and D. Baker. De novo computational design of retro-aldol enzymes. *Science*, 319(5868):1387–91, 2008.
- [98] D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, and D. Baker. Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192):190–5, 2008.
- [99] J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, and D. Baker. Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. *Science*, 329(5989):309–13, 2010.

- [100] J. J. Havranek and P. B. Harbury. Automated design of specificity in molecular recognition. *Nat Struct Biol*, 10(1):45–52, 2003.
- [101] X. I. Ambroggio and B. Kuhlman. Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J Am Chem Soc*, 128(4):1154–61, 2006.
- [102] V. Ovchinnikov, K. Nam, and M. Karplus. A simple and accurate method to calculate free energy profiles and reaction rates from restrained molecular simulations of diffusive processes. *J Phys Chem B*, 120(33):8457–72, 2016.
- [103] S. Ovchinnikov, H. Kamisetty, and D. Baker. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife*, 3:e02030, 2014.
- [104] S. Hahn, O. Ashenberg, G. Grigoryan, and A. E. Keating. Identifying and reducing error in cluster-expansion approximations of protein energies. *J Comput Chem*, 31(16):2900–14, 2010.
- [105] X. Yang and D. Suash. Cuckoo search via levy flights. *2009 World Congress on Nature and Biologically Inspired Computing*, pages 210–214, 2009.

3 Deep Mutational Scanning to Resolve Highly Active Single-Point Mutants of a Model Enzyme

3.1 Introduction

A central tenet of protein biophysics is that sequence gives rise to structure which in turn gives rise to function. This understanding has largely arisen from careful examination of individual protein sequences, their structure, and their corresponding function. Further insights have been drawn from small sets of point mutations that represent small deviations from the starting sequence. New functional proteins have been engineered and evolved by capitalizing on the ability to manipulate a protein's function by manipulating a small percentage of its sequence[1].

While we know that large regions of sequence space — sparsely sampled by homologous proteins — are redundant for a particular function, single point mutations may eliminate function, enable new function, or have no impact. The challenge in distinguishing a functional protein sequence with 30% identity to a known protein and non-functional single point mutant with the same model highlights the complexity of sequence-structure-function relationships. It is a continuing challenge to identify which mutations will have a thermodynamic impact that disrupts protein fold and/or function. Accordingly, we do not know where the bounds of functional sequence space are. Mapping the mutational landscape of functional proteins under conditions that are carefully calibrated to report on a molecular function would allow us to address an elusive question about how sequence space is shaped by the thermodynamic constraints on functional proteins.

In this chapter, we sought to first accurately measure a mutational landscape for a model system, and then analyze that landscape to make inferences about how the structure and sequence

are constrained by the requirements for function. To accomplish this, we calibrated a selection assay for DHFR and screened a library of all possible single point mutants using deep mutational scanning (DMS).

3.2 Background

3.2.1 Deep Mutational Scanning

Currently, DMS is the preeminent method for measuring mutational landscapes. In DMS, mapping the mutational landscape begins with generating a library that represents the sequence space of interest. A standard library is all possible single point mutants of the target protein sequence. Selection is then performed on this library, generally in a competitive growth assay[2–5], phage display[6], or by using fluorescence activated cell sorting (FACS)[7–9] or fluorescence activated droplet sorting (FADS)[10], depending on the selectable trait.

3.2.1.1 Scientific Insights from Deep Mutational Scanning Deep mutational scanning studies have provided insights into the evolution of new protein functions[4, 7, 11], protein design[9, 12], functional trade-offs[13, 14], and adaptation to altered environments[15]. Despite these lessons, we have not learned why protein natural protein sequences are fixed, or any overarching rules that govern protein architecture and function. With a few exceptions[2, 4, 15, 16], the consistent picture from DMS studies is a general tolerance to mutation for residues outside of active sites and binding interfaces[6, 11, 13, 17, 18]. In almost all cases, it is not clear if this tolerance is the result of the specific conditions of the selection experiment or if it represents a general property of proteins. This uncertainty arises from a lack of calibration of the selection conditions in many of these experiments.

3.2.1.2 Technical Details of Deep Mutational Scanning From the DMS selection experiment, the allelic population over timepoints (for growth) or sorting gates (for FACS/FADS) is measured using standard deep sequencing. The Δ (selection population – unselected population) or linear regression slope over allele frequencies is used to calculate a selection coefficient (**Figure 3.1**, page 56). Since the initial development of the DMS methodology, it has become more common to use experimental set-ups that allow for a regression to be calculated over multiple data measurements. Because DMS libraries are often generated from degenerate codons, selection coefficients can be calculated for each allele (unique DNA sequence) or for each mutation (unique protein sequence) with averaging over redundant alleles. Given the multiple noise factors, each DMS experiment should be performed with multiple biological replicates for quantifying data quality and justifying the data analysis and interpretation.

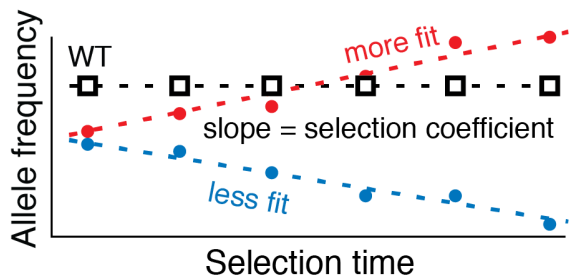


Figure 3.1: The selection coefficient from DMS for each mutant is the slope of the linear regression of allele frequency over time. The wild-type (squares) value is normalized to zero. Advantageous (red) mutations increase and disadvantageous (blue) mutations decrease in frequency.

Timepoints The choice of timepoints has been examined as a critical factor in setting up a DMS experiment[19]. While the depth of sequencing has a negative linear relationship with the mean squared error (MSE) for selection coefficients, increasing the number of timepoints exponentially decreases the error. The experimenter should also consider the spacing of the timepoints in the experiment to better observe mutational impacts. Mutational impacts are

measurable differences in an *in vivo* or *in vitro* assay for activity, and increasing the number of timepoints early in selection will improve resolution of the strongest mutational impacts, and increasing the number of timepoints late in selection will improve the resolution of weaker mutational impacts. Furthermore, the latest timepoints will have the greatest change in signal, and some components of the noise will scale correspondingly. Thus, these points can have outsized impacts on linear regression. Having tightly spaced timepoints at the end of the experiment will decrease variation for a broad range of mutational impacts. Noise from the deep sequencer will be added to the noise from selection, and this technical noise can only be addressed by increasing the depth of sequencing or changing the sequencing platform to one with better average Q30 rates. In general, it is advisable to optimize the amount of library sequencing such that sufficient timepoints can be collected to resolve the relevant selection regime with ~1,000 counts per allele at each timepoint, and sequencing platforms with 4-color laser optics yield higher quality sequencing data than sequencing platforms with 2-color LED optics.

Calibration Another critical factor is calibrating the DMS selection such that the selection coefficient is an interpretable reporter. For the earliest deep mutational scanning studies, a transfer function for mapping selection coefficients to biochemical properties was not determined[6], but more recent studies have explicitly calibrated selection using mutants with known enzymatic properties [4, 7, 20] Comparing the selection coefficient to the biochemical properties of a panel of well-characterized point mutants is the gold standard for interpreting selection coefficients. While this analysis can be done after selection, identifying and characterizing a point mutant panel before performing the full selection can allow for careful calibration of the selection conditions.

3.3 Results

We began by applying DMS to map the mutational landscape of our model enzyme, DHFR (**Chapter 1.1**, page 2). Specifically, we sought to test the hypothesis that the DHFR mutational landscape is a restrictive mutational landscape constrained by multiple required functional conformations, we performed DMS over all single point mutants of DHFR.

3.3.1 Preliminary Selection on DHFR at UT Southwestern

In Kim Reynold's lab at the University of Texas, Southwestern (UTSW), we performed selection on mutations to the first 40 positions in DHFR. We made initial assumptions for several of the experiment parameters, including selection time, growth medium and supplements, temperature, minimum number of sequencing counts per timepoint (see **Methods, Chapter 3.5.0.7**, page 113). Only one biological replicate was performed in M9 minimal medium with 5 $\mu\text{g}/\text{mL}$ supplemented thymidine, and unlike the previous study[20], selection was performed for 48 hours in a turbidostat (**Figure 3.2**, page 59) to maintain culture density in early Log-phase growth ($\text{OD}_{600} < 0.15$). With growth rate controlled by a fixed OD_{600} , the change in Log_2 of population frequencies over the selection timepoints can be fit to a linear model (**Figure 3.1**, page 56). Furthermore, recent results have shown that the growth rate of *E. coli* can be modulated by changing the maximum OD_{600} , thus tuning the selection pressure (Kim Reynolds and Thuy Nguyen, correspondence). Results from this experiment using a protocol adapted from previous experiments[20] could then be used to inform a more optimized experimental protocol.

Timepoint samples from the selection experiment were analyzed by deep sequencing as described in **Methods (Chapter 3.5.0.7**, page 113). The median number of counts per

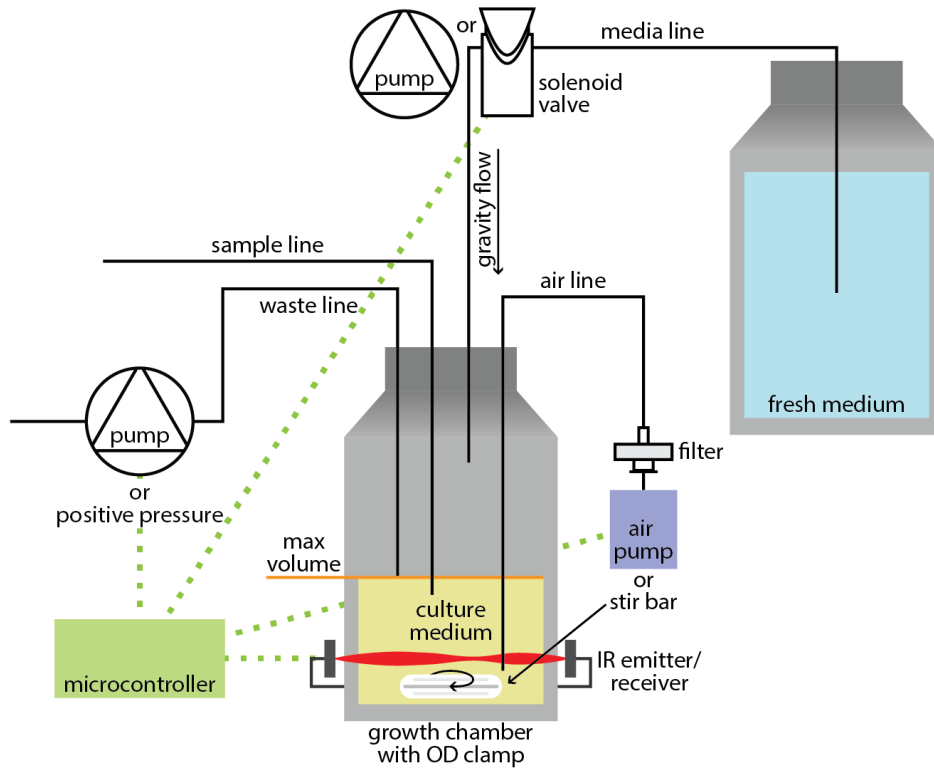


Figure 3.2: A turbidostat is a continuous culture device that maintains a constant cell density.

mutant over all time points was 500 (**Figure 3.3**, page 60). The distribution of frequencies was a normal distribution over $\text{Log}_{10}(\text{counts})$ and did not show a strong bias, except that the counts for the WT sequence were an order of magnitude higher than those for any other sequence. This is an expected artifact from the PCR-based construction of the library (see **Methods, Chapter 3.5.0.2**, page 108). We determined that the test library was of usable quality.

Selection coefficients were fit to the allele frequencies in the sequencing data (see **Methods, Chapter 3.5.0.7**, page 113). We observed that the Pearson correlation coefficient (R) from the fitting was not strongly correlated to the allele frequencies in the library (y-axis in **Figure 3.4**, page 61), but it was correlated with the magnitude of the selection coefficient (point size in **Figure 3.4**, page 61). Because the selection coefficient is the slope of the linear regression of $\text{Log}_2(\text{frequency}_{\text{mutation}}/\text{frequency}_{\text{WT}})$, it is expected that selection coefficients near zero will

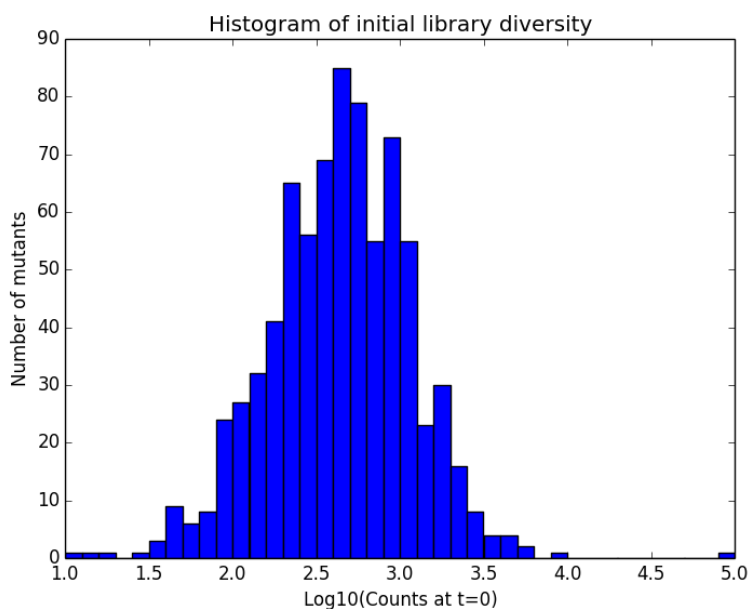


Figure 3.3: Counts per mutant at t=0 during preliminary selection on positions 2-40 of DHFR.

have poor correlation coefficients even because the noise is higher than the signal. We concluded that the sequencing data was of sufficient quality, but a better method for analyzing error in the data would be needed for the final selection experiment.

To see if our selection experiment could resolve a broad range of mutational impacts, we plotted the selection coefficients in a histogram. We observed a bimodal distribution of selection coefficients with $\sim 3/4$ of the population of mutants near WT-level fitness (**Figure 3.5**, page 61). This distribution is qualitatively similar to those from previous DMS experiments[3, 4, 7, 15, 21].

To examine if biochemically intuitive patterns emerged based on the position being mutated, we plotted the selection data as a heatmap. Within the heatmap, most positions show a general tolerance to mutation. In contrast, positions that lie within the active site such as I14, G15, W22, D27, F31, and T35 are generally intolerant to mutation. Together with the distribution of all selection coefficients, we interpreted these data as indicating that the selection pressure was not sufficiently stringent to resolve highly active mutants, and mutations at conserved positions



Figure 3.4: Analysis of the error in preliminary selection coefficients. The Pearson correlation coefficient (R) value from linear regression in selection coefficient calculated is plotted against the number of initial counts for mutant. The radius of each dot is |selection coefficient|. Colors are assigned to points randomly to make points visually distinguishable.

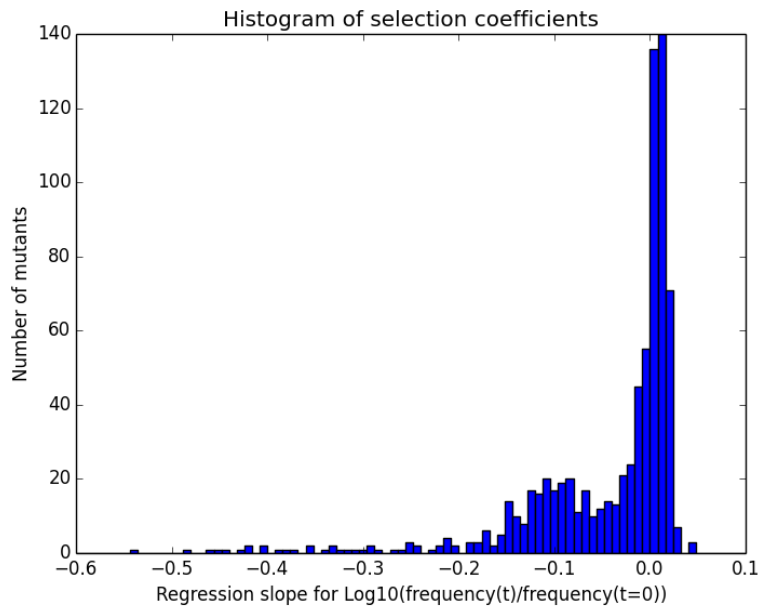


Figure 3.5: Distribution of selection coefficients from preliminary selection experiment.

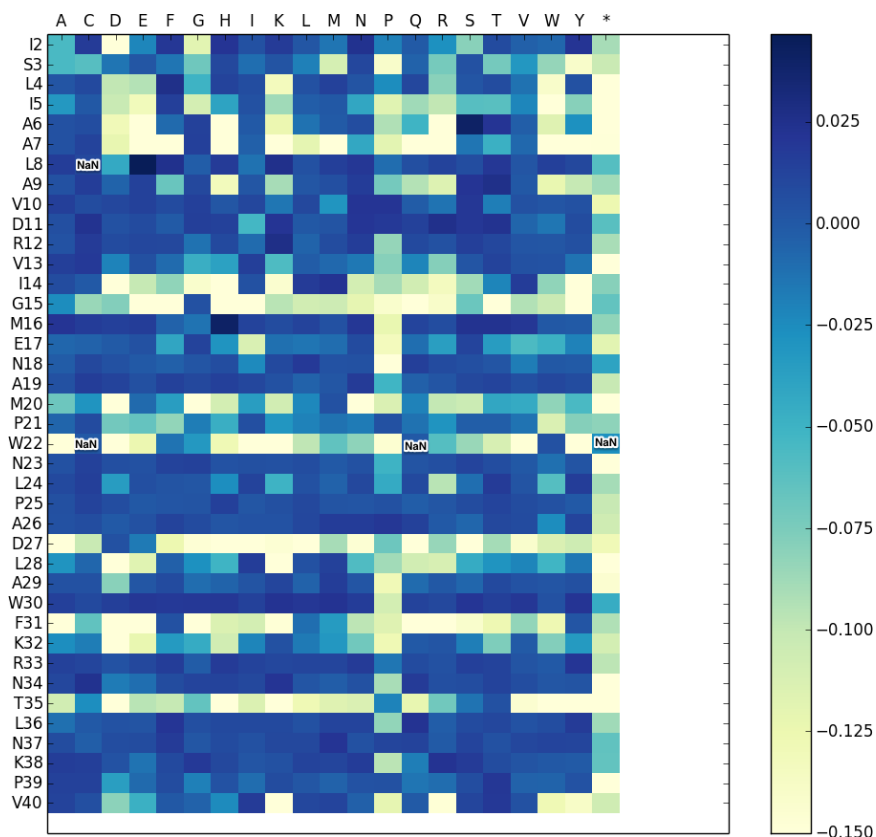


Figure 3.6: Heatmaps of selection coefficients from preliminary selection experiment. Amino acid positions (rows) are listed by the one-letter code of the WT amino acid and their position in the peptide chain. Amino acid substitutions (columns) are listed by the one-letter amino acid code. Stop codons are represented by an asterisk. Each mutant in the heatmap is colored according to its selection coefficient (colormap, right). The mutant frequencies at each timepoint are normalized to the WT frequency, so the WT sequence has a selection coefficient of 0 (see **Figure 3.1**, page 56).

with strong deleterious impacts on function will have selection coefficients that fall in the peak centered around -0.1.

To visualize the amino acid preferences at each position and examine if these preferences corresponded to a preference for the WT amino acid or amino acids that appear in DHFR orthologues, we represented the selection data as sequence profiles by calculating amino acid frequencies using the selection coefficients as energies in Boltzmann-weighted averaging (**Equation 2.1**, page 20). While the amino acid frequencies positions within the active site (15, 20, 22, 27, 31, and 35) match the amino acid frequencies for these conserved positions, the vast majority of positions do not have sequences preferences in selection that are consistent with

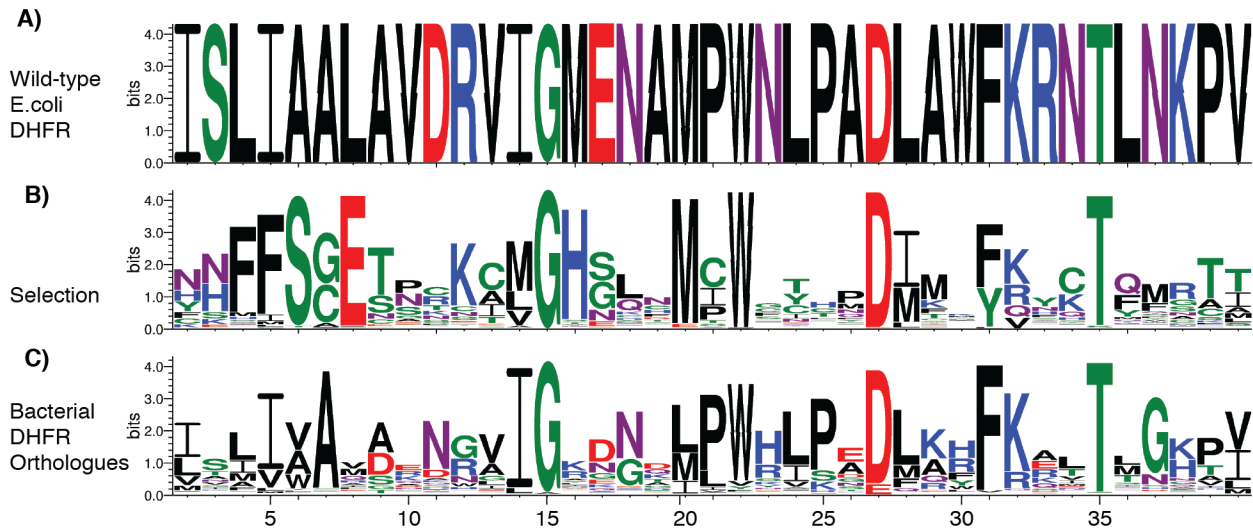


Figure 3.7: Amino acid preferences represented as sequence profiles for **A)** WT *E. coli* DHFR, **B)** preliminary selection, and **C)** a multiple sequence alignment of DHFR orthologues[22].

WT or orthologous DHFR sequences. We cannot determine from these data if this is a result of very different selection pressures in nature and in our experiment, or if the difference is a result of noise in selection coefficients that are very close to WT-level under these permissive selection conditions.

3.3.1.1 Conclusions from Preliminary Selection on DHFR From this experiment, we came to several conclusions that directed the optimization of our selection experiment. First, the selection pressure in the preliminary selection experiment is permissive and needs to be increased by modifying **1)** the supplements in the growth conditions, **2)** the expression levels for the *folA* and *thyA* genes on the library construct, or **3)** the baseline growth rate of the selection strain. Second, metrics for evaluating the error in the selection experiment are limited with only one repeat experiment, and analysis of error in the regression analysis is only indicative of the precision in the measurements. To assess the accuracy of our selection coefficients, multiple biological replicates need to be performed. Third, thresholds for considering a mutant present need to be treated as a function of frequency over all counts for each timepoint to make the

threshold independent of sequencing depth. Forth, the minimum number of timepoints over which a selection coefficient can be calculated should be evaluated based on how the minimum number of timepoints impacts the error between biological replicates. Finally, an average of 500 counts per timepoint was sufficient to measure large changes in allele frequency over the course of the population. The number of timepoints should be maximized while aiming for an average sequencing coverage depth of >500 counts per allele at each timepoint. We concluded that the selection pressure could be optimized using monoculture experiments with a focused panel of mutants and that issues in the data analysis could be addressed with turbidostat experiments under the optimized selection conditions.

3.3.2 Optimizing the Selection Pressure

We aimed to make DHFR activity limiting in the selection assay and increase selection pressure such that selection would resolve 2- to 3-fold changes in k_{cat}/K_M relative to the WT-level DHFR value of $7.1 \mu\text{M}^{-1}\text{s}^{-1}$.

3.3.2.1 Identifying Highly Active DHFR Variants To test our ability to resolve near WT-level activities, we needed a panel of mutants for calibration. The original assay had been calibrated against a panel of deleterious mutants[20], so we reasoned that this calibration could be expanded to include more highly fit DHFR mutants. Previously, hyperactive variants of DHFR had been identified [23, 24], and we were able to identify 4 highly active DHFR single point mutations – L28F, L28Y, L54F, and T113V – that all have k_{cat} values that are near or greater than that for WT DHFR and added them to the calibration set (**Table 3.1**, page 65).

Table 3.1: Michaelis-Menten kinetics for the set of DHFR mutants used to calibrate the selection are reported together with the reference from which the values were taken.

VARIANT	k_{cat} (s^{-1})	K_M (μM)	Reference
WT	7.95	1.1	[20]
W22H	1.89	18	[20]
L28F	18.5	9.9	this work
L28Y	19.2	21.2	this work
F31V	8.65	108	[20]
F31Y	20.61	80	[20]
M42F	79.2	13	[20]
L54F	6.3	0.7	[25]
L54I	7.88	35	[20]
T113V	32.9	21.4	[26]
G121V	0.3	6.1	[20]
F131Y/L54I	1.94	168.3	[20]
F31Y/G121V	0.13	90.6	[20]
M42F/G121V	0.4	71.8	[20]
L54I/G121V	0.22	73	[20]

3.3.2.2 Comparing DHFR Selection with Two TYMS Promoters To further increase the selection pressure in our assay, we switched to a second construct generated by our collaborator, Kim Reynolds, that replaced the lac operator upstream of TYMS with a TET promoter (**Table 5.2**, page 174). We examined the relative expression levels of DHFR and TYMS in this new pTET vector (SMT201) as compared to the original pACYC vector and pACYC variant with no His₆ tag (SMT101). From the western-blots, we determined that ratio of TYMS expression to DHFR expression is several-fold higher in two pTET vectors (SMT201, SMT205) as compared to that in the pACYC vectors. Further competitive growth experiments performed in the plate reader under induction of TYMS with anhydrotetracycline (aTC) (see **Methods, Chapter 3.5.0.4**, page 110) did not change the growth rates (**Figure 3.9**, page 66), so no further optimization of TYMS expression was attempted.

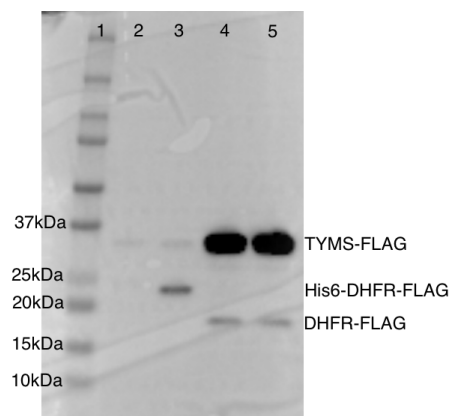


Figure 3.8: Expression of FLAG-tagged DHFR and TYMS off the selection plasmid. The relative expression levels of DHFR and TYMS was observed by anti-FLAG western blot of selection plasmids with modifications to the TYMS promoter and DHFR upstream sequence. Lane identities are marked by a number (top): **1**) Ladder, **2-5**) lysates from ER2566 $\Delta folA/\Delta thyA$ transformed with **(2)** pACYC(WT-DHFR, WT-TYMS) (SMT101), **(3)** pACYC(his₆WT-DHFR, WT-TYMS), **(4)** pTET(WT-DHFR, WT-TYMS) (SMT201), **(5)** pTET("AATGAG" RBS, WT-DHFR, WT-TYMS) (SMT205) see **Table 5.2**, page 174. Molecular weights to the ladder markers (labels left) and molecular identities (labels right) are indicated in label text.

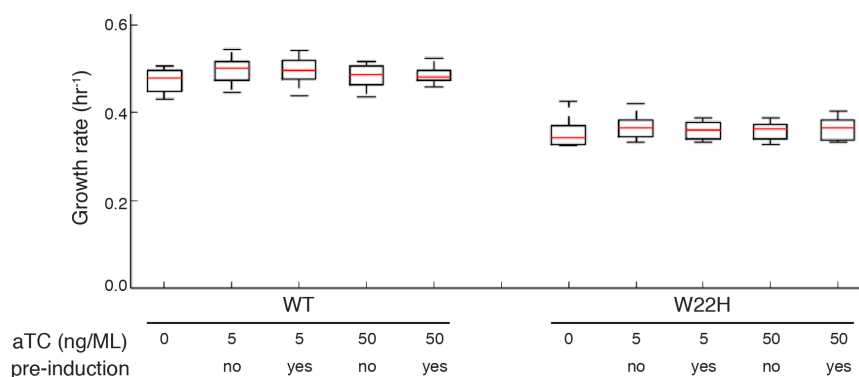


Figure 3.9: Selection strain growth rate with TYMS induction off pTET plasmid. Growth rates from plate reader experiments (see **Methods, Chapter 3.5.0.4**, page 110) are displayed as box plots for 8 independent replicates of plate reader growth experiments using ER2566 $\Delta folA/\Delta thyA$ transformed with SMT201 (**Table 5.2**, page 174) expressing TYMS and either WT or W22H DHFR (**Table 3.1**, page 65). Cultures were induced with 0, 5, or 50 ng/mL aTC and grown with and without 4 hours of pre-induction (labels, bottom). An orange line marks the median. Box edges mark the first and third quartiles. Whiskers mark the maximum and minimum.

3.3.2.3 Testing Selection with a Mutated Ribosome Binding Site for DHFR To further increase selection pressure, we next sought to decrease DHFR expression. In the selection assay, DHFR is not induced. DHFR is produced through expression under glucose inhibition of the lac operon on the selection plasmid. We therefore modified the DHFR ribosome binding site (RBS) from the Shine-Dalgarno consensus sequence (AAGGAG) to "AACGAG" to reduce DHFR expression (see SMT203 in **Table 5.2**, page 174). With this construct, we performed a second preliminary selection experiment. We observed that the selection coefficients with AACGAG as the DHFR RBS (SMT203) correlated poorly both with results from our first preliminary selection (**Figure 3.10**, page 68) and with *in vitro* velocities for characterized mutants (**Figure 3.11**, page 69). We also observed that selection coefficients did not change with the addition of 0.5 $\mu\text{g}/\text{mL}$ thymidine to the selection media. From this, we concluded that selection with the SMT203 plasmid was not reporting on DHFR activity, but we were unable to determine conclusively if the poor correlation to previous experiments was a result of the "AACGAG" RBS sequence, an off-target mutation in the selection plasmid, or an unknown problem specific to this particular experiment.

3.3.2.4 Predicting Low-expression Ribosome Binding Sites for DHFR We hypothesized that we had decreased DHFR expression to a level where expression became stochastic. Therefore, our final optimization of the selection construct was to test for maximum resolution of mutant activities over a small panel of RBS strengths between the extremes of the strong "AAGGAG" (SMT201) sequence and the weak "AACGAG" (SMT203) sequence. We first calculated a predicted translation rate for "AAGGAG", "AACGAG" and 8 additional RBS sequences in the context of the upstream region of the DHFR gene in our selection plasmid (**Table 3.2**, page 70)[27, 28]. From this set, we selected two additional RBS sequences, "AAGGAA" (STM202) and "AATGAG" (SMT205), for experimental characterization of expression level and DHFR

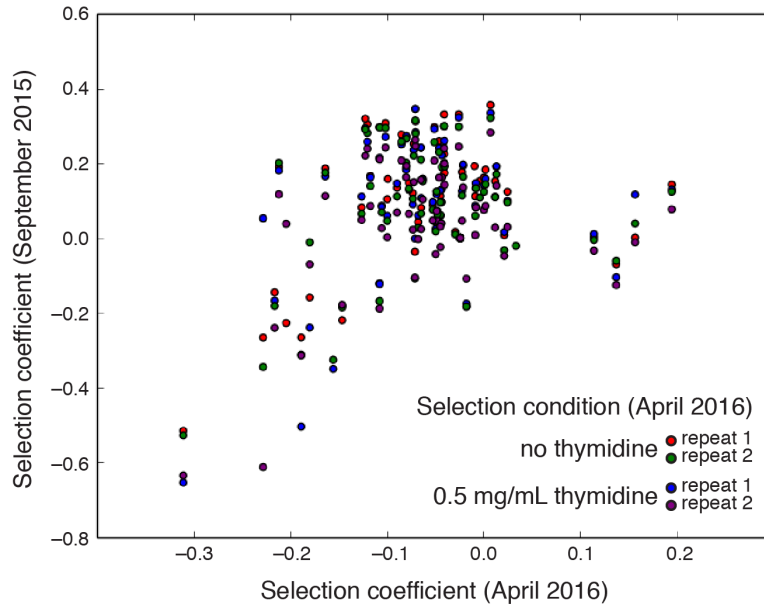


Figure 3.10: Selection using an "AACGAG" mutated RBS sequence for DHFR yields selection coefficients that correlate poorly with preliminary selection. Selection coefficients for selection with ER2566 $\Delta foIA/\Delta thyA$ transformed with SMT203 containing an RBS mutated from "AAGGAG to "AACGAG" with addition of **A**) no or **B**) 0.5 mg/mL thymide are plotted against selection coefficients from preliminary selection (see **Chapter 3.3.1**, page 58). Points on the scatterplot are colored by the conditions in the second round (April 2016) of selection (legend, bottom right).

mutant resolution.

3.3.2.5 Comparing WT DHFR Expression from Selection Plasmid and Endogenous Gene

To determine if the mutated RBS sequences have decreased expression, we performed a western blot with anti-DHFR polyclonal antibodies, see **Methods (Chapter 3.5.0.11, page 117)** We observed that the selection plasmids with the "AAGGAG" (SMT201) and "AAGGAA" (SMT202) RBS sequences appear to have total DHFR expression that is greater than the endogenous expression level. The total expression level with the "AATGAG" (SMT205) and "AACGAG" (SMT203) RBS sequences appear to be much closer to the endogenous expression level. For both these mutated RBS sequences, however, DHFR expression was decreased relative to the expression from the pACYC selection plasmid used in the preliminary selection experiment (see **Chapter 3.3.1, page 58**).

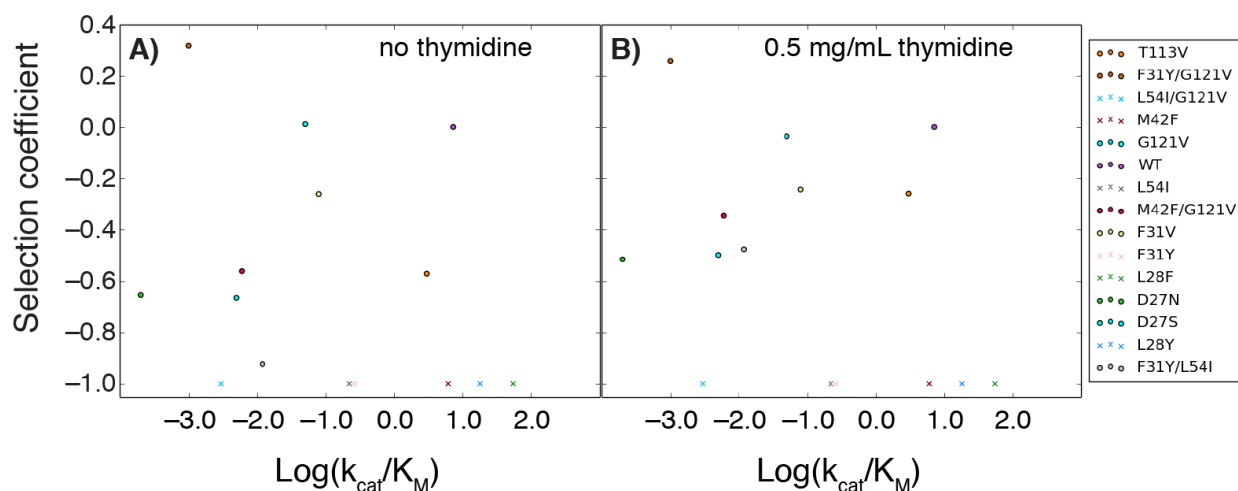


Figure 3.11: Selection using an "AACGAG" mutated RBS sequence for DHFR yields selection coefficients that correlate poorly with *in vitro* kinetics. Selection coefficients for selection with ER2566 $\Delta folA/\Delta thyA$ transformed with SMT203 containing an "AACGAG" mutated RBS sequence with addition of **A)** no or **B)** 0.5 mg/mL thymidine are plotted against $\text{Log}(k_{cat}/K_M)$ from *in vitro* measurements with purified enzyme. Points on the scatterplot are colored by mutant identity (legend, right). Mutants for which a selection coefficient could not be calculated are marked with an X in the plot and are assigned the lowest value.

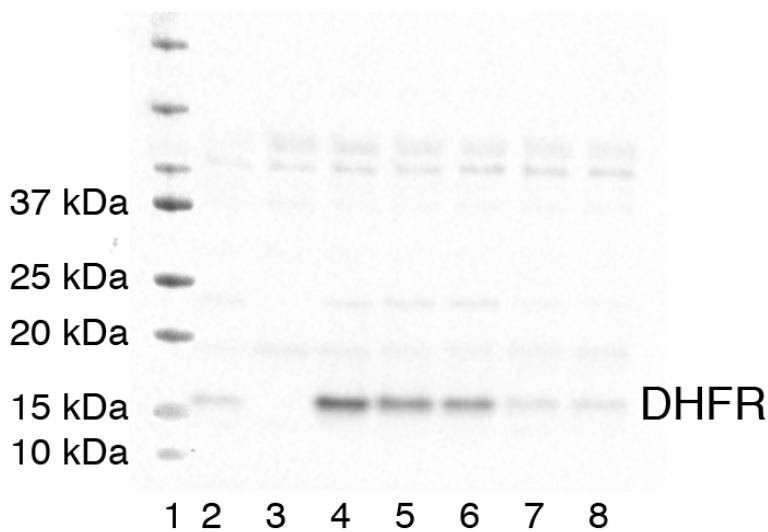


Figure 3.12: Total DHFR expression from the optimized selection plasmid is qualitatively decreased relative to endogenous expression. The expression of DHFR was observed by anti-DHFR western blot for expression from plasmids and the endogenous gene. Lane identities are marked by a number (bottom): **1)** Ladder; **2-3)** lysates from ER2566 (**2)**, or ER2566 $\Delta folA/\Delta thyA$ (**3**); **4-8)** lysates from ER2566 $\Delta folA/\Delta thyA$ transformed with (**4**) SMT101 (pACYC, AAGGAG), (**5**) SMT201 (pTET, AAGGAG), (**6**) SMT202 (pTET, AAGGAA), (**7**) SMT205 (pTET, AATGAG), (**8**) SMT203 (pTET, AACGAG) see **Table 5.2**, page 174 and **Table 3.2**, page 70. Molecular weights to the ladder markers (labels left) and molecular identities (labels right) are indicated in label text.

Table 3.2: Predicted strength for DHFR ribosome binding sites and construct names from **Table 5.2** (page 174). Previously tested RBS (**Chapter 3.3.2.3**, page 67) sequences are in bold. RBS sequence selected for further characterization are in bold italics.

RBS sequence	Predicted translation rate	Construct
AAGGAG	212828.20	SMT101, SMT201
AAGGAA	66047.10	SMT202
AAAGAG	9537.00	
<i>AATGAG</i>	9117.30	SMT205
AAGTAG	4055.60	
AACGAG	3095.80	SMT203
AAAGAG	2959.60	
AAGTAA	2704.90	
AAGCAG	2064.80	
AAGCAA	1377.10	

Table 3.3: Soluble expression of DHFR measured from lysate activity for heterologous expression from plasmid and for endogenous expression as described and plotted in **Figure 3.13**, page 71. Standard deviations are from three independent experiments for velocity and three biological replicates for lysate activity.

Variant	Molecules per cell	Standard deviation
WT (chromosomal)	550.0	9.2
WT (SMT102)	240.0	33.5
WT (SMT201)	330.5	17.2
WT (SMT202)	262.0	5.5
WT (SMT205)	52.0	9.2

3.3.2.6 Lysate Activity to Quantify Expression for Endogenous and Plasmid-borne WT DHFR

Because the western blots for DHFR expression in the previous section (**Chapter 3.3.2.5**, page 68) were not quantitative, we next sought to quantitatively measure the soluble expression of DHFR using a lysate activity assay adapted from previous studies [29, 30]. Comparing the DHFR activity in lysates to *in vitro* velocities from purified WT DHFR (see **Methods, Chapter 3.5.0.14**, page 122), we were able to calculate DHFR expression. From these results, we observed that our measurement for endogenous expression of DHFR is consistent with some previous measurements of ~400-500 molecules/cell[31], but our measurements were higher than other previous measurements for endogenous DHFR expression in MG1566 *E. coli*[29] (**Figure 3.13**, page 71, **Table 3.3**, page 70). For the selection plasmids, soluble expression was

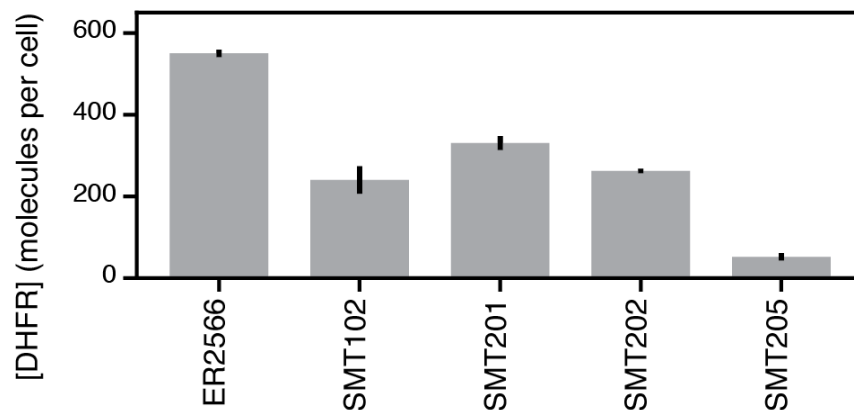


Figure 3.13: Soluble DHFR expression from the optimized selection plasmid is decreased relative to endogenous expression. ER2566 is the parental strain. SMT102, SMT201, SMT202, SMT205 denote plasmid constructs with altered promoters and ribosome binding sites (see **Table 5.2**, page 174). DHFR expression for plasmids was measured from DHFR activity in lysates of the selection strain, ER2566 $\Delta folA/\Delta thyA$. Error bars represent the cumulative percent error (standard deviation) from three independent experiments for velocity and three biological replicates for lysate activity.

approximately half that of the endogenous expression level except with SMT205. This measurement for soluble DHFR expression was lower than expected based on the total apparent expression from the selection plasmids as observed on the anti-DHFR western blot (**Figure 3.12**, page 69). Nevertheless, expression decreased as expected with one mutation to the RBS. The SMT205 plasmid with an "AATGAG" RBS sequence resulted in ~10% of the endogenous expression level (52.0 ± 9.2 molecules/cell).

3.3.2.7 Plate-reader Experiment to Measure the Resolution of Selection To examine how the optimization of the selection plasmid impacted the ability of selection to resolve highly active DHFR point mutants, we generated 11 point mutations to the DHFR sequence in plasmid constructs with three RBS sequences (see **Methods, Chapter 3.5.0.2**, page 110). We then measured growth rates for our selection strain transformed with this plasmid panel (see **Methods, Chapter 3.5.0.4**, page 110). We observed a linear relationship with k_{cat}/K_M (**Table 3.1**, page 65) only to growth rates measured in transformants of the SMT205 plasmid with an "AATGAG" RBS (**Figure 3.14(A-C)**, page 73, **Table 3.4**, page 72). From this experiment

Table 3.4: Growth rates for the ER2566 $\Delta folA/\Delta thyA$ selection strain with a panel of 12 DHFR mutants and 3 DHFR RBS sequences, see **Methods (Chapter 3.5.0.4, page 110)**. SMT201, SMT202, SMT205 denote pTET plasmid constructs with altered ribosome binding sites (see **Table 5.2, page 174**). ND indicates that growth was not observed. NM indicates that the measurement was not taken. Standard deviations are measured over 8 independent experiments.

Variant	SMT201		SMT202		SMT205	
	Growth rate (hr^{-1})	Standard deviation	Growth rate (hr^{-1})	Standard deviation	Growth rate (hr^{-1})	Standard deviation
WT	0.45	0.02	0.42	0.02	0.30	0.02
W22H	0.40	0.01	0.17	0.01	0.16	0.01
L28F	0.47	0.3	0.40	0.01	0.31	0.01
L28Y	0.48	0.02	0.37	0.01	0.24	0.02
F31V	0.39	0.01	0.27	0.01	0.21	0.01
F31Y	0.45	0.01	0.36	0.01	0.16	0.01
M42F	0.48	0.01	0.41	0.01	0.38	0.01
L54F	NM	NM	0.41	0.01	0.29	0.01
L54I	0.43	0.02	0.26	0.01	0.19	0.01
T113V	0.44	0.02	0.38	0.01	0.35	0.02
G121V	0.34	0.01	0.14	0.03	0.12	0.02
F31Y/L54I	NM	NM	0.14	0.01	ND	ND

forward, we used the SMT205 plasmid as the optimized selection plasmid.

Because the relationship between growth rate and k_{cat}/K_M was linear but poorly resolved for mutants that caused faster growth rates than did WT, we examined other kinetic properties that might better describe the observed growth phenotypes. We anticipated that some of the mutants will be saturated by the DHF substrate at cellular concentrations of low 10s of μM [32], so we plotted the growth rates against DHFR reaction velocities calculated from k_{cat} and K_M at 20 μM DHF. In the plots of velocity calculated for 20 μM DHF versus growth rate (**Figure 3.14(D-F)**, page 73), we observed a more consistent correlation for mutants with near-WT level velocities for the SMT205 plasmid compared to results with the other two plasmids. This demonstrated that our selection had been optimized to report linearly on a broad range of DHFR activities, including better than WT-level activities.

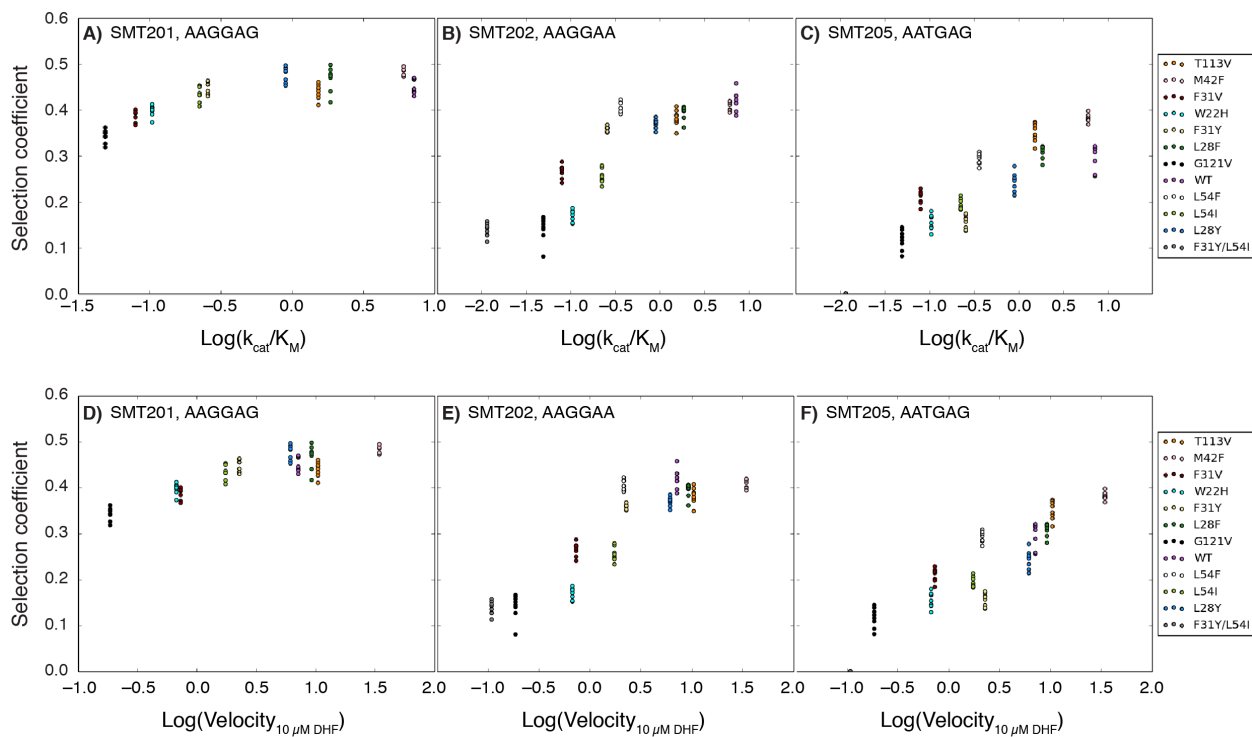


Figure 3.14: Correlation between *in vitro* kinetics and growth rates for DHFR point mutants shows increasing selection pressure. Growth rates for ER2566 $\Delta folA/\Delta thyA$ transformed with one of a panel of point mutants in plasmids with predicted RBS strength decreasing from left to right: SMT201, AAGGAG (**A,D**); SMT202, AAGGAA (**B,E**); SMT205, AATGAG (**C,F**), see **Table 5.2**, page 174 and **Table 3.2**, page 70. Plot points are colored by mutant identity (legend, right). Growth rates from 8 independent experiments are plotted against $\text{Log}_{10}(k_{cat}/K_M)$ (**A-C**) or $\text{Log}_{10}(\text{Velocity}_{10 \mu\text{M DHF}})$ (**D-F**).

3.3.3 Developing Non-selective Growth Conditions

With optimized and more stringent selection conditions established, we next turned to optimizing non-selective conditions for post-transformation rescue and out-growth steps. To prevent large fractions of the library population from dropping out before the first timepoint on the turbidostat, we sought to identify supplements to M9 minimal medium that would keep the growth rates of highly-fit and unfit mutants within a factor of 2. With this criterion and a WT growth rate in the range of 0.3-0.5 hr⁻¹, the relative frequency of alleles would not change by more than 10- to 20-fold between transformation and the first timepoint on the turbidostat. Additionally, because post-transformation rescue in SOB medium is necessary for sufficient transformation efficiency, we aimed for non-selective out-growth conditions in M9 medium to allow the selection strain to adapt to glucose as a carbon source. We first attempted to use a previously reported supplement mix ("*folA* mix") of 100 μg/mL (1 mM) adenine, 250 μg/mL (1 mM) thymidine, 1 μg/mL (2 μM) pantothenate, 38 μg/mL (500 μM) glycine, and μg/mL (250 μM) methionine[33]. From plate reader growth experiments with these supplements, we observed that the unfit F31Y/L54I mutant (see **Table 3.1**, page 65) had a measureable growth rate that was within 2-fold of that for WT DHFR (**Figure 3.15**, page 75), but the selection strain displayed a significant lag phase both for moderately active DHFR mutants (**Figure 3.16**, page 75) and for a monoculture of WT DHFR in the turbidostat (**Figure 3.17**, page 76). From these results, we concluded that the individual components of this published supplement should be optimized individually.

We examined the supplement components and their concentration, and hypothesized that the high concentration of nucleotides could cause toxicity by perturbing the nucleotide pools[?]. We therefore decreased the nucleotide concentrations, resulting in a supplement mix of 20 μg/mL (75 μM) adenosine, 50 μg/mL (200 μM) thymidine, 1 μg/mL (2 μM) pantothenate, 38

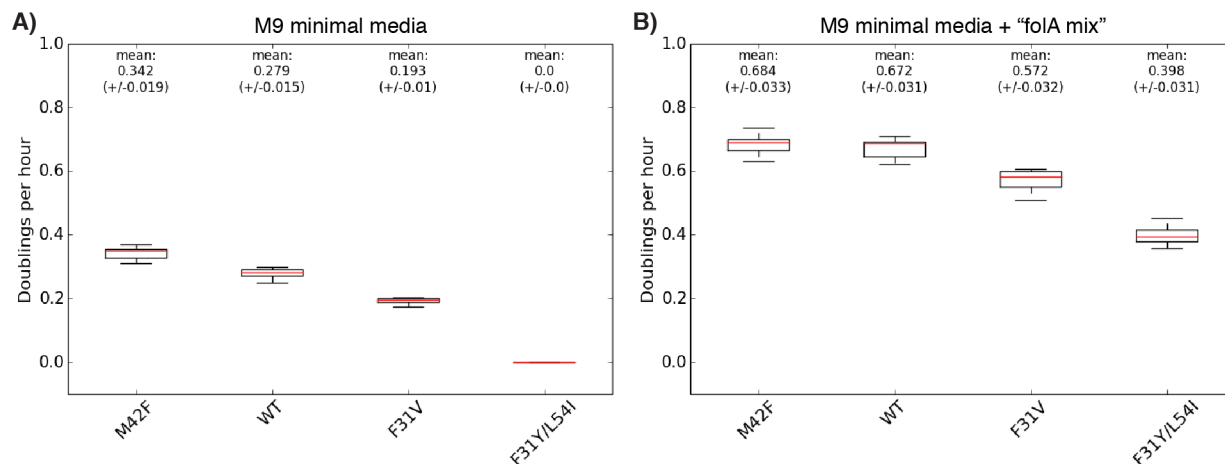


Figure 3.15: Selection strain growth rates in M9 minimal medium without and with the previously published "foIA" supplement mix are shown as box plots. ER2566 $\Delta foIA/\Delta thyA$ is complemented with one of 4 DHFR variants (labels bottom, **Table 3.1**, page 65) on the SMT205 optimized selection plasmid (**Table 5.2**, page 174) and grown in M9 minimal medium. For each DHFR variant, box plots represent 8 individual experiments on the plate reader (see **Methods, Chapter 3.5.0.4**, page 110). An orange line marks the median. Box edges mark the first and third quartiles. Whiskers mark the maximum and minimum.

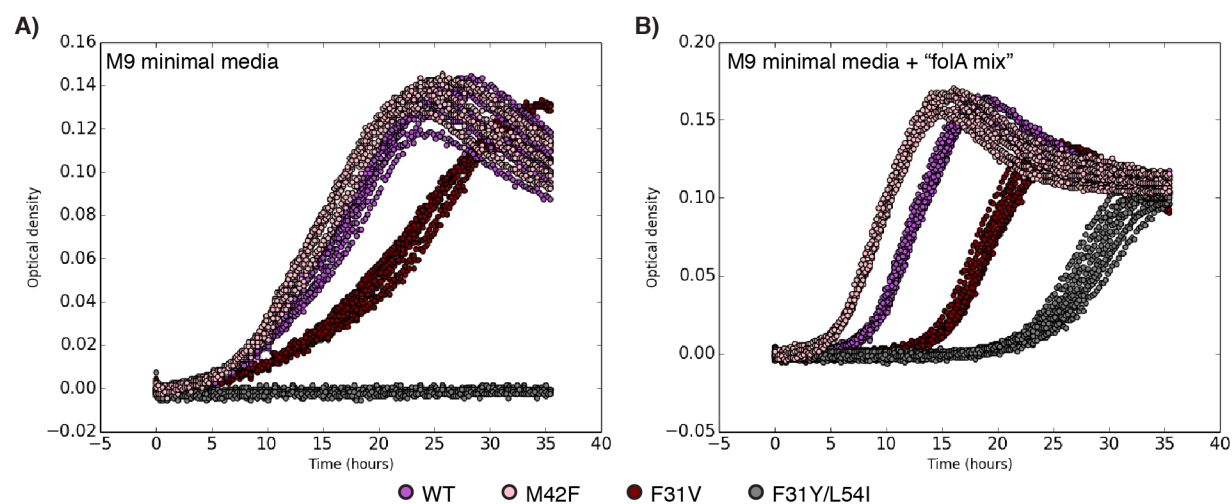


Figure 3.16: Selection strain growth curves in M9 minimal medium without and with the previously published "foIA" supplement mix. ER2566 $\Delta foIA/\Delta thyA$ is complemented with one of 4 DHFR variants (labels bottom, **Table 3.1**, page 65) on the SMT205 optimized selection plasmid (**Table 5.2**, page 174) and grown in M9 minimal medium. For each DHFR variant, growth curves are shown for 8 individual experiments on the plate reader (see **Methods, Chapter 3.5.0.4**, page 110).

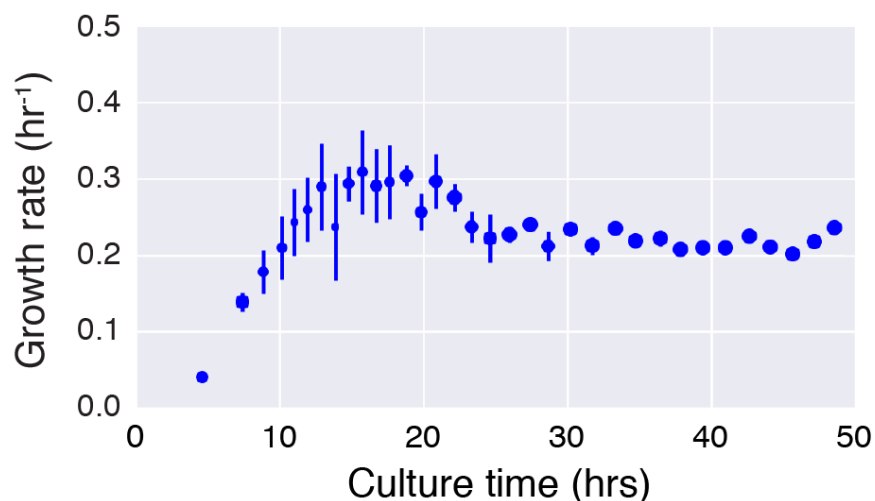


Figure 3.17: Growth rates in the turbidostat after overnight growth in M9 minimal medium supplemented with the "folA mix". ER2566 $\Delta folA/\Delta thyA$ was transformed with WT DHFR in the SMT205 selection plasmid (Table 5.2, page 174) and grown in M9 minimal medium under turbidostat clamp after overnight growth and 4 hours out-growth in M9 media with 100 $\mu\text{g}/\text{mL}$ adenine, 250 $\mu\text{g}/\text{mL}$ thymidine, 1 $\mu\text{g}/\text{mL}$ pantothenate, 38 $\mu\text{g}/\text{mL}$ glycine, and 38 $\mu\text{g}/\text{mL}$ methionine. Points represent the growth rate calculated as the linear regression slope of $\text{Log}_2(\text{ABS600})$. Error bars represent the standard error from linear regression, and the size of the marker is scale by the Pearson R^2 value.

$\mu\text{g}/\text{mL}$ (500 μM) glycine, and 38 $\mu\text{g}/\text{mL}$ (250 μM) methionine. In plate reader growth experiments (see **Methods, Chapter 3.5.0.4**, page 110) with this supplemented M9, we observed a measurable growth rate for the F31Y/L54I mutant that was not observed in the absence of thymidine or adenosine (**Figure 3.18**, page 77). Furthermore, the growth rate with F31Y/L54I was reproducibly 60% of the growth rate with WT when using the fully supplemented M9 medium. The doubling period with WT was 1.7 hours, and the doubling period with F31Y/L54I was 2.5 hours (**Figure 3.19**, page 77). During an estimated 16-hours of rescue and overnight out-growth, the selection strain complemented with WT DHFR would go through 9.4 generations versus 6.4 generations with F31Y/L54I. The difference in 3 generations would result in an 8-fold change in relative frequency in the library. Finally, a monoculture of the selection strain in the turbidostat after out-growth in supplemented M9 medium did not show an initial lag phase (**Figure 3.20**, page 78). These results fulfilled our criteria for non-selective conditions.

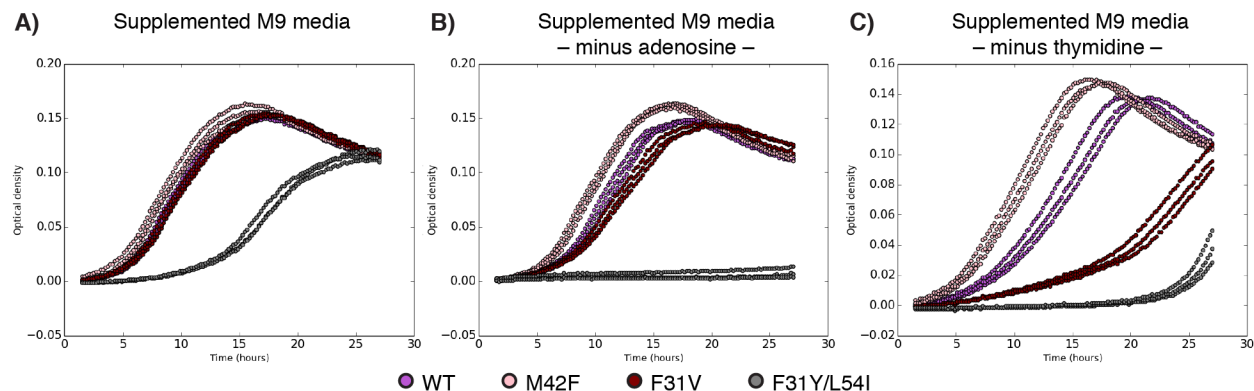


Figure 3.18: Selection strain growth curves in supplemented M9 minimal medium without and with adenosine and thymidine drop-outs to the medium. ER2566 $\Delta foIA/\Delta thyA$ is complemented with one of 4 DHFR variants (labels bottom, **Table 3.1**, page 65) on the SMT205 optimized selection plasmid (**Table 5.2**, page 174) and grown in M9 minimal medium with **A)** 20 $\mu\text{g}/\text{mL}$ adenosine, 50 $\mu\text{g}/\text{mL}$ thymidine, 1 $\mu\text{g}/\text{mL}$ pantothenate, 38 $\mu\text{g}/\text{mL}$ glycine, and 38 $\mu\text{g}/\text{mL}$ methionine; **B)** the supplements from A) minus adenosine; or **C)** the supplements from A) minus thymidine. For each DHFR variant, growth curves are shown for 3 individual experiments on the plate reader (see **Methods, Chapter 3.5.0.4**, page 110).

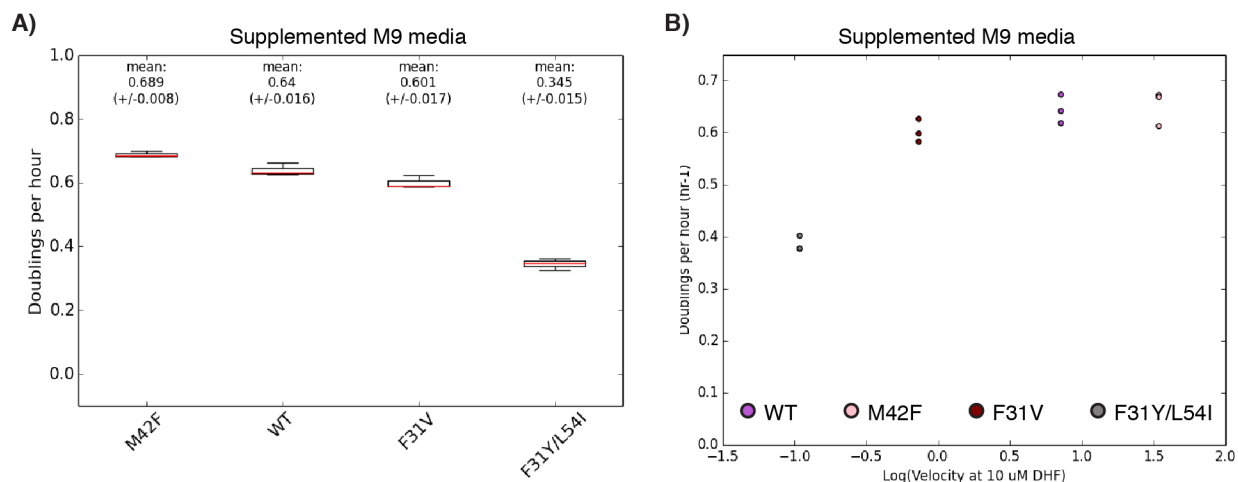


Figure 3.19: Selection strain growth rates in M9 minimal medium. **A)** Growth rates for ER2566 $\Delta foIA/\Delta thyA$ complemented with one of 4 DHFR variants (labels bottom, **Table 3.1**, page 65) on the SMT205 optimized selection plasmid (**Table 5.2**, page 174) and grown in M9 minimal medium with 20 $\mu\text{g}/\text{mL}$ adenosine, 50 $\mu\text{g}/\text{mL}$ thymidine, 1 $\mu\text{g}/\text{mL}$ pantothenate, 38 $\mu\text{g}/\text{mL}$ glycine, and 38 $\mu\text{g}/\text{mL}$ methionine. For each DHFR variant, box plots represent 3 individual experiments on the plate reader (see **Methods, Chapter 3.5.0.4**, page 110). An orange line marks the median. Box edges mark the first and third quartiles. Whiskers mark the maximum and minimum. **B)** Correlation between growth rates from A) and *in vitro* enzyme velocity at 10 μM DHF calculated from Michaelis-Menten kinetics (**Table 3.1**, page 65). Plot points are colored by mutant identity (legend, right).

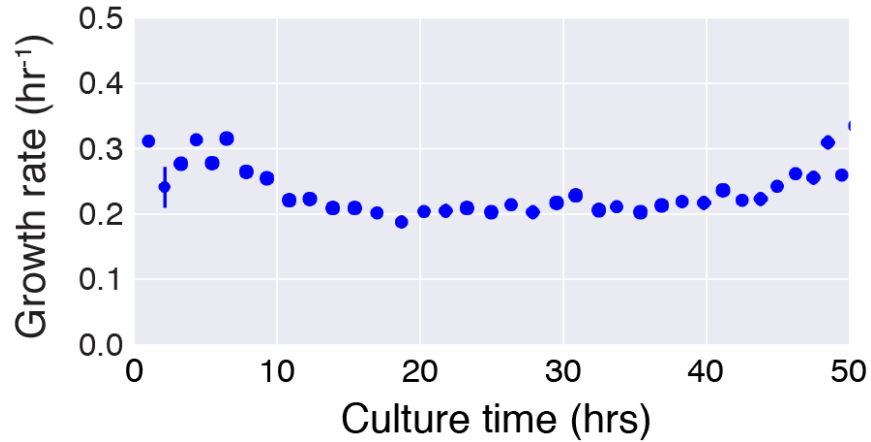


Figure 3.20: Growth rates in the turbidostat after overnight growth in supplemented M9 minimal medium. ER2566 $\Delta folA/\Delta thyA$ was transformed with WT DHFR in the SMT205 selection plasmid (Table 5.2, page 174) and grown in M9 minimal medium under turbidostat clamp after overnight growth and 4 hours out-growth in M9 media with 20 $\mu\text{g}/\text{mL}$ adenosine, 50 $\mu\text{g}/\text{mL}$ thymidine, 1 $\mu\text{g}/\text{mL}$ pantothenate, 38 $\mu\text{g}/\text{mL}$ glycine, and 38 $\mu\text{g}/\text{mL}$ methionine. Points represent the growth rate calculated as the linear regression slope of $\text{Log}_2(ABS600)$. Error bars represent the standard error from linear regression, and the size of the marker is scale by the Pearson R^2 value.

3.3.4 Building a DMS Library of DHFR Single Point Mutants

We next constructed a library of all single point mutants to DHFR in our optimized SMT205 selection plasmid.

3.3.4.1 Strategy for Screening DHFR Single Point Mutants We planned the construction of our library around the technical limitations of the deep sequencing process. The DHFR gene is 477 bp, which is longer than the standard 300 bp amplicon for Illumina sequencing platforms with capacity for extended amplicon read length such as the MiSeq and the NextSeq. We considered two options to address this, 1) barcoding the library and 2) building the library as a collection of smaller sublibraries. The barcoding strategy has notable drawbacks [correspondence, Michael Schnebly and Bill Russ, University of Texas, Southwestern]. Strategies for barcoding after the library has been generated can result in additional technical noise arising from many barcodes assigned to the same allele. This effectively raises the frequency threshold that must be applied

to consider a mutant present at any time point. Using library bottlenecking methods is only partially effective because low-frequency alleles will be filtered out of the library entirely while high-frequency alleles are still associated with many redundant barcodes. We determined that until libraries can be synthesized with explicit barcode-allele matching, library barcoding methods are most useful when the aim of selection is sub-sampling a much larger library sequence space. We therefore chose to pursue direct sequencing of individual sublibraries.

We chose to split the library into 4 sublibraries of mutations to no more than 40 positions each: Any 2 of these sublibraries can be pooled, screened, and sequenced together using a 300 bp amplicon. Each sublibrary was generated using the robust method of performing inverse PCR for each position in DHFR. For each position and reaction, a unique set of primers was designed and synthesized that replaced the WT codon with an NNS codon (see **Methods, Chapter 3.5.0.2**, page 108). All reactions were examined with agarose gel electrophoresis. If no product band was visible, the PCR reaction for that position was optimized until a visible product band appeared. Finally, reactions were pooled stoichiometrically into sublibraries by position: positions 2-40 in sublibrary (SL)1, positions 41-80 in SL2, positions 81-120 in SL3, and positions 121-159 in SL4. This yielded a version of the library that could then be subjected to quality control.

3.3.4.2 Library Quality Control To check the completeness of the library and the number of off-target mutations, we deep sequenced each of the 4 sublibraries (see **Methods, Chapter 3.5.0.8**, page 114, **Chapter 3.5.0.9**, page 114). Amplicons were generated for each library separately, but each amplicon covered positions in two sublibraries: SL1-SL2 or SL3-SL4. For each sublibrary, the error could be estimated from the region where no mutations are expected. We examined the background for each library as a function of the frequency-based cut-off for considering a mutation present in the library. We observed that approximately 94% of the library

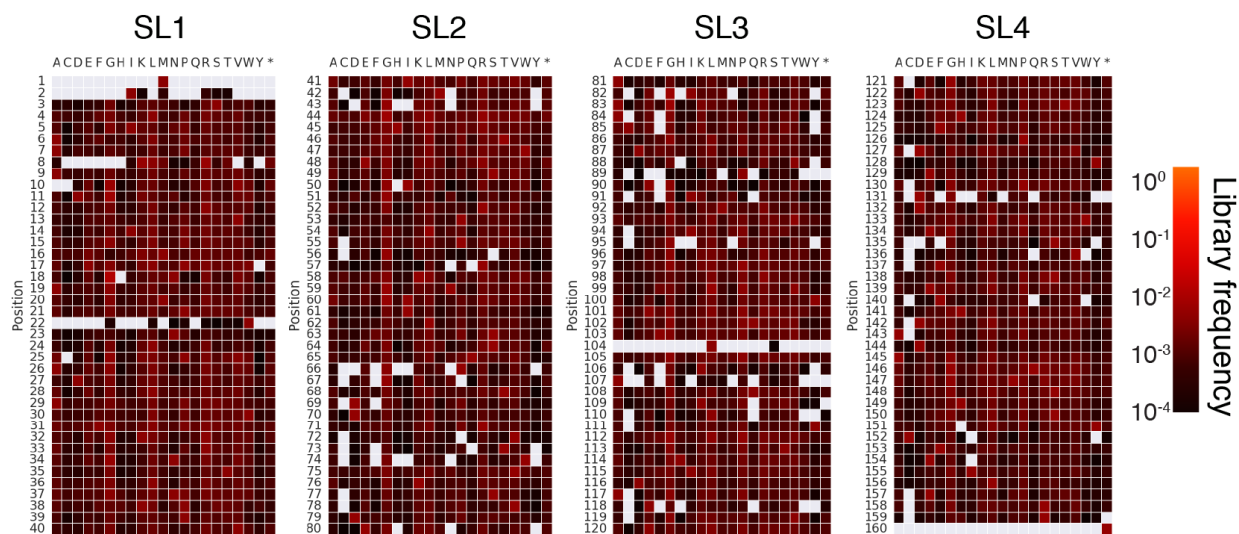


Figure 3.21: Preliminary mutant frequencies in DHFR point mutant sublibraries. Mutant frequencies from deep sequencing counts for 4 sublibraries covering all single point mutants are displayed as heatmaps with a cut off of 1×10^{-4} . Rows are labeled by the position in DHFR. Columns are labeled by the amino acid identity. The matrix is split into 4 groups by sublibrary. The matrix is colored by allele frequency according to the heatmap (right).

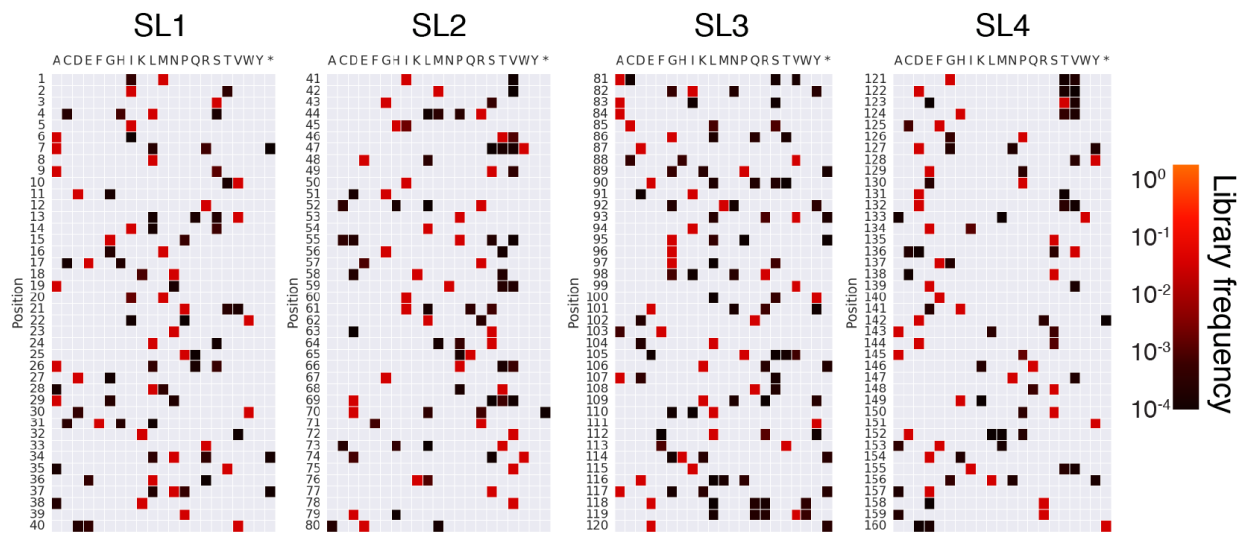


Figure 3.22: Preliminary background frequencies in DHFR point mutant sublibraries. Background frequencies from deep sequencing counts for 4 sublibraries covering all single point mutants are displayed as heatmaps with a cut off of 1×10^{-4} . Background is assigned based on location the same amplicon reading frame. SL1 is background for SL2, SL2 for SL1, SL3 for SL4, and SL4 for SL3. Rows are labeled by the position in DHFR. Columns are labeled by the amino acid identity. The matrix is split into 4 groups by sublibrary. The matrix is colored by allele frequency according to the heatmap (right).

was covered, with missing mutations generally located at common positions (**Figure 3.21**, page 80), which is consistent with a failure at the PCR step. In contrast, the background was randomly distributed throughout the library (**Figure 3.22**, page 80), which is consistent with noise in the sequencing and mutations introduced during PCR. In the deep sequencing, we observed that mutations appeared at higher frequency in their sublibrary than they did as background mutations in another sublibrary (**Figure 3.23**, page 82). Finally, we observed that quality control from two different sequencing runs resulted in almost no overlap in background mutations. From these data, we determined that the overall quality of the library is sufficient to confirm the presence of >90% of all possible single point mutations, that the background mutations are the result of technical noise, and that deficiencies in the library could be addressed by supplementing in DNA from optimized reactions for positions that were de-enriched in the library.

We then supplemented de-enriched positions by optimizing PCR reactions for positions 2, 8, 22, 42, 43, 66, 67, 73, 74, 80, 82, 84, 89, 95, 104, 107, 118, 131, 135, 136, 140 and spiking each reaction into its respective sublibrary proportional to the total number of positions represented in the sublibrary. The supplemented library was deep sequenced and analyzed as described above. With the supplemented library, we observed more complete coverage over the library (**Figure 3.25**, page 83) and similar background as compared to the first round library (**Figure 3.26**, page 84), which was consistent with the expectation that background noise was due primarily to technical noise from the sequencer and from the amplicon generation process. We then calculated statistics for coverage (% mutations observed in library) and error (% mutations observed in background) in the library based on the deep sequencing (**Table 3.5**, page 85).

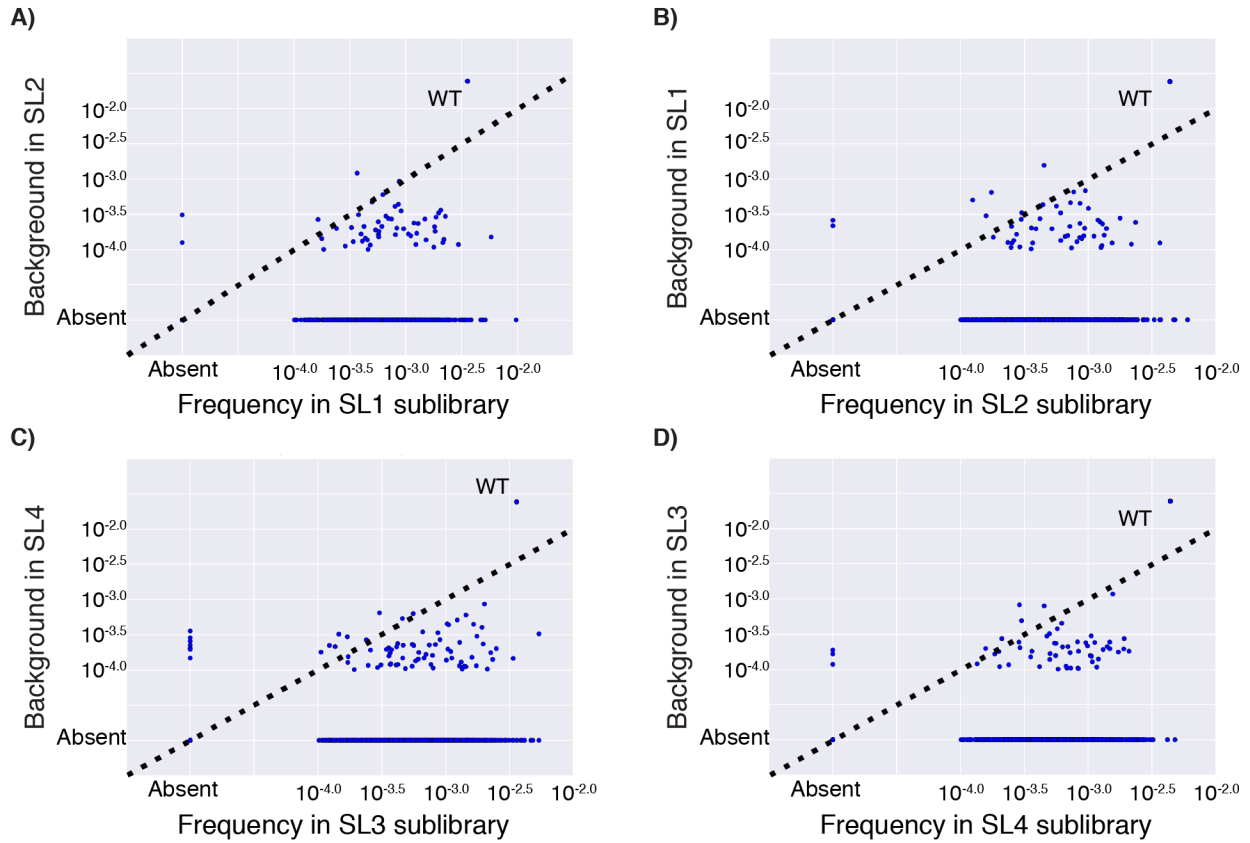


Figure 3.23: Comparison of mutant frequencies in sublibraries and as background. Mutant frequencies in their respective sublibraries from **Figure 3.21** (page 80) versus mutant frequencies as background from Figure 3.22 (page 80) are plotted as scatter plots for **A)** SL1, **B)** SL2, **C)** SL3, **D)** SL4. Each point represents a single mutant. The point representing WT is labeled on each plot. The dotted lines represents $y = x$ for each plot.

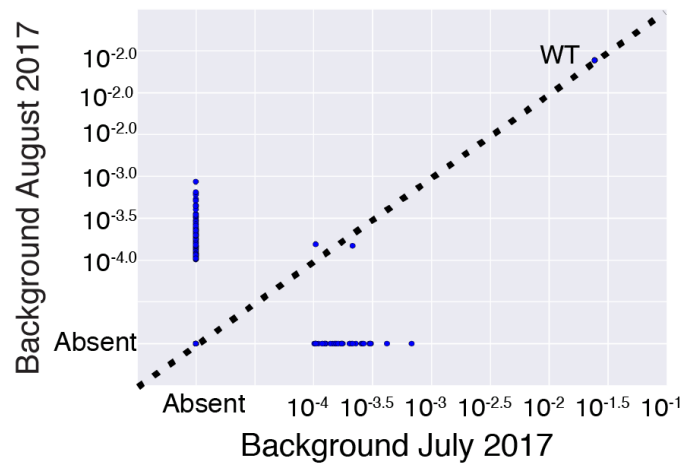


Figure 3.24: Comparison of background mutant frequencies in two independent sequencing experiments. Mutant background frequencies from **Figure 3.22** (page 80) are plotted as a scatter plot against mutant background frequencies from an independent deep sequencing experiment. Each point represents a single mutant. The point representing WT is labeled on each plot. The dotted lines represents $y = x$ for each plot.

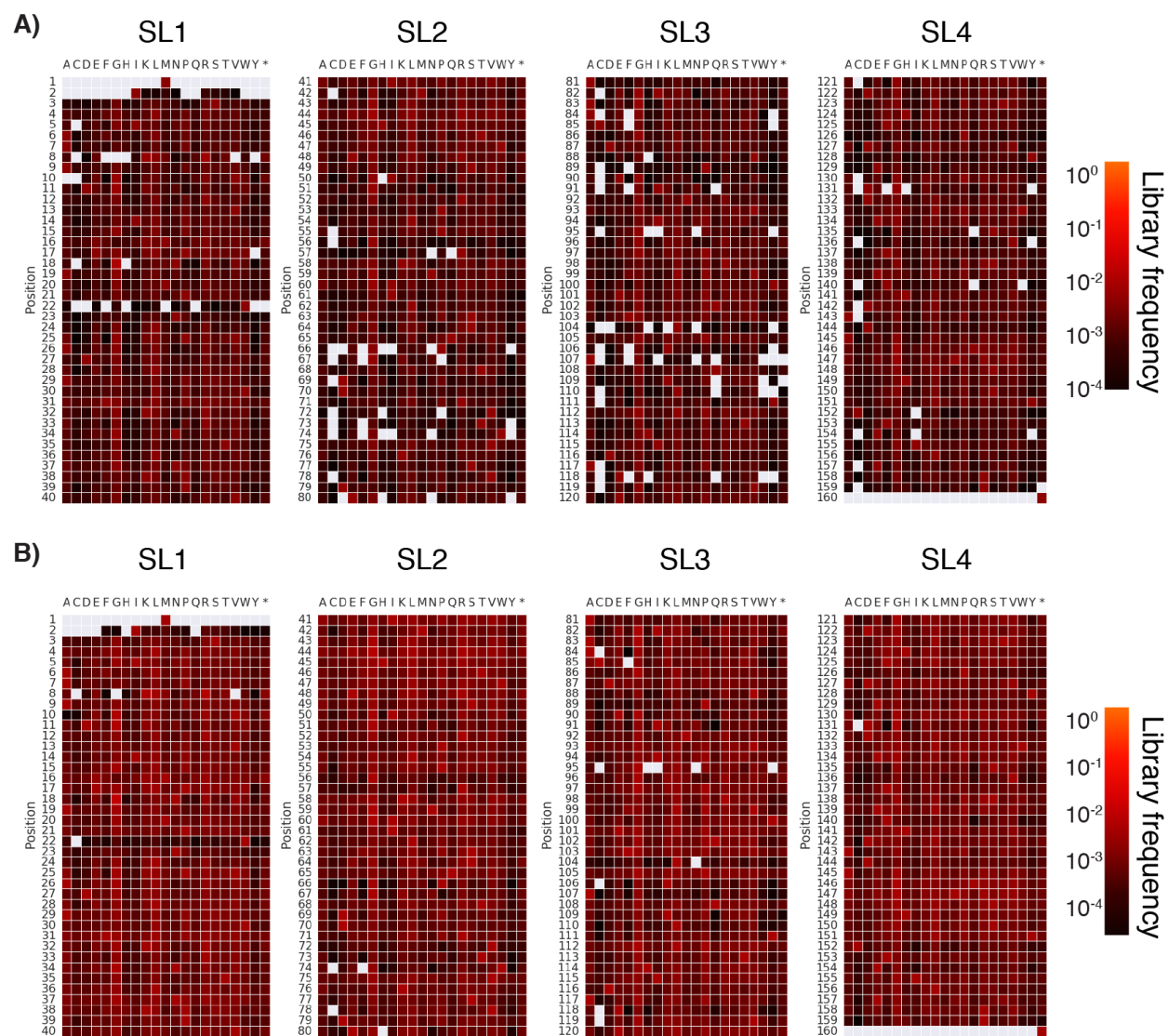


Figure 3.25: Final mutant frequencies in DHFR point mutant sublibraries. Mutant frequencies from deep sequencing counts for 4 sublibraries covering all single point mutants are displayed as heatmaps with a cut off of **A)** 1×10^{-4} or **B)** 1×10^{-5} . Rows are labeled by the position in DHFR. Columns are labeled by the amino acid identity. The matrix is split into 4 groups by sublibrary. The matrix is colored by allele frequency according to the heatmap (right).

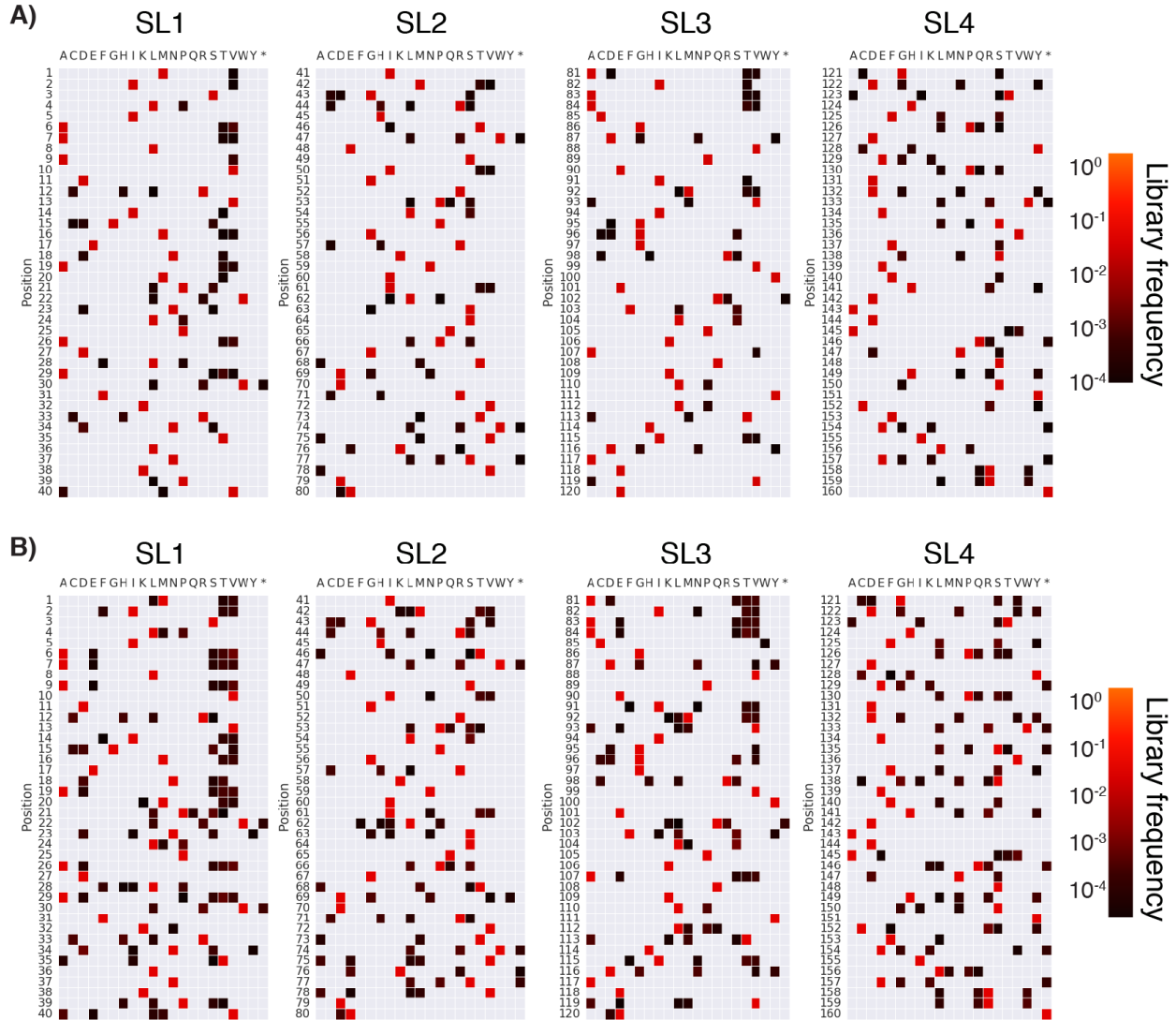


Figure 3.26: Final background frequencies in DHFR point mutant sublibraries. Background frequencies from deep sequencing counts for 4 sublibraries covering all single point mutants are displayed as heatmaps with a cut off of **A)** 1×10^{-4} or **B)** $1 \times 10^{-4.5}$. Background is assigned based on location the same amplicon reading frame. SL1 is background for SL2, SL2 for SL1, SL3 for SL4, and SL4 for SL3. Rows are labeled by the position in DHFR. Columns are labeled by the amino acid identity. The matrix is split into 4 groups by sublibrary. The matrix is colored by allele frequency according to the heatmap (right).

Table 3.5: Single point mutant library coverage and background.

Sublibrary	Frequency $\geq 1 \times 10^{-4.0}$		Frequency $\geq 1 \times 10^{-4.5}$	
	Coverage (%)	Error (%)	Coverage (%)	Error (%)
1	95.7	6.1	98.6	9.7
2	95.5	6.4	99.4	9.2
3	93.4	5.1	98.5	10.1
4	97.2	7.8	99.8	12.1

3.3.5 Construction of an In-house Turbidostat

We next aimed to build our own turbidostat based off designs and a microcontroller gifted from Victor Salinas in Rama Ranganathan's lab (University of Chicago, formerly UTSW). This turbidostat design uses sterile air flow both to oxygenate and mix the culture and to remove waste medium as fresh medium is added (**Figure 3.2**, page 59). Medium is provided by gravity flow from a 2-5 L source bottle that is clamped by a solenoid valve. The solenoid valve is triggered by a microcontroller that measures the optical density of the growth culture using the voltage across an IR emitter/receiver pair housed in a plastic ring. The turbidostat is housed in an incubator for temperature control.

We encountered two issues when building this turbidostat. First, we observed that the signal from the IR emitter/receiver pair is highly sensitive to the position of the growth chamber. To address this sensitivity, we developed a system to minimize direct contact with the turbidostat while in operation. A laser-cut acrylic sheet secures the sensor ring in position in the incubator, and a sampling loop is used to collect samples without opening the incubator during operation. Second, the positive pressure from the air-pump was greater than the hydrostatic pressure from the medium source. To address the loss of flow, we installed a peristaltic pump to provide a constant flow of fresh medium when the microcontroller triggered the pump. This also allowed us to optimize the duration of pump activation to produce dilutions from which growth rates

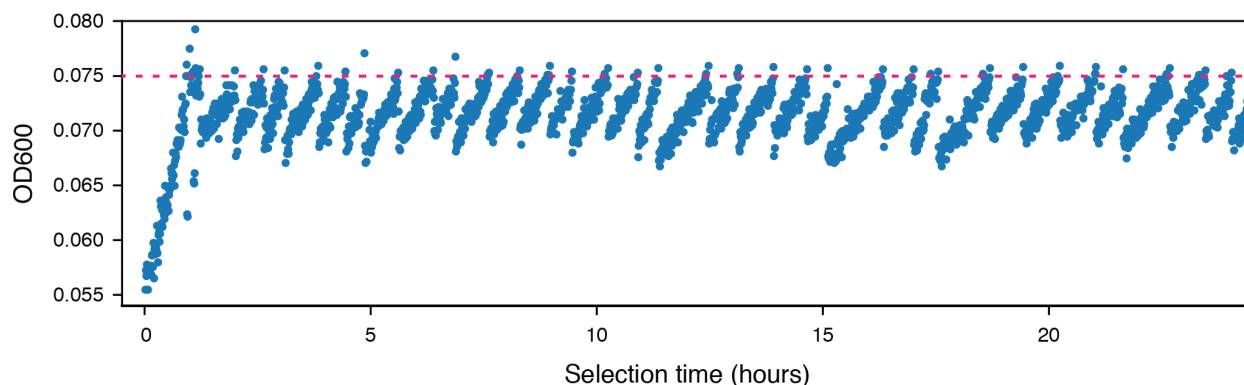


Figure 3.27: Example OD measurements from the in-house turbidostat during a selection experiment on a library of DHFR single point mutants. The OD600 value inferred from the voltage across an IR emitter-receiver pair is plotted as a function of time. The “clamp” OD value (0.075) is shown as a dashed red line. Decreases in OD correspond to dilution from automatic addition of M9 medium.

could be reliably fit (**Figure 3.27**, page 86). The final turbidostat was mechanically different from the original design, but adequately functional (**Figure 3.28**, page 87). Taken together with the optimization of selection and construction of the library, we concluded that we were ready to perform DMS on DHFR.

3.3.6 DMS on DHFR Under Optimized Conditions

We next performed deep mutational scanning using the calibrated selection conditions to determine growth effects in biological triplicate for a library of all possible DHFR single point mutants.

3.3.6.1 Selection on all DHFR Single Point Mutants To maintain the culture in early Log phase growth, we performed selections with a turbidostat at a clamp OD of 0.075. To quantify the effects of DHFR mutations on growth, we calculated selection coefficients^[35] from the change in allele frequency over time by deep sequencing of timepoint samples. Under these controlled selection conditions (see **Methods, Chapter 3.5.0.6**, page 111) the selection coefficients for these point mutants correlate linearly with growth rates measured in a plate reader as is expected

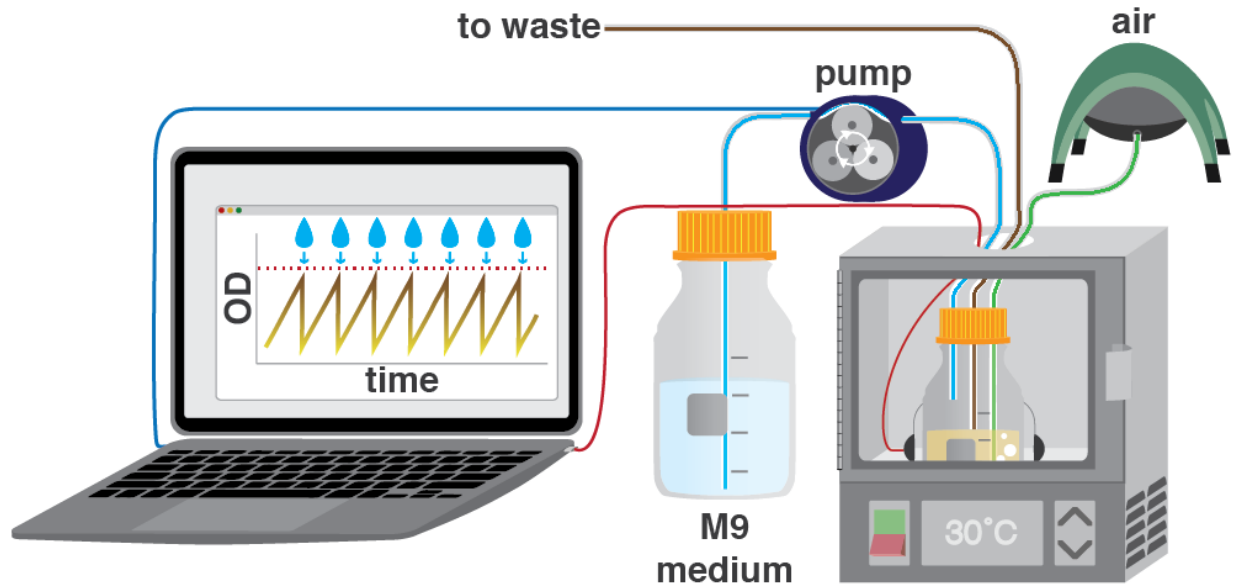


Figure 3.28: Schematic for the in-house turbidostat. The turbidostat is comprised of a 250 mL pyrex bottle as a growth chamber that is housed in an incubator. The voltage IR emitter/receiver pair measures the optical density of the culture. This value is then used to calculate the OD600 equivalent based on a pre-experiment calibration with samples of known OD600. When the OD600 value crosses a user-set threshold, the controlling software triggers a peristaltic pump to add fresh M9 medium. The culture is mixed and oxygenated by input sterile air, which also provides positive pressure for the removal of excess medium.

because the selection coefficient mathematically relates to the cell doubling rate by a scalar factor (Figure 3.29A, page 88, Table 3.4, page 72). Furthermore, we observed a linear relationship between selection coefficient and *in vitro* activity at cytosolic substrate concentrations[36] for a panel of 14 DHFR mutants (Figure 3.30A, page 89, Table 3.1, page 65). These results confirm that selection coefficients between -1.5 and 1.0 in our experiment are correlated with DHFR activity over approximately 3 orders of magnitude, and that selection can resolve mutants with higher turnover than wild-type level activity.

3.3.6.2 Analyzing Noise in Selection Coefficients We next analyzed the deep mutational scanning data for all possible DHFR single point mutants under the calibrated selection conditions. All pairwise replicates were related with a Pearson correlation R^2 value of 0.70 and the median standard deviation between replicates for selection coefficients was 0.2 (Figure

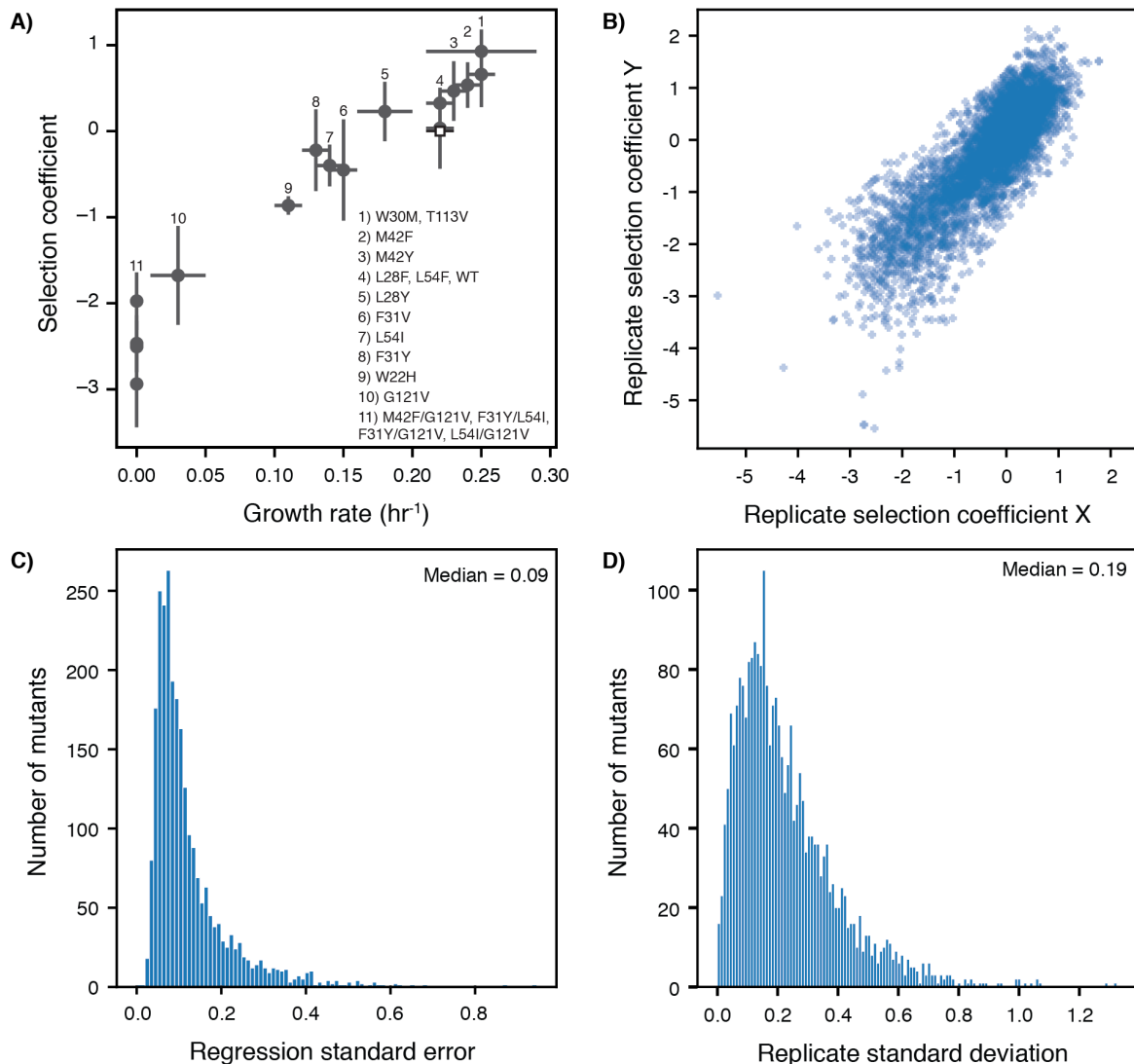


Figure 3.29: Error and reproducibility in DHFR DMS. **A)** Comparison of selection coefficients from **Figure 3.30C** (page 89) with growth rates measured in a plate reader for monocultures of the selection strain transformed with the SMT205 selection plasmid (**Table 5.2**, page 174) encoding a single DHFR variant (identified by number, in-Figure caption). Error bars reflect the standard deviation over at least three biological replicates. **B)** Comparison of all pairwise replicates for selection coefficients from triplicate deep mutational scanning on DHFR. The Pearson correlation R^2 value from linear regression was 0.70. **C)** Distribution of standard errors for individual selection coefficients from a single replicate. Selection coefficients are the slope from a linear regression of allele frequency as a function of time in selection. The standard error here is the mean square of residuals. **D)** Distribution of standard deviations of selection coefficients for individual point mutants over replicate experiments. Each mutant had a measured selection coefficient in at least 2 of the 3 replicates. The median of this distribution of standard deviations over all alleles was 0.2 and was used to determine the cut-offs for advantageous and disadvantageous mutations in **Figure 3.30B** (page 89).

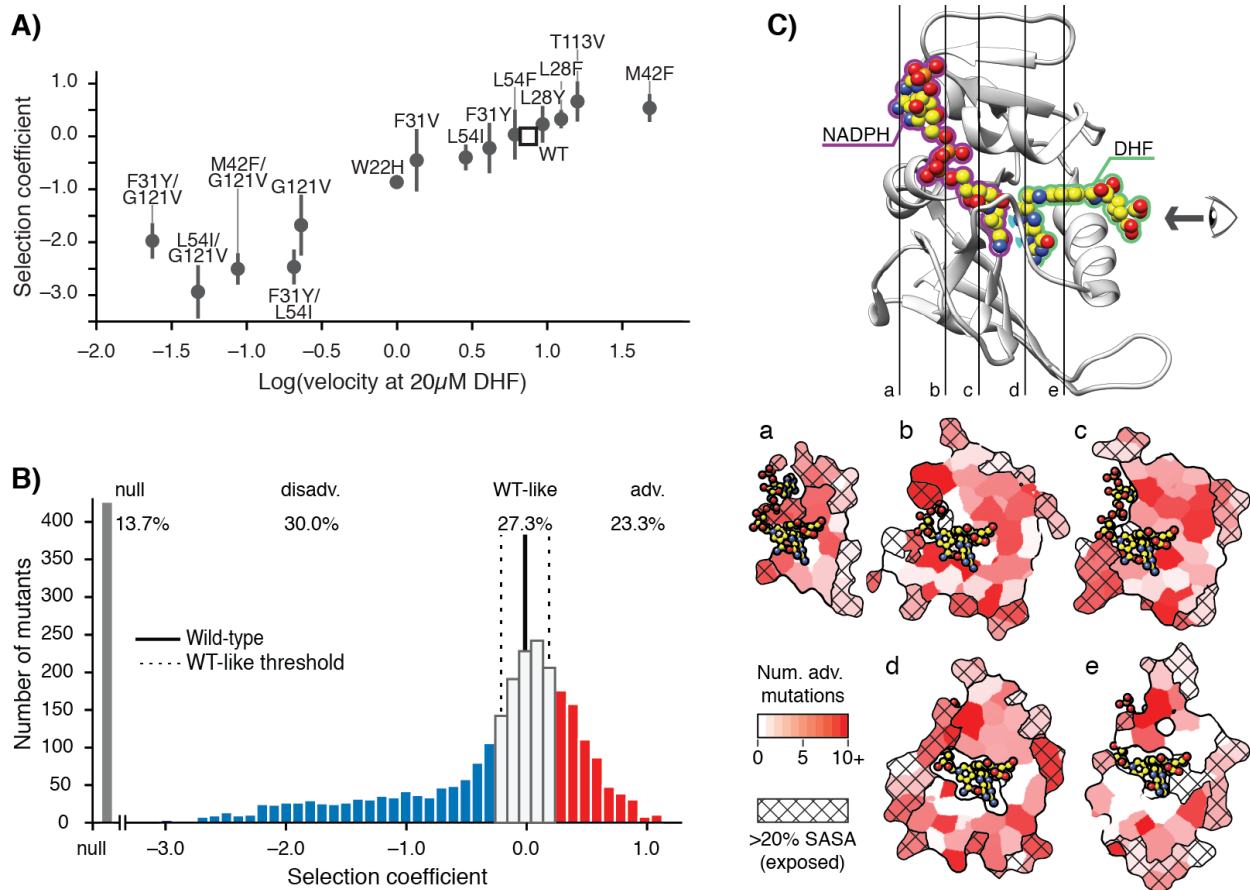


Figure 3.30: *E. coli* DHFR deep mutational scanning uncovers many advantageous mutations. **A)** Selection coefficients from deep mutational scanning as a function of enzymatic velocity for purified DHFR point mutants measured *in vitro*. Velocities at 20 μ M DHF were calculated from Michaelis-Menten parameters. Error bars reflect the standard deviation from 3 biological replicates. **B)** Histogram of selection coefficients. The wild-type value is indicated with a vertical black line. The median standard deviation over all mutations is the cut-off for WT-like behavior (see **Methods, Chapter 3.5.0.10**, page 115, **Figure 3.29**, page 88, **Figure 3.31**, page 90) and is indicated with dashed lines. Mutations are colored as advantageous (red), disadvantageous (blue), WT-like (white), or null (grey). **C)** Structural model of DHFR (PDB ID: 3QL3) with cross-section slices (a-e) indicated. The DHF substrate (green) and the NADPH cofactor (purple) are represented by spheres (yellow carbons and heteroatom coloring). An arrow indicates the perspective for each slice. **a-e)** 5 cross-section slices. Color scale indicates numbers of advantageous mutations at each position. Crosshatching indicates residues with >20% solvent accessible surface area.

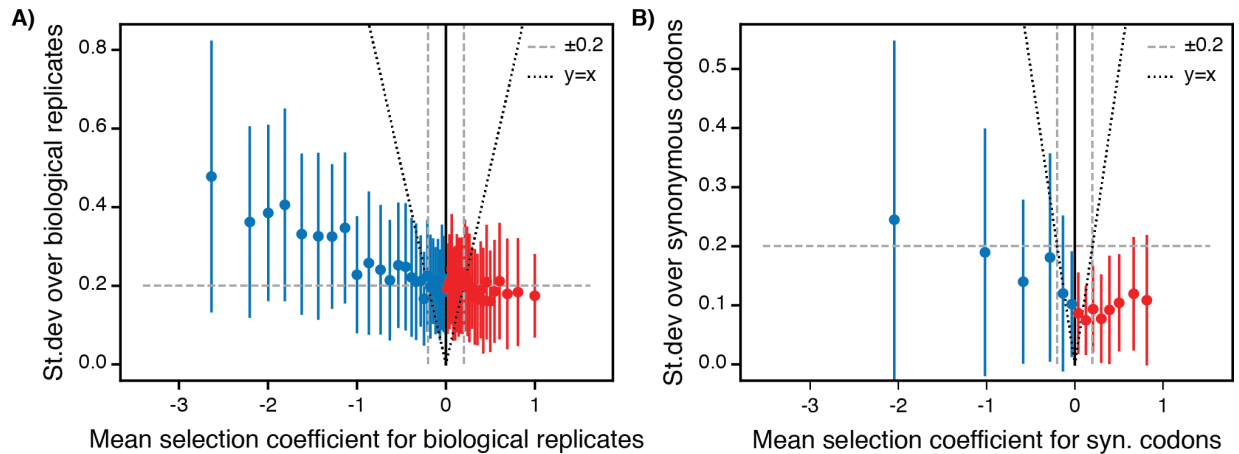


Figure 3.31: Variation in selection coefficients for DHFR DMS. **A)** Standard deviation of selection coefficients over biological replicates. The data were plotted as a function of a sliding window over all single point mutants sorted by selection coefficient. Each point represents the mean error (biological replicate standard deviation) over 50 consecutive selection coefficients (after sorting by value) and the error bars represent ± 1 standard deviation of the error. The dashed line represents median error over the entire dataset, which was used to determine the for WT-like behavior in **Figure 3.30B** (page 89). The dotted line represents $y = x$ for comparison between the magnitude of the error relative to the magnitude of the selection coefficient. **B)** Standard deviation over synonymous codons coding for the same sequence, plotted as in A).

3.31BC, page 90). The standard deviation of selection coefficients for WT synonymous codons was 0.12. From these data, we define DHFR mutations with selection coefficients of < -0.2 and > 0.2 as disadvantageous and advantageous, respectively (**Figure 3.30B**, page 89). This cutoff corresponds to the error in the regime of WT-like alleles with selection coefficients > -0.2 and < 0.2 (**Figure 3.31**, page 90).

3.3.7 Functionally Characterized Positions in DHFR are Enriched for Disadvantageous and Null Mutations

Mutations that were depleted during overnight growth in supplemented M9 (see **Methods**, **Chapter 3.5.0.6**, page 111) were assigned a null phenotype. As expected, mutations at DHFR positions that are known to be functionally important (M20, W22, D27, L28, F31, T35, M42, L54, R57, T113, G121, D122, and S148) were generally disadvantageous or null mutations (**Figure 3.32**, page 91). These results indicate that our selection assay is a sensitive reporter of

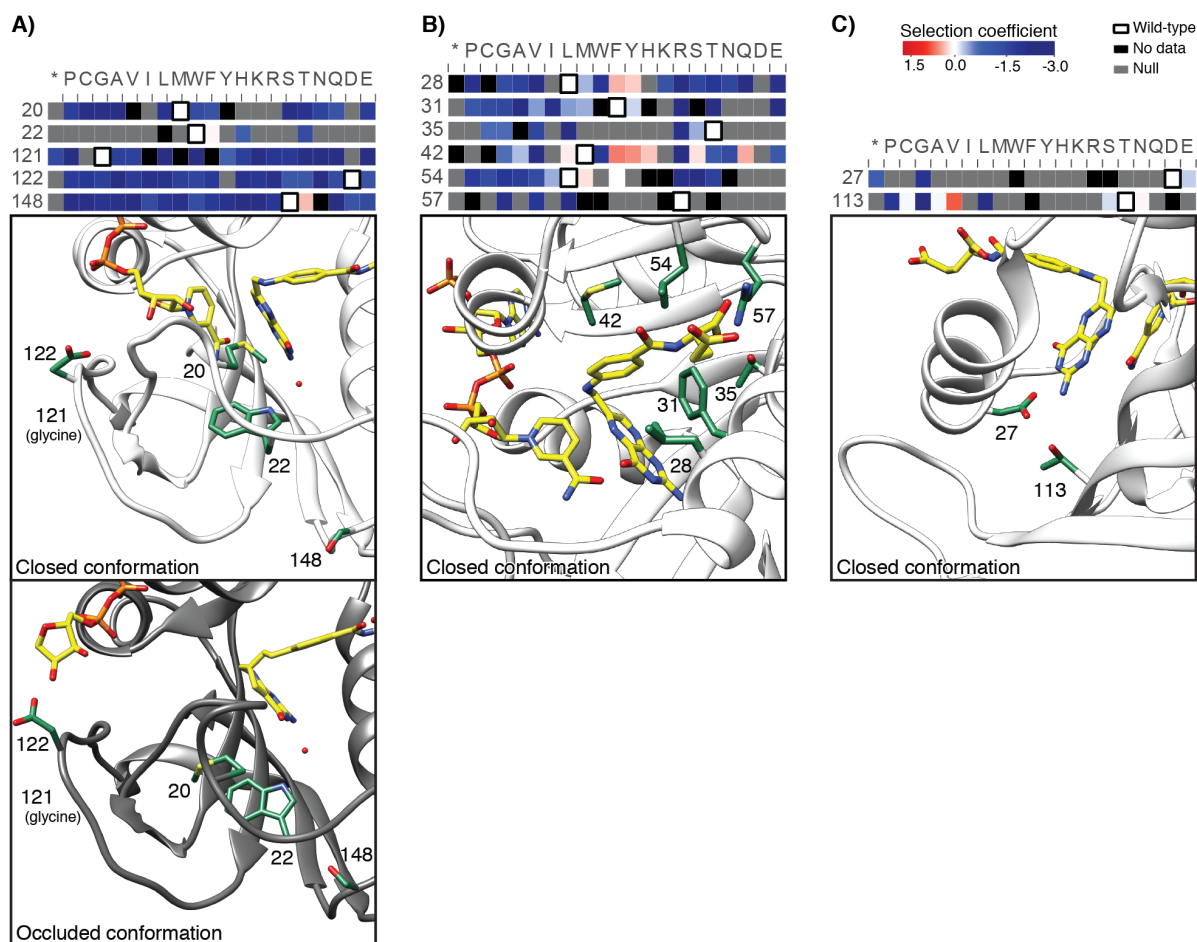


Figure 3.32: Residues previously shown to have a functional role displayed on the DHFR structure. **A-C)** Functionally important residues are colored green, labeled, and shown with slices of the $-L_{on}$ heatmap (heatmap coloring by selection coefficient is as in Figure 2). The wild-type residue is outlined in black. Positions 22, 27, 35, 57, and 113 are Intolerant, and positions 20, 28, 31, 42, 54, 121, 122, and 148 are Deleterious. In A) the closed (upper, white, PDB ID: 3QL3) and occluded (lower, grey, PDB ID: 1RX4) conformation are shown to illustrate alternate stabilization of the two conformations by D122 (closed)[37] and S148 (occluded)[38]. For all other panels, only the closed conformation is shown.

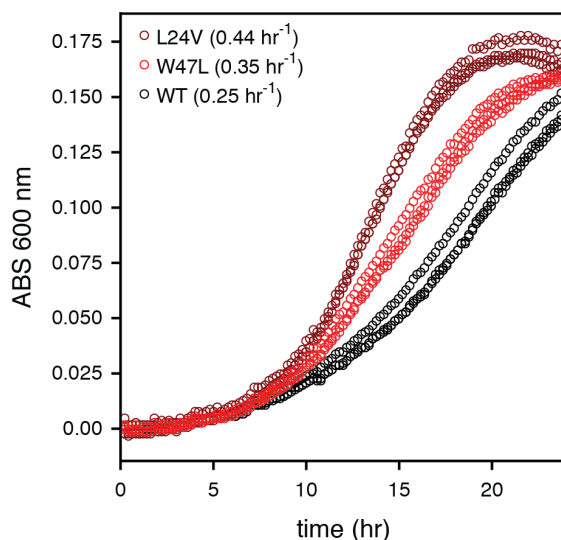


Figure 3.33: Growth curves for top advantageous mutations. The absorbance (ABS) at 600 nm was monitored in 96-well plate format for monocultures of the selection strain transformed with strong advantageous mutants (L24V in dark red, W47L in bright red, wild-type in black). The doubling rates (top left in plot) were calculated from the early exponential phase of growth (see **Methods, Chapter 3.5.0.4**, page 110). All growth curves are shown as sets of three biological replicates.

functionally important residues and that our results are consistent with previous biochemical characterization of DHFR.

3.3.8 Advantageous Mutations to DHFR

In contrast, the observation of a large fraction of advantageous mutations was unexpected: 737 of 3161 possible variants were advantageous mutations (23.3%), and wild-type DHFR only ranked 1205th. In direct measurements of individual growth rates under our selection conditions, the top two DHFR variants (W47L and L24V) led to increases in growth rate of 40 and 76%, respectively, when compared to wild-type DHFR (**Figure 3.33**, page 92). Advantageous mutations were widely distributed over 127 of the 159 positions of DHFR (**Figure 3.30C**, page 89). Furthermore, when we examined the DHFR structure, many of the advantageous mutations appeared to disrupt key side-chain interactions, for example by disrupting atomic packing interactions or surface salt-bridges (**Figure 3.34**, page 93).

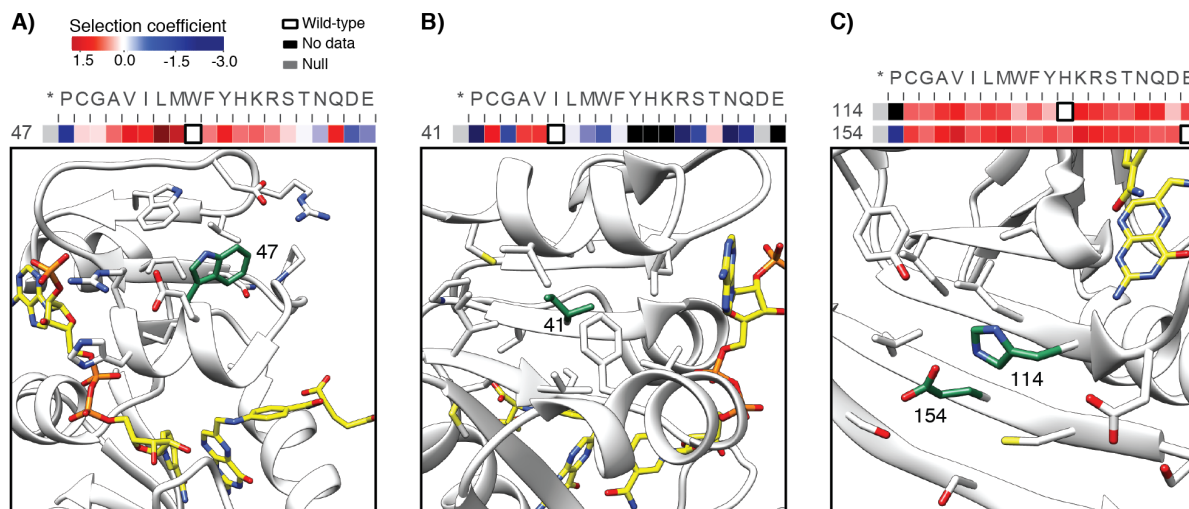


Figure 3.34: Example positions with multiple advantageous mutations hypothesized to be destabilizing. **A-C)** Wild-type residues are colored in green on the DHFR structure (PDB ID: 3QL3) and depicted with slices of the $-L_{on}$ heatmap (heatmap coloring is as in Figure 2). The wild-type residue is outlined in black on the heatmap. Positions 47, 114, and 154 are in the Beneficial category, and position 41 is in the Deleterious category. In the examples here, advantageous mutations appear to disrupt core packing and a surface salt bridge.

3.3.9 Characterizing Top Advantageous Mutations to DHFR

We next sought to test whether advantageous mutations make DHFR a faster enzyme. To select specific mutations for *in vitro* tests, we considered all positions with more than one mutation in the top 100 most advantageous mutations. We describe these positions by their location (**Figure 3.35**, page 94) in one of four structural regions that appear to be hot-spots for the top advantageous mutations: 1) exchanges between hydrophobic residues at core positions, 2) disruptions of surface residues on the beta-sheet near the active site, 3) disruptions of polar interactions with the adenine ring of NADPH, or 4) mutations to the active site or M20 loop that controls access to the active site. At these positions, we selected 24 strongly advantageous mutations for *in vitro* characterization. (**Figure 3.36**, page 95)

3.3.9.1 Measuring DHFR Point Mutant Activities in Lysates The DHFR activity in ER2566 $\Delta folA/\Delta thyA$ lysates was uniformly higher for these advantageous mutants relative to the

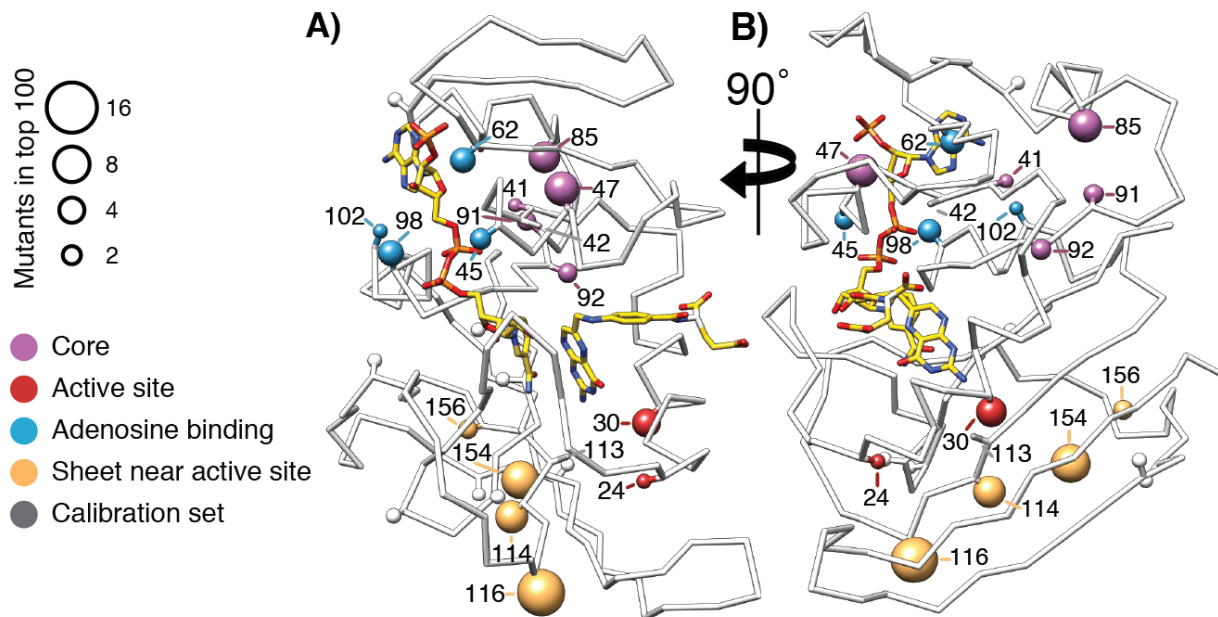


Figure 3.35: DHFR structure with mutational hot-spots. **A)** Positions with 2 or more top 100 advantageous mutations displayed on the DHFR structure. The beta carbon is depicted as a sphere scaled according to the number of top mutations. For mutants selected for *in vitro* characterization, the beta carbon is colored according to its location in the DHFR structure: core (purple), surface beta-sheet (gold), proximal to the adenine ring on NADPH (blue), or proximal to the active site and M20 loop (red). Positions for advantageous mutants from the calibration set are depicted in dark grey. **B)** The structure from A) rotated 90° clockwise.

activity for WT DHFR (**Figure 3.37**, page 96). This lysate activity assay reports on both kinetic and expression effects as the product of $[\text{DHFR}] \bullet k_{cat}$. These results are consistent with the results from *in vivo* selection.

3.3.9.2 *In vitro* Velocities for Purified DHFR Point Mutants We then examined the *in vitro* velocities of these mutants (**Figure 3.38**, page 97). We observed that approximately 2/3 of the mutants have velocities equal an up to three-fold higher than that of WT at cytosolic concentrations of DHF. In contrast, approximately 1/3 of the mutants have velocities as much as two-fold lower than that of WT in the same DHF concentration regime.

3.3.9.3 Compound Contributions from Multiple Molecular Properties Lead to Increased Growth with Advantageous Mutations We therefore examined the soluble expression level of these mutants

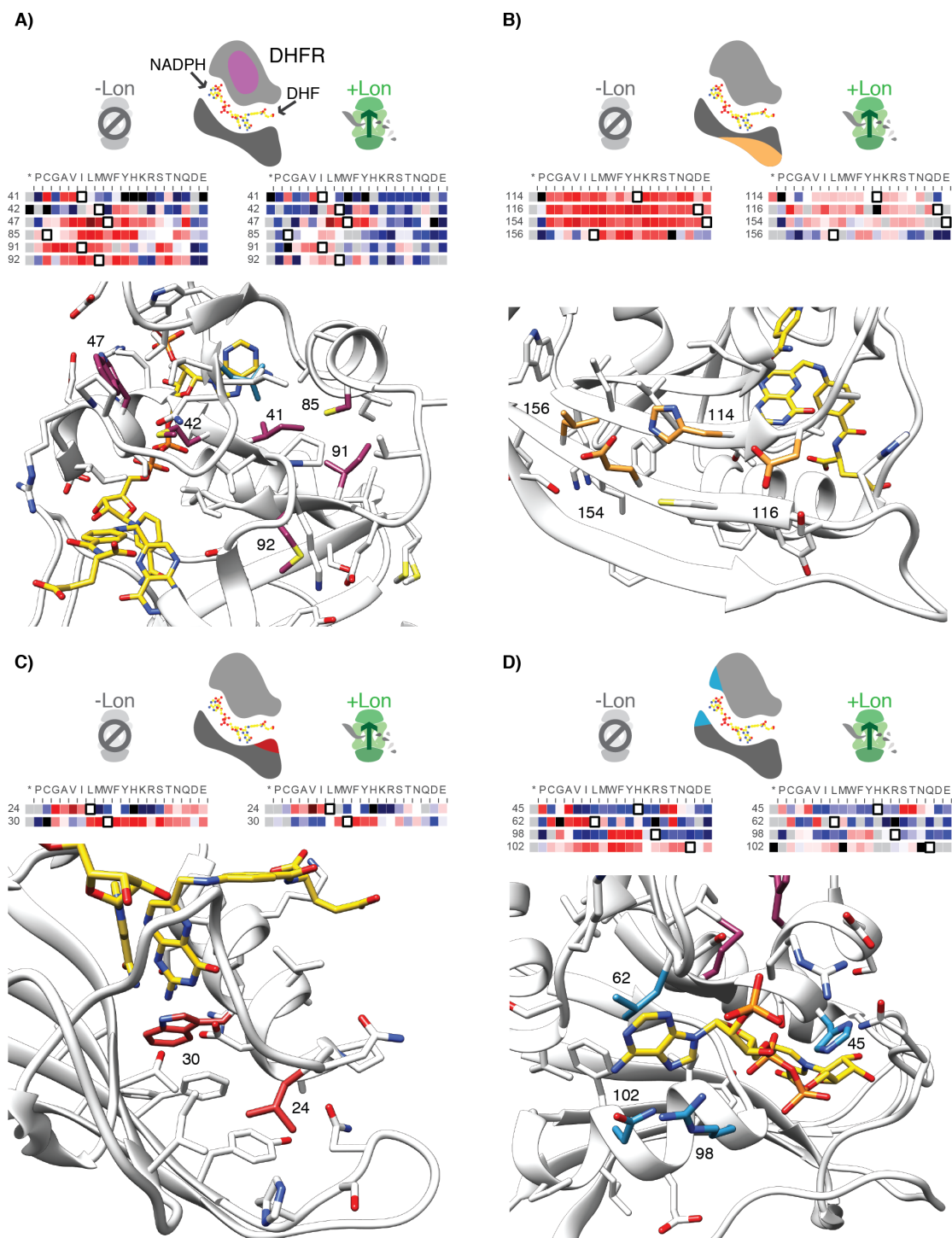


Figure 3.36: Structural context for hotspot residues from **Figure 3.35** (page 94). **A-D)** For each panel, the hot spot region is indicated on a cartoon of DHFR: globular core in purple (A), the beta-sheet surface below the active site in gold (B), the base of the M20 loop in red (C) and the adenosine binding site in blue (D). Slices of the -Lon and +Lon heatmaps are shown for each position within the hot spot region (heatmap coloring is as in Figure 2). The wild-type residue is outlined in black. Positions 30, 47, 85, 102, 114, 116, 154 are in the Beneficial category. Position 24, 25, 62, 91, 92, 156 are in the Mixed category. Positions 41, 42, and 98 are in the Deleterious category. For A-C) the structure shown is PDBID: 3QL3, and for D) the structure shown is PDB ID: 1RX1. In 1RX1 (as in 1RX4), R98 is in proximity to the adenine ring. In 3QL3, R98 extends into bulk solvent. Residues within the hot spot cluster are labeled with their residue number.

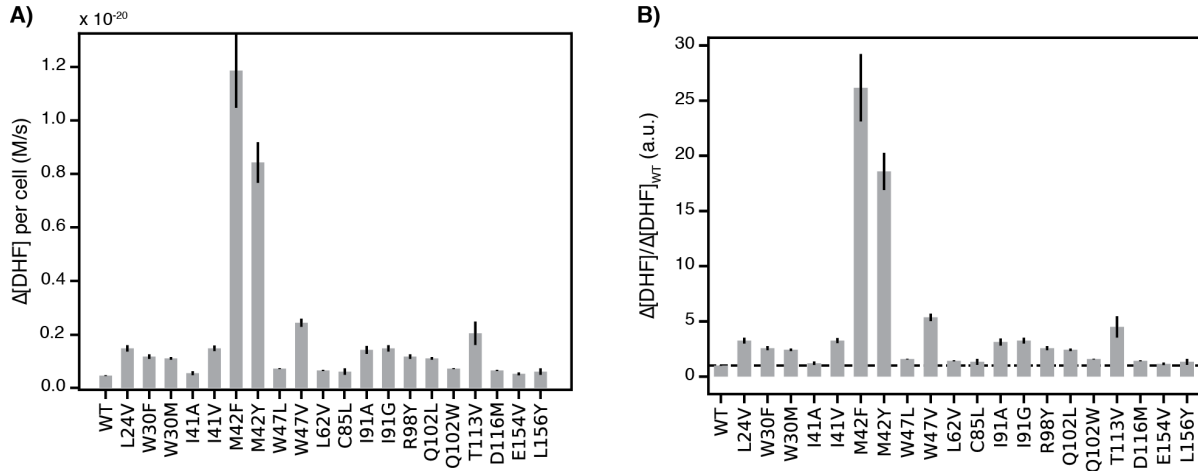


Figure 3.37: Lysate activity for DHFR wild-type and point mutants on the selection plasmid. **A)** Lysate activity for DHFR variants under selection growth conditions (see **Methods, Chapter 3.5.0.14**, page 122) plotted as the rate of change in DHF concentration as a function of time monitored over the window of DHF concentration from $30 \mu\text{M}$ to $20 \mu\text{M}$ in ER2566 $\Delta folA/\Delta thyA$ lysates. Error bars represent ± 1 standard deviation from three biological replicates. **B)** Relative lysate activities for DHFR variants. Lysate activities from A) normalized by WT-level of activity.

and observed that mutant expression levels relative to WT levels varied widely from an 8-fold decrease to a 20-fold (**Figure 3.39**, page 100). Combining the data on abundance and velocity, the expected total DHFR activity ($[\text{DHFR}] \bullet \text{velocity}$) is not a strong quantitative predictor of the advantageous mutants in selection (**Figure 3.40**, page 101, **Figure 3.41**, page 102). We attribute discrepancies at least in part to the difficulty of accurately quantifying rather small differences in activity and abundance, in addition to other potential complicating factors such as differential activity of cellular chaperones for different DHFR variants[39], and feedback regulation that could affect cellular concentrations of the substrate DHF[32, 40]. Nevertheless, our *in vitro* measurements are in qualitative agreement with the *in vivo* selection. Specifically, the majority of advantageous mutants show increased activity compared to WT when considering both velocity and abundance (positions above the line indicating WT activity in **Figure 3.42**, page 103). Taken together, these results suggest that increased selection coefficients arise from an interplay of effects of the mutations on cellular abundance and catalytic activity, and that each parameter

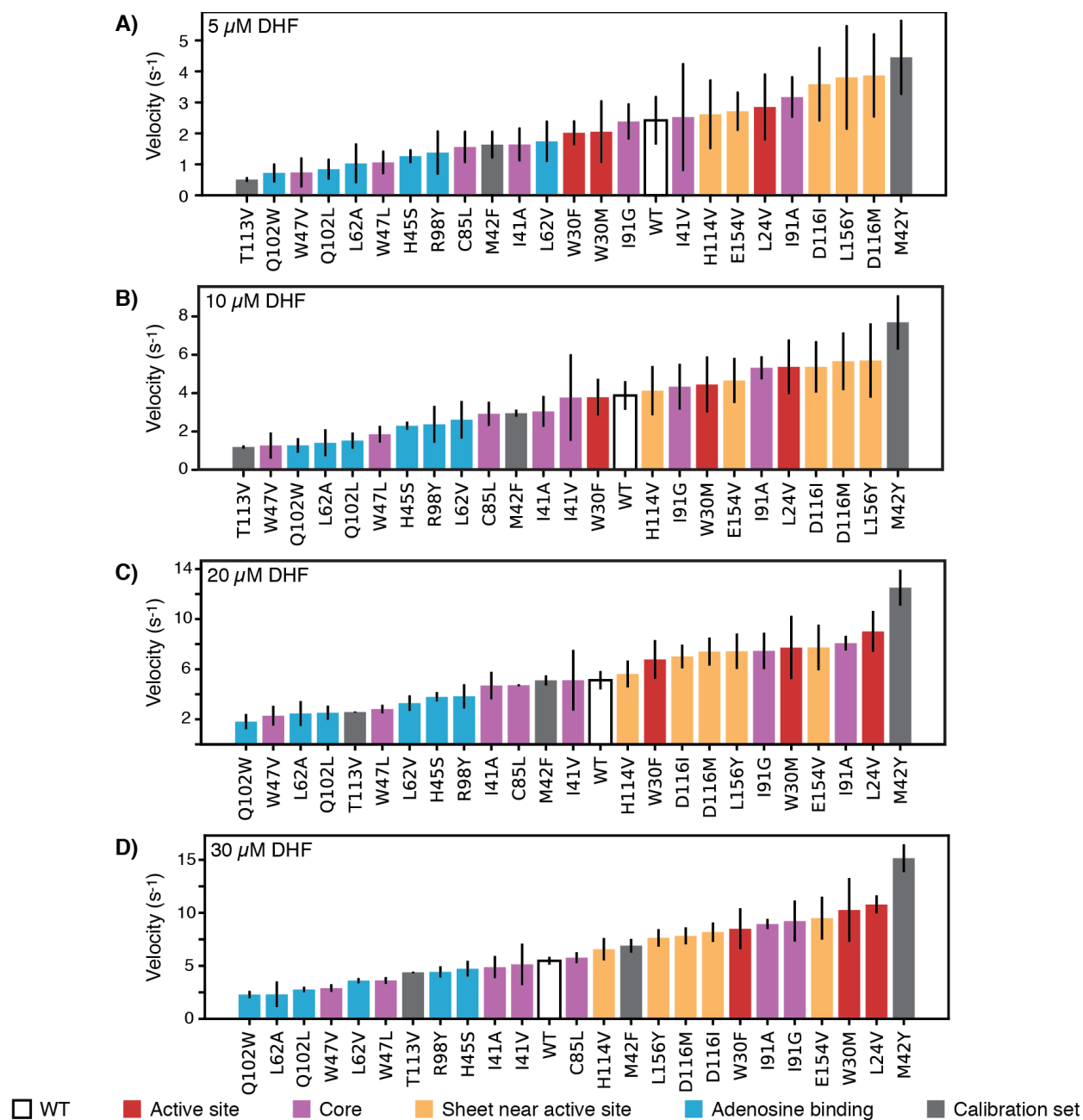


Figure 3.38: *In vitro* velocities of purified DHFR wild-type and point mutants. Velocities were measured at **A)** 5, **B)** 10, **C)** 20, and **D)** 30 μM DHF (Table 3.6, page 98). For each mutant, the bar is colored by the mutation's location within the hot spots from Figure 3.35 (page 94) and Figure 3.36 (page 95). Error bars represent ± 1 standard deviation from three independent experiments.

Table 3.6: *In vitro* velocity for selected advantageous mutations measured as described in **Methods (Chapter 3.5.0.13, page 121)** at multiple concentrations of DHF are reported with the standard deviation over three independent experiments.

Variant	5 μ M DHF		10 μ M DHF		20 μ M DHF		30 μ M DHF	
	Velocity	Std. dv.	Velocity	Std. dv.	Velocity	Std. dv.	Velocity	Std. dv.
WT	2.42	0.75	3.88	0.75	5.12	0.74	5.48	0.37
L24V	2.85	1.05	5.37	1.43	9.01	1.64	10.79	0.86
W30F	2.02	0.37	3.79	0.96	6.77	1.56	8.50	1.94
W30M	2.05	0.98	4.45	1.46	7.73	2.53	10.27	3.03
I41A	1.64	0.52	3.05	0.81	4.69	1.11	4.88	1.06
I41V	2.52	1.71	3.77	2.27	5.11	2.44	5.13	1.96
M42F	3.08	1.40	4.62	1.68	6.42	1.24	6.90	0.64
M42Y	4.45	1.18	7.69	1.41	12.51	1.43	15.16	1.32
H45S	1.26	0.19	2.29	0.21	3.79	0.38	4.73	0.74
W47L	1.06	0.35	1.85	0.44	2.81	0.35	3.62	0.32
W47V	0.74	0.45	1.27	0.68	2.28	0.79	2.90	0.37
L62A	1.03	0.61	1.41	0.71	2.46	1.00	2.30	1.22
L62V	1.74	0.64	2.61	0.64	3.29	0.62	3.60	0.26
C85L	1.56	0.49	2.92	0.63	4.71	0.09	5.77	0.52
I91A	3.17	0.64	5.32	0.60	8.08	0.59	8.94	0.49
I91G	2.38	0.55	4.33	1.20	7.46	1.46	9.22	1.95
R98Y	1.38	0.68	2.38	0.97	3.83	0.97	4.43	0.53
Q102L	0.84	0.31	1.52	0.43	2.52	0.57	2.77	0.26
Q102W	0.72	0.28	1.27	0.38	1.81	0.61	2.29	0.35
T113V	0.51	0.05	1.19	0.09	2.58	0.04	4.38	0.08
H114V	2.61	1.09	4.12	1.28	5.62	1.08	6.57	1.06
D116I	3.59	1.17	5.37	1.34	7.01	0.95	8.18	0.87
D116M	3.87	1.33	5.66	1.50	7.40	1.21	7.83	0.80
E154V	2.71	0.60	4.66	1.16	7.73	1.82	9.50	2.02
L156Y	3.80	1.66	5.70	1.94	7.43	1.41	7.64	0.82

alone is insufficient to explain the majority of the advantageous mutations.

3.3.9.4 Measuring T_M Values for DHFR Point Mutants We then measured apparent melting temperature (T_m) values from non-reversible thermal denaturation monitored by circular dichroism spectroscopy (**Figure 3.43, page 104**), which revealed that many of the advantageous mutations considerably destabilized the protein, and no mutant proteins were more thermostable than WT DHFR.

Table 3.7: Soluble DHFR abundance levels in molecules per cell. Abundances were measured from lysate activity assays as described in **Methods (Chapter 3.5.0.14, page 122)**. All values are for the SMT205 plasmid transformed into the ER2566 $\Delta folA \Delta thyA$. NM, not measured.

Variant	Molecules per cell	Standard deviation
WT	52.0	9.2
L24V	83.6	6.8
W30F	81.3	2.7
W30M	76.9	0.3
I41A	65.6	5.9
I41V	177.5	72.8
M42F	1199.1	17.6
M42Y	360.9	2.6
H45S	101.9	10.1
W47L	558.3	94.5
W47V	NM	NM
L62A	NM	NM
L62V	112.0	9.4
C85L	67.2	11.6
I91A	98.5	3.8
I91G	110.7	12.8
R98Y	116.1	16.2
Q102L	239.5	27.0
Q102W	194.9	42.4
T113V	418.3	19.1
H114V	NM	NM
D116I	NM	NM
D116M	49.8	4.4
E154V	37.0	5.1
L156Y	47.3	3.1

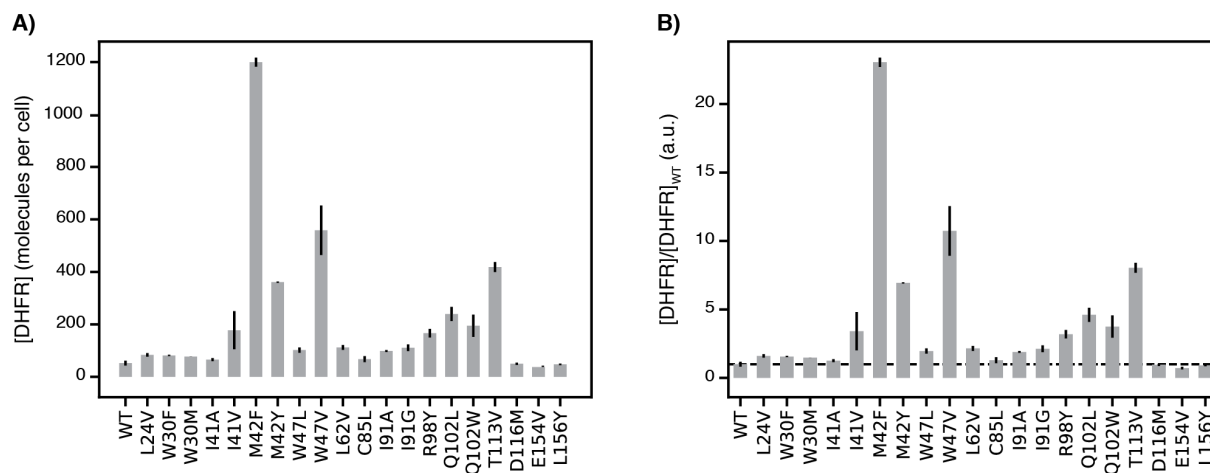


Figure 3.39: Soluble cellular abundance for DHFR wild-type and point mutants on the selection plasmid. **A)** DHFR cellular abundance calculated from the lysate DHFR activity in **Figure 3.37** (page 96) and *in vitro* kinetics with purified enzyme (see **Methods, Chapter 3.5.0.14**, page 122). Error bars represent the cumulative percent error (standard deviation) from three independent experiments for velocity and three biological replicates for lysate activity. **B)** Relative expression of DHFR variants. DHFR abundances from A) normalized by WT-level of abundance. Error bars represent the cumulative percent error as in A).

Table 3.8: Apparent T_m values from thermal denaturation experiments monitored by CD signal at 225 nm.

Variant	T_M
WT	57.4
L24V	54.8
W30F	48.7
W30M	48.4
I41A	39.0
I41V	53.3
M42F	47.9
M42Y	44.2
H45S	55.1
W47L	50.6
W47V	41.9
L62A	42.0
L62V	53.3
C85L	42.4
I91A	55.2
I91G	49.8
R98Y	54.8
Q102L	57.0
Q102W	52.8
T113V	52.5
H114V	47.3
D116I	50.3
D116M	50.6
E154V	53.9
L156Y	47.8

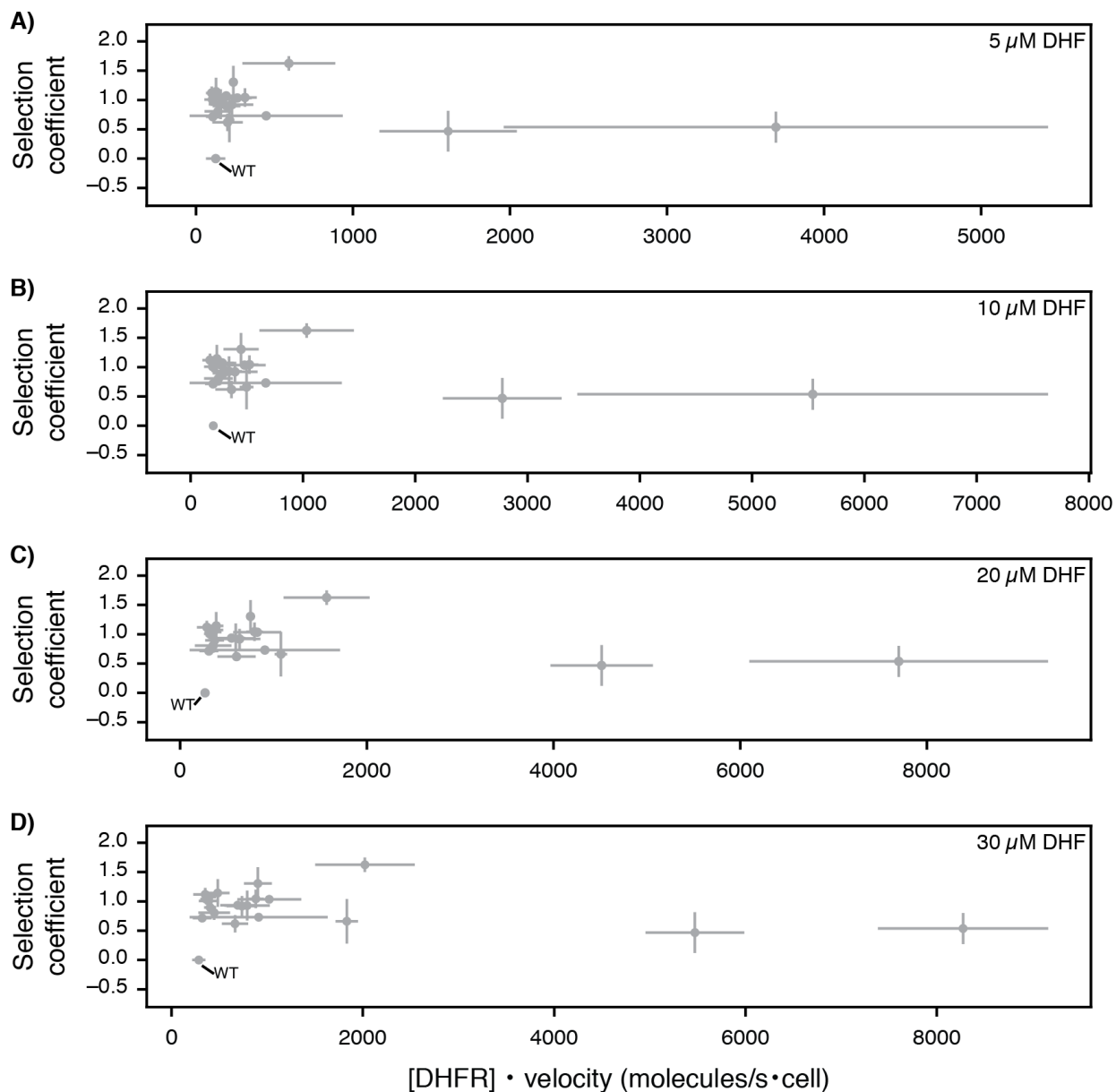


Figure 3.40: Selection coefficient compared to predictions of WT DHFR and point mutant activity from cellular abundance and *in vitro* velocity measurements. Selection coefficients for selection plotted against DHFR activity calculated as cellular abundance of DHFR times *in vitro* velocities of purified DHFR variants ($[\text{DHFR}] \cdot \text{velocity}[\text{DHF}]$) measured at **A)** 5, **B)** 10, **C)** 20, and **D)** 30 μM DHF (see **Figure 3.38**, page 97, **Table 3.6**, page 98, **Figure 3.39**, page 100, **Table 3.7**, page 99). X-axis error bars represent the cumulative percent error (standard deviation) from three measurements of DHFR concentration and three independent experiments for velocity (see **Methods, Chapter 3.5.0.13**, page 121, **Chapter 3.5.0.14**, page 122). Y-axis error bars represent ± 1 standard deviation from biological replicate selection experiments (see **Methods, Chapter 3.5.0.10**, page 115).

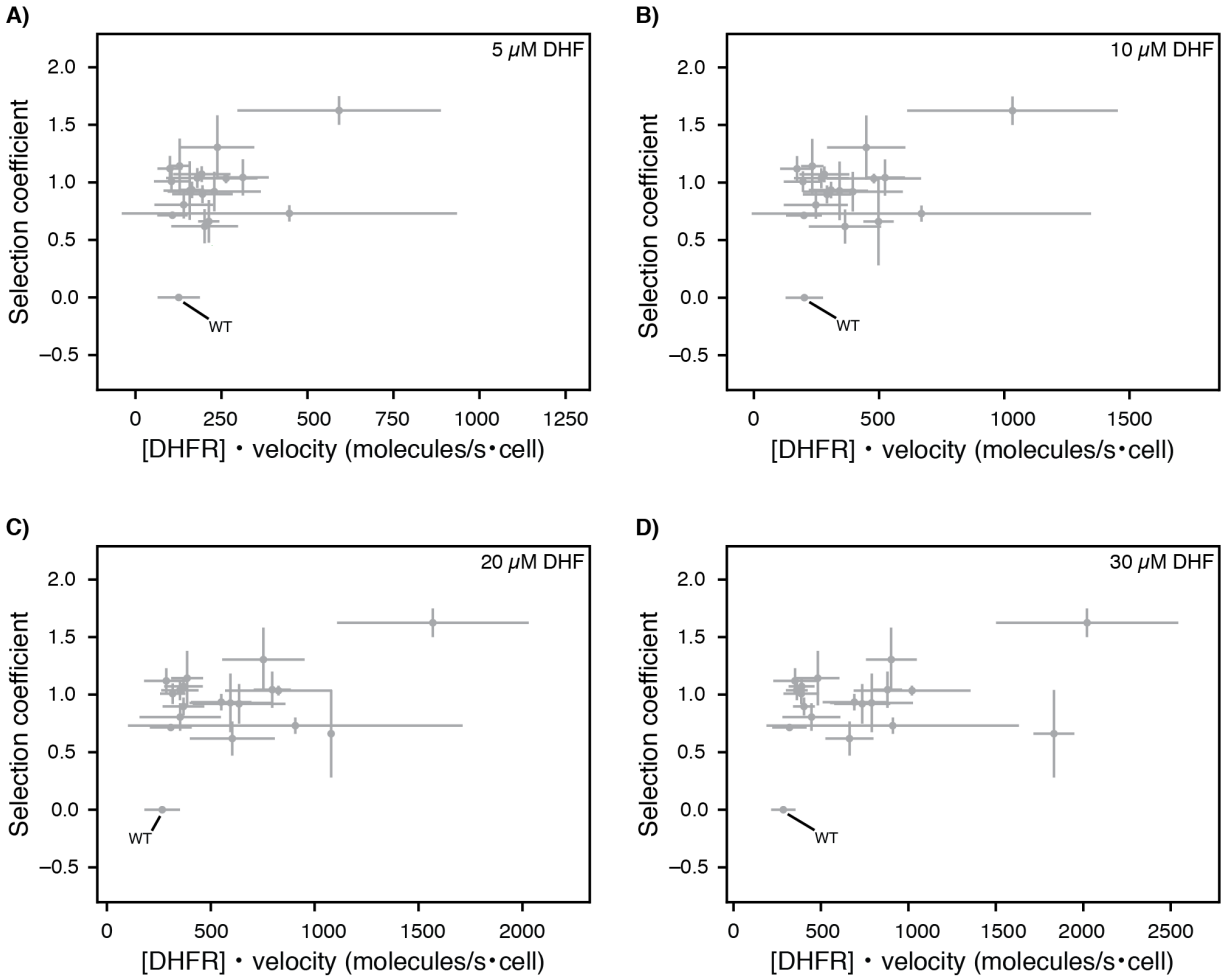


Figure 3.41: Zoom in for selection coefficient compared to predictions of WT DHFR and point mutant activity from cellular abundance and *in vitro* velocity measurements. Selection coefficients for selection plotted against DHFR activity calculated as cellular abundance of DHFR times *in vitro* velocities of purified DHFR variants ($[\text{DHFR}] \cdot \text{velocity}[\text{DHF}]$) measured at **A)** 5, **B)** 10, **C)** 20, and **D)** 30 μM DHF (see **Figure 3.38**, page 97, **Table 3.6**, page 98, **Figure 3.39**, page 100, **Table 3.7**, page 99). X-axis error bars represent the cumulative percent error (standard deviation) from three measurements of DHFR concentration and three independent experiments for velocity (see **Methods, Chapter 3.5.0.13**, page 121, **Chapter 3.5.0.14**, page 122). Y-axis error bars represent ± 1 standard deviation from biological replicate selection experiments (see **Methods, Chapter 3.5.0.10**, page 115).

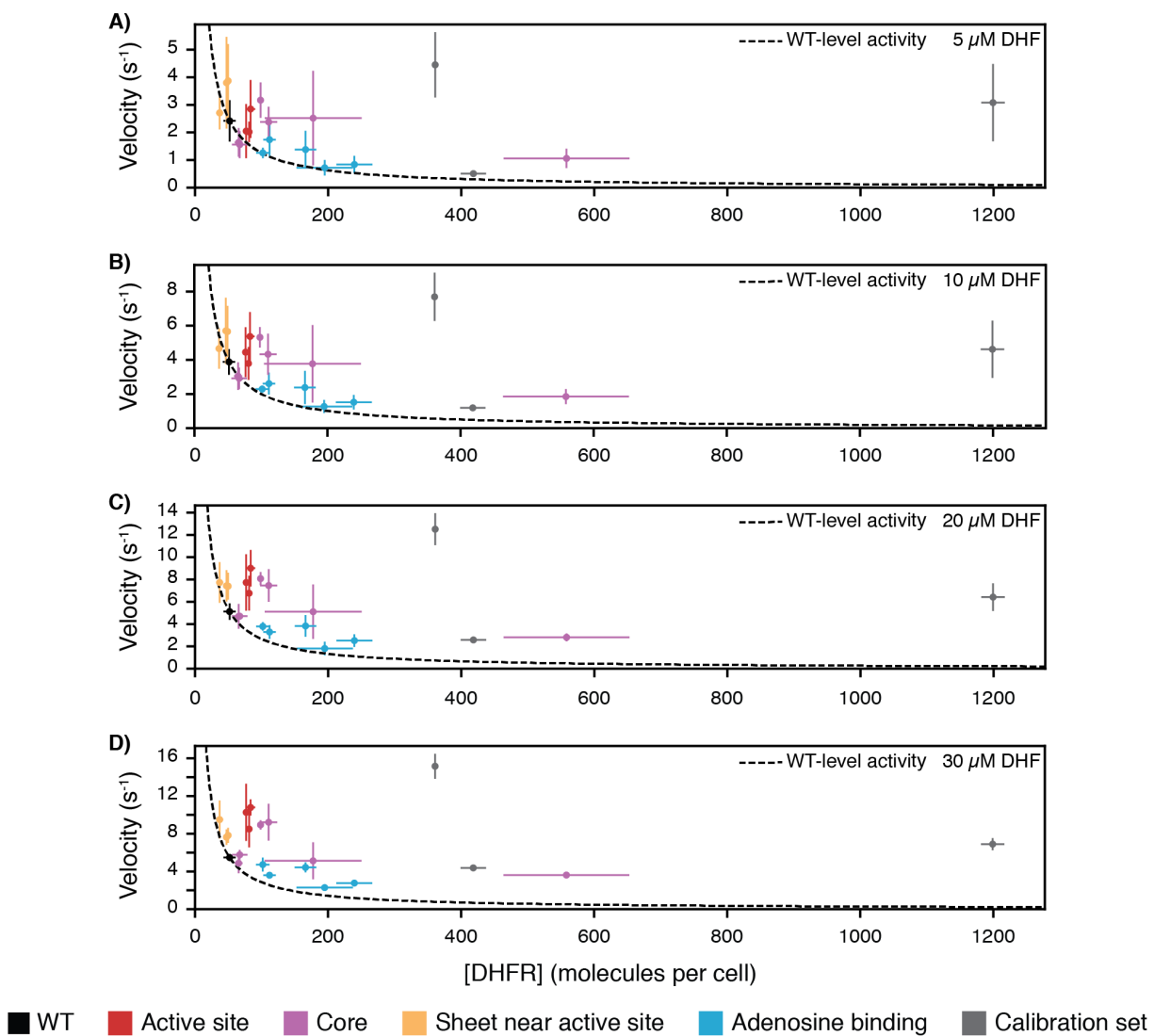


Figure 3.42: Cellular abundance versus *in vitro* velocity for DHFR wild-type and point mutants. Cellular abundance of DHFR vs. *in vitro* velocities of purified DHFR measured at **A)** 5, **B)** 10, **C)** 20, and **D)** 30 μM DHF (see **Figure 3.38**, page 97, **Table 3.6**, page 98, **Figure 3.39**, page 100, **Table 3.7**, page 99). Points are colored by the mutation's location within the hot spots (**Figure 3.35**, page 94, **Figure 3.36**, page 95). Error bars represent ± 1 standard deviation from three independent experiments (see **Methods, Chapter 3.5.0.13**, page 121, **Chapter 3.5.0.14**, page 122). The dashed line represents WT equivalent DHFR activity, where $[\text{DHFR}]_{WT} \cdot \text{velocity}_{WT} = [\text{DHFR}]_{mut} \cdot \text{velocity}_{mut}$.

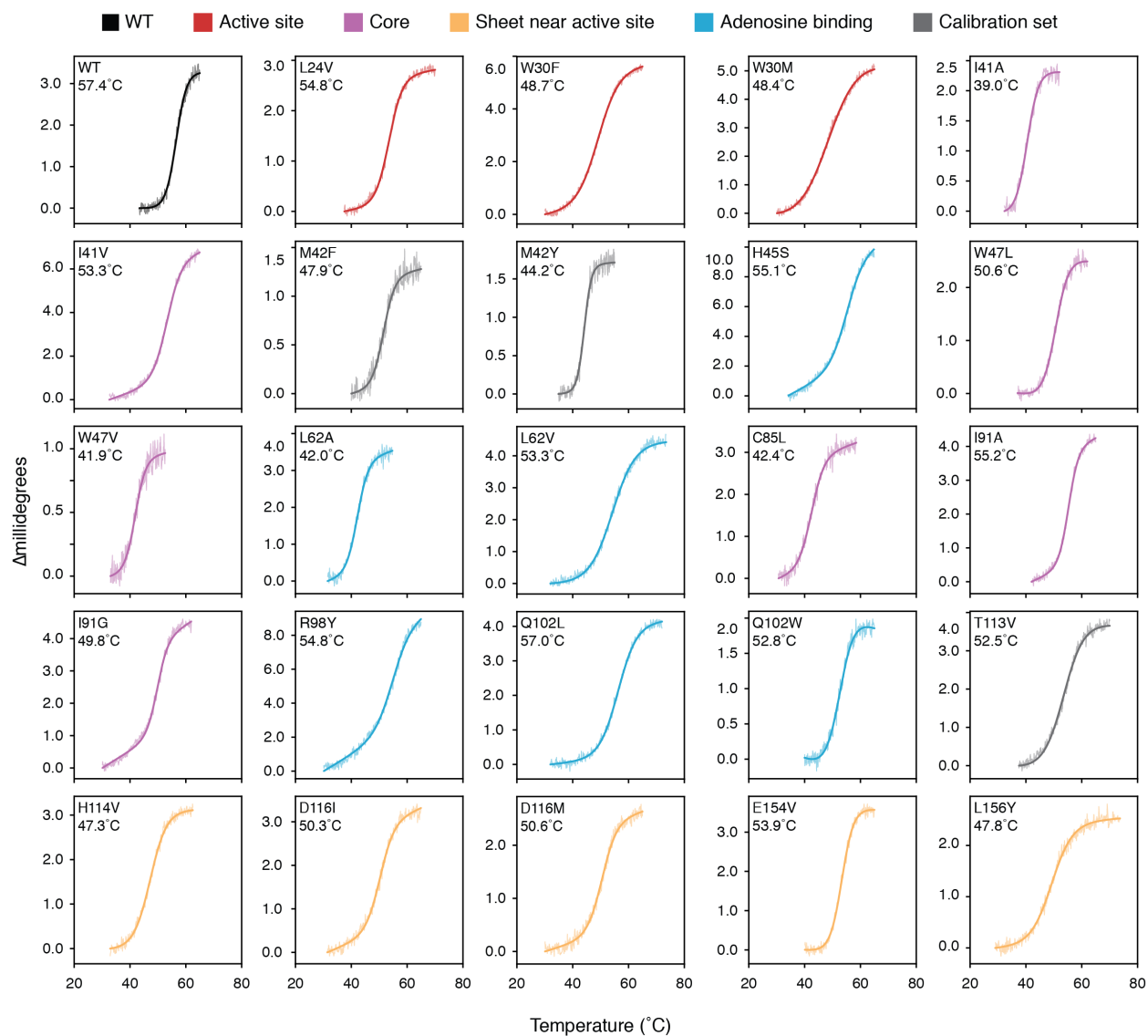


Figure 3.43: Thermal denaturation curves monitored by CD signal at 225 nm for selected hotspot mutants. The curves are colored by the mutation's location within the hot spots (**Figure 3.35**, page 94, **Figure 3.36**, page 95). The raw data are shown with thin lines and the fitted curves are shown as thick lines. For each plot, the mutant identity and apparent T_m value are listed in the top left corner.

3.4 Discussion

The experiments in this chapter show the optimization of a selection assay for DHFR activity that can resolve 2-3 fold changes in velocity at cytosolic concentrations of DHF. Selection under these conditions revealed many advantageous mutations to DHFR. We observed that these advantageous mutations show a selection advantage in monoculture and higher lysate activity which results from a combination of expression level effects and changes to the kinetic parameters of the enzyme. We also show that a representative sample of the strongest advantageous mutations show a wide-range of destabilizing impacts.

It is common to conceptualize mutational impacts in terms of activity-stability trade-offs, but it is not immediately evident that destabilization is causal to DHFR activation. It is expected that the set of all single point mutations to DHFR gives rise to a broad range of thermodynamic impacts in which destabilizing mutations more likely. Therefore, selection advantage and stability may be independent properties, and the advantageous mutations we characterized may only represent the underlying global distribution of thermodynamic impacts. There may be no correlation between fold stability and selection coefficients if we measure T_m values for all possible single mutations. Alternatively, protein fold thermostability is by definition a very low resolution metric. It is possible that destabilizing mutations in specific regions (e.g. the hot spots for advantageous mutations) preferentially impact the kinetic parameters of the enzyme. Within these hotspots, there may be more correlation between stability and selection advantage. While the advantageous mutations are spread throughout DHFR (**Figure 3.30C**, page 89), we do not know if these mutations act to increase DHFR activity by a consistent method and if they have similar destabilizing effects. It would be useful to have a broad picture of which mutations destabilize DHFR and how that pattern relates to the pattern of observed advantageous mutations. We will describe experiments

performed to address questions around activity stability trade-offs in **Chapter 4**.

3.4.1 Future Directions

To date, however, we have not used our well-calibrated assay to perform selection on extensive libraries of higher order mutations. There are several interesting questions that we can approach with this powerful tool. One question is, can a predictive model of epistasis be developed? Recent results from the lab of Rama Ranganathan were used to support a model where mutational impacts of single and double point mutants averaged over a large number of combinatorial backgrounds were more predictive than the measurements of those mutational impacts in the wild-type background alone [41]. What was not clear from this study was whether that increase in predictive power resulted from a fundamental principle of epistasis, or if averaging corrected for noise in the original measurements. Using combinatorial libraries in *E. coli* DHFR and in the background of DHFR orthologues could be a compelling experiment for examining how to predict mutational impacts with both many and few changes to the background sequence.

Taking the line of inquiry on mutational impact, the extent of functional sequence space, and the possibility of predictive models further, we can ask where the pareto front for DHFR activity lies in sequence space. How far can a sequence differ from DHFR orthologue sequences and still have wild-type level DHFR activity? In **Chapter 2** (page 11), we described and tested a computational method for designing sequences that conformed to multiple structural and energetic constraints. Computational methods may allow us to predict and test sequences on the pareto front for DHFR activity in sequence space. Screening libraries of designed DHFR sequences could also be very powerful for understanding where computational design models work and fail in producing functional DHFR sequences.

Finally, we can expand the number of dimensions when we consider the constraints on

functional sequence space by including perturbations to the environment around DHFR. DHFR sequences that are functional and able to support *E. coli* growth under our current selection conditions may be non-functional in perturbed conditions such as low-folate stress from the antibiotic sulfathiazole or inhibition of TYMS from 5-fluorouracil. Carbon source limitations and anaerobic conditions may greatly change the redox potential within the cell. Temperature will affect expression levels, $\Delta\Delta G_{\text{folding}}$, and kinetic rates in many cellular processes. All of these resulting stresses may reshape the DHFR mutational landscape. These types of experiments, and how different mutational landscapes are integrated over time should yield new insights into how organisms adapt and protein sequences evolve.

3.5 Methods

All plasmid and primer sequences are listed in **Table 5.2** (page 174). Key plasmids were deposited in the Addgene plasmid repository and are available at https://www.addgene.org/Tanja_Kortemme. All code and python scripts are available at https://github.com/keleayon/2019_DHFR_Lon.git with key input files and example command lines.

3.5.0.1 Generation of Plasmids for *In Vivo* Selection The vector bearing DHFR and TYMS for *in vivo* selection (SMT205) was derived from the pACYC-Duet vector described by Reynolds et al [20]. The lac operon upstream of the TYMS gene was replaced with a Tet-inducible promoter. A Tet promoter fragment had been generated with overlap extension polymerase chain reaction (PCR) and cloned into the pACYC vector (SMT101) at unique AflII/BglII sites to produce SMT201. Selection conditions that resolved increased-fitness mutations were obtained with the SMT205 plasmid where the DHFR “AAGGAG” RBS was replaced with “AATGAG” based on

prediction from the RBS calculator (Salis et al., 2009) using inverse PCR. Briefly, PCR reactions were set up using 2x Q5 mastermix (NEB, cat# M0492), 10 ng of plasmid template, and 500 nM forward and reverse primers. PCR was performed in the following steps: 1) 98°C for 30 seconds, 2) 98°C for 10 seconds, 3) 57-63°C for 30 seconds, 4) 72°C for 2 minutes, 5) return to step 2 for 22 cycles, 6) 72°C for 5 minutes. As needed, the annealing temperature (step 3) was optimized in the range of 57-63°C. 25 μ L of RBS reaction was mixed with 1 μ L of DpnI (NEB, cat# R0176), 1 μ L of T4 PNK (NEB, cat# M0201), 1 μ L of T4 ligase (NEB, cat# M0202), and 3.1 μ L of T4 ligase buffer (NEB, cat# B0202) at 37°C for 2-4 hours. The reactions were then transformed into chemically competent Top10 cells and plated on LB agar plates with 35 μ g/mL chloramphenicol (Fisher BioReagents, BP904, CAS: 56-76-7, 35 mg/mL in ethanol). The plates were incubated overnight at 37°C. Single colonies were picked and used to inoculate 5 mL of LB medium (10 g Bacto-tryptone (Fisher BioReagent, cat# BP1415, CAS: 73049-73-7), 5 g Bacto-yeast extract (BD Difco, cat# 212720, CAS: 8013-01-2), 10 g NaCl (Fisher BioReagents, cat# BP358, CAS 7647-14-5), 0.186 g KCl (Sigma, cat# P9541, CAS: 7447-40-7), volume brought to 1 L with MilliQ water, autoclaved) + 35 mg/mL chloramphenicol. Cultures were incubated overnight in 14 mL plastic culture tubes (Falcon, cat# 352059) at 37°C under 225 rpm shaking. Pellets were collected by centrifugation at 3500 rpm for 10 minutes at 4°C in a swinging-bucket centrifuge (Beckman Coulter, Allegra X-12R) and miniprep (Qiagen, cat# 27104). Constructs were confirmed by Sanger sequencing (Quintara Biosciences) by alignment to the template sequence in ClustalOmega.

3.5.0.2 Generation of Plasmid Libraries Four sublibraries were generated to cover the entire mutational space of *E. coli* DHFR: positions 1-40 (sublibrary1, SL1), positions 41-80 (sublibrary2, SL2), positions 81-120 (sublibrary3, SL3), and positions 121-159 (sublibrary4,

SL4). The single point mutant library was performed by multiple parallel inverse RBS reactions to substitute an NNS degenerate codon at every codon in DHFR. RBS primers (**Table 5.3**, page 174) were phosphorylated in a 20 μL reaction with 1 μL T4 polynucleotide kinase and 1x T4 ligase buffer. Inverse PCR reactions were performed as described above, followed by RBS clean-up (Qiagen, cat# 28104). The cleaned RBS reactions were incubated for 4 hours with 1 μL DpnI, 1 μL of T4 ligase, and 3 μL of T4 ligase buffer. RBS reactions were analyzed by gel electrophoresis using a 1% agarose gel in TAE buffer (20 mM acetic acid (Sigma Aldrich, cat#, 695092), 2 mM ethylenediaminetetraacetic acid (EDTA) (ACROS Organics, cat# AC118432500, CAS: 60-00-4), 40 mM Tris, pH 8.5) with 0.01% v/v GelRed (Biotium, cat# 41003), and the product amount was quantified using gel densitometry in the FIJI image processing software package[42]. Samples were pooled stoichiometrically, cleaned once with a gel extraction kit (Qiagen, cat# 28115), and again with a RBS clean-up kit. The pooled and cleaned ligation products were transformed into *E. coli* Top10 cells by electroporation (BioRad GenePulser Xcell, 1 mm path length cuvette (cat# 165-2089), 1.8 kV, time constant ~ 5 ms) using ~ 5 μL to obtain a minimum of 10⁷ transformants as measured by dilution plating on LB-agar plates with 35 $\mu\text{g}/\text{mL}$ chloramphenicol. The transformed cells were rescued in SOB medium (20 g Bacto-tryptone, 5 g Bacto-yeast extract, 0.584 g NaCl, 0.186 g KCl, 800 mL MilliQ water, pH 7.0, volume brought to 1 L with MilliQ water, autoclaved) without antibiotics for 45 minutes at 37°C before culturing overnight in 10 mL SOB medium with 35 $\mu\text{g}/\text{mL}$ chloramphenicol. In the morning, glycerol stocks were made by mixing 500 μL of saturated culture with 500 μL of sterile filtered 50% (v/v) glycerol. 5 mL of the culture was used to miniprep the transformed library with a Qiagen miniprep kit.

Generation of Individual Point Mutant Plasmids Point mutants in all DHFR-containing plasmids were generated via inverse PCR as described above for the generation of SMT205 except that the appropriate antibiotic was matched with the plasmid (**Table 5.2**, page 174). Library primer sequences (**Table 5.3**, page 174) were used except that the “NNS” sequence on the forward primer was replaced with the desired codon.

3.5.0.3 Generation of ER2566 $\Delta folA/\Delta thyA$ –Lon The ER2566 $\Delta folA/\Delta thyA$ –Lon strain was generated as previously described[20] and a gift from Prof. Stephen Benkovic.

3.5.0.4 Plate Reader Assay for *E. coli* Growth Growth rates for the selection strains bearing individual DHFR mutants were measured in 96-well plate growth assays as described for one individual mutant. The SMT205 plasmid was transformed via heat shock into chemically competent ER2566 $\Delta folA/\Delta thyA$ \pm Lon cells and plated on an LB-agar plate with 30 $\mu\text{g}/\text{mL}$ chloramphenicol plus 50 $\mu\text{g}/\text{mL}$ thymidine and incubated overnight at 37°C. On the second day, 2 mL M9 medium (1x M9 salts (BD Difco, cat# 248510), 0.4% glucose w/v (Fisher Chemical, cat# D16, CAS: 50-99-7), 2 mM MgSO₄ (Sigma Aldrich, cat# 63138, CAS:10034-99-8)) with supplements for deficient folate metabolism (50 $\mu\text{g}/\text{mL}$ thymidine (Sigma Aldrich, cat# T1895, CAS: 50-89-5), 22 $\mu\text{g}/\text{mL}$ adenosine (Sigma Aldrich, cat# A9251, CAS: 56-61-7), 1 $\mu\text{g}/\text{mL}$ calcium pantothenate (TCI, cat# P0012, CAS: 137-08-6), 38 $\mu\text{g}/\text{mL}$ glycine (Fisher BioReagents, cat# BP381, CAS: 56-40-6), and 37.25 $\mu\text{g}/\text{mL}$ methionine (Fisher BioReagents, cat# BP388, CAS 63-68-3)) and 30 $\mu\text{g}/\text{mL}$ chloramphenicol in a 14 ml culture tube was inoculated with 5-10 colonies scraped from the plate and incubated at 37°C at 225 rpm shaking for 12-14 hours. Biological replicates were obtained from separate inoculations at this step and run on the same plate. All assays were run from fresh transformations. Then, 20 - 50 μL of the

previous culture was used to inoculate 5 mL of M9 medium (no supplements) with 30 $\mu\text{g}/\text{mL}$ chloramphenicol in a 14 ml culture tube. This fresh culture was incubated for 6 hours at 30°C at 225 rpm shaking. Meanwhile 2 mL of M9 medium with 30 $\mu\text{g}/\text{mL}$ chloramphenicol and a transparent 96-well plate were pre-warmed at 30°C. After the 6-hour incubation, the optical density at 600 nm (OD600) of the culture was measured on a Cary 50 spectrophotometer over a path of 1 cm. This early log-phase culture was diluted to an OD600 = 0.005 in the 2 mL aliquot of warmed M9. 200 μL of the dilute culture was pipetted into a well in the 96-well plate. Technical replicates were obtained by dispensing the same dilute culture into multiple wells. Wells were covered with 50 μL of mineral oil (Sigma Aldrich, cat# M5904, CAS: 8042-47-5) using the reverse pipetting technique. The plate was then incubated for 20-48 hours at 30°C in a Victor X3 multimode plate reader (Perkin Elmer). Every 10 minutes, the plate was shaken for 30 seconds with an orbital diameter of 1.8 mm under the “normal” speed setting. Then, the absorbance at 600 nm (ABS600) was measured for each well. Growth rates were calculated from the slope of $\text{Log}_2(\text{ABS600} - \text{ABS600}_{t=0})$ for ΔABS600 in the range of 0.015 – 0.04 using an in-house python script.

3.5.0.5 Construction of a Turbidostat

3.5.0.6 Deep Mutational Scanning Selection Assay Competitive growth under selection for DHFR activity was performed in a continuous culture turbidostat (gift of Rama Ranganathan) as described below for a single sublibrary. Sublibraries of DHFR single point mutants were transformed via electroporation as described above into electrocompetent ER2566 $\Delta\text{folA}/\Delta\text{thyA} \pm \text{Lon}$ cells using approximately 50 ng of plasmid DNA and 80 μL of competent cells with a transformation efficiency of 1×10^8 cfu/ng (based on testing with 10 ng of pACYC

plasmid DNA). Immediately after electroporation, the cells were rescued with 2 mL of SOB medium with 50 $\mu\text{g}/\text{mL}$ thymidine warmed to 37°C. The rescue culture was incubated at 37°C for 45 minutes at 225 rpm shaking. After the rescue step, 4 μL of the rescue medium (1/500 of the rescue volume) was serially diluted in 10-fold increments. Half the volume of each dilution (1/1000 – 1/10⁷ of the rescue volume) was plated on an LB-agar plate with 30 $\mu\text{g}/\text{mL}$ chloramphenicol plus 50 $\mu\text{g}/\text{mL}$ thymidine and incubated overnight at 37°C. The colonies were counted the following morning to check for a minimum of 1,000x oversampling of the theoretical diversity in the library (~10⁶ transformants for each sublibrary). Meanwhile, the larger portion of the rescue medium was mixed with 4 mL of SOB medium with 45 $\mu\text{g}/\text{mL}$ chloramphenicol (1.5x) plus 50 $\mu\text{g}/\text{mL}$ thymidine warmed to 37°C. This 6 mL culture was incubated for 5-6 hours at 37°C at 225 rpm shaking in a 14 mL culture tube. After incubation, the culture was pelleted by centrifuging for 5 minutes at 3000 rpm at room temperature in a swinging bucket centrifuge. The cells were resuspended in 50 mL of supplemented M9 medium + 30 $\mu\text{g}/\text{mL}$ chloramphenicol and incubated for 12-14 hours at 37°C at 225 rpm shaking in a 250 mL flask. In the morning, 150 mL of supplemented M9 medium + 30 $\mu\text{g}/\text{mL}$ chloramphenicol in a 1 L flask was inoculated with 15 mL of the overnight culture. This pre-culture was incubated at 30°C for 4 hours at 225 rpm shaking. After 4 hours, the pre-culture was centrifuged at 3000 rpm for 5 min at room temperature in a swinging bucket centrifuge, and the OD600 was measured to ensure that the culture did not grow beyond early-mid log phase (OD600 ~0.3). The supernatant was decanted, and the pellet was resuspended in 30 mL of M9 medium. Pelleting and resuspension were repeated for a total of 3 washes to remove the supplemented medium. After 3 washes, the OD600 was measured for the resuspended pellet using a 10-fold dilution to stay in the linear range of the spectrophotometer.

The washed pellet was then transferred to the growth chamber of the turbidostat (a 250 mL pyrex bottle) containing 150 mL of M9 medium with 50 $\mu\text{g}/\text{mL}$ chloramphenicol. Selection experiments were performed with 2 of the 4 sublibraries at a time (two repeats of SL1-SL2 and SL3-SL4, and one repeat of SL1+SL3 and SL2+SL4 for a net of biological triplicates for every codon in the gene), and the resuspended pellet from each library was diluted in the initial culture to an $\text{OD}_{600} = 0.035$. Mixing and oxygenation was provided by sterile filtered air from an aquarium pump. Every 60 seconds, the aquarium pump was stopped, and the optical density of the culture was read by an infrared emitter-receiver pair. The ADC (analog-to-digital converter) of the voltage over the receiver was calibrated against a spectrophotometer to convert the signal into an approximate OD_{600} . The cells were grown at 30°C with an OD_{600} threshold of 0.075. When the OD_{600} of the selection culture exceeded the threshold, the selection culture was diluted to $\text{OD}_{600} \sim 0.065$ with 25 mL of M9 medium with 50 $\mu\text{g}/\text{mL}$ chloramphenicol, and the additional culture volume was driven through a waste line by the positive pressure of the aquarium pump.

At timepoints of $t=0, 2, 4, 6, 8, 12, 16,$ and 18 hours, 6 mL of the selection culture in 2 mL centrifuge tubes was pelleted at 5000 rpm for 5 minutes at 4°C in a microcentrifuge (Eppendorf, 5242R). The supernatant was removed except for the last $\sim 200 \mu\text{L}$, and the tubes were again pelleted at 5000 rpm for 5 minutes at 4°C in a microcentrifuge, and all the supernatant was carefully removed from the pellet. The pellets were stored at -20°C until sequencing.

3.5.0.7 Preliminary Deep Mutational Scanning Assay Using the Turbidostat in the Reynolds Lab at

UT Southwestern Preliminary selection on a library of NNS degenerate codon substitutions at the first 40 positions of DHFR was performed essentially as described above in section "Deep Mutational Scanning Selection Assay" (**Chapter 3.5.0.6**, page 111) but with a few modifications.

3.5.0.8 Amplicon Generation Amplicons were generated by two rounds of PCR. The first round of PCR amplifies a portion of the DHFR gene from the pACYC plasmid containing 2-3 sublibraries. For quality control templates were 1 ng/ μ L plasmid solutions and the amplicons covered SL1-SL2 or SL3-SL4. Round 1 PCR reactions were set up using 1 μ L of template, 1% v/v Q5 hotstart polymerase (NEB, cat# M0493), 1x Q5 Reaction Buffer, 1x Q5 High GC Enhancer, 200 μ M dNTPs, and 500 nM forward and reverse primers. PCR was performed in the following steps: 1) 98°C for 30 seconds, 2) 98°C for 10 seconds, 3) 57°C for 30 seconds, 4) 72°C for 12 seconds, 5) return to step 2 for 16 cycles, 6) 72°C for 2 minutes.

The Round 2 PCR uses primers that attach the Illumina adapters and the i5 (reverse) and i7 (forward) barcodes for sample identification and demultiplexing. Round 2 PCR reactions were set up and run identically to Round 1 reactions except that the template was 1 μ L of Round 1 PCR. Round 2 reactions were analyzed by gel electrophoresis using a 1% TAE-agarose gel in TAE buffer with 0.01% v/v GelRed, and the product amount was quantified using gel densitometry in FIJI. Samples were pooled stoichiometrically and cleaned with a gel extraction kit (Qiagen). Because of the risk of contamination from small primer dimers, gel extraction was performed with very dilute samples. Only 20 μ L of sample was loaded onto a 50 mL TAE-agarose gel (OWL EasyCast, B1A) with 8 of the 10 wells combined into a single well. The pooled amplicons were then cleaned again with a PCR clean-up kit (Zymogen, cat# D4013) to allow for small volume elution. The final amplicon concentration was measured with a NanoDrop One UV spectrophotometer and by Picogreen assay (Thermo Scientific, cat# P11496).

3.5.0.9 Sequencing for Deep Mutational Scanning Experiments Templates for amplicon PCR were prepared from the frozen pellets. The pellets were resuspended in 20 μ L of autoclaved MilliQ water and incubated on ice for 10 minutes. The samples were then centrifuged at 15,000 rpm

for 10 minutes at 4°C in a benchtop microcentrifuge. 1 μ L of the supernatant was used as template in the amplicon generation protocol for sublibraries described above. The amplicons were sequenced on an Illumina NextSeq using a 300-cycle 500/550 high-output kit. Because of the limitations in the number of sequencing cycles on the Illumina NextSeq, the full amplicon was not sequenced for amplicons containing non-adjacent sublibraries (SL1+SL3, and SL2+SL4). Reads were demultiplexed into their respective selection experiment and timepoint using their TruSeq barcodes. Paired end reads were joined using FLASH[43]. For amplicons with adjacent sublibraries (SL1-SL2 and SL3-SL4), the joined reads were kept. For amplicons with distal sublibraries (SL1+SL3 and SL2+SL4), the unjoined reads were kept. Reads from all lanes of the Illumina chip were concatenated and raw counts of DHFR mutants were obtained from these reads.

Reads on the Illumina NextSeq (two-color chemistry, LED optics) generally have lower quality scores than reads from the Illumina MiSeq (four-color chemistry, laser optics). This lower quality leads to a background signal. This background was estimated from a WT sample. The median + 1 standard deviation value of background count was subtracted from every allele and the alleles were translated into the amino acid sequence, combining synonymous sequences. Counts at each timepoint were only reported for an allele if its frequency was above 2.0×10^{-5} .

3.5.0.10 Analysis of Deep Mutational Scanning Data Mutant counts were used to generate selection coefficients on our background-subtracted count files with Enrich2 using unweighted linear regression[35]. The raw Enrich2 values for each unique selection experiment were combined with a post-processing script. Enrich2 does not calculate selection coefficients for mutants that have no counts at a timepoint, so some selection coefficients were recalculated using only the timepoints before the counts for that allele fell below the cutoff frequency of

2×10^{-5} . Individual selection coefficients were evaluated based on two criteria: noise and number of timepoints. Individual selection coefficients were discarded 1) if the standard error from regression was greater than $0.5 + 0.5 \times |\text{selection coefficient}|$ or 2) if there were fewer than 4 timepoints reporting on the mutant. The regression for the fitness value of the mutants from replicate selection experiments to the average values across all experiments was calculated and the fitness values in each replicate were scaled to correct for linear differences in the selection values between replicates. These normalized values were then averaged for the final fitness value. Averaged selection coefficients values were evaluated based on two criteria: the standard deviation of the averaged selection coefficients and the number of replicates. Averaged selection coefficients were discarded 1) if the standard deviation over the normalized replicates was greater than $0.5 + 0.25 \times |\text{selection coefficient}|$ or 2) if there were fewer than 2 replicates. In **Table 5.5** (page 179) the fitness is reported as the mean normalized fitness, the standard error is reported as the combined Enrich2 standard error (from linear regression of timepoints), and the standard deviation is reported as the standard deviation of the biological replicates.

Selection was evaluated by comparing selection coefficients to DHFR velocity from reported Michaelis-Menten kinetics at cytosolic concentrations of DHF[36]. Kinetic values are listed in **Table 3.1** (page 65)

Based on this calibration, differences between selection coefficients below ~ -2.5 were not considered interpretable, and a floor value of -2.5 was applied to all selection coefficients for the purpose of analysis.

For subtraction to calculate Δ selection coefficients, null selection coefficients in +Lon selection were substituted with the lowest measured selection coefficient. Mutations with a null selection coefficient in -Lon selection were assigned a Δ selection coefficient of "No data" (colored black).

Mutations with “No data” value in either selection condition were also assigned a Δ selection coefficient of “No data” here.

For clustering of positions, an in-house Python script was used for K-means clustering of positions into categories based on general mutational response at a position (i.e discarding the amino acid identities of the mutants). Spatial clustering was performed based on selection coefficient with distance between two positions calculated in the following steps: 1) sorting the vectors of selection coefficients for each position, 2) trimming the vectors to match vector lengths after discarding “no data” values, 3) calculating a Δ vector by subtracting the two sorted and trimmed vectors, and finally calculating the distance as the mean of the absolute value of the Δ vector. For the first round, categories were seeded with virtual positions that have prototypical mutational profiles for the 5 categories (Beneficial, Tolerant, Mixed, Restricted, and Intolerant). From this first round, all positions in DHFR were categorized into initial clusters. In subsequent rounds, the virtual positions were removed and candidate positions were compared to the non-self positions populating each cluster. The distance between a candidate position and a cluster of positions is calculated as the average of the distance between the candidate position and the three closest non-self positions in the cluster. Clustering was performed over 10 rounds following the initial seeded round, and convergence was confirmed by observing that 5 repetitions gave identical clusters.

3.5.0.11 Western Blot for Untagged and FLAG-tagged DHFR Constructs LB agar plates with 50 $\mu\text{g}/\text{mL}$ thymidine were streaked from frozen stocks and incubated overnight at 37°C. The next day, 2 mL of LB medium with 50 $\mu\text{g}/\text{mL}$ thymidine were inoculated with a single colony picked from each plate and cultured overnight at 37°C. The next day, 20 μL of overnight culture was used to inoculate two cultures 6 mL of LB medium with 50 $\mu\text{g}/\text{mL}$ thymidine. This fresh culture

was incubated for ~4 hours at 30°C or 42°C at 225 rpm shaking until the OD600 value was between 0.3 and 0.5 on a Cary 50 spectrophotometer over a path of 1 cm. 2 mL of each culture was pelleted by centrifugation at 5000 rpm for 5 minutes at 4°C in a benchtop microfuge, and the supernatant was carefully removed from the pellet, and the pellet was stored at -80°C until the next step.

The pellets were then resuspended with a 1x dilution of 4x Laemmli buffer (BioRad, cat#161-0747) with 10% beta-mercaptoethanol. Volumes for lysis were calculated according to the formula: lysis volume = (volume of culture for pellet) × (OD600 of culture for pellet) × (150 μL Laemmli buffer). The resuspended samples were then boiled at 95°C for 10 minutes and then allowed to cool to room temperature. 2 μL of Precision Plus Dual Color Standards (BioRad, cat# 1610374) 5 μL of each sample was then loaded onto a precast 12-well 4-20% tris-glycine PAGE gel (BioRad, cat# 4561095) and developed at 100 V for 90 minutes. Proteins were then transferred to the blot using a wet transfer protocol. The transfer was run in a buffer of (25 mM Tris base, 200 mM glycine with 20% v/v methanol in water) with a voltage source set to limits of 75 V/150 mA for 150 minutes. Once transfer was complete and confirmed by ladder transfer, the nitrocellulose blot was placed in a clear plastic dish, and 20 mL of TBS-T with 3% dry milk buffer solution was added. The nitrocellulose blot was blocked for 1 hour at room temperature on a rocker. The buffer solution was then decanted and replaced with a fresh 20 mL buffer solution. 2 μL of (1:10,000) of polyclonal rabbit anti-DHFR antibody (gift of Kimberly Reynolds) or monoclonal rabbit anti-FLAG (DYKDDDDK) antibody (Cell Signalling, cat #14793S) was added to the buffer and incubated overnight at 4°C on a rocker. Then, the nitrocellulose blot was washed three times with 20 mL TBS-T buffer, with 5 minutes of incubation at 4°C on a rocker for the last two washes. The TBS-T was decanted

and replaced with 20 mL of TBS-T with 3% dry milk. 2 μ L of (1:10,000) of monoclonal mouse anti-rabbit antibody-HRP conjugate (Millipore, cat# MAB201P) was added to the buffer and incubated for 1 hour at 4°C on a rocker. The Super Signal™ West Pico stable peroxide and enhancer solution were equilibrated at room temperature while protected from light exposure. Then, the nitrocellulose blot was washed three times with 20 mL TBS-T buffer, with 5 minutes of incubation at 4°C on a rocker for the last two washes. 4 mL of the Pico substrate solution was prepared according to instructions, and the nitrocellulose membrane was incubated with the substrate for 5 minutes at room temperature. The nitrocellulose blot was then wrapped in plastic wrap and the chemiluminescence was imaged on a BioRad ChemImager for 6 seconds at normal sensitivity (2 \times 2 downsampling). A second whitefield image was collected for superposing.

3.5.0.12 Purification of His₆-tagged DHFR DHFR variants were expressed from pHis8 plasmids (KR101/SMT301) for nickel affinity purification as described for one DHFR variant. The plasmid bearing the his-tagged DHFR mutant was transformed via heat shock into chemically competent ER2566 Δ *folA*/ Δ *thyA* -Lon cells, then the cells were plated on LB-agar plates containing 50 μ g/mL kanamycin (AMRESCO, cat# 0408, CAS: 25389-94-0, 50 mg/mL in ethanol) and 50 mg/mL thymidine. The plates were incubated overnight at 37°C. The next day 2 mL of LB medium with 50 μ g/mL kanamycin was inoculated with a single colony. This culture was incubated overnight at 37°C at 225 rpm shaking. The next day, 25 mL of TB medium (12 g Bacto-tryptone, 24 g Bacto-yeast extract, 0.4% glycerol v/v (Sigma Aldrich, cat# G7893, CAS: 56-81-5), brought to 900 mL with MilliQ water, autoclaved, cooled, mixed with 100 mL sterile filtered buffered phosphate (0.17 M KH₂PO₄ (Sigma Aldrich, cat# P0662, CAS: 7778-77-0), 0.72 M K₂HPO₄ (Sigma Aldrich, cat# P550, CAS: 16786-57-1))) with 50

$\mu\text{g}/\text{mL}$ kanamycin in a 50 mL conical tube was inoculated with 100 μL of the overnight culture. The culture was grown at 37°C until the OD600 reached 0.5-0.6. Then, the culture was induced with 0.25 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) (Gold Biotechnology, cat# I2481C100, CAS: 367-93-1, 1M in autoclaved water, sterile filtered) and incubated for 18 hours at 18°C at 225 rpm shaking. The cultures were pelleted by centrifugation at 3000 rpm for 5 minutes at 4°C in a swinging-bucket centrifuge, the supernatant was discarded, and the pellet was resuspended by pipetting in 4 mL/g-pellet of B-PER (ThermoScientific, cat# 78266) with 1 mM phenylmethylsulfonyl fluoride (PMSF) (Millipore Sigma, cat# 7110, CAS: 329-98-6, 100 mM in ethanol), 10 $\mu\text{g}/\text{mL}$ leupeptin (VWR Chemicals, cat# J583, CAS: 26305-03-3, 5 mg/mL in water), and 2 $\mu\text{g}/\text{mL}$ pepstatin (VWR Chemicals, cat# J580, CAS: 103476-89-7, 2 mg/mL in water). The lysates were incubated at room temperature for 30 minutes on a rocker and clarified by centrifugation at 3000 rpm for 5 minutes at 4°C in a swinging-bucket centrifuge. The lysate supernatant was then transferred to a fresh 50 mL conical tube and incubated for 30 minutes with 20 μL of NiNTA resin pre-equilibrated in Nickel Binding Buffer (50 mM Tris base (Fisher BioReagents, cat# BP152, CAS: 77-86-1) pH 8.0, 500 mM NaCl, 10 mM imidazole (Fisher Chemical, cat# 03196, CAS: 288-32-4), and then supernatant was removed by pipetting. The resin was washed 3 times for 5 minutes with 1 mL of Nickel Binding Buffer. Then the protein was eluted into 200 μL of Nickel Elution Buffer (100 mM Tris pH 8.0, 1 M NaCl, 400 mM imidazole) and dialyzed against DHFR Storage Buffer (50 mM Tris pH 8.0, 300 mM NaCl, 1% glycerol v/v) in 3000 Da MW cut-off Slidalyzer dialysis cups (Thermo Scientific, cat# 88401) at 4°C. After 4 changes of dialysis buffer over 24 hours, the protein was aliquoted, flash frozen in liquid nitrogen, and stored at -80°C. Proteins were purified to ~90-95% purity as judged from PAGE gel analysis.

3.5.0.13 In Vitro Assay for DHFR Velocity and Michaelis-Menten Kinetics *In vitro* measurements of DHFR velocity were carried out by monitoring the change in UV absorbance. For each mutant screened, a purified enzyme aliquot was thawed and centrifuged at 15,000 rpm for 5 minutes at 4°C in a benchtop microcentrifuge. The soluble enzyme was then transferred to a fresh tube, and the concentration was measured by UV absorption on a Nanodrop. Molar concentration of DHFR was calculated using an extinction coefficient of 33585 M⁻¹ cm⁻¹ at 280 nm for all variants with the following exceptions: 28085 (W30F/M, W47L/M), 35075 (M42Y, R98Y, L165Y), or 39085 (Q102W) M⁻¹ cm⁻¹. The enzyme was diluted to 555 nM in DHFR storage buffer. A pre-reaction mixture was prepared in MTEN buffer (5 mM 2-ethanesulfonic acid (MES) (Sigma Aldrich, cat# 69889, CAS: 145224-94-8), 25 mM ethanolamine (Sigma Aldrich, cat# E6133, CAS: 2002-24-6), 100 mM NaCl, 25 mM Tris base, pH to 7.0) with 55.5 nM enzyme, 111 μM NADPH (Sigma Aldrich, cat# N7505, CAS: 2646-71-1) and 5 mM dithiothreitol (DTT) (GoldBio, cat# DTT25, CAS: 27565-41-9, 1M in water, sterile filtered). The pre-reaction mixture and a micro quartz cuvette (Fisher Scientific, cat# 14-958-103, 10 mm path length, 2 mm window width) were pre-incubated at 30°C. The reaction was started by adding 20 μl of 500 μM DHF (Sigma Aldrich, cat# D7006, CAS: 4033-27-6) in MTEN with 5 mM DTT to 180 μL of pre-reaction mixture. The substrate solution was made fresh from a sealed ampule on the day of the experiment. The reaction was briefly mixed by pipetting and then the reaction was monitored by reading the absorbance at 340 nm with an interval of 0.1 seconds in a Cary 50 spectrophotometer with the Peltier temperature set to 30°C. The reactions were allowed to run to completion to establish the baseline, which was subtracted from the absorbance values. The real-time concentration of DHF was calculated by dividing the normalized absorbance values by the decrease in absorbance at 340 nm for the reaction, 0.0132 μM⁻¹cm⁻¹, the velocity of the reaction was calculated as the

slope of linear regression to a 30 second window with a mean DHF concentration equal to 5, 10, 20, or 30 μM . Final velocities were normalized to enzyme concentration.

Michaelis-Menten kinetics were performed as described above using 1-5 μM DHFR for concentrations of DHFR from 0.5-100 μM . Initial velocities were estimated from linear regression to the absorbance divided by the decrease in absorbance at 340 nm for the reaction, and then they were fit to the Michaelis-Menten equation using the non-linear least squares method in R.

3.5.0.14 Determining DHFR Activity and Concentration in Cell Lysates The cellular activity of DHFR was measured in cell lysates, and then used to calculate DHFR cellular abundance using a method adapted from previous studies[29, 30]. For each characterized DHFR variant, a plasmid (WT DHFR in plasmids SMT102, SMT201, SMT202 and SMT205 with modified promoters and RBSs or DHFR single point mutants in the final selection plasmid SMT205, see **Table 5.2**, page 174) was transformed via heat shock into chemically competent ER2566 $\Delta folA/\Delta thyA$ \pm Lon cells, which were plated on an LB-agar plate with 30 $\mu\text{g}/\text{mL}$ chloramphenicol plus 50 $\mu\text{g}/\text{mL}$ thymidine and incubated overnight at 37°C. On the second day, 2 mL M9 medium with supplements for deficient folate metabolism (50 $\mu\text{g}/\text{mL}$ thymidine, 22 $\mu\text{g}/\text{mL}$ adenosine, 1 $\mu\text{g}/\text{mL}$ calcium pantothenate, 38 $\mu\text{g}/\text{mL}$ glycine, and 37.25 $\mu\text{g}/\text{mL}$ methionine) and 30 $\mu\text{g}/\text{mL}$ chloramphenicol in a 14 ml culture tube was inoculated with a single colony scraped from the plate and incubated at 37°C at 225 rpm shaking for 12-14 hours. Three biological replicates were obtained from separate single colonies at this step, and all biological replicates were processed in parallel for subsequent steps. All assays were run from fresh transformations. 20-50 μL of the previous culture were used to inoculate 20 mL of M9 medium (no supplements) with 30 $\mu\text{g}/\text{mL}$ chloramphenicol in a 50 ml conical tube. This fresh culture was incubated for 12-18

hours at 30°C at 225 rpm shaking until the OD600 value was between 0.3 and 0.5 on a Cary 50 spectrophotometer over a path of 1 cm. The cultures were pelleted by centrifugation at 3000 rpm for 5 minutes at 4°C in a swinging-bucket centrifuge, the supernatant was discarded, and the pellet was thoroughly resuspended in 1.1 mL of M9 medium. 1 mL of the resuspension was transferred to a 1.5 mL Eppendorf tube, and the sample was pelleted at 5000 rpm for 5 minutes at 4°C in a microcentrifuge. The supernatant was carefully removed from the pellet, and the pellet was stored at -80°C until the next step. The remained 100 μ L of resuspended pellet was mixed with 900 μ L and the OD600 value was measured for each pellet to determine the number of cells in the pellet, with a conversion factor of 8×10^8 cells/mL at OD600 = 1.0. Pellets for positive (ER2566) and negative (ER2566 Δ *folA*/ Δ *thyA* \pm Lon) control samples were collected in a similar fashion, except that antibiotics were not used and initial plates were streaked from glycerol stocks. Additionally, the M9 medium for ER2566 Δ *folA*/ Δ *thyA* \pm Lon contained folate supplements in every step.

Cell pellets were lysed in B-PER with 1 mM PMSF, 10 μ g/mL leupeptin, and 2 μ g/mL pepstatin. Volumes for lysis were calculated to have consistent lysate concentration according to the formula: lysis volume = (volume of resuspended pellet) \times (OD600 of resuspended pellet) \times (30 μ L BPER lysis buffer). Pellets were resuspended by pipetting in the calculated volume, and the lysates were incubated at room temperature for 30 minutes on a rocker. The lysates were then clarified by centrifuged at 15,000 rpm for 5 minutes at 4°C in a benchtop microcentrifuge. Lysates were kept on ice while the reactions were prepared.

Measurements of DHFR activity in lysates were carried out by monitoring the change in UV absorbance in a BioTek Synergy H1 multimode plate reader. A 180 μ L pre-reaction mixture was prepared with MTEN buffer (5 mM MES, 25 mM ethanolamine, 100 mM NaCl, 25 mM Tris

base, pH to 7.0), 111 μM NADPH, 5 mM DTT, and containing 20 μL lysate. The pre-reaction mixtures in a UV transparent 96-well plate (Grenier Bio-One, cat# 655809) were pre-incubated at 30°C for 10 minutes. The substrate solution of 500 μM DHF in MTEN with 5 mM DTT was made freshly from a sealed ampule of DHF on the day of the experiment. The reaction was started by automatic injection of 20 μl of 500 μM DHF in MTEN with 5 mM DTT into each well with pre-reaction mixture. The plate was then orbital shaken for 1 minute at 365 rpm with a 2 mm amplitude. The reaction was briefly mixed by pipetting and then the reaction was monitored by reading the absorbance at 340 nm with an interval 1 minute for 2 hours while incubating at 30°C. To establish a baseline for accurate calculation of DHF concentration in each well, 50 μL of 1 μL WT DHFR in DHFR storage buffer was injected into each well, the plate was then orbital shaken for 1 minute at 365 rpm with a 2 mm amplitude, and the reactions were allowed to run to completion over 10 minutes, before a final reading of absorbance at 340 nm was taken. In processing, this baseline value was subtracted from the absorbance values for each well. The real-time concentration of DHF was calculated by dividing the normalized absorbance values by the decrease in absorbance at 340 nm for the reaction, $0.0132 \mu\text{M}^{-1}\text{cm}^{-1}$, times a correction factor of 1.5 for calibration between the plate reader and the absorbance at 340 nm using a Cary 50 spectrophotometer with a 1 cm pathlength quartz cuvette. The velocity of the reaction was calculated as the slope of linear regression for DHF concentration as a function of time over a window of DHF concentration from 20 to 30 μM . The mean slope of the negative control wells (untransformed ER2566 $\Delta\text{folA}/\Delta\text{thyA} \pm \text{Lon}$) was subtracted from all wells as a baseline. The linear regression of *in vitro* DHFR reactions using purified enzyme over the same window of DHF concentration from 20 to 30 μM was calculated from measurements described above (**Chapter 3.5.0.13**, page 121, **Table 5.6**, page 243, and the DHFR abundance in each well was calculated

from the ratio of activity_{lysate}/velocity_{purifiedenzyme}. The number of DHFR molecules per cell was then calculated by dividing the total number of DHFR molecules in each 200 μ L of reaction by the number of cells in 20 μ L of lysate based on the OD600 measurements.

3.5.0.15 CD Spectroscopy Samples for circular dichroism (CD) spectroscopy were prepared at a concentration of 10 μ M in a buffer of 150 mM NaCl and 50 mM Tris, pH 8.0. CD spectra acquisition and thermal denaturation was carried out in a Jasco J-715 CD spectrometer using a cuvette with a 2 mm pathlength (Starna Cell Inc., cat# 21-Q-2). For each DHFR variant, a pre-denaturation spectra was recorded between 207 nm and 280 nm where the high tension voltage was below 600 V. Thermal denaturation data was collected at 225 nm with a bandwidth of 2 nm, a response time of 8s, and a resolution of 0.1°C during heating at a rate of 1°C/min. When the curve flattened, the sample was removed from the CD spectrometer and the system was returned to 30°C. The sample was returned to the chamber and allowed to equilibrate for 10 minutes. A post-denaturation spectrum was recorded after equilibration. Between samples, the cuvette was cleaned with sonication in Hellmanex III (Hellma, cat# 2805939) followed by washing with 50% concentrated nitric acid. Thermal denaturation was found to be only partially reversible based on comparisons of spectra recorded before and after denaturation. Thermal denaturation curves were fit to a sigmoidal model for the calculation of an approximate apparent T_m for all mutants as previously reported[44].

3.5.0.16 Structural Representation of DHFR All images of the DHFR structure were prepared with UCSF Chimera, and volumetric representations were prepared using the MSMS package[45]. 4 crystal structures of DHFR (1RX1, 3QL3, 1RX4, and 1RX5) representing different states in DHFR's catalytic cycle were downloaded from PDB_REDO[46].

References

- [1] F. H. Arnold. Directed evolution: Bringing new chemistry to life. *Angew Chem Int Ed Engl*, 57(16):4143–4148, 2018.
- [2] P. Bandaru, N. H. Shah, M. Bhattacharyya, J. P. Barton, Y. Kondo, J. C. Cofsky, C. L. Gee, A. K. Chakraborty, T. Kortemme, R. Ranganathan, and J. Kuriyan. Deconstruction of the ras switching cycle through saturation mutagenesis. *Elife*, 6, 2017.
- [3] E. Firnberg, J. W. Labonte, J. J. Gray, and M. Ostermeier. A comprehensive, high-resolution map of a gene’s fitness landscape. *Mol Biol Evol*, 31(6):1581–92, 2014.
- [4] M. A. Stiffler, D. R. Hekstra, and R. Ranganathan. Evolvability as a function of purifying selection in tem-1 beta-lactamase. *Cell*, 160(5):882–892, 2015.
- [5] A. D. Garst, M. C. Bassalo, G. Pines, S. A. Lynch, A. L. Halweg-Edwards, R. Liu, L. Liang, Z. Wang, R. Zeitoun, W. G. Alexander, and R. T. Gill. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat Biotechnol*, 35(1):48–55, 2017.
- [6] C. L. Araya, D. M. Fowler, W. Chen, I. Muniez, J. W. Kelly, and S. Fields. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc Natl Acad Sci U S A*, 109(42):16858–63, 2012.
- [7] Jr. McLaughlin, R. N., F. J. Poelwijk, A. Raman, W. S. Gosal, and R. Ranganathan. The spatial architecture of protein function and adaptation. *Nature*, 491(7422):138–42, 2012.
- [8] A. S. Raman, K. I. White, and R. Ranganathan. Origins of allostery and evolvability in proteins: A case study. *Cell*, 166(2):468–480, 2016.

- [9] T. A. Whitehead, A. Chevalier, Y. Song, C. Dreyfus, S. J. Fleishman, C. De Mattos, C. A. Myers, H. Kamisetty, P. Blair, I. A. Wilson, and D. Baker. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol*, 30(6):543–8, 2012.
- [10] P. A. Romero, T. M. Tran, and A. R. Abate. Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc Natl Acad Sci U S A*, 112(23):7159–64, 2015.
- [11] E. E. Wrenbeck, L. R. Azouz, and T. A. Whitehead. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat Commun*, 8:15695, 2017.
- [12] C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard, and D. Baker. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*, 501(7466):212–216, 2013.
- [13] J. R. Klesmith, J. P. Bacik, E. E. Wrenbeck, R. Michalczyk, and T. A. Whitehead. Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proc Natl Acad Sci U S A*, 114(9):2265–2270, 2017.
- [14] B. Steinberg and M. Ostermeier. Shifting fitness and epistatic landscapes reflect trade-offs along an evolutionary pathway. *J Mol Biol*, 428(13):2730–43, 2016.
- [15] R. T. Hietpas, C. Bank, J. D. Jensen, and D. N. A. Bolon. Shifting fitness landscapes in response to altered environments. *Evolution*, 67(12):3512–22, 2013.
- [16] L. Jiang, P. Mishra, R. T. Hietpas, K. B. Zeldovich, and D. N. Bolon. Latent effects of hsp90 mutants revealed at reduced expression levels. *PLoS Genet*, 9(6):e1003600, 2013.

- [17] J. I. Boucher, D. N. Bolon, and D. S. Tawfik. Quantifying and understanding the fitness effects of protein mutations: Laboratory versus nature. *Protein Sci*, 25(7):1219–26, 2016.
- [18] B. P. Roscoe, K. M. Thayer, K. B. Zeldovich, D. Fushman, and D. N. Bolon. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J Mol Biol*, 425(8):1363–77, 2013.
- [19] Sebastian Matuszewski, Marcel E. Hildebrandt, Ana-Hermina Ghenu, Jeffrey D. Jensen, and Claudia Bank. A statistical guide to the design of deep mutational scanning experiments. *Genetics*, 204:77–87, 2016.
- [20] K. A. Reynolds, R. N. McLaughlin, and R. Ranganathan. Hot spots for allosteric regulation on protein surfaces. *Cell*, 147(7):1564–75, 2011.
- [21] R. Hietpas, B. Roscoe, L. Jiang, and D. N. Bolon. Fitness analyses of all possible point mutations for regions of genes in yeast. *Nat Protoc*, 7(7):1382–96, 2012.
- [22] S. Ovchinnikov, H. Kamisetty, and D. Baker. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife*, 3:e02030, 2014.
- [23] M. Iwakura, K. Maki, H. Takahashi, T. Takenawa, A. Yokota, K. Katayanagi, T. Kamiyama, and K. Gekko. Evolutional design of a hyperactive cysteine- and methionine-free mutant of escherichia coli dihydrofolate reductase. *J Biol Chem*, 281(19):13234–46, 2006.
- [24] C. T. Liu, P. Hanoian, J. B. French, T. H. Pringle, S. Hammes-Schiffer, and S. J. Benkovic. Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proc Natl Acad Sci U S A*, 110(25):10159–64, 2013.

- [25] Z. Huang, C. R. Wagner, and S. J. Benkovic. Nonadditivity of mutational effects at the folate binding site of escherichia coli dihydrofolate reductase. *Biochemistry*, 33(38):11576–85, 1994.
- [26] C. A. Fierke and S. J. Benkovic. Probing the functional role of threonine-113 of escherichia coli dihydrofolate reductase for its effect on turnover efficiency, catalysis, and binding. *Biochemistry*, 28(2):478–86, 1989.
- [27] A. Espah Borujeni, A. S. Channarasappa, and H. M. Salis. Translation rate is controlled by coupled trade-offs between site accessibility, selective rna unfolding and sliding at upstream standby sites. *Nucleic Acids Res*, 42(4):2646–59, 2014.
- [28] H. M. Salis, E. A. Mirsky, and C. A. Voigt. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol*, 27(10):946–50, 2009.
- [29] J. V. Rodrigues, S. Bershtein, A. Li, E. R. Lozovsky, D. L. Hartl, and E. I. Shakhnovich. Biophysical principles predict fitness landscapes of drug resistance. *Proc Natl Acad Sci U S A*, 113(11):E1470–8, 2016.
- [30] R. F. Guerrero, S. V. Scarpino, J. V. Rodrigues, D. L. Hartl, and C. B. Ogbunugafor. Proteostasis environment shapes higher-order epistasis operating on antibiotic resistance. *Genetics*, 212(2):565–575, 2019.
- [31] A. Schmidt, K. Kochanowski, S. Vedelaar, E. Ahrne, B. Volkmer, L. Callipo, K. Knoop, M. Bauer, R. Aebersold, and M. Heinemann. The quantitative and condition-dependent escherichia coli proteome. *Nat Biotechnol*, 34(1):104–10, 2016.
- [32] Y. K. Kwon, W. Lu, E. Melamud, N. Khanam, A. Bogner, and J. D. Rabinowitz. A domino effect in antifolate drug action in escherichia coli. *Nat Chem Biol*, 4(10):602–8, 2008.

- [33] S. Bershtein, A. W. Serohijos, S. Bhattacharyya, M. Manhart, J. M. Choi, W. Mu, J. Zhou, and E. I. Shakhnovich. Protein homeostasis imposes a barrier on functional integration of horizontally transferred genes in bacteria. *PLoS Genet*, 11(10):e1005612, 2015.
- [34] M. W. Levine Ra Fau Taylor and M. W. Taylor. Mechanism of adenine toxicity in escherichia coli. (0021-9193 (Print)).
- [35] A. F. Rubin, H. Gelman, N. Lucas, S. M. Bajjalieh, A. T. Papenfuss, T. P. Speed, and D. M. Fowler. A statistical framework for analyzing deep mutational scanning data. *Genome Biol*, 18(1):150, 2017.
- [36] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nat Chem Biol*, 5(8):593–9, 2009.
- [37] G. P. Miller and S. J. Benkovic. Strength of an interloop hydrogen bond determines the kinetic pathway in catalysis by escherichia coli dihydrofolate reductase. *Biochemistry*, 37(18):6336–42, 1998.
- [38] G. P. Miller, D. C. Wahnnon, and S. J. Benkovic. Interloop contacts modulate ligand cycling during catalysis by escherichia coli dihydrofolate reductase. *Biochemistry*, 40(4):867–75, 2001.
- [39] Y. Cho, X. Zhang, K. F. Pobre, Y. Liu, D. L. Powers, J. W. Kelly, L. M. Gierasch, and E. T. Powers. Individual and collective contributions of chaperoning and degradation to protein homeostasis in e. coli. *Cell Rep*, 11(2):321–33, 2015.
- [40] S. Bershtein, J. M. Choi, S. Bhattacharyya, B. Budnik, and E. Shakhnovich. Systems-level

response to point mutations in a core metabolic enzyme modulates genotype-phenotype relationship. *Cell Rep*, 11(4):645–56, 2015.

- [41] F. J. Poelwijk, M. Socolich, and R. Ranganathan. Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nat Commun*, 10(1):4213, 2019.
- [42] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: an open-source platform for biological-image analysis. *Nat Methods*, 9(7):676–82, 2012.
- [43] T. Magoc and S. L. Salzberg. Flash: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21):2957–63, 2011.
- [44] C. A. Smith, C. A. Shi, M. K. Chroust, T. E. Bliska, M. J. S. Kelly, M. P. Jacobson, and T. Kortemme. Design of a phosphorylatable pdz domain with peptide-specific affinity changes. *Structure*, 21(1):54–64, 2013.
- [45] M. F. Sanner, A. J. Olson, and J. C. Spohner. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, 38(3):305–20, 1996.
- [46] R. P. Joosten, F. Long, G. N. Murshudov, and A. Perrakis. The pdb redo server for macromolecular structure model optimization. *IUCrJ*, 1(Pt 4):213–20, 2014.

4 Altered Expression of Lon Protease Reshapes the Mutational Landscape of a Model Enzyme

4.1 Introduction

The tenet of protein sequence-structure-function relationships in biophysics is clearly useful for understanding mutational impacts on protein function, but it is also an incomplete understanding of protein function in native cellular environments. In **Chapter 3** (page 54), we attempted to examine how the mutational landscape of DHFR is shaped by intrinsic structural properties and constraints on function. In optimizing our selection assay to map the mutational landscape of DHFR, we observed that changes to extrinsic properties (e.g. DHFR expression level) and environmental factors (e.g. TYMS expression, media conditions) can tune the selection pressure and reshape the mutational landscape. Our selection under stringent conditions in minimal medium and with DHFR expressed at ~10% of the endogenous level revealed advantageous mutations not observed at higher expression levels. Thus, our experiments report on sequence-structure-function relationships that are further convoluted by environment. In **Chapter 3**, we initially considered only sequence-structure-function relationships. Under this view, the mutational landscape is something fixed and intrinsic, shaped by thermodynamic and structural constraints on a molecular function. Now, by examining how environmental factors shape the mutational impacts, we consider sequence-structure-function-environment relationships. With the consideration of environmental pressures, we can view the mutational landscape as something that can be purposefully modulated.

Many experiments such as genetic screens and reverse genetic screens have examined how

changes to the cellular environment change phenotypes. These studies are powerful, but they often examine only a small number of variants that survive a selection experiment or they examine more coarse-grained changes such as gene deletions. Even careful experiments to follow-up on large library screens are generally limited to a handful of mutations due to low experimental throughput. Therefore, we sought to advance our understanding of how changes to the cellular environment reshape sequence-function relationships by quantifying a large mutational landscape (thousands of mutations) before and after modulating a key cellular factor.

In this chapter, we repeated our optimized DMS assay for DHFR from **Chapter 3** with a perturbation to a key component of protein homeostasis in the selection strain. We then examined how this perturbation impacted the population of advantageous mutations from **Chapter 3** by examining changes in the distribution of selection coefficient, changes in the general mutational response at a position, and changes in expression level. Finally, we examined how well the *in vitro* measurements recapitulated the measurements from the *in vivo* selection assay.

4.2 Background

The protein-structure-function paradigm is overly simplistic because it ignores the complex environment that proteins fold in and the competing kinetic processes along the folding pathway. From the synthesis of the nascent peptide chain, the alternative pathways of unfolding, misfolding, aggregation, and degradation can shunt expressed protein away from the native fold[1, 2].

4.2.1 Protein Homeostasis in Cells

Cells have evolved complex protein quality control machinery to recycle non-native proteins by refolding or degradation. Accordingly, the two major components of protein quality control

machinery are chaperones and proteases. Chaperones intervene at every possible stage of a protein's life — from nascent chain to misfolding under stress conditions — to restore the native fold of a protein. Certain chaperones can act passively (e.g. DnaJK, Trigger Factor) or utilize ATP as they unfold and refold proteins (e.g. Hsp90, GroEL, MmCpn). Proteases cleave peptide chains, and proteases with a role in degradation (e.g. Lon, ClpXP) recycle the constituent amino acids of a protein. This can also occur at many different stages in the lifetime of a protein, including degradation during stalled translation and degradation of protein aggregates.

4.2.2 Interplay between Protein Homeostasis in Cells and Protein Function

Seminal studies have examined the role of protein chaperones on mutational impacts observed at the cellular level. As one example, Hsp90 has been shown to buffer mutational impacts in a wide range of eukaryotic organisms, serving as a buffer of mutational impact and a capacitor for phenotypic variation[3–6]. Additionally, GroEL overexpression has been demonstrated to dampen the stability penalty against mutations during directed evolution [7, 8]. Using a model selection experiment, sequences with ~10-fold higher catalytic efficiencies were obtained with GroEL overexpression compared to selection with basal GroEL expression. Furthermore, endogenous GroEL expression levels have been shown to be linked to the rate of genetic diversification[9]. Notably, none of the studies we identified focused on the role of degradation machinery with the exception of studies interrogating the role of Lon protease in changing tolerance to genetic diversification [10–12], and no study examined if the perturbations to protein quality control cause identifiable patterns within the changes in an exhaustively mapped mutational landscape.

4.2.3 Lon Protease

Lon protease is a hexameric AAA+ ATP-dependent serine-protease that is responsible for degrading misfolded proteins. Lon is broadly conserved in all domains of life and is critical for mitochondrial function in higher organisms. In *E. coli*, Lon has a specific role in the cell cycle for degrading the checkpoint inhibitor SulaA[13], but it is not generally believed to behave in a client specific manner[14, 15]. Rather, Lon has been demonstrated to bind and degrade proteins with exposed hydrophobic patches. Lon can be targeted with recombinant fusion tags of hydrophobic amino acids or by specific regulatory tags in some organisms[14], such as the *ssrA* tag that is fused to peptides stalled in translation[16]. Moreover, Lon had previously been implicated in degrading unstable DHFR variants in *E. coli*[1, 17], and deleting Lon in an MG1655 strain of *E. coli* masked the deleterious impact of 2 destabilizing mutations out of a panel of 21 mutants tested in growth experiments[17]. As with our competitive growth experiments (**Chapter 3.5.0.4**, page 110), these experiments were performed in minimal medium at 30°C.

4.2.3.1 B-strain *E. coli* Lon was identified in genetic screens of K12 *E. coli* because Lon depletion results in the inhibition of septation and cell division after exposure to UV radiation. When Lon is not present to degrade the checkpoint inhibitor SulaA after exposure, exposed cells remain in an extended quiescent state[13]. This molecular mechanism was ultimately found to be the driver for the UV sensitivity of B strain *E. coli*, which are naturally deficient in Lon.[18] Furthermore, due to this deficiency in protein degradation machinery, B strain *E. coli* have been used for heterologous expression (e.g. BL21 cells), for many directed evolution experiments, and for the adaptation experiments of the *E. coli* Long Term Evolution Experiment (LTEE)[19–22]. Furthermore, our selection strain[23] is a B strain of *E. coli* that is naturally deficient in Lon protease [14, 15, 24] due to an insertion of IS186 in the *lon* promoter region[18].

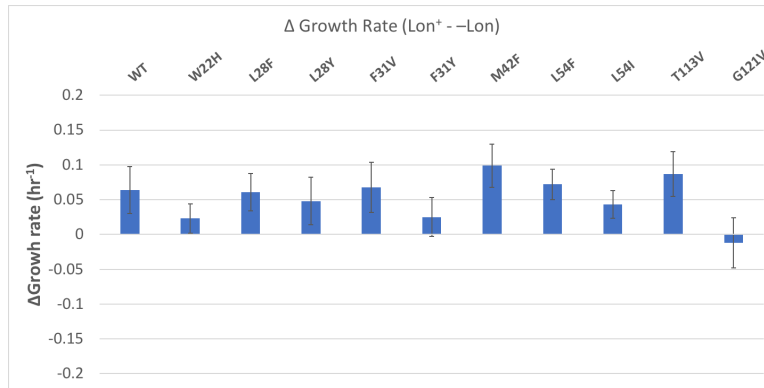


Figure 4.1: Change in DHFR mutant dependent growth rates in the selection strain and a selection strain with a restored endogenous Lon promoter. The difference in growth rates from plate reader experiments (see **Methods, Chapter 3.5.0.4**, page 110) for ER2566 $\Delta folA/\Delta thyA$ with and without the endogenous Lon promoter over a panel of 11 DHFR variant (labels, top) are plotted as bars. The error bars represent the sum of errors for growth rates in the two cell lines, where error is calculated as the standard deviation over 8 independent experiments.

4.3 Results

To understand the origins of the counter-intuitive preference for mutation that we observed in DMS on DHFR (**Chapter 3.3.6.1**, page 86), we identified Lon as a cellular factor that potentially affects our mutational landscape for DHFR. To test the hypothesis that the permissive DHFR mutational landscape for DHFR can be made more restrictive with increased penalties against destabilizing mutations through expression of Lon protease, we performed a DMS over all single point mutants of DHFR in a variant of our selection strain that constitutively expresses Lon.

4.3.1 Generating a Lon-expressing Selection Strain

To test the impact of Lon expression on the mutational landscape of DHFR, we aimed to generate a variant of the ER2566 $\Delta folA/\Delta thyA$ selection strain that had expression levels of Lon that were similar to those for Lon in K12 *E. coli*. To change the expression level of Lon, we aimed to reverse the IS186 insertion in the Lon promoter and restore the native Lon promoter.

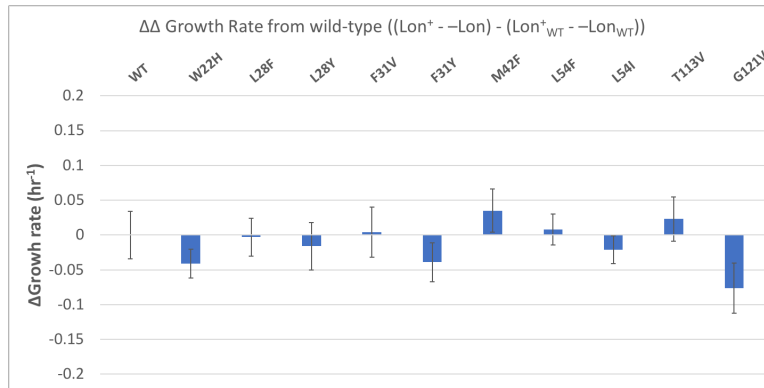


Figure 4.2: Change relative to WT in DHFR mutant-dependent growth rates in the selection strain and a selection strain with a restored endogenous Lon promoter. The difference in growth rates from plate reader experiments (see **Figure 4.1**, page 136) for ER2566 $\Delta folA/\Delta thyA$ with and without the endogenous Lon promoter over a panel of 11 DHFR variant (labels, top) and normalized to the change in growth rate for WT DHFR are plotted as bars. The error bars represent the sum of errors for growth rates in the two cell lines, where error is calculated as the standard deviation over 8 independent experiments.

4.3.1.1 Reversing the IS186 Insertion by Restoring the Native Promoter Using Lambda Red recombination (see **Methods, Chapter 4.5.0.1**, page 163) we first generated a counter-selectable ER2566 $\Delta folA/\Delta thyA$ strain with a Kan-SacB selection/counter-selection cassette replacing the region upstream of Lon that contains the promoter. In a second round of Lambda Red recombination, we replaced the counter selectable marker with the native genome sequence for the region upstream of Lon protease.

In both selection and counter selection steps, colonies were first screened on the selection plates and then checked with colony PCR using primers that bind 50 bp upstream and downstream of the re-inserted region. When amplified from the *E. coli* circular chromosome with primers for the C-terminus of the ClpXP gene and the N-terminus of Lon, the Kan-SacB cassette is ~3 kb, the B-strain Lon promoter with the IS186 insertion is ~1.5 kb, and the K12-strain Lon promoter is 250 bp. These size differences allows all three constructs to be identified on a gel. PCR products from positive hits were then sequenced using the same forward and reverse primers from amplification to check for scarring at the recombination sites.

We then tested our re-engineered selection strain for expression of Lon protease. In Western blots of with anti-Lon antibodies, we did not see differences in Lon expression between our selection strain, the engineered +Lon strain, and a K-12 positive control (unpublished data). These results were further compounded by problems with a poor-affinity primary antibody. Importantly, plate reader growth experiments (see **Methods, Chapter 3.5.0.4**, page 110) with 11 DHFR variants did not show a difference in growth rates between the two cell lines (**Figure 4.1**, page 136). Compared to the change in the growth rate with the new cellular background for WT DHFR, the change in the growth rate for the 11 DHFR variants was within error of 0.0 for all but 3 (**Figure 4.2**, page 137). From this, we concluded that restoring the endogenous promoter would not measurably change selection pressure in the selection assay.

4.3.1.2 Reversing the IS186 Insertion with a Constitutive Promoter Because Lon is reported to be a heat-shock protein[25], we hypothesized that its expression under the endogenous promoter would be repressed at the selection temperature of 30°C. To generate selection conditions that would measure the impact of Lon expression on the mutational landscape, we aimed to engineer a ER2566 $\Delta folA/\Delta thyA$ selection strain variant with Lon constitutively expressed at a level that is consistent with heat-shock in *E. coli*. Using Lambda Red recombination (see **Methods, Chapter 4.5.0.1**, page 163), we generated two ER2566 $\Delta folA/\Delta thyA$ selection strains with Lon expressed under the Anderson Set consensus promoter[26]. One construct has the native Lon RBS and one has an engineered RBS predicted by the Salis Lab RBS Calculator[27, 28]. We performed an anti-Lon western blot on lysate from multiple cell lines grown at 30°C and 42°C (**Figure 4.3**, page 139). Although the polyclonal antibody showed multiple bands, we observed a band at the expected molecular weight for the Lon monomer, and the additional bands could be labeling of degradation products. Overall, we saw a marked increase in Lon expression relative to B-strain

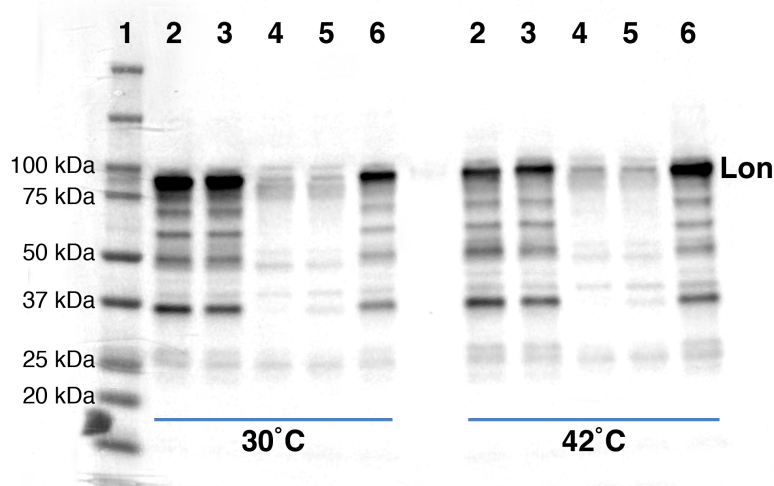


Figure 4.3: Expression of Lon under a constitutive promoter in the re-engineered selection strain. The expression level of Lon was observed by anti-Lon western blot for multiple strains of *E. coli*. Lane identities are marked by a growth temperature (bottom) and a number (top): **1**) Ladder, **2**) engineered selection strain with constitutive Lon promoter, **3**) engineered selection strain with constitutive Lon promoter and optimized RBS (ER2566 $\Delta folA/\Delta thyA$ +Lon), **4**) ER2566 $\Delta folA/\Delta thyA$ -Lon, **5**) B-strain BL21, **6**) K12 strain Top 10, see **Table 5.1**, page 174. Molecular weights to the ladder markers (labels left) and molecular identities (labels right) are indicated in label text.

negative controls for our re-engineered +Lon selection strain and for a K12 strain positive control.

We observed that the expression of Lon at 30°C with the constitutive promoter was qualitatively similar to endogenous Lon expression in a K12 strain under heat shock at 42°C. Furthermore, plate reader growth experiments showed a significant deleterious growth impact from Lon expression on a series of DHFR point mutants with multiple advantageous mutations (**Figure 4.4**, page 140). The impact of Lon on growth rates was generally of a greater magnitude with point mutants of DHFR as compared to WT (**Figure 4.5**, page 140). From this, we concluded that selection in our engineered +Lon selection strain would display measurable changes on the mutational landscape resulting from Lon expression.

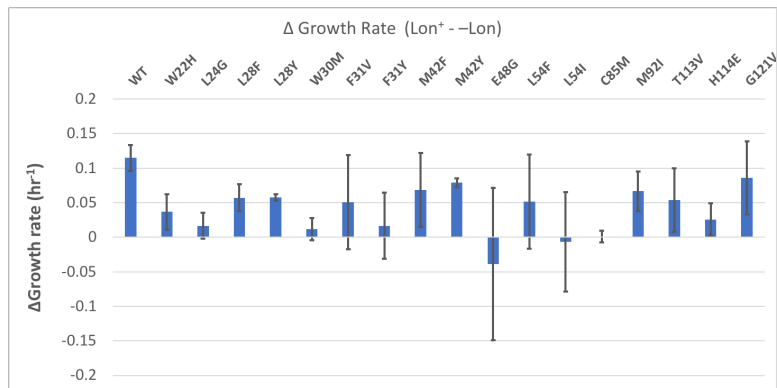


Figure 4.4: Change in DHFR mutant-dependent growth rates in the selection strain \pm Lon. The difference in growth rates from plate reader experiments (see **Methods, Chapter 3.5.0.4**, page 110) for ER2566 $\Delta folA/\Delta thyA \pm$ Lon over a panel of 18 DHFR variant (labels, top) are plotted as bars. The error bars represent the sum of errors for growth rates in the two cell lines, where error is calculated as the standard deviation over 8 independent experiments.

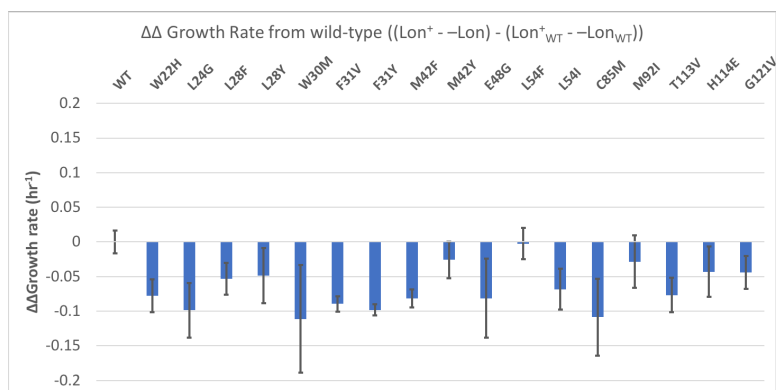


Figure 4.5: Change relative to WT in DHFR mutant-dependent growth rates in the selection strain \pm Lon. The difference in growth rates from plate reader experiments (see **Figure 4.1**, page 136) for ER2566 $\Delta folA/\Delta thyA \pm$ Lon over a panel of 18 DHFR variant (labels, top) and normalized to the change in growth rate for WT DHFR are plotted as bars. The error bars represent the sum of errors for growth rates in the two cell lines, where error is calculated as the standard deviation over 8 independent experiments.

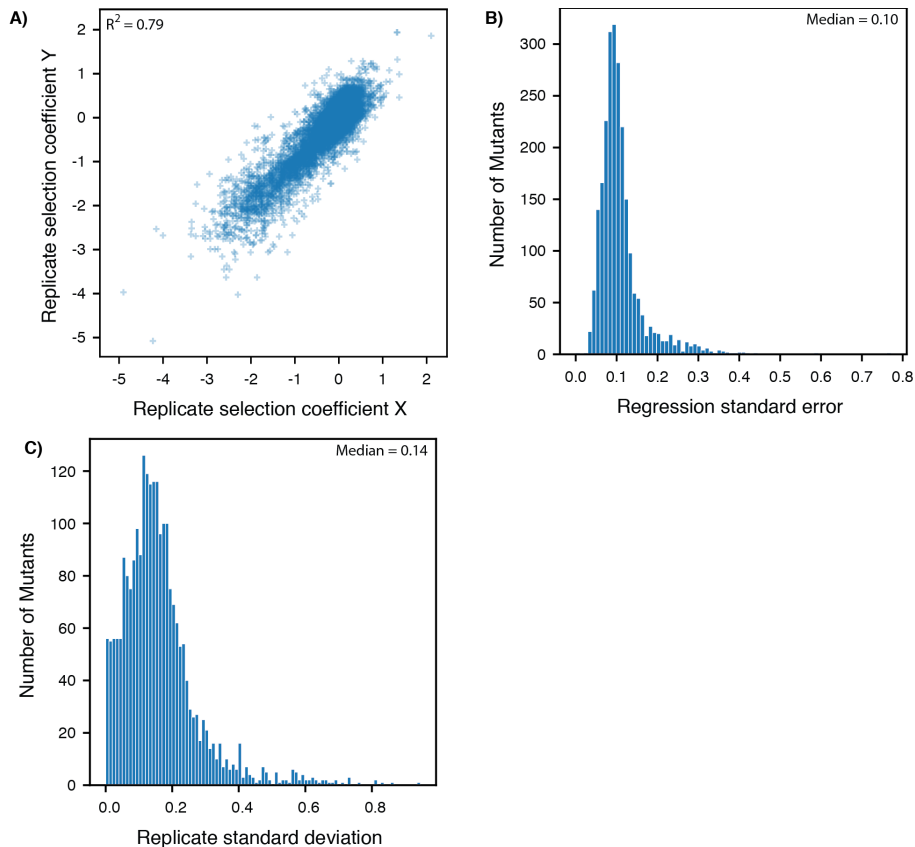


Figure 4.6: Quality of the selection under +Lon conditions. **A)** Comparison of all pairwise replicates for +Lon selection coefficients from triplicate deep mutational scanning on DHFR. The Pearson correlation R^2 value from linear regression was 0.70. **B)** Distribution of standard errors for individual +Lon selection coefficients from a single replicate. Selection coefficients are the slope from a linear regression of allele frequency as a function of time in selection. The standard error here is the mean square of residuals. **C)** Distribution of standard deviations of selection coefficients for individual point mutants over replicate experiments. Each mutant had a measured selection coefficient in at least 2 of the 3 replicates.

4.3.2 Repeating Selection with a Lon-expressing Strain

To test this prediction about Lon's impact on the mutational landscape, we then repeated deep mutational scanning in biological triplicate (see **Methods, Chapter 3.5.0.6**, page 111). We refer to the two regimes as +Lon (**Chapter 4**) and -Lon (**Chapter 3**) selection.

4.3.2.1 Analyzing Noise in Selection Coefficients The quality of +Lon selection (**Figure 4.6**, page 141) was comparable to that of -Lon selection (see **Figure 3.29**, page 88). The R^2 value between biological replicates was 0.79 and the median standard deviation between biological

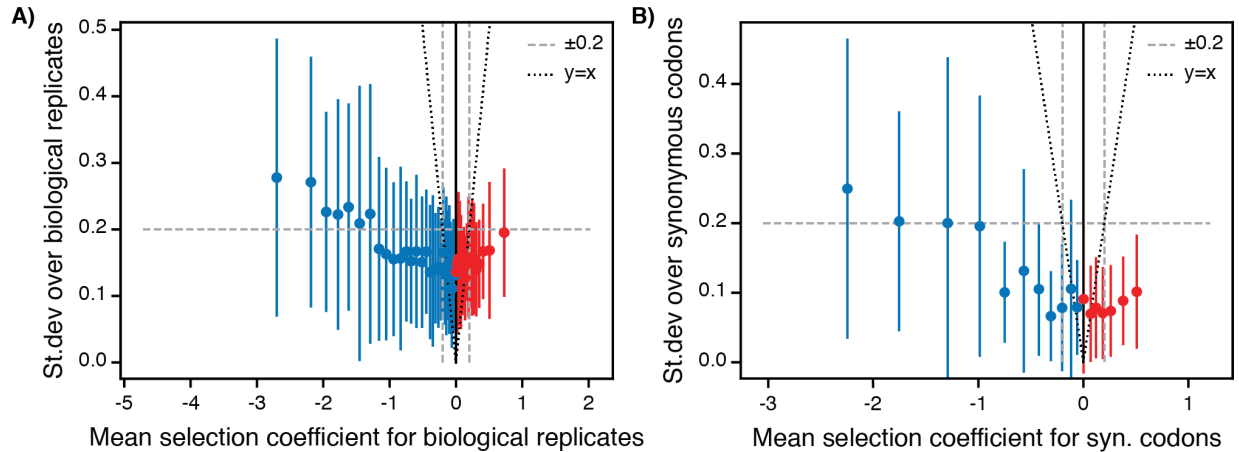


Figure 4.7: Relationship between error and selection coefficient for +Lon selection. **A)** Standard deviation of selection coefficients over biological replicates. The data were plotted as a function of a sliding window over all single point mutants sorted by selection coefficient. Each point represents the mean error (biological replicate standard deviation) over 50 consecutive selection coefficients (after sorting by value) and the error bars represent ± 1 standard deviation of the error. The dashed line represents median error over the entire dataset, which was used to determine the for WT-like behavior in **Figure 3.30A** (page 89). The dotted line represents $y = x$ for comparison between the magnitude of the error relative to the magnitude of the selection coefficient. **B)** Standard deviation over synonymous codons coding for the same sequence, plotted as in A.

replicates was 0.14. The error in the assay (**Figure 4.7**, page 142) also behaved similarly to that for $-$ Lon selection (see **Figure 3.31**, page 90) as a function of selection coefficient. Because the quality metrics were better for +Lon selection, we kept the thresholds for advantageous and disadvantageous mutations.

4.3.3 Population-level Mutational Analysis of Selection Data

We then analyzed the change in selection with Lon expression using multiple metrics. Consistent with our hypothesis about the impact of Lon, we observed that the number of advantageous mutations after reintroducing Lon decreased from 737 in $-$ Lon selection to 384 in +Lon selection. We also observed a selective decrease in the population of advantageous mutations and subsequent increase in the number of disadvantageous mutations (**Figure 4.8C**, page 143). When we compare selection coefficients for selection \pm Lon, we see a consistent pattern that selection coefficients are lower in selection with Lon, meaning that more mutations are disfavored relative

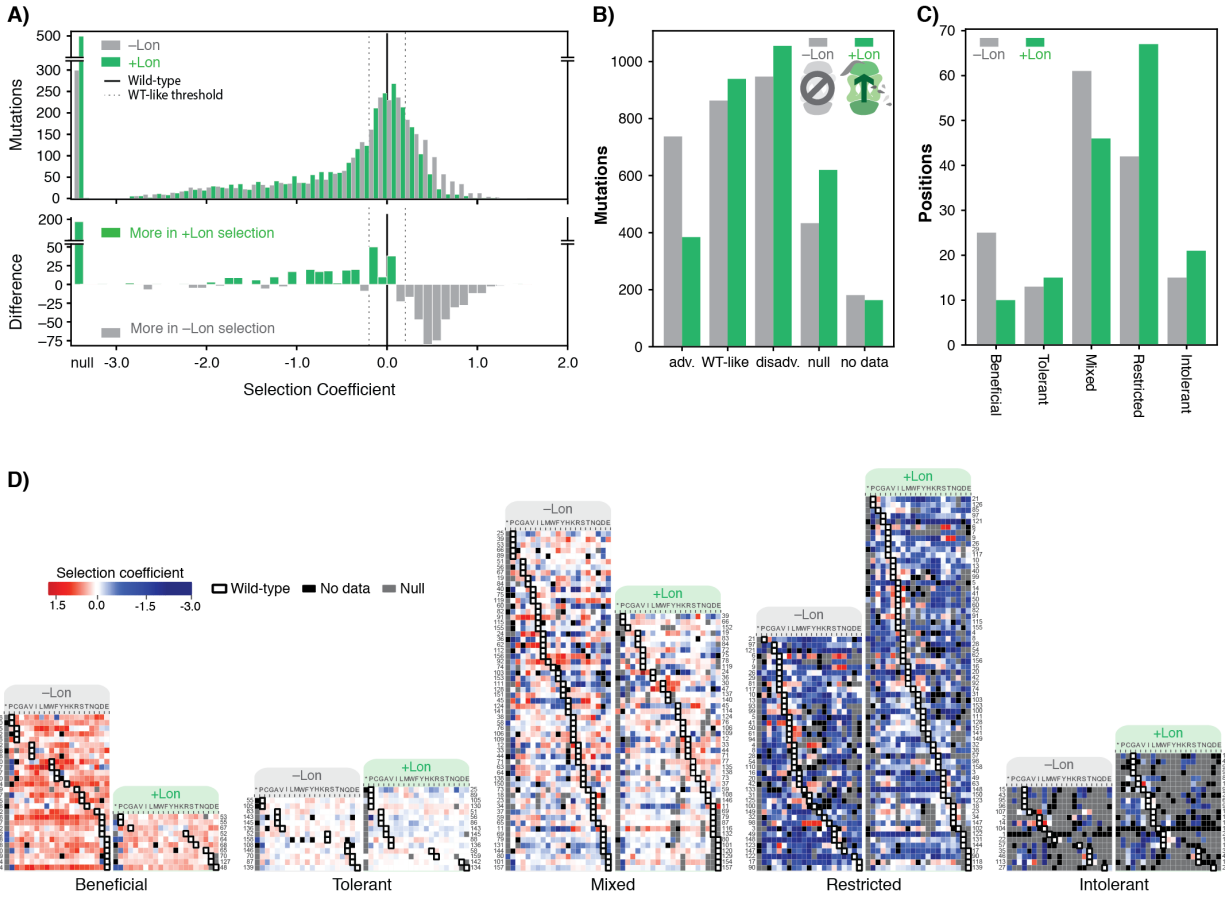


Figure 4.8: Lon protease expression reshapes the mutational landscape. **A)** Histogram of selection coefficients for mutations (top) in $-Lon$ (grey) and $+Lon$ selection (green). The difference of the histograms (bottom) is shown with grey indicating more mutants for $-Lon$ selection and green indicating more mutants for $+Lon$ selection. The threshold for classification for advantageous and disadvantageous mutations is as in Figure 1 and indicated with dashed lines. **B)** Distribution of mutations classified by selection coefficients: $0.2 \leq$ advantageous (adv.), $0.2 > WT\text{-like} > -0.2$, $-0.2 \geq$ disadvantageous (disadv.), null, and no data (a mutant was not detected in the library after transformation into the selection strain). Grey bars: $-Lon$ selection; green bars: $+Lon$ selection. **C)** Distribution of sequence positions into the 5 mutational response categories: Beneficial, Tolerant, Mixed, Deleterious, Intolerant. Grey bars: $-Lon$ selection; green bars: $+Lon$ selection. **D)** Heatmap of DHFR selection coefficients in the $-Lon$ and $+Lon$ strains, showing details of the distributions shown in C) (dotted border). Positions (rows) are grouped by their mutational response category for $-Lon$ and $+Lon$ as in C) and sorted by the wild-type amino acid. Amino acid residues (columns) are organized by physiochemical similarity and indicated by their one-letter amino acid code. An asterisk indicates a stop codon. Advantageous mutations are shown in shades of red, disadvantageous mutations in shades of blue, Null mutations in grey and “No data” as defined in A) in black. Wild-type amino acid residues are outlined in black.

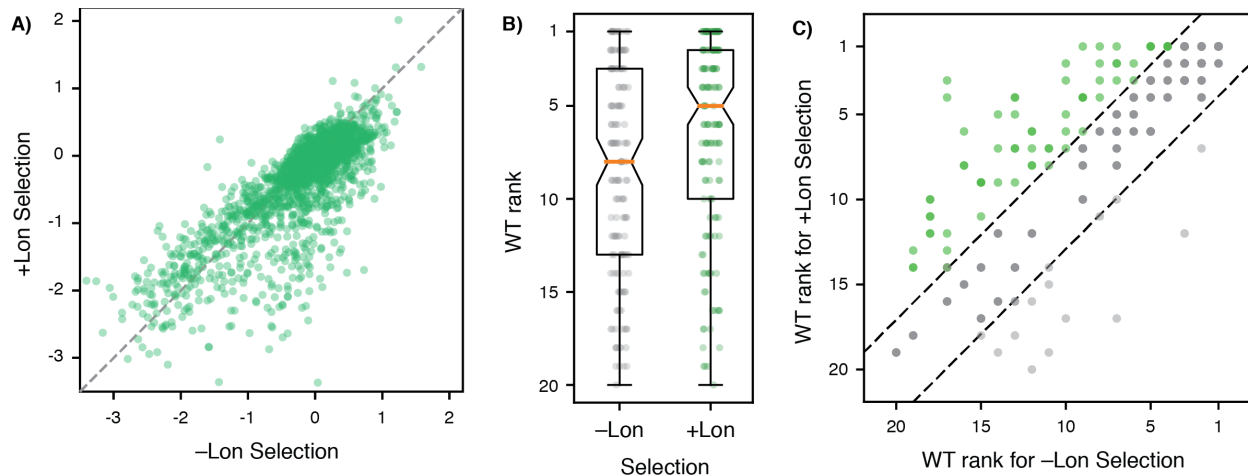


Figure 4.9: Comparison of selection coefficients \pm Lon. **A)** Scatterplot comparing selection coefficients in $-$ Lon and $+$ Lon selection, showing that mutations are generally repressed by Lon activity. Despite this general trend, we note that some top advantageous mutations are not impacted by Lon activity. **B)** Boxplot showing the distribution of wild-type amino acid residue rankings for $-$ Lon (grey) and $+$ Lon (green) selection. The wild-type amino acid residue ranking at each position is also shown as a distribution of points. Box plots show the median (orange bar) and upper/lower quartiles. The median wild-type amino acid residue rank is 8 for $-$ Lon selection and 5 for $+$ Lon selection. **C)** Wild-type amino acid residue rankings from $-$ Lon selection plotted against wild-type amino acid residue rankings from $+$ Lon selection. Dashed lines show ± 1 standard deviation for the change in rank between $-$ Lon and $+$ Lon selection.

to wild type (**Figure 4.9A**, page 144). These largest impacts are seen for mutations that are disadvantageous in both selection experiments, but the vast majority number of mutations that are advantageous in $-$ Lon selection negatively impacted by Lon expression. In total, the mean selection coefficient for advantageous mutations decreased from 0.47 to 0.37, and the rank of the wild-type sequence increased by 340 to 865th. Finally, the median rank of the wild-type residue over all positions decreased from 8 in $-$ Lon selection to 5 in $+$ Lon selection (**Figure 4.9BC**, page 144).

4.3.4 Categorizing Positions by Mutational Impact and Position-based Analysis

The variability in the rank of the wild-type amino acid lead us to ask if there were patterns in the general mutational responses of different position. To examine in more detail how the mutational response of individual residues changes between selection \pm Lon, we used a K-means clustering

algorithm (see **Methods, Chapter 4.5.0.3**, page 164) to group all DHFR sequence positions into 5 categories: positions where mutations were generally advantageous (Beneficial), generally WT-like (Tolerant), variably advantageous and disadvantageous (Mixed), generally disadvantageous (Restricted), and generally null (Intolerant). Grouping was performed separately for –Lon and +Lon selection (**Figure 4.8CD**, page 143 **Table 4.1**, page 146). Comparing the distributions of DHFR positions in –Lon and +Lon conditions illustrates the extensive reshaping of the mutational landscape by Lon. For –Lon selection, 28 positions (17.6%) were classified as Beneficial, where nearly every mutation was preferred over the wild-type residue. In comparison, the number of Beneficial positions decreased to 10 in +Lon selection, with only 3 surface-exposed positions (E48, T68, D127) common between the two Beneficial sets (**Figure 4.8D**, page 143, **Table 4.1**, page 146). Simultaneously, the number of Restricted positions increased from 42 to 67 with the reintroduction of Lon into the selection strain. The number of intolerant positions remained largely the same as many mutations are too deleterious for selection coefficient to be measured in either condition. These results support the conclusion that Lon activity broadly penalizes mutations, including a large subset of the advantageous mutations. Overall, the changes upon modulating Lon activity lead to a model in which upregulating Lon increases constraints on DHFR, and the mutational landscape changes from being permissive when Lon is absent to being more restricted when Lon is present.

4.3.5 Lon Impact as Δ Selection Coefficient

To analyze the constraints imposed by Lon on the DHFR mutational landscape in structural detail, we defined a Δ selection coefficient for each amino acid residue at each position as the difference between the +Lon and –Lon selections (**Figure 4.10A**, page 147). The Δ selection coefficient values were most negative at positions in the Beneficial category and at positions with a native

Table 4.1: Mutational response category and burial classification for DHFR positions. Positions are classified into mutational response categories of Beneficial ("B"), Tolerant ("T"), Mixed ("M"), Deleterious ("D"), and Intolerant ("I") for both -Lon and +Lon selection as described in **Methods (Chapter 4.5.0.3, page 164)**. Positions are categorized in burial states of buried ("b") and exposed ("e") based on Getarea server calculations as described in **Methods (Chapter 4.5.0.4, page 165)**.

Pos.	Lon			Pos.	Lon			Pos.	Lon			Pos.	Lon		
	-	+	burial		-	+	burial		-	+	burial		-	+	burial
1	I	I	b	41	D	D	b	81	D	I	b	121	D	D	b
2	I	I	b	42	D	D	b	82	M	D	e	122	D	D	e
3	D	D	b	43	I	I	b	83	T	M	e	123	D	D	e
4	D	D	b	44	M	M	b	84	M	M	b	124	M	M	e
5	D	D	b	45	M	M	e	85	B	D	b	125	D	I	b
6	D	D	b	46	I	I	b	86	B	T	e	126	B	D	b
7	D	D	b	47	B	M	b	87	T	M	e	127	B	B	e
8	D	D	b	48	B	B	e	88	B	T	e	128	M	D	b
9	D	D	b	49	D	D	b	89	M	T	e	129	B	M	e
10	D	D	e	50	D	D	b	90	D	D	b	130	B	T	e
11	M	M	e	51	M	T	e	91	M	D	b	131	M	D	e
12	M	M	e	52	T	B	e	92	M	D	b	132	B	M	e
13	D	D	b	53	M	B	e	93	D	I	b	133	D	I	b
14	I	D	b	54	D	D	b	94	D	I	b	134	B	T	e
15	I	I	b	55	T	B	e	95	I	I	b	135	B	M	e
16	D	D	e	56	M	T	e	96	I	I	e	136	T	T	e
17	D	D	e	57	I	D	b	97	D	D	b	137	B	M	e
18	M	D	e	58	M	T	e	98	D	D	e	138	M	M	e
19	M	M	e	59	M	M	b	99	D	D	b	139	T	D	e
20	D	D	b	60	M	D	b	100	D	D	b	140	B	M	e
21	D	D	b	61	D	I	b	101	M	M	e	141	M	D	e
22	I	I	b	62	M	D	b	102	B	D	e	142	B	T	e
23	M	D	e	63	M	D	b	103	M	D	b	143	T	T	e
24	M	M	b	64	M	B	e	104	I	I	b	144	M	D	e
25	M	T	e	65	M	B	e	105	T	T	e	145	T	T	e
26	D	D	e	66	M	M	e	106	M	M	b	146	T	M	e
27	I	I	b	67	M	B	e	107	I	I	b	147	D	D	b
28	D	D	e	68	B	B	e	108	T	M	e	148	D	D	e
29	D	D	e	69	M	M	b	109	M	M	b	149	D	D	e
30	B	M	b	70	T	B	e	110	D	I	b	150	M	D	e
31	D	D	e	71	M	M	e	111	M	D	b	151	M	D	e
32	D	D	e	72	B	M	b	112	M	I	b	152	B	M	b
33	M	M	e	73	M	M	e	113	I	I	b	153	M	D	b
34	M	D	b	74	M	D	b	114	B	M	e	154	B	M	b
35	I	I	b	75	M	M	b	115	M	D	b	155	M	D	b
36	M	M	e	76	M	M	e	116	B	M	b	156	M	D	b
37	M	M	e	77	B	M	e	117	D	D	b	157	M	M	e
38	M	D	e	78	B	M	b	118	B	D	e	158	T	D	e
39	M	M	b	79	M	M	e	119	M	M	e	159	B	T	e
40	M	D	b	80	M	M	e	120	B	M	e				

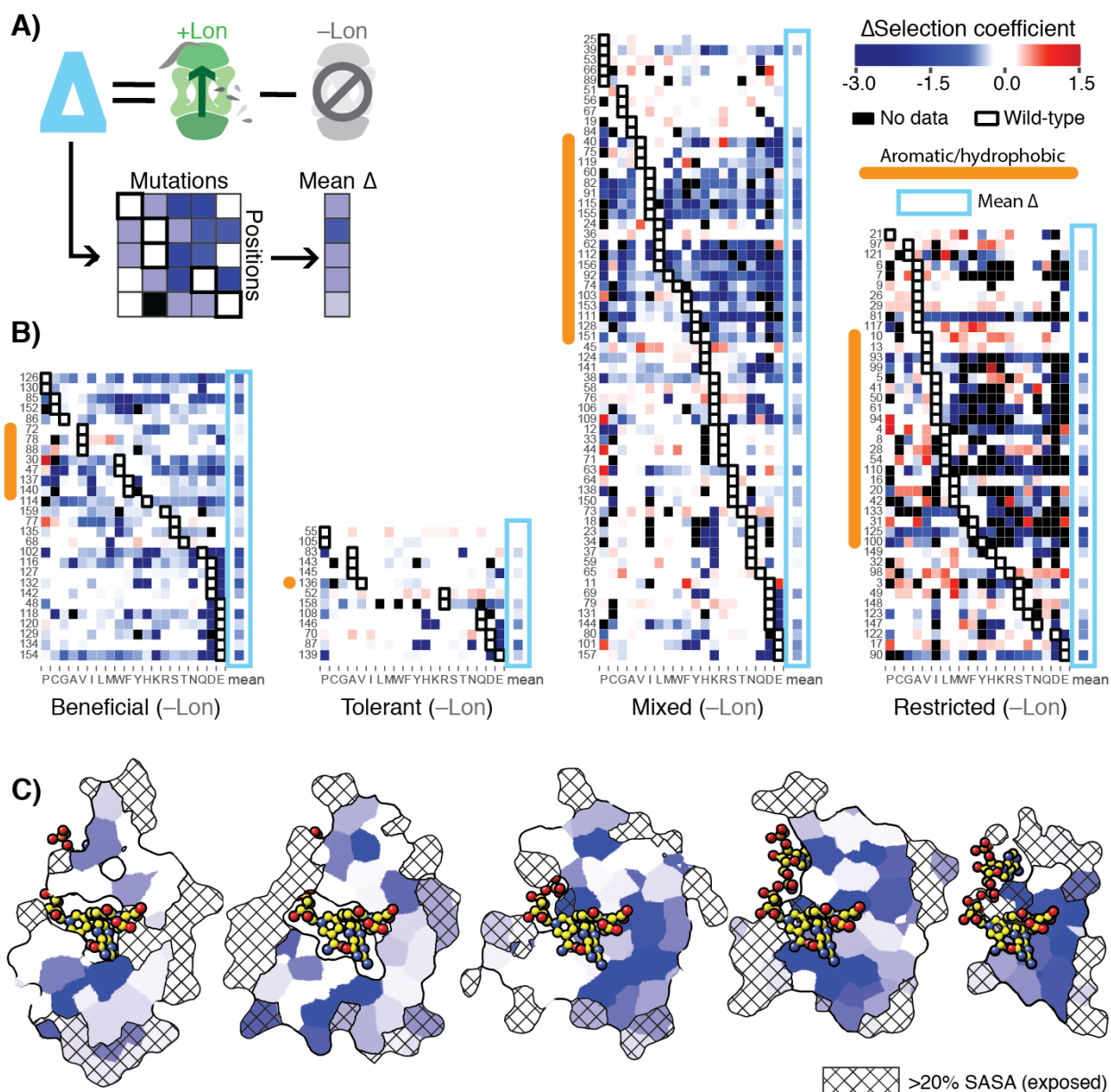


Figure 4.10: Δ selection coefficients show Lon impact. **A)** Conceptual diagram of Δ selection coefficients, calculated as the +Lon selection coefficient minus the -Lon selection coefficient (see **Methods, Chapter 3.5.0.10**, 115). **B)** Heatmap of Δ selection coefficient values for all positions not classified as Intolerant. Δ selection coefficients values between -0.2 and 0.2 are shown in white; Δ selection coefficients >0.2 are in shades of red and Δ selection coefficients <-0.2 in shades of blue. Amino acid residues (columns) are organized by physiochemical similarity and indicated by their one-letter amino acid code. The mean Δ selection coefficient (avg) at each position is shown as a separate column and outlined with a light blue box. Positions (rows) are sorted by the wild-type amino acid and grouped by their mutational response category from the -Lon selection in **Figure 4.8CD** (page, 143). Positions with a native VILMWF or Y amino acid are indicated with an orange bar to the left. **C)** Per-position mean Δ selection coefficient displayed on the structural model of DHFR. The 5 cross-section slices of the DHFR structure are displayed as in **Figure 3.30C** (page, 89), and the color scale is as in B).

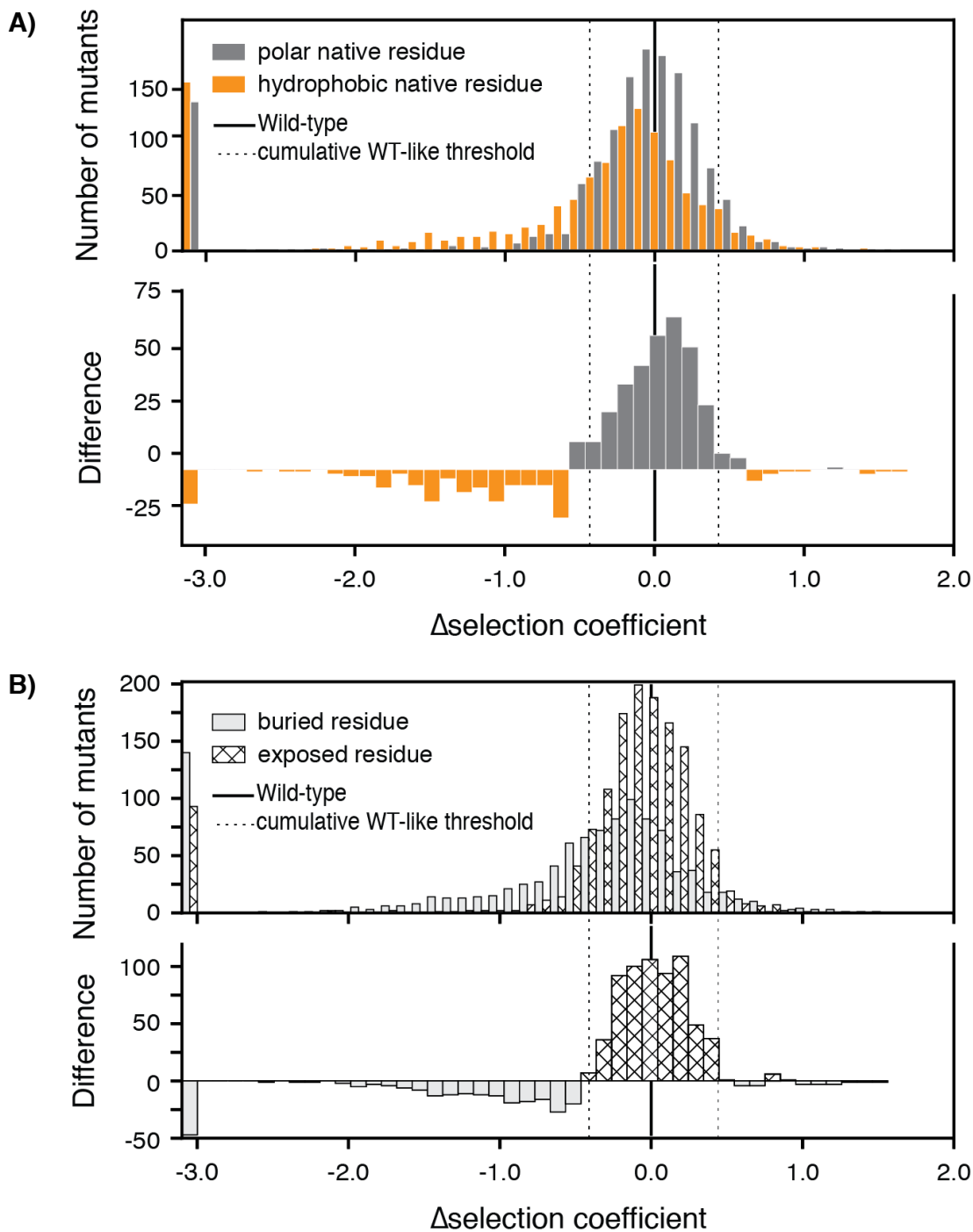


Figure 4.11: Δ selection coefficients show Lon impact on natively buried and hydrophobic residues. **A)** Histogram of Δ selection coefficients (top) with mutants at positions with hydrophobic (AVILMWFY) wild-type amino acid residues in orange and at positions with polar (HKRSTNQDE) wild-type amino acid residues in grey. Selection coefficients for positions with a wild-type P, G, or C residue are not included. The difference of the histograms (bottom) is shown with grey indicating more mutants to positions with a wild-type polar residue and orange indicating more mutants to positions with a wild-type hydrophobic residue. Dotted lines indicate twice the median of standard deviations from (Figure 3.30B, page 89). **B)** Histogram of Δ selection coefficients (top) with mutants at buried positions (solid) and at exposed positions (hatched) as listed in Table S4. Selection coefficients for positions that were Intolerant in $-$ Lon selection are not included. The difference of the histograms (bottom) is shown with solid indicating more mutants to buried positions and hatched indicating more mutants to exposed positions.

VILMWF or Y amino acid residue (**Figure 4.10B**, page 147). This pattern for deleterious is most measurable in the Mixed category. This pattern is also clearly observed in the Deleterious category, except that it is not readily measurable because many mutations go from being disadvantageous to null. Notably, most mutations to Tolerant positions do not have Δ selection coefficient with large magnitudes. When we analyze over the whole dataset of Δ selection coefficients, mutations at positions with native hydrophobic residues are enriched for negative Δ selection coefficients (**Figure 4.11**, page 148). Strikingly, the mean Δ selection coefficient was -0.71 for the 65 buried positions with $< 20\%$ side-chain solvent accessible surface area, compared to -0.27 for the 79 exposed positions (see **Table 4.1**, page 146). This may explain, in part, why the Tolerant positions are not greatly impacted by Lon. This result is also visible on the DHFR structure, which makes it clear that many of the buried residues that are not impacted by Lon are the residues that are in the active site and are already classified as Deleterious or Intolerant (**Figure 4.10C**, page 147). These results show that Lon has a broad impact on the mutational landscape throughout the DHFR structure but imposes particularly strong constraints in the DHFR core.

4.3.5.1 Correlation between Δ Selection Coefficient and T_M Moreover and as expected, the Δ selection coefficients between +Lon and -Lon selection are correlated with T_m , except for mutations near the active site (**Figure 4.12**, page 150). Strikingly, when we compare different mutations at the same position, the change in Δ selection coefficients (i.e. Lon sensitivity) correlates with the change in T_m values. These results indicate that the selected advantageous mutations are typically destabilizing, and that destabilization is correlated with Lon sensitivity. One possible explanation for the selection advantage of destabilizing mutations that have an increase in velocity (e.g. L24V, W30F/M, M42F/Y, H114V, D116I/M, E154V) is that these mutations promote breathing motions that accelerate product release, which is rate limiting for

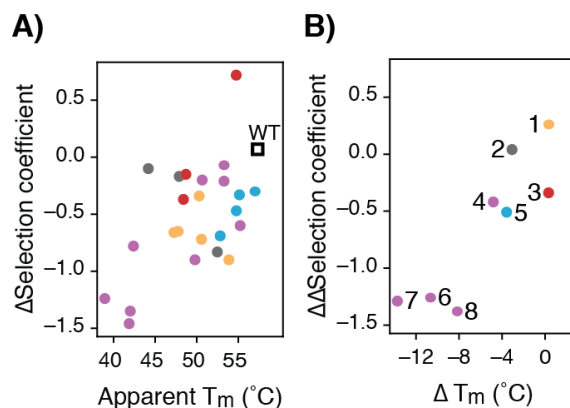


Figure 4.12: Correlation of Δ selection coefficients to T_m . **A)** Correlation between *in vitro* T_m values and *in vivo* Δ selection coefficients for DHFR wild-type and characterized mutants. Each point represents a single point mutations (see **Table 3.8**, page 100). Points are colored according to the position's location in the DHFR structure: core (purple), surface beta-sheet (gold), proximal to the adenine ring on NADPH (blue), or proximal to the active site and M20 loop (red). **B)** ΔT_m values and $\Delta\Delta$ selection coefficient for mutations at the same position. Points representing comparison between mutants are numbered as follows: **1)** D116I-M, **2)** M42Y-F, **3)** W30M-F, **4)** I91G-A, **5)** Q102W-L, **6)** L62A-V, **7)** I41A-V, **8)** W47V-L.

wild-type DHFR at neutral pH[29] and for a hyperactive DHFR mutant with a 7-fold increase in k_{cat} [30].

4.3.5.2 Lon Expression Decreases *In Vivo* Expression of DHFR To examine the impact of Lon further, we measured the soluble expression in ER2566 $\Delta folA/\Delta thyA$ +Lon cells (see **Methods, Chapter 3.5.0.14**, page 122, **Table 4.2**, page 151) Importantly, expression measurably decreased for all mutants in the presence of Lon (**Figure 4.13A**, page 152). Furthermore, the change in expression level has moderate correlation with the change in selection coefficient between the two regimes (**Figure 4.13B**, page 152). While these data are not sufficient to support a direct interaction between DHFR and Lon, we can conclude that Lon can suppress a broad range of mutations with multiple mechanisms underlying their advantageous selection impact.

Table 4.2: Soluble DHFR abundance levels in molecules per cell. Abundances were measured from lysate activity assays as described in **Methods (Chapter 3.5.0.14, page 122)**. All values are for the SMT205 plasmid transformed into the ER2566 $\Delta folA/\Delta thyA$ +Lon. NM, not measured.

Variant	Molecules per cell	Standard deviation
WT	71.4	23.9
L24V	65.5	13.4
W30F	79.0	9.3
W30M	63.7	11.0
I41A	9.2	8.4
I41V	218.9	108.5
M42F	290.6	5.9
M42Y	162.3	4.9
H45S	69.6	13.8
W47L	349.2	46.7
W47V	NM	NM
L62A	NM	NM
L62V	104.7	15.3
C85L	63.6	2.6
I91A	56.0	14.5
I91G	86.8	20.7
R98Y	161.5	29.4
Q102L	198.4	57.5
Q102W	133.1	27.6
T113V	212.2	27.5
H114V	NM	NM
D116I	NM	NM
D116M	46.5	7.0
E154V	36.2	0.8
L156Y	44.8	6.6

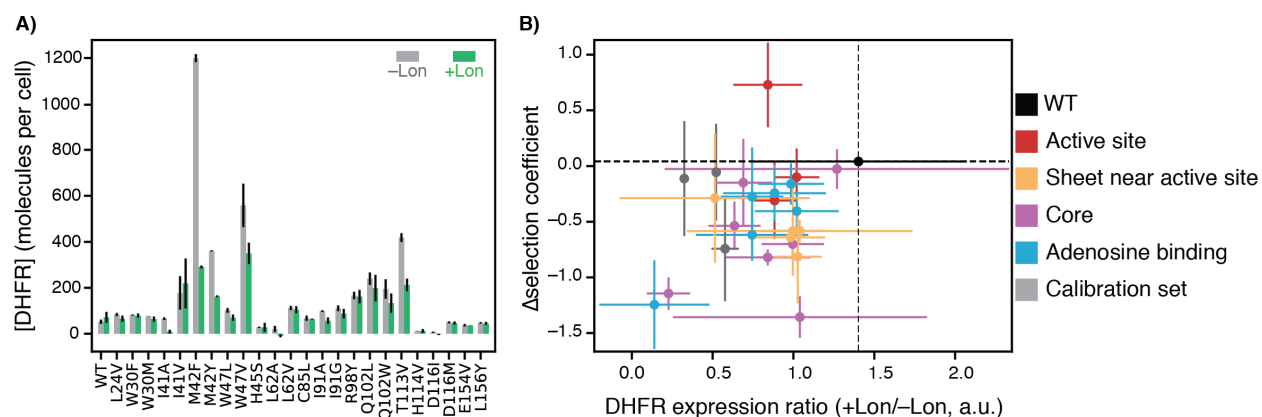


Figure 4.13: Correlation of Δ selection coefficients to change in cellular abundance \pm Lon. **A)** DHFR cellular abundance calculated from the lysate DHFR activity and *in vitro* kinetics with purified enzyme (see **Methods, Chapter 3.5.0.13**, page 121). Error bars represent the cumulative percent error (standard deviation) from three independent experiments for velocity and three biological replicates for lysate activity. Data are shown in both the -Lon (light grey) and +Lon (green) conditions. **B)** Correlation between the ratio of cellular DHFR abundance in A) and *in vivo* Δ selection coefficients \pm Lon for DHFR wild-type and point mutants. Each point represents a mutation and is colored according to the position's location in the DHFR structure: core (purple), surface beta-sheet (gold), proximal to the adenine ring on NADPH (blue), or proximal to the active site and M20 loop (red). X-axis error bars represent the cumulative percent error (standard deviation) from three measurements of DHFR concentration with and without Lon (Methods). Y-axis error bars the cumulative error (standard deviation) from three biological replicates for selection with and without Lon (Methods).

4.3.6 Structural Patterns in DHFR Selection

Taken together, our data indicate that the observed widespread changes in the mutational landscape of DHFR can be explained by a penalty for destabilizing mutations from Lon expression, leading to extensive activity – stability tradeoffs for advantageous mutations. The effect of these two selection pressures is directly observable in the structural arrangement of the mutational response categories. In -Lon conditions, mutational responses are arranged in shells around the hydride transfer site[31] (**Figure 4.14A**, page 153) where the proportion of advantageous mutations increases with increasing distance (**Figure 4.14B**, page 153, **Figure 4.15B**, page 154). This same spatial pattern also holds for +Lon selection (**Figure 4.14C**, page 153), but it is now superimposed with the additional pressure against destabilizing mutations such that there are no Beneficial positions in the core(**Figure 4.16**, page 155). In contrast, the mutational responses as a function of distance to other DHFR sites (e.g. C5 of

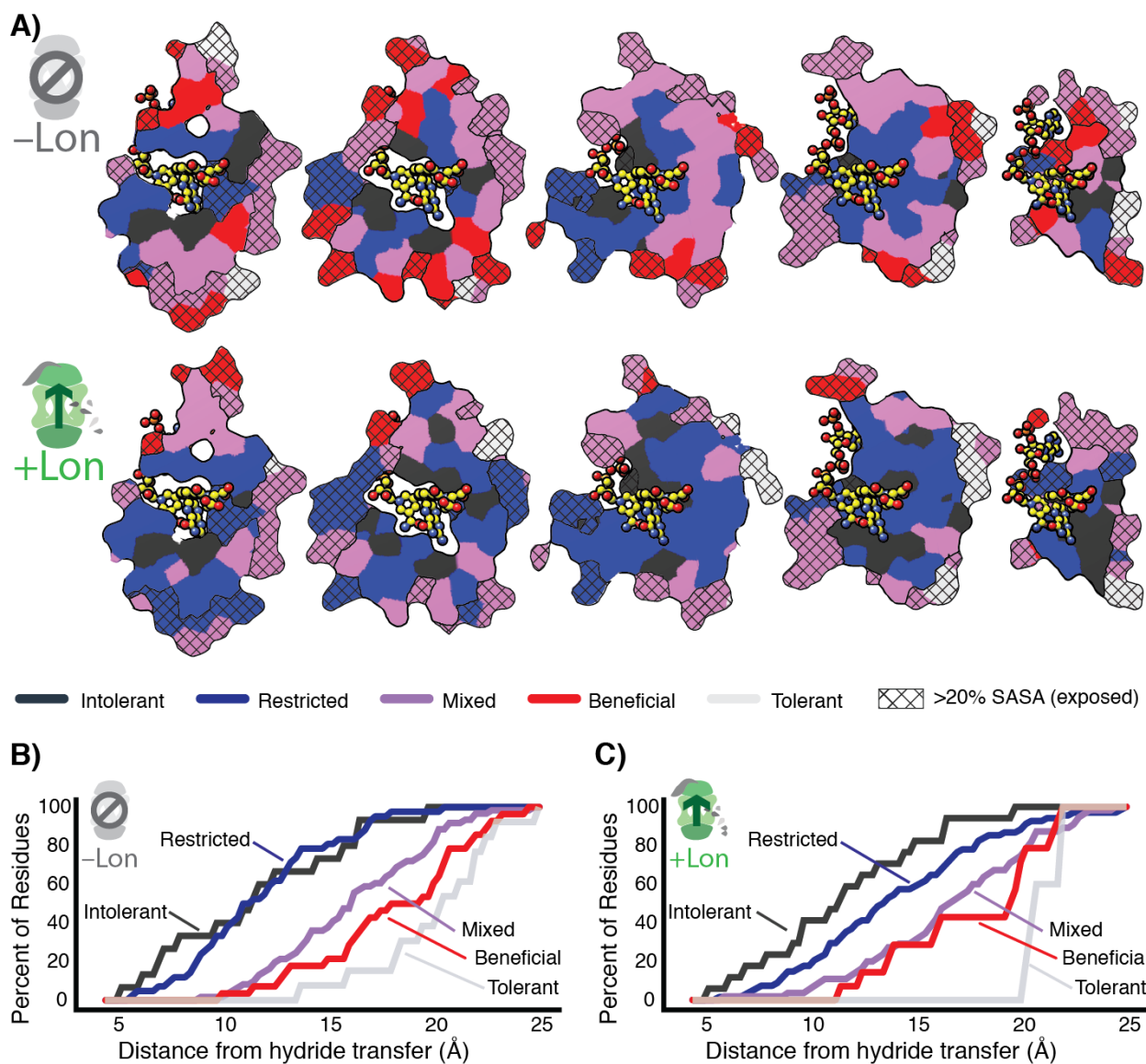


Figure 4.14: Structural characterization of multiple constraints on the DHFR mutational landscape. **A)** Mutational response categories from $-Lon$ selection (top, categories in **Figure 4.8CD**, page 143) and $+Lon$ selection (bottom, categories as in **Figure 4.8CD**, page 143) colored onto residues and displayed on slices as in **Figure 3.30C** (page 89). **B)** Relationship between mutational response and distance from hydride transfer for $-Lon$ selection. The percent of positions from each mutational response category are plotted as a function of distance from the site of hydride transfer. Each category colored as in A, top). **C)** Relationship between mutational response and distance from hydride transfer for $+Lon$ selection. Each category colored as in A, bottom).

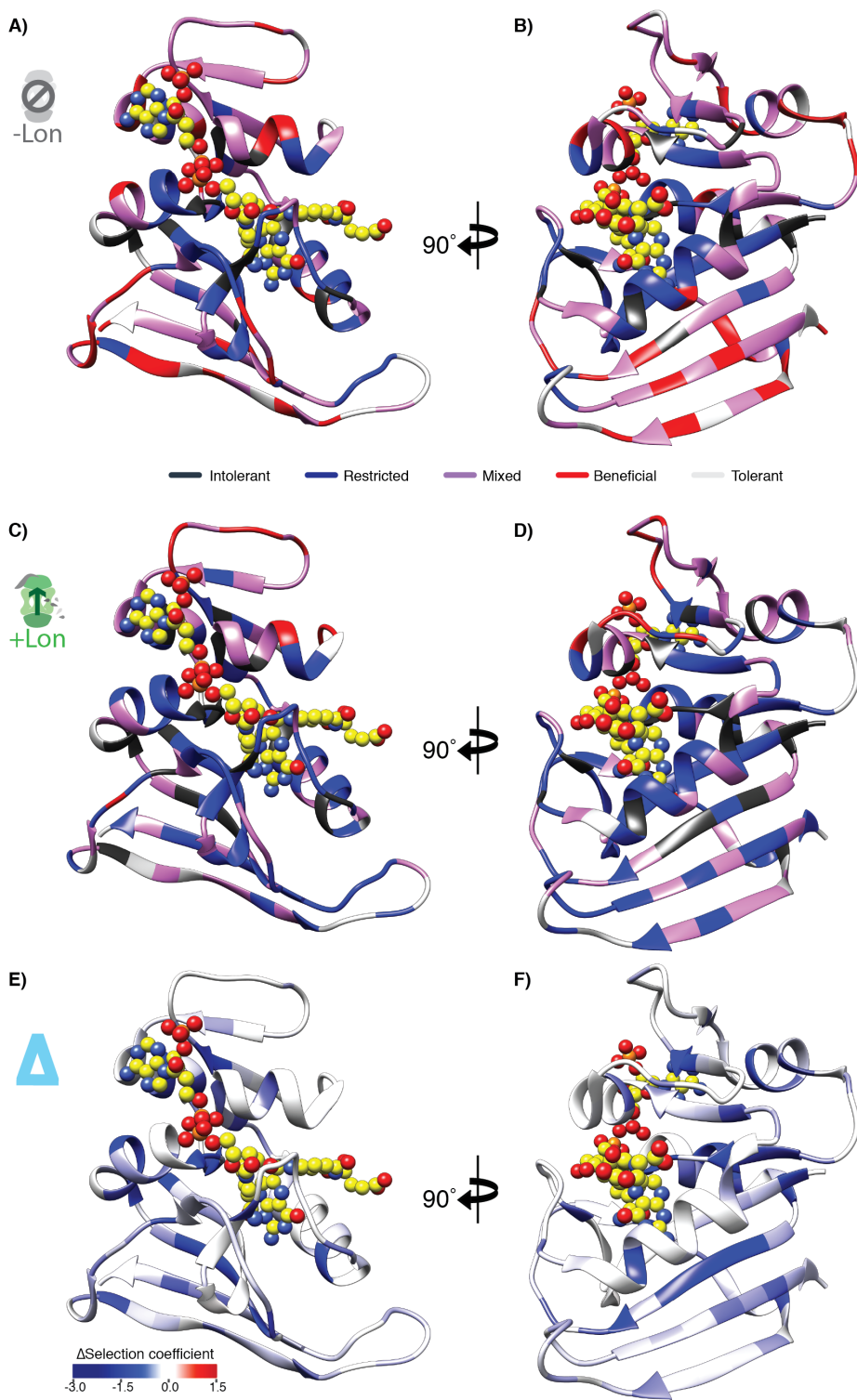


Figure 4.15: Selection coefficients under the two Lon expression regimes mapped on the DHFR structure. Structural model of DHFR (PDB ID: 3QL3) in ribbon representation with the DHF substrate and the NADPH cofactor represented by spheres (yellow carbon and heteroatom coloring). The residues are colored in **A,B**) by mutational response category from **Figure 4.8CD** (page 143) for -Lon selection, in **C,D**) by mutational response category from **Figure 4.8CD** (page 143) for +Lon selection, or in **E,F**) by the per-position mean Δ selection coefficient from **Figure 4.10** (page 147).

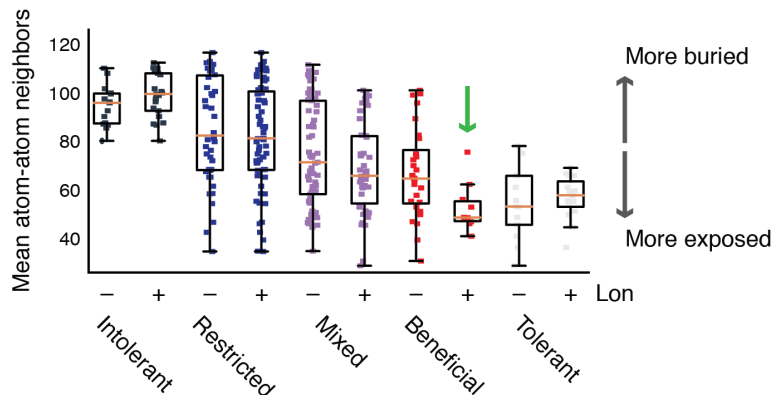


Figure 4.16: Burial of residues within each mutation response category reported as the mean number of atomic neighbors. Each point represents one amino acid side chain, and the y-axis reports the average number of heavy atom neighbors within an 8 Å shell for all heavy atoms in that side chain. Box plots are overlaid on the distribution to show the median (orange bar) and upper/lower quartiles. Mutational response categories are shown for both -Lon and +Lon selection. The green arrow highlights the absence of buried Beneficial positions in +Lon selection.

the NADPH adenine ring) do not show as strong of a relationship (**Figure 4.17**, page 156).

These findings illustrate how the contributions from two constraints – one structural (distance from hydride transfer) and one dependent on cellular context (Lon) – can be distinguished in structural patterns in the mutational landscape.

4.3.7 Comparing Sequence Preferences in Selection and Evolution

We then asked how constraints in our assay compare to the constraints that have shaped natural DHFR sequences. We first examined the performance of the WT sequence in the selection assay. As mentioned previously, there are 737 advantageous mutation in -Lon selection and 384 advantageous mutation in +Lon selection, and the median rank of the wild-type residue over all positions was 8 in -Lon selection and 5 in +Lon selection (**Figure 4.9**, page 144). Thus, the addition of Lon makes the WT sequence generally more favorable in selection. To further investigate how our selection assay recapitulates natural selection pressures, we used profile similarity metrics to (**Chapter 4.5.0.5**, page 165). Briefly, profile similarity is an informatics metric that quantifies the similarity of two distributions. When we look at the similarity of

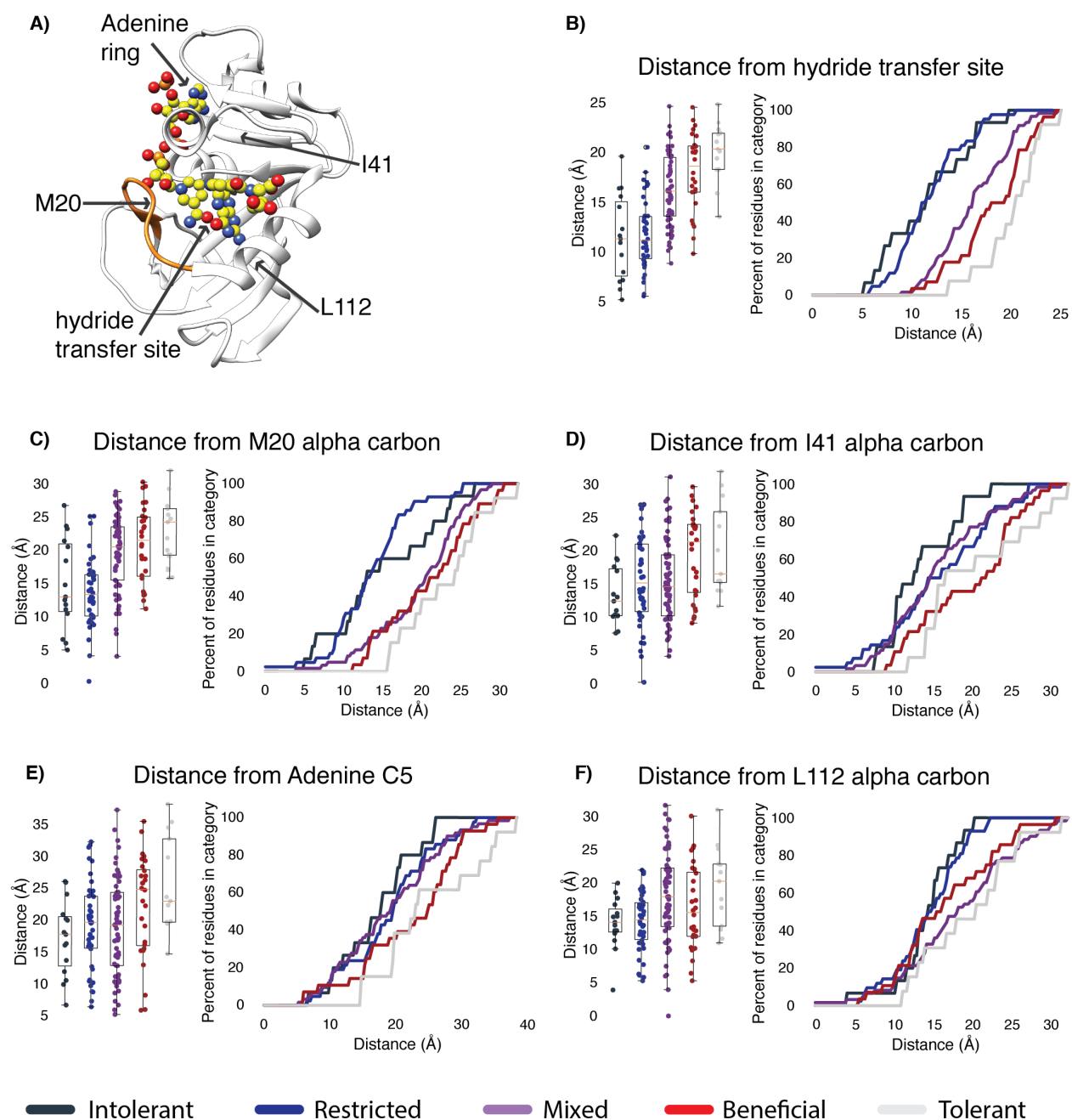


Figure 4.17: Residues in mutational response categories in the $-Lon$ selection as a function of distance from several sites in the DHFR structure. **A)** Location of hydride transfer site, the M20 residue on the M20 loop (orange), and hot spot sites from Figure 4 (the core of the globular domain represented by I41, the beta-sheet surface below the active site represented by L112, and the adenine ring on NADPH) indicated on the DHFR structure (PDB ID: 3QL3). **B-F)** The distance relationships between each site and the residues in each mutational response category in the $-Lon$ selection are shown (left) as boxplots with points representing the individual mutants and (right) as curves showing the percent of sequence positions in each mutational response category as a function of distance from the site. Boxplots and curves are colored by mutational response categories from $-Lon$ selection as in **Figure 4.14** (page 153).

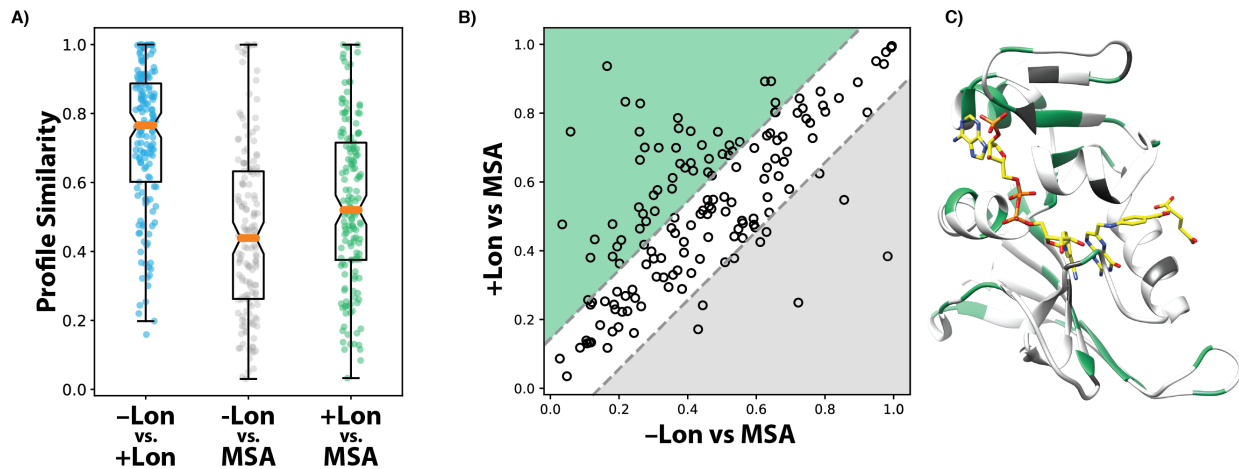


Figure 4.18: Comparison of DHFR per-position sequence preferences. **A)** Profile similarity (see Methods) was calculated to compare the per-position distribution of amino acid frequencies between selection \pm Lon (blue), between $-$ Lon selection and an MSA of DHFR orthologues (grey), and between $+$ Lon selection and the MSA. Each point represents a single position in DHFR. A profile similarity value of 1.0 indicates identical distributions at that position, and a value of 0 represents no overlap in the distributions. The box plot shows the median (orange line), the interval between the first and third quartiles (box), and the maximum and minimum (whiskers). **B)** Scatterplot comparing the similarity of amino acid preferences in the MSA to selection \pm Lon. Each dot represents a single position in the DHFR sequence. X-axis values represent the profile similarity score between the MSA and $-$ Lon selection for amino acid preferences at each position. Y-axis values represent the profile similarity score between the MSA and $+$ Lon selection for amino acid preferences at each position. The grey dashed lines represent $y = x \pm$ one standard deviation for $|\text{Similarity}(-\text{Lon vs. MSA})_{\text{position}} - \text{Similarity}(+\text{Lon vs. MSA})_{\text{position}}|$. Positions in the green region have amino acid preferences more similar to the MSA for $+$ Lon selection, and positions in the grey region have amino acid preferences more similar to the MSA for $-$ Lon selection. **C)** Crystal structure model of DHFR (PDB ID: 3QL3) with positions colored by their location in the green, grey, and white regions from panel B).

tolerated amino acid distributions from selection \pm Lon and a MSA, we see that the addition of Lon expression into selection broadly improves similarity between the experiment and naturally observed DHFR sequences (**Figure 4.18**, page 157). As with WT rank, the addition of Lon to selection appears to reconcile about half of the differences between the assay selection pressure and natural selection pressure(s) as measured by these metrics, although the behavior of these metrics is non-linear. Therefore, we interpret this result as measurable, but incomplete improvement in recapitulating natural selection pressures. Furthermore, we can see that the improvements in the profile similarity and WT rank metrics are not localized, but broadly spread throughout the protein. These improvements are largely observed at buried positions, which is expected because the Δ selection coefficient values are largest at buried positions.

4.3.8 Examining the Lon Locus in *E. coli* Genome from Seminal Experiments

For the sake of hypothesis generation, we next asked what seminal experiments had been performed in B strain *E. coli*, and therefore might be impacted by Lon expression. Because we anticipated that the impact of Lon would be more obvious over long selection times, we examined the *E. coli* LTEE. A B-strain of *E. coli* was indeed used for the LTEE [19], and the published representative genome for the LTEE strain did indeed have the IS186 insertion into the Lon promoter[?]. Considering the lessons learned from the LTEE experiment about the rate and mechanism of adaptation in *E. coli*, this result opens questions about Lon's impact on how fast adaptation occurs, what pathways are accessible, and what adapted states are incompatible with more stringent protein quality control regimes.

4.4 Discussion

In this chapter, we examined how the cellular state of protein quality control can reshape the mutational landscape of DHFR by changing the penalties for destabilizing mutations. Selection on DHFR point mutants in the study illustrates how a structural constraint from hydride transfer can shape a mutational landscape and how an environmental constraint from Lon protease reshapes that landscape. The impact of structural and environmental constraints have been considered previously [32–38], and our work allows us to see how multiple constraints combine to create alternate mutational landscapes. In the results presented here, we observe distinct structural patterns arising from both constraints.

More can be learned about the basics of enzyme function by examining the rich mutational dataset we generated for DHFR. Metabolomics experiments with various mutants will help deconvolute changes in the kinetic properties of DHFR from changes in the cellular environment. Mechanistic studies on the origin of increased DHFR activity via stopped flow kinetics, NMR, and hydrogen-deuterium exchange mass spectroscopy (HDX) may reveal how the destabilization of different regions of the DHFR protein increases velocity by affecting different microscopic constants. These experiments may clarify why so many advantageous mutations are distal to the active site.

We can propose a mechanistic explanation for the spatial patterning within the mutational landscapes from this work. The positive mutations we characterized in this study showed moderate decreases in stability coupled with mild increases in turnover. Many of the observations can be explained if these destabilizing mutations allow access to excited conformational states. This kind of molecular mechanism for rate acceleration would be analogous to that observed in recent work by Hilser and colleagues for surface glycine mutations to *E. coli* adenylate kinase [39, 40].

The rate limiting step of DHFR is product release, which is coupled to the dynamics of the M20 loop that lies over the active site[29, 41]. Excited conformational states could promote faster conformational exchange in the M20 loop or potentially enable the hinge-like motions observed in the human DHFR orthologue and in “humanized” *E. coli* DHFR [42], both of which have a higher k_{cat} than the wild-type. This would explain the observation that there are many positive mutations throughout the DHFR structure.

We also note here that the pattern of advantageous mutations throughout the protein is similar to the observation of many distal mutations becoming fixed in directed evolution. A number of directed evolution experiments have been performed in BL21 [43, 44], although many directed evolution projects use K12 strains or proprietary strains of *E. coli*. It could be informative to examine if the Lon expression in the selection strain changes the likelihood that destabilizing mutations are fixed during directed evolution and if the number of fixed mutations distal to the active site is decreased. Similarly, more detailed bioinformatics analysis may be able to detect evolutionary signal from Lon, but even the naïve analyses here support the conclusion that more stringent selection for stability makes selection more consistent with natural protein sequences. It remains to be seen if this is true in general or only for specific model systems and selection conditions. One potential caveat to this is that most endogenous enzyme expression or induced expression from a plasmid is much higher than tens of DHFR copies per cell that are expressed in our optimized selection conditions. It is possible that changes in client expression level may change the impact of Lon on the mutational landscape (i.e. over-saturating Lon).

We anticipate that mutational landscapes are subject to the interplay between different components of protein quality control. Our work allows us to consider the engineering potential in exploiting opposing pressures for precisely tuning mutations landscapes, where protease

activities add constraints and chaperones relieve them. To illustrate the engineering potential in tuning refolding and degradation activities, we will discuss seminal works about the impact on mutational landscapes from two different chaperones: GroEL and Hsp90. The first comparison is with Danny Tawfik and Nobu Tokuriki's use of GroEL overexpression to dampen activity-stability trade-offs during directed evolution [8, 45]. In that study using a model selection experiment, sequences with 10-fold higher activities were obtained with GroEL overexpression compared to selection with basal GroEL expression. If GroEL over-expression is generally useful for traversing through active but destabilizing regions of sequence space, Lon over-expression may also be useful in later polishing rounds of selection and may help identify active and well-behaved sequences. The second comparison is with Susan Lindquist and colleagues' study that showed deleterious phenotypes from mutations are buffered by Hsp90, a major chaperone in eukaryotic cells [3–6]. Considering their observations with Hsp90 and our observations with Lon, protein homeostasis appears to act as a set of antagonistic forces on mutational landscapes.

4.4.1 Discussion

We envision a general method of engineering landscapes to respond to external switches. The possibility of engineering mutational landscapes has not yet been explicitly considered to our knowledge. This work is a preliminary example of how a mutational landscape can be tuned along two dimensions: activity and stability, but we anticipate that many other tunable pressures will be identified. We anticipate that further study of antagonistic pressures on mutational landscapes will yield a new toolkit for the fine control of the mutational landscapes that guide movements through sequence space.

This concept could also be tested directly by performing proof of principle experiments using

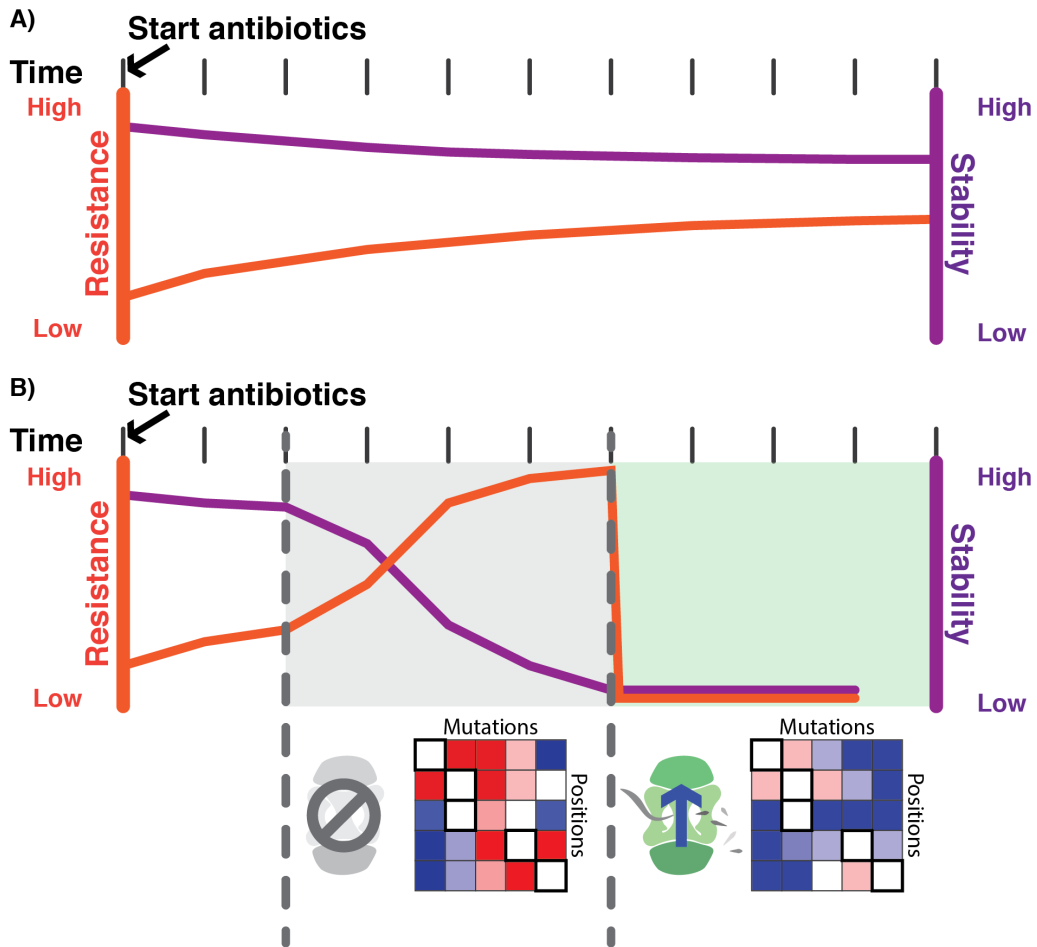


Figure 4.19: Concept for modulating the mutational landscape with Lon to combat antibiotic resistance. **A)** Resistance (orange) to an antibiotic mediated by a protein is a function of trade-offs between activity and stability (purple) in that protein. Resistance mutations arise in the population at the expense of stability. **B)** Resistance to an antibiotic is mediated as in A), but Lon protease expression is manipulated by a Lon inhibitor (grey) and a Lon agonist (green) in a time-dependent fashion. This manipulation of the mutational landscape causes over adaptation under permissive protein quality control and elimination of the resistant bacteria under stringent protein quality control.

directed evolution while gradually increasing Lon expression. These experiments are similar in concept to directed evolution with upregulated GroEL, and they could be combined with early expression of chaperones to allow the directed evolution experiment to more effectively search functional sequence space in early selection rounds and optimize for a well-folded, thermostable protein in later rounds. Directed evolution with titrated protein quality control is one example of the more general idea of applying the concept of engineering switchable behavior into the mutational landscape.

Another potential application for the principle of engineering entire mutational landscapes is combating antibiotic resistance (**Figure 4.19**, page 162). Antibiotic treatment could be combined with a Lon inhibitor, to incentivize pathogens to adapt via destabilizing mutations. Then, later in the time course, the Lon inhibitor would be replaced with a Lon agonist, strongly selecting against the pathogen. Further experimentation is needed to establish a proof of principle and to determine the optimal time-periods for the different phases of treatment.

4.5 Methods

4.5.0.1 Generation of ER2566 $\Delta folA \Delta thyA$ +Lon The ER2566 $\Delta folA/\Delta thyA$ +Lon strain was generated from ER2566 $\Delta folA/\Delta thyA$ -Lon by lambda red recombination using Support Protocol I from Thomason et al. 2014[46]. The pSim6 plasmid bearing the Lamda red genes linked to a temperature sensitive promoter and the pIB279 plasmid bearing the Kan-SacB positive-negative selection marker[47] were gifts from Carol Gross. The Kan-SacB cassette was amplified with 2 rounds of PCR using primers with 5' homology arms for the region upstream of the Lon gene (**Table 5.3**, page 174). The insertion fragment containing the Anderson consensus promoter[26] with homology arms for the region upstream of Lon in the ER2566 genome was amplified from primers using overlap extension PCR.

4.5.0.2 Western Blot with anti-Lon Polyclonal Antibody Western blots to detect Lon expression were performed as described in **Chapter 2 – Methods (Chapter 3.5.0.11, page 117)**, except that the lysate volume was calculated as (volume of culture for pellet) \times (OD600 of culture for pellet) \times (45 μ L Laemmli buffer), and 2 μ L (1:10,000) of polyclonal rabbit anti-Lon antibody (Biorbyte, cat #orb231326) was used as the secondary antibody.

4.5.0.3 Clustering of DHFR Positions by Mutational Response For clustering of positions, an in-house Python script was used for K-means clustering of positions into categories based on general mutational response at a position (i.e discarding the amino acid identities of the mutants). Spatial clustering was performed based on selection coefficient with distance between two positions calculated in the following steps: **1)** sorting the vectors of selection coefficients for each position, **2)** trimming the vectors to match vector lengths after discarding “no data” values, **3)** calculating a Δ vector by subtracting the two sorted and trimmed vectors, and finally calculating the distance as the mean of the absolute value of the Δ vector. For the first round, categories were seeded with virtual positions that have prototypical mutational profiles for the 5 categories (Beneficial, Tolerant, Mixed, Restricted, and Intolerant). From this first round, all positions in DHFR were categorized into initial clusters. In subsequent rounds, the virtual positions were removed and candidate positions were compared to the non-self positions populating each cluster. The distance between a candidate position and a cluster of positions is calculated as the average of the distance between the candidate position and the three closest non-self positions in the cluster. Clustering was performed over 10 rounds following the initial seeded round, and convergence was confirmed by observing that 5 repetitions gave identical clusters.

4.5.0.4 Structural Analysis of DHFR Selection The distance between the positions within each mutational response category and sites within the DHFR structure (hydride transfer site, M20 loop, core of the globular domain, and the beta-sheet surface beneath the active site) were determined using a model of the transition state provided by Phil Hanoian[31]. The representative atom for the hydride transfer site is the hydride atom in the transition state model. The representative atom for the adenine ring is C5 (C18 in the pdb). The representative atom for the core of the globular domain is the alpha carbon of I41. The representative atom for the beta sheet region is the alpha carbon of D114. For all cases, the distance is defined as the distance between the representative atom and the alpha carbon of the target position.

Mean atom neighbors for each residue on a structure were calculated using an in-house python script. The number of non-hydrogen atoms within an 8 Å shell of each non-hydrogen atom in the structure were counted and averaged for all non-hydrogen atoms at each side chain. These values we calculated for 4 crystal structures of DHFR (PDB IDs: 1RX1, 3QL3, 1RX4, 1RX5) and averaged over the set.

Solvent accessible surface accessible surface area (SASA) was calculated using the Getarea server[48] for 4 crystal structures of DHFR (1RX1, 3QL3, 1RX4, and 1RX5) representing different states in DHFR's catalytic cycle. For all positions in DHFR, if the residue had <20% SASA in any structure, the residue was classified as buried. All other residues were classified as exposed.

4.5.0.5 Profile Similarity Analysis We downloaded the DHFR alignment from OpenSeq.org[49], selected all bacterial DHFR sequences, and aligned the *E. coli* DHFR sequence to the MSA using MUSCLE[50]. Frequencies for each amino acid at each sequence position in the MSA were calculated from counts in each column, with absent amino acids given an arbitrarily low frequency of 0.0001. To compare the amino acid frequencies from the MSA to the selection

coefficients, we multiplied the selection coefficients by -1 and back-calculated frequencies using Boltzmann weighting using a temperature (5.44 for -Lon selection, and 6.95 for +Lon selection) that resulted in the mean sequence entropy to be within ± 0.01 of that of the MSA (0.50). Then, profile similarity at each sequence position was calculated as $1 -$ the Jensen-Shannon Divergence of the amino acid frequencies. Profile similarity was determined over columns corresponding to positions 2 -158 because the DHFR library begins at position 2 and the DHFR MSA cuts off after position 158.

References

- [1] Y. Cho, X. Zhang, K. F. Pobre, Y. Liu, D. L. Powers, J. W. Kelly, L. M. Gierasch, and E. T. Powers. Individual and collective contributions of chaperoning and degradation to protein homeostasis in e. coli. *Cell Rep*, 11(2):321–33, 2015.
- [2] E. T. Powers, D. L. Powers, and L. M. Gierasch. Foldeco: a model for proteostasis in e. coli. *Cell Rep*, 1(3):265–76, 2012.
- [3] S. L. Rutherford and S. Lindquist. Hsp90 as a capacitor for morphological evolution. *Nature*, 396(6709):336–42, 1998.
- [4] C. Queitsch, T. A. Sangster, and S. Lindquist. Hsp90 as a capacitor of phenotypic variation. *Nature*, 417(6889):618–24, 2002.
- [5] D. F. Jarosz and S. Lindquist. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. *Science*, 330(6012):1820–4, 2010.
- [6] N. Rohner, D. F. Jarosz, J. E. Kowalko, M. Yoshizawa, W. R. Jeffery, R. L. Borowsky, S. Lindquist, and C. J. Tabin. Cryptic variation in morphological evolution: Hsp90 as a capacitor for loss of eyes in cavefish. *Science*, 342(6164):1372–5, 2013.
- [7] N. Tokuriki and D. S. Tawfik. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol*, 19(5):596–604, 2009.
- [8] K. T. Wyganowski, M. Kaltenbach, and N. Tokuriki. Groel/es buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. *J Mol Biol*, 425(18):3403–14, 2013.

- [9] N. A. Moran. Accelerated evolution and muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A*, 93(7):2873–8, 1996.
- [10] S. Bershtein, W. Mu, and E. I. Shakhnovich. Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc Natl Acad Sci U S A*, 109(13):4857–62, 2012.
- [11] S. Bershtein, A. W. Serohijos, S. Bhattacharyya, M. Manhart, J. M. Choi, W. Mu, J. Zhou, and E. I. Shakhnovich. Protein homeostasis imposes a barrier on functional integration of horizontally transferred genes in bacteria. *PLoS Genet*, 11(10):e1005612, 2015.
- [12] J. V. Rodrigues, S. Bershtein, A. Li, E. R. Lozovsky, D. L. Hartl, and E. I. Shakhnovich. Biophysical principles predict fitness landscapes of drug resistance. *Proc Natl Acad Sci U S A*, 113(11):E1470–8, 2016.
- [13] S. Mizusawa and S. Gottesman. Protein degradation in escherichia coli: the lon gene controls the stability of sula protein. *Proc Natl Acad Sci U S A*, 80(2):358–62, 1983.
- [14] E. Gur and R. T. Sauer. Recognition of misfolded proteins by lon, a aaa(+) protease. *Genes Dev*, 22(16):2267–77, 2008.
- [15] E. Gur and R. T. Sauer. Degrons in protein substrates program the speed and operating efficiency of the aaa+ lon proteolytic machine. *Proc Natl Acad Sci U S A*, 106(44):18503–8, 2009.
- [16] C. K. Smith, T. A. Baker, and R. T. Sauer. Lon and clp family proteases and chaperones share homologous substrate-recognition domains. *Proc Natl Acad Sci U S A*, 96(12):6678–82, 1999.
- [17] S. Bershtein, W. Mu, A. W. Serohijos, J. Zhou, and E. I. Shakhnovich. Protein quality control

- acts on folding intermediates to shape the effects of mutations on organismal fitness. *Mol Cell*, 49(1):133–44, 2013.
- [18] L. saiSree, M. Reddy, and J. Gowrishankar. Is186 insertion at a hot spot in the lon promoter as a basis for lon protease deficiency of escherichia coli b: identification of a consensus target sequence for is186 transposition. *J Bacteriol*, 183(23):6943–6, 2001.
- [19] Richard E Lenski, Michael R Rose, Suzanne C Simpson, and Scott C Tadler. Long-term experimental evolution in escherichia coli. i. adaptation and divergence during 2,000 generations. *The American Naturalist*, 138(6):1315–1341, 1991.
- [20] M. J. Wisner, N. Ribeck, and R. E. Lenski. Long-term dynamics of adaptation in asexual populations. *Science*, 342(6164):1364–7, 2013.
- [21] O. Tenaillon, J. E. Barrick, N. Ribeck, D. E. Deatherage, J. L. Blanchard, A. Dasgupta, G. C. Wu, S. Wielgoss, S. Cruveiller, C. Medigue, D. Schneider, and R. E. Lenski. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature*, 536(7615):165–70, 2016.
- [22] B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai. The dynamics of molecular evolution over 60,000 generations. *Nature*, 551(7678):45–50, 2017.
- [23] B. P. Anton, A. Fomenkov, E. A. Raleigh, and M. Berkmen. Complete genome sequence of the engineered escherichia coli shuffle strains and their wild-type parents. *Genome Announc*, 4(2), 2016.
- [24] E. Gur, M. Vishkautzan, and R. T. Sauer. Protein unfolding and degradation by the aaa+ lon protease. *Protein Sci*, 21(2):268–78, 2012.

- [25] T. A. Phillips, R. A. VanBogelen, and F. C. Neidhardt. Ion gene product of escherichia coli is a heat-shock protein. *J Bacteriol*, 159(1):283–7, 1984.
- [26] 2006.
- [27] A. Espah Borujeni, A. S. Channarasappa, and H. M. Salis. Translation rate is controlled by coupled trade-offs between site accessibility, selective rna unfolding and sliding at upstream standby sites. *Nucleic Acids Res*, 42(4):2646–59, 2014.
- [28] H. M. Salis, E. A. Mirsky, and C. A. Voigt. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol*, 27(10):946–50, 2009.
- [29] D. Oyen, R. B. Fenwick, P. C. Aoto, R. L. Stanfield, I. A. Wilson, H. J. Dyson, and P. E. Wright. Defining the structural basis for allosteric product release from e. coli dihydrofolate reductase using nmr relaxation dispersion. *J Am Chem Soc*, 139(32):11233–11240, 2017.
- [30] M. Iwakura, K. Maki, H. Takahashi, T. Takenawa, A. Yokota, K. Katayanagi, T. Kamiyama, and K. Gekko. Evolutional design of a hyperactive cysteine- and methionine-free mutant of escherichia coli dihydrofolate reductase. *J Biol Chem*, 281(19):13234–46, 2006.
- [31] C. T. Liu, P. Hanoian, J. B. French, T. H. Pringle, S. Hammes-Schiffer, and S. J. Benkovic. Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proc Natl Acad Sci U S A*, 110(25):10159–64, 2013.
- [32] C. L. Araya, D. M. Fowler, W. Chen, I. Muniez, J. W. Kelly, and S. Fields. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc Natl Acad Sci U S A*, 109(42):16858–63, 2012.
- [33] P. Bandaru, N. H. Shah, M. Bhattacharyya, J. P. Barton, Y. Kondo, J. C. Cofsky, C. L. Gee,

- A. K. Chakraborty, T. Kortemme, R. Ranganathan, and J. Kuriyan. Deconstruction of the ras switching cycle through saturation mutagenesis. *Elife*, 6, 2017.
- [34] E. Firnberg, J. W. Labonte, J. J. Gray, and M. Ostermeier. A comprehensive, high-resolution map of a gene's fitness landscape. *Mol Biol Evol*, 31(6):1581–92, 2014.
- [35] R. T. Hietpas, C. Bank, J. D. Jensen, and D. N. A. Bolon. Shifting fitness landscapes in response to altered environments. *Evolution*, 67(12):3512–22, 2013.
- [36] L. Jiang, P. Mishra, R. T. Hietpas, K. B. Zeldovich, and D. N. Bolon. Latent effects of hsp90 mutants revealed at reduced expression levels. *PLoS Genet*, 9(6):e1003600, 2013.
- [37] B. P. Roscoe, K. M. Thayer, K. B. Zeldovich, D. Fushman, and D. N. Bolon. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J Mol Biol*, 425(8):1363–77, 2013.
- [38] M. A. Stiffler, D. R. Hekstra, and R. Ranganathan. Evolvability as a function of purifying selection in tem-1 beta-lactamase. *Cell*, 160(5):882–892, 2015.
- [39] Harry G. Saavedra, James O. Wrabl, Jeremy A. Anderson, Jing Li, and Vincent J. Hilser. Dynamic allostery can drive cold adaptation in enzymes. *Nature*, 558(7709):324–328, 2018.
- [40] T. P. Schrank, D. W. Bolen, and V. J. Hilser. Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. *Proc Natl Acad Sci U S A*, 106(40):16984–9, 2009.
- [41] D. Oyen, R. B. Fenwick, R. L. Stanfield, H. J. Dyson, and P. E. Wright. Cofactor-mediated conformational dynamics promote product release from escherichia coli dihydrofolate reductase via an allosteric pathway. *J Am Chem Soc*, 137(29):9459–68, 2015.

- [42] G. Bhabha, D. C. Ekiert, M. Jennewein, C. M. Zmasek, L. M. Tuttle, G. Kroon, H. J. Dyson, A. Godzik, I. A. Wilson, and P. E. Wright. Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat Struct Mol Biol*, 20(11):1243–9, 2013.
- [43] R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon, and G. W. Huisman. Improving catalytic function by prosar-driven enzyme evolution. *Nat Biotechnol*, 25(3):338–44, 2007.
- [44] G. J. Williams, C. Zhang, and J. S. Thorson. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nat Chem Biol*, 3(10):657–62, 2007.
- [45] N. Tokuriki and D. S. Tawfik. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*, 459(7247):668–73, 2009.
- [46] L. C. Thomason, J. A. Sawitzke, X. Li, N. Costantino, and D. L. Court. Recombineering: genetic engineering in bacteria using homologous recombination. *Curr Protoc Mol Biol*, 106:1.16.1–39, 2014.
- [47] I. C. Blomfield, V. Vaughn, R. F. Rest, and B. I. Eisenstein. Allelic exchange in escherichia coli using the bacillus subtilis sacb gene and a temperature-sensitive psc101 replicon. *Mol Microbiol*, 5(6):1447–57, 1991.
- [48] Robert Fraczekiewicz and Werner Braun. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *Journal of computational chemistry*, 19(3):319–333, 1998.
- [49] S. Ovchinnikov, H. Kamisetty, and D. Baker. Robust and accurate prediction of residue-

residue interactions across protein interfaces using evolutionary information. *Elife*, 3:e02030, 2014.

- [50] R. C. Edgar. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32(5):1792–7, 2004.

5 Appendix

Table 5.1: *E. coli* strains

Designation	Source or reference	Identifiers
ER2566	New England Biolabs	Cat# C2566I
ER2566 $\Delta folA/\Delta thyA$ (-Lon)	Reynolds et al. <i>Cell</i> 2011 [1]	
ER2566 $\Delta folA/\Delta thyA$ (+Lon)	This work (Chapter 4.5.0.1)	

Table 5.2: Plasmids

Designation	Source or reference	Additional information
SMT101	This work	Dual expression of DHFR and TYMS, <i>in vivo</i> assays, chloramphenicol (35 $\mu\text{g}/\text{mL}$ conc. <i>f</i>)
SMT201	This work	SMT101 with TET promoter for TYMS, <i>in vivo</i> assays, chloramphenicol (35 $\mu\text{g}/\text{mL}$ conc. <i>f</i>)
SMT202	This work	SMT201 with "AAGGAA" RBS for DHFR, <i>in vivo</i> assays, chloramphenicol (35 $\mu\text{g}/\text{mL}$ conc. <i>f</i>)
SMT203	This work	SMT201 with "AACGAG" RBS for DHFR, <i>in vivo</i> assays, chloramphenicol (35 $\mu\text{g}/\text{mL}$ conc. <i>f</i>)
SMT205	This work	SMT201 with "AATGAG" RBS for DHFR, <i>in vivo</i> assays, chloramphenicol (35 $\mu\text{g}/\text{mL}$ conc. <i>f</i>)
SMT215	This work	SMT205 with DHFR-FLAG-tag, western blot, chloramphenicol (35 $\mu\text{g}/\text{mL}$ conc. <i>f</i>)
KR101/SMT301	Reynolds et al. <i>Cell</i> 2011	His ₈ -tag, NiNTA purification, Heterologous expression, kanamycin (50 $\mu\text{g}/\text{mL}$ conc. <i>f</i>)
pSIM6	Blomfield, Vaughn, Rest, and Eisenstein, 1991	Lambda Red recombinase expression, temperature-sensitive promoter, ampicillin/carbenicilin (100 $\mu\text{g}/\text{mL}$ conc. <i>f</i>)
pIB279	Blomfield, Vaughn, Rest, and Eisenstein, 1991	KAN-SacB cassette for positive/negative selection, ampicillin/carbenicilin (100 $\mu\text{g}/\text{mL}$ conc. <i>f</i>)

Table 5.3: Primers

Designation	Forward primer	Reverse primer
WT_DHFR_pos2	catggtatatctccttattaaagttaaa	catggtatatctccttattaaagttaaa
WT_DHFR_pos22	catggtatatctcattattaaagttaaac	catggtatatctcattattaaagttaaac

(Continued on next page)

Designation	Forward primer	Reverse primer
WT_DHFR_pos3	gatcatggtatatctccttattaaagtt	gatcatggtatatctccttattaaagtt
WT_DHFR_pos4	actgatcatggtatatctccttattaa	actgatcatggtatatctccttattaa
WT_DHFR_pos5	cagactgatcatggtatatctccttatt	cagactgatcatggtatatctccttatt
WT_DHFR_pos6	aatcagactgatcatggtatatctcctt	aatcagactgatcatggtatatctcctt
WT_DHFR_pos7	cgcaatcagactgatcatggtatatct	cgcaatcagactgatcatggtatatct
WT_DHFR_pos8	cgccgcaatcagactgatc	cgccgcaatcagactgatc
WT_DHFR_pos82	cgccgcaatcagactgatc	cgccgcaatcagactgatc
WT_DHFR_pos9	taacgccgcaatcagactga	taacgccgcaatcagactga
WT_DHFR_pos10	cgctaacgccgcaatcag	cgctaacgccgcaatcag
WT_DHFR_pos11	taccgctaacgccgcaat	taccgctaacgccgcaat
WT_DHFR_pos12	atctaccgctaacgccgc	atctaccgctaacgccgc
WT_DHFR_pos13	gcgatctaccgctaacgc	gcgatctaccgctaacgc
WT_DHFR_pos14	aacgcgatctaccgctaac	aacgcgatctaccgctaac
WT_DHFR_pos15	gataacgcgatctaccgctaac	gataacgcgatctaccgctaac
WT_DHFR_pos16	gccgataacgcgatctacc	gccgataacgcgatctacc
WT_DHFR_pos17	catgccgataacgcgatctac	catgccgataacgcgatctac
WT_DHFR_pos18	ttccatgccgataacgcg	ttccatgccgataacgcg
WT_DHFR_pos19	gttttccatgccgataacgc	gttttccatgccgataacgc
WT_DHFR_pos20	ggcgttttccatgccgataacg	ggcgttttccatgccgataacg
WT_DHFR_pos21	catggcgttttccatgcc	catggcgttttccatgcc
WT_DHFR_pos22	cggcatggcgttttccat	cggcatggcgttttccat
WT_DHFR_pos222	cggcatggcgttttccatg	cggcatggcgttttccatg
WT_DHFR_pos23	ccacggcatggcgttttc	ccacggcatggcgttttc
WT_DHFR_pos24	gttccacggcatggcgtt	gttccacggcatggcgtt
WT_DHFR_pos25	caggttccacggcatggc	caggttccacggcatggc
WT_DHFR_pos26	aggcaggttccacggcat	aggcaggttccacggcat
WT_DHFR_pos27	ggcaggcaggttccacgg	ggcaggcaggttccacgg
WT_DHFR_pos28	atcggcaggcaggttcca	atcggcaggcaggttcca
WT_DHFR_pos29	gagatcggcaggcaggtt	gagatcggcaggcaggtt
WT_DHFR_pos30	ggcagatcggcaggcag	ggcagatcggcaggcag
WT_DHFR_pos31	ccaggcagatcggcagg	ccaggcagatcggcagg
WT_DHFR_pos32	aaaccaggcagatcggc	aaaccaggcagatcggc
WT_DHFR_pos33	ttaaaccaggcagatcgg	ttaaaccaggcagatcgg
WT_DHFR_pos34	gcgttaaaccaggcagat	gcgttaaaccaggcagat
WT_DHFR_pos35	gttgcgttaaaccagcga	gttgcgttaaaccagcga
WT_DHFR_pos36	ggtgttgcgttaaaccagg	ggtgttgcgttaaaccagg
WT_DHFR_pos37	taaggtgttgcgttaaaccagg	taaggtgttgcgttaaaccagg
WT_DHFR_pos38	atttaaggtgttgcgttaaaccagg	atttaaggtgttgcgttaaaccagg
WT_DHFR_pos39	tttattaaggtgttgcgttaaaccagg	tttattaaggtgttgcgttaaaccagg
WT_DHFR_pos40	gggtttattaaggtgttgcgttaaacc	gggtttattaaggtgttgcgttaaacc
WT_DHFR_pos41	cacgggtttattaaggtgttgcgt	cacgggtttattaaggtgttgcgt
WT_DHFR_pos42	aatcacgggtttattaaggtgttgc	aatcacgggtttattaaggtgttgc
WT_DHFR_pos422	aatcacgggtttattaaggtgttgc	aatcacgggtttattaaggtgttgc
WT_DHFR_pos43	cataatcacgggtttattaaggtgttg	cataatcacgggtttattaaggtgttg
WT_DHFR_pos432	cataatcacgggtttattaaggtgttg	cataatcacgggtttattaaggtgttg
WT_DHFR_pos44	gcccataatcacgggtttattaaggg	gcccataatcacgggtttattaaggg
WT_DHFR_pos45	gcggccataatcacggg	gcggccataatcacggg
WT_DHFR_pos46	atggcgccataatcac	atggcgccataatcac
WT_DHFR_pos47	ggtatggcgccataatc	ggtatggcgccataatc
WT_DHFR_pos48	ccaggtatggcgccata	ccaggtatggcgccata
WT_DHFR_pos49	ttcccaggtatggcgcc	ttcccaggtatggcgcc
WT_DHFR_pos50	tgattcccaggtatggcgcc	tgattcccaggtatggcgcc

(Continued on next page)

Designation	Forward primer	Reverse primer
WT_DHFR_pos51	gattgattcccaggtatggcgg	gattgattcccaggtatggcgg
WT_DHFR_pos52	accgattgattcccaggtatg	accgattgattcccaggtatg
WT_DHFR_pos53	acgaccgattgattcccag	acgaccgattgattcccag
WT_DHFR_pos54	cggacgaccgattgattcc	cggacgaccgattgattcc
WT_DHFR_pos55	caacggacgaccgattgattc	caacggacgaccgattgattc
WT_DHFR_pos56	tggcaacggacgaccgat	tggcaacggacgaccgat
WT_DHFR_pos57	tcctggcaacggacgacc	tcctggcaacggacgacc
WT_DHFR_pos58	gcgtcctggcaacggacg	gcgtcctggcaacggacg
WT_DHFR_pos59	tttgcgtcctggcaacgg	tttgcgtcctggcaacgg
WT_DHFR_pos60	atTTTTgcgtcctggcaac	atTTTTgcgtcctggcaac
WT_DHFR_pos61	aatatTTTTgcgtcctggcaac	aatatTTTTgcgtcctggcaac
WT_DHFR_pos62	gataatatTTTTgcgtcctggcaac	gataatatTTTTgcgtcctggcaac
WT_DHFR_pos63	gaggataatatTTTTgcgtcctggc	gaggataatatTTTTgcgtcctggc
WT_DHFR_pos64	gctgaggataatatTTTTgcgtcctg	gctgaggataatatTTTTgcgtcctg
WT_DHFR_pos65	actgctgaggataatatTTTTgcgtcct	actgctgaggataatatTTTTgcgtcct
WT_DHFR_pos66	ttgactgctgaggataatatTTTTgcg	ttgactgctgaggataatatTTTTgcg
WT_DHFR_pos662	ttgactgctgaggataatatTTTTgc	ttgactgctgaggataatatTTTTgc
WT_DHFR_pos67	cggttgactgctgaggataatatTTTTg	cggttgactgctgaggataatatTTTTg
WT_DHFR_pos672	cggttgactgctgaggataatatTTTTg	cggttgactgctgaggataatatTTTTg
WT_DHFR_pos68	accggttgactgctgag	accggttgactgctgag
WT_DHFR_pos69	cgtaccggttgactgct	cgtaccggttgactgct
WT_DHFR_pos70	gtccgtaccggttgact	gtccgtaccggttgact
WT_DHFR_pos71	atcgtccgtaccggttg	atcgtccgtaccggttg
WT_DHFR_pos72	gcgatcgtccgtaccgg	gcgatcgtccgtaccgg
WT_DHFR_pos73	tacgcgatcgtccgtacc	tacgcgatcgtccgtacc
WT_DHFR_pos732	tacgcgatcgtccgtacc	tacgcgatcgtccgtacc
WT_DHFR_pos74	cgttacgcgatcgtccgt	cgttacgcgatcgtccgt
WT_DHFR_pos742	cgttacgcgatcgtccgtac	cgttacgcgatcgtccgtac
WT_DHFR_pos75	ccacgttacgcgatcgtc	ccacgttacgcgatcgtc
WT_DHFR_pos76	caccacgttacgcgatc	caccacgttacgcgatc
WT_DHFR_pos77	cttcaccacgttacgcg	cttcaccacgttacgcg
WT_DHFR_pos78	cgacttcaccacgttacg	cgacttcaccacgttacg
WT_DHFR_pos79	caccgacttcaccacgttac	caccgacttcaccacgttac
WT_DHFR_pos80	atccaccgacttcaccacgttac	atccaccgacttcaccacgttac
WT_DHFR_pos802	atccaccgacttcaccac	atccaccgacttcaccac
WT_DHFR_pos81	ttcatccaccgacttcacca	ttcatccaccgacttcacca
WT_DHFR_pos82	ggcttcatccaccgacttcac	ggcttcatccaccgacttcac
WT_DHFR_pos822	ggcttcatccaccgacttcac	ggcttcatccaccgacttcac
WT_DHFR_pos83	gatggcttcatccaccgac	gatggcttcatccaccgac
WT_DHFR_pos84	cgcgatggcttcatccac	cgcgatggcttcatccac
WT_DHFR_pos842	cgcgatggcttcatccac	cgcgatggcttcatccac
WT_DHFR_pos85	cgccgcatggcttcatc	cgccgcatggcttcatc
WT_DHFR_pos86	acacgccgcatggcttc	acacgccgcatggcttc
WT_DHFR_pos87	accacacgccgcatggc	accacacgccgcatggc
WT_DHFR_pos88	gtcaccacacgccgcat	gtcaccacacgccgcat
WT_DHFR_pos89	tacgtcaccacacgccg	tacgtcaccacacgccg
WT_DHFR_pos892	tacgtcaccacacgccg	tacgtcaccacacgccg
WT_DHFR_pos90	tggtagctcaccacacgc	tggtagctcaccacacgc
WT_DHFR_pos91	ttctgtagctcaccacacgc	ttctgtagctcaccacacgc
WT_DHFR_pos92	gatttctgtagctcaccacacg	gatttctgtagctcaccacacg
WT_DHFR_pos93	catgatttctgtagctcaccacac	catgatttctgtagctcaccacac
WT_DHFR_pos94	caccatgatttctgtagctcacc	caccatgatttctgtagctcacc

(Continued on next page)

Designation	Forward primer	Reverse primer
WT_DHFR_pos95	aatcacatgatttctggtacgtca	aatcacatgatttctggtacgtca
WT_DHFR_pos952	aatcacatgatttctggtacgtc	aatcacatgatttctggtacgtc
WT_DHFR_pos96	gccaatcacatgatttctggtac	gccaatcacatgatttctggtac
WT_DHFR_pos97	gccgcaatcacatgattt	gccgcaatcacatgattt
WT_DHFR_pos98	accgcccaatcacatgatttc	accgcccaatcacatgatttc
WT_DHFR_pos99	gcgaccgcccaatcac	gcgaccgcccaatcac
WT_DHFR_pos100	aacgcgaccgcccaat	aacgcgaccgcccaat
WT_DHFR_pos101	ataaacgcgaccgcc	ataaacgcgaccgcc
WT_DHFR_pos102	ttcataaacgcgaccgcc	ttcataaacgcgaccgcc
WT_DHFR_pos103	ctgttcataaacgcgaccg	ctgttcataaacgcgaccg
WT_DHFR_pos104	gaactgttcataaacgcgacc	gaactgttcataaacgcgacc
WT_DHFR_pos1042	gaactgttcataaacgcgaccg	gaactgttcataaacgcgaccg
WT_DHFR_pos105	caagaactgttcataaacgcgac	caagaactgttcataaacgcgac
WT_DHFR_pos106	tggcaagaactgttcataaacgc	tggcaagaactgttcataaacgc
WT_DHFR_pos107	ttttggcaagaactgttcataaacg	ttttggcaagaactgttcataaacg
WT_DHFR_pos1072	ttttggcaagaactgttcataaacg	ttttggcaagaactgttcataaacg
WT_DHFR_pos108	cgcttttggcaagaactgttcataaa	cgcttttggcaagaactgttcataaa
WT_DHFR_pos109	ttgcgcttttggcaagaact	ttgcgcttttggcaagaact
WT_DHFR_pos110	tttttgcgcttttggcaagaac	tttttgcgcttttggcaagaac
WT_DHFR_pos111	cagttttgcgcttttggcaag	cagttttgcgcttttggcaag
WT_DHFR_pos112	atacagttttgcgcttttggcaa	atacagttttgcgcttttggcaa
WT_DHFR_pos113	cagatacagttttgcgcttttgg	cagatacagttttgcgcttttgg
WT_DHFR_pos114	cgtcagatacagttttgcgctttt	cgtcagatacagttttgcgctttt
WT_DHFR_pos115	atgcgctcagatacagttttgcg	atgcgctcagatacagttttgcg
WT_DHFR_pos116	gatatgcgctcagatacagttttgcg	gatatgcgctcagatacagttttgcg
WT_DHFR_pos117	gtcgatagcgctcagatacagttttg	gtcgatagcgctcagatacagttttg
WT_DHFR_pos118	tgcgctgatagcgctcagata	tgcgctgatagcgctcagata
WT_DHFR_pos1182	tgcgctgatagcgctcagatac	tgcgctgatagcgctcagatac
WT_DHFR_pos119	ttctgcgctgatagcgctca	ttctgcgctgatagcgctca
WT_DHFR_pos120	cacttctgcgctgatagcg	cacttctgcgctgatagcg
WT_DHFR_pos121	ttccacttctgcgctgatag	ttccacttctgcgctgatag
WT_DHFR_pos122	gccttccacttctgcgctc	gccttccacttctgcgctc
WT_DHFR_pos123	gtcgcttccacttctgc	gtcgcttccacttctgc
WT_DHFR_pos124	ggtgtcgcttccacttc	ggtgtcgcttccacttc
WT_DHFR_pos125	atgggtgtcgcttccac	atgggtgtcgcttccac
WT_DHFR_pos126	gaaatgggtgtcgcttcc	gaaatgggtgtcgcttcc
WT_DHFR_pos127	cgggaaatgggtgtcgcc	cgggaaatgggtgtcgcc
WT_DHFR_pos128	atccgggaaatgggtgtc	atccgggaaatgggtgtc
WT_DHFR_pos129	gtaatccgggaaatgggtgtc	gtaatccgggaaatgggtgtc
WT_DHFR_pos130	ctcgtaatccgggaaatggg	ctcgtaatccgggaaatggg
WT_DHFR_pos131	cggctcgtaatccgggaa	cggctcgtaatccgggaa
WT_DHFR_pos1312	cggctcgtaatccgggaaatg	cggctcgtaatccgggaaatg
WT_DHFR_pos132	atccggctcgtaatccgg	atccggctcgtaatccgg
WT_DHFR_pos133	gtcatccggctcgtaatcc	gtcatccggctcgtaatcc
WT_DHFR_pos134	ccagtcacccggctcgta	ccagtcacccggctcgta
WT_DHFR_pos135	ttccagtcacccggctc	ttccagtcacccggctc
WT_DHFR_pos1352	ttccagtcacccggctc	ttccagtcacccggctc
WT_DHFR_pos136	cgattccagtcacccgg	cgattccagtcacccgg
WT_DHFR_pos1362	cgattccagtcacccggc	cgattccagtcacccggc
WT_DHFR_pos137	taccgattccagtcacccg	taccgattccagtcacccg
WT_DHFR_pos138	gaataccgattccagtcaccc	gaataccgattccagtcaccc
WT_DHFR_pos139	gctgaataccgattccagtc	gctgaataccgattccagtc

(Continued on next page)

Designation	Forward primer	Reverse primer
WT_DHFR_pos140	ttcgtgtaataaccgattccca	ttcgtgtaataaccgattccca
WT_DHFR_pos1402	ttcgtgtaataaccgattcccag	ttcgtgtaataaccgattcccag
WT_DHFR_pos141	gaattcgctgaataaccgattccc	gaattcgctgaataaccgattccc
WT_DHFR_pos142	gtggaattcgctgaataaccgattc	gtggaattcgctgaataaccgattc
WT_DHFR_pos143	atcgtggaattcgctgaataacc	atcgtggaattcgctgaataacc
WT_DHFR_pos144	agcatcgtggaattcgctg	agcatcgtggaattcgctg
WT_DHFR_pos145	atcagcatcgtggaattcgc	atcagcatcgtggaattcgc
WT_DHFR_pos146	cgcatcagcatcgtggaatt	cgcatcagcatcgtggaatt
WT_DHFR_pos147	ctgcatcagcatcgtg	ctgcatcagcatcgtg
WT_DHFR_pos148	gttctgcatcagcatc	gttctgcatcagcatc
WT_DHFR_pos149	agagttctgcatcagc	agagttctgcatcagc
WT_DHFR_pos150	gtgagagttctgcatcag	gtgagagttctgcatcag
WT_DHFR_pos151	gctgtgagagttctgcatc	gctgtgagagttctgcatc
WT_DHFR_pos152	atagctgtgagagttctgcatc	atagctgtgagagttctgcatc
WT_DHFR_pos153	gcaatagctgtgagagttctgc	gcaatagctgtgagagttctgc
WT_DHFR_pos154	aaagcaatagctgtgagagttctg	aaagcaatagctgtgagagttctg
WT_DHFR_pos155	ctcaaagcaatagctgtgagagttc	ctcaaagcaatagctgtgagagttc
WT_DHFR_pos156	aatctcaaagcaatagctgtgagagttc	aatctcaaagcaatagctgtgagagttc
WT_DHFR_pos157	cagaatctcaaagcaatagctgtgagag	cagaatctcaaagcaatagctgtgagag
WT_DHFR_pos158	ctccagaatctcaaagcaatagctg	ctccagaatctcaaagcaatagctg
WT_DHFR_pos159	ccgctccagaatctcaaagc	ccgctccagaatctcaaagc

Designation	Primer sequence
TetDuet1_sense	ccgcttaagtgcgaacagaaagtaatcgtattgtacatccctac
TetDuet2_anti	gatagggatgtcaatctctatcactgatagggatgtacaatac
TetDuet3_sense	agagattgacatccctatcagtgatagagatactgagcacatcag
TetDuet4_anti	cttaaatgaattcggctcagtcgctcctgctgatgtgctcagatctc
TetDuet5_sense	cactgaccgaattcattaagaggagaaaggtaccatattggc
TetDuet_5flanking	ccgcttaagtgcgaacagaaag
TetDuet_3flanking	cggagatctgcatatggacc
SL1_amplicon_fwd	cactctttccctacacgacgctcttccgatctnnnactttaataacgagataaccatg
SL1_amplicon_rev	tgactggagttcagacgtgtgctcttccgatctnnnngatggcgccataat
SL2_amplicon_fwd	cactctttccctacacgacgctcttccgatctnnnacacctaataaaccgctg
SL2_amplicon_rev	tgactggagttcagacgtgtgctcttccgatctnnnncacgacgctgatggc
SL3_amplicon_fwd	cactctttccctacacgacgctcttccgatctnnnntgaagtcgggtgatgaa
SL3_amplicon_rev	tgactggagttcagacgtgtgctcttccgatctnnnngaagtcgggtgctgcc
SL4_amplicon_fwd	cactctttccctacacgacgctcttccgatctnnnncgacgcagaagtgaa
SL4_amplicon_rev	tgactggagttcagacgtgtgctcttccgatctnnnngcttgcgacgctg
Illumina adapter D501	aatgatacggcgaccaccgagatctacacatagaggcacactctttccctacacgac
Illumina adapter D502	aatgatacggcgaccaccgagatctacacatagaggcacactctttccctacacgac
Illumina adapter D503	aatgatacggcgaccaccgagatctacacacatagaggcacactctttccctacacgac
Illumina adapter D504	aatgatacggcgaccaccgagatctacacggctctgaacactctttccctacacgac
Illumina adapter D505	aatgatacggcgaccaccgagatctacacaggcgaagacactctttccctacacgac
Illumina adapter D506	aatgatacggcgaccaccgagatctacactaatcttaacactctttccctacacgac
Illumina adapter D507	aatgatacggcgaccaccgagatctacaccaggacgtacactctttccctacacgac
Illumina adapter D508	aatgatacggcgaccaccgagatctacacgactgacacactctttccctacacgac
Illumina adapter D701	caagcagaagacggcatacagagatcgagtaatgtgactggagttcagacgtg
Illumina adapter D702	caagcagaagacggcatacagagattctcggagtgactggagttcagacgtg
Illumina adapter D703	caagcagaagacggcatacagagataatgagcgggtgactggagttcagacgtg
Illumina adapter D704	caagcagaagacggcatacagagatggaatctcgtgactggagttcagacgtg
Illumina adapter D705	caagcagaagacggcatacagagatttgaatgtgactggagttcagacgtg

(Continued on next page)

Designation	Primer sequence
Illumina adapter D706	caagcagaagacggcatacagatacgaattcgtgactggagttcagacgtg
Illumina adapter D707	caagcagaagacggcatacagatagcttcaggtgactggagttcagacgtg
Illumina adapter D708	caagcagaagacggcatacagatgacgattagtgactggagttcagacgtg
Illumina adapter D709	caagcagaagacggcatacagatcatagccggtgactggagttcagacgtg
Illumina adapter D710	caagcagaagacggcatacagatctcgcggagtgactggagttcagacgtg
Illumina adapter D711	caagcagaagacggcatacagatgacgagagtgactggagttcagacgtg
Illumina adapter D712	caagcagaagacggcatacagatctatcgctgtgactggagttcagacgtg
KanSacb_round1_fwd	caggcatctggtgaataatccttttatgattttctatcaaaaaagagg
KanSacb_round1_rev	tcaatgcttcagaacgctcaggattcatgcttggtcggtcatttcgaac
KanSacb_round2_fwd	gtcaaaagcaaacctgtgctgattatggcaagccggaagcgcaacaggcatctggtgaataa
and Anderson_promoter_outer_fwd	
KanSacb_round2_rev	ccaccacatcgcgcagcggcaatacggggatttcaatgcttcagaacgctcaggattcat
and Anderson_promoter_outer_rev	
Anderson_promoter_outer_fwd	same as KanSacb_round2_fwd and Anderson_promoter_outer_fwd
and KanSacb_round2_fwd	
Anderson_promoter_inner_fwd	cctaggactgagctagctgtcaacgtcagtatatggggatgtttcccc
Anderson_promoter_inner_rev	gctagctcagtcctaggtataatgctagcaggatacctggcgaaattaaactaagagag
Anderson_promoter_outer_rev	same as KanSacb_round2_rev and Anderson_promoter_outer_rev
and KanSacb_round2_rev	

Table 5.5: Selection coefficients \pm Lon.

Mutation	Selection coefficient	-Lon		+Lon		
		Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
WT	0.0	0.0	0.0	0.0	0.0	0.0
M1A	ND	ND	ND	ND	ND	ND
M1C	ND	ND	ND	ND	ND	ND
M1D	ND	ND	ND	ND	ND	ND
M1E	ND	ND	ND	ND	ND	ND
M1F	ND	ND	ND	ND	ND	ND
M1G	ND	ND	ND	ND	ND	ND
M1H	ND	ND	ND	ND	ND	ND
M1I	Null	nan	nan	Null	nan	nan
M1K	Null	nan	nan	Null	nan	nan
M1L	Null	nan	nan	Null	nan	nan
M1N	ND	ND	ND	ND	ND	ND
M1P	ND	ND	ND	ND	ND	ND
M1Q	ND	ND	ND	ND	ND	ND
M1R	Null	nan	nan	Null	nan	nan
M1S	ND	ND	ND	ND	ND	ND
M1T	Null	nan	nan	Null	nan	nan
M1V	Null	nan	nan	Null	nan	nan
M1W	ND	ND	ND	ND	ND	ND
M1Y	ND	ND	ND	ND	ND	ND
M1*	ND	ND	ND	ND	ND	ND
I2A	Null	nan	nan	Null	nan	nan
I2C	Null	nan	nan	Null	nan	nan

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
I2D	Null	nan	nan	Null	nan	nan
I2E	Null	nan	nan	Null	nan	nan
I2F	Null	nan	nan	Null	nan	nan
I2G	Null	nan	nan	Null	nan	nan
I2H	ND	ND	ND	ND	ND	ND
I2K	0.216337	0.321471	0.105432	0.216337	0.321471	0.105432
I2L	Null	nan	nan	Null	nan	nan
I2M	-0.210279	0.586243	0.123327	-0.210279	0.586243	0.123327
I2N	Null	nan	nan	Null	nan	nan
I2P	Null	nan	nan	Null	nan	nan
I2Q	Null	nan	nan	Null	nan	nan
I2R	-2.157047	0.323759	0.284249	-2.157047	0.323759	0.284249
I2S	Null	nan	nan	Null	nan	nan
I2T	-1.567543	0.03467	0.207969	-1.567543	0.03467	0.207969
I2V	Null	nan	nan	Null	nan	nan
I2W	Null	nan	nan	Null	nan	nan
I2Y	Null	nan	nan	Null	nan	nan
I2*	Null	nan	nan	Null	nan	nan
S3A	ND	ND	ND	ND	ND	ND
S3C	-2.516449	0.745068	0.471011	-2.516449	0.745068	0.471011
S3D	-1.170709	0.465269	0.31078	-1.170709	0.465269	0.31078
S3E	-0.719255	0.409749	0.155627	-0.719255	0.409749	0.155627
S3F	-1.704388	0.361193	0.320096	-1.704388	0.361193	0.320096
S3G	-2.577865	0.447471	0.296933	-2.577865	0.447471	0.296933
S3H	ND	ND	ND	ND	ND	ND
S3I	-0.509124	0.267101	0.176759	-0.509124	0.267101	0.176759
S3K	0.024459	0.200856	0.078928	0.024459	0.200856	0.078928
S3L	-2.29468	0.315031	0.192154	-2.29468	0.315031	0.192154
S3M	-3.030999	0.773323	0.224491	-3.030999	0.773323	0.224491
S3N	0.401406	0.0535	0.067945	0.401406	0.0535	0.067945
S3P	Null	nan	nan	Null	nan	nan
S3Q	-0.895832	0.583995	0.153871	-0.895832	0.583995	0.153871
S3R	-2.73797	1.061132	0.685487	-2.73797	1.061132	0.685487
S3T	ND	ND	ND	ND	ND	ND
S3V	-2.13557	0.195311	0.214164	-2.13557	0.195311	0.214164
S3W	ND	ND	ND	ND	ND	ND
S3Y	Null	nan	nan	Null	nan	nan
S3*	Null	nan	nan	Null	nan	nan
L4A	-1.161436	0.490408	0.163341	-1.161436	0.490408	0.163341
L4C	-1.393341	0.337111	0.178913	-1.393341	0.337111	0.178913
L4D	Null	nan	nan	Null	nan	nan
L4E	-2.051432	0.694628	0.261317	-2.051432	0.694628	0.261317
L4F	1.141956	0.275313	0.105774	1.141956	0.275313	0.105774
L4G	-1.908804	0.296116	0.217841	-1.908804	0.296116	0.217841
L4H	0.391125	0.403592	0.097044	0.391125	0.403592	0.097044
L4I	0.392574	0.308739	0.183329	0.392574	0.308739	0.183329
L4K	ND	ND	ND	ND	ND	ND
L4M	Null	nan	nan	Null	nan	nan
L4N	-0.256804	0.188434	0.100904	-0.256804	0.188434	0.100904
L4P	-2.371455	0.53885	0.35189	-2.371455	0.53885	0.35189

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
L4Q	ND	ND	ND	ND	ND	ND
L4R	-1.647299	0.036419	0.498675	-1.647299	0.036419	0.498675
L4S	-0.701789	0.128076	0.138075	-0.701789	0.128076	0.138075
L4T	-0.546932	0.483893	0.120469	-0.546932	0.483893	0.120469
L4V	-0.960204	0.153566	0.1331	-0.960204	0.153566	0.1331
L4W	Null	nan	nan	Null	nan	nan
L4Y	0.750328	0.160922	0.093864	0.750328	0.160922	0.093864
L4*	Null	nan	nan	Null	nan	nan
I5A	-2.451661	0.186218	0.326731	-2.451661	0.186218	0.326731
I5C	-1.386073	0.454757	0.252096	-1.386073	0.454757	0.252096
I5D	Null	nan	nan	Null	nan	nan
I5E	-1.131798	0.393274	0.467673	-1.131798	0.393274	0.467673
I5F	-0.339196	0.053035	0.137193	-0.339196	0.053035	0.137193
I5G	-1.218013	0.24563	0.269344	-1.218013	0.24563	0.269344
I5H	-2.034677	0.239848	0.239041	-2.034677	0.239848	0.239041
I5K	-1.993555	0.326319	0.295294	-1.993555	0.326319	0.295294
I5L	-0.486626	0.039484	0.088369	-0.486626	0.039484	0.088369
I5M	-1.95295	0.223797	0.409923	-1.95295	0.223797	0.409923
I5N	-1.848083	0.244222	0.219494	-1.848083	0.244222	0.219494
I5P	ND	ND	ND	ND	ND	ND
I5Q	-2.302987	0.236346	0.480967	-2.302987	0.236346	0.480967
I5R	ND	ND	ND	ND	ND	ND
I5S	-2.5282	0.250386	0.494141	-2.5282	0.250386	0.494141
I5T	-2.582328	0.207108	0.255231	-2.582328	0.207108	0.255231
I5V	-2.15771	0.399273	0.332901	-2.15771	0.399273	0.332901
I5W	Null	nan	nan	Null	nan	nan
I5Y	-1.425945	0.122684	0.265415	-1.425945	0.122684	0.265415
I5*	Null	nan	nan	Null	nan	nan
A6C	0.241983	0.582579	0.142028	0.241983	0.582579	0.142028
A6D	Null	nan	nan	Null	nan	nan
A6E	Null	nan	nan	Null	nan	nan
A6F	0.557178	0.154437	0.08855	0.557178	0.154437	0.08855
A6G	-0.1111	0.230774	0.095983	-0.1111	0.230774	0.095983
A6H	ND	ND	ND	ND	ND	ND
A6I	0.939539	0.245201	0.118827	0.939539	0.245201	0.118827
A6K	Null	nan	nan	Null	nan	nan
A6L	-0.641696	0.151687	0.091163	-0.641696	0.151687	0.091163
A6M	0.316078	0.401527	0.245254	0.316078	0.401527	0.245254
A6N	0.540301	0.124123	0.083828	0.540301	0.124123	0.083828
A6P	Null	nan	nan	Null	nan	nan
A6Q	-1.542807	0.184192	0.267474	-1.542807	0.184192	0.267474
A6R	Null	nan	nan	Null	nan	nan
A6S	0.613944	0.200642	0.097189	0.613944	0.200642	0.097189
A6T	1.223413	0.297294	0.185485	1.223413	0.297294	0.185485
A6V	-0.002717	0.020712	0.124471	-0.002717	0.020712	0.124471
A6W	-2.984097	0.755598	0.374835	-2.984097	0.755598	0.374835
A6Y	0.494292	0.127158	0.212107	0.494292	0.127158	0.212107
A6*	ND	ND	ND	ND	ND	ND
A7C	0.136871	0.155831	0.170931	0.136871	0.155831	0.170931
A7D	Null	nan	nan	Null	nan	nan

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
A7E	Null	nan	nan	Null	nan	nan
A7F	Null	nan	nan	Null	nan	nan
A7G	-2.078709	0.258743	0.311049	-2.078709	0.258743	0.311049
A7H	ND	ND	ND	ND	ND	ND
A7I	-0.766823	0.446111	0.174069	-0.766823	0.446111	0.174069
A7K	Null	nan	nan	Null	nan	nan
A7L	-2.142025	0.574031	0.517863	-2.142025	0.574031	0.517863
A7M	-2.238888	0.014214	0.60888	-2.238888	0.014214	0.60888
A7N	-1.003936	0.134326	0.225629	-1.003936	0.134326	0.225629
A7P	ND	ND	ND	ND	ND	ND
A7Q	Null	nan	nan	Null	nan	nan
A7R	Null	nan	nan	Null	nan	nan
A7S	-1.049421	0.153562	0.16128	-1.049421	0.153562	0.16128
A7T	-0.818897	0.095212	0.16487	-0.818897	0.095212	0.16487
A7V	-1.463825	0.276595	0.198458	-1.463825	0.276595	0.198458
A7W	Null	nan	nan	Null	nan	nan
A7Y	Null	nan	nan	Null	nan	nan
A7*	Null	nan	nan	Null	nan	nan
L8A	-0.271751	0.090074	0.105893	-0.271751	0.090074	0.105893
L8C	ND	ND	ND	ND	ND	ND
L8D	ND	ND	ND	ND	ND	ND
L8E	0.006935	0.047428	0.035936	0.006935	0.047428	0.035936
L8F	Null	nan	nan	Null	nan	nan
L8G	-1.051835	0.039429	0.069601	-1.051835	0.039429	0.069601
L8H	-0.466469	0.323214	0.357106	-0.466469	0.323214	0.357106
L8I	ND	ND	ND	ND	ND	ND
L8K	-0.002378	0.093398	0.044722	-0.002378	0.093398	0.044722
L8M	-0.269702	0.088026	0.068231	-0.269702	0.088026	0.068231
L8N	-0.639517	0.716025	0.232531	-0.639517	0.716025	0.232531
L8P	-1.93959	0.233481	0.275879	-1.93959	0.233481	0.275879
L8Q	-1.027435	0.147561	0.175299	-1.027435	0.147561	0.175299
L8R	-0.371908	0.059623	0.067549	-0.371908	0.059623	0.067549
L8S	-0.90975	0.252685	0.113297	-0.90975	0.252685	0.113297
L8T	0.219171	0.152804	0.072058	0.219171	0.152804	0.072058
L8V	-1.04921	0.173473	0.124077	-1.04921	0.173473	0.124077
L8W	-0.483802	0.136972	0.080396	-0.483802	0.136972	0.080396
L8Y	ND	ND	ND	ND	ND	ND
L8*	-1.118843	0.283026	0.555126	-1.118843	0.283026	0.555126
A9C	-0.017291	0.076154	0.181286	-0.017291	0.076154	0.181286
A9D	-0.943036	0.117996	0.178323	-0.943036	0.117996	0.178323
A9E	-1.402183	0.535841	0.327389	-1.402183	0.535841	0.327389
A9F	-0.895022	0.147989	0.181844	-0.895022	0.147989	0.181844
A9G	-0.184594	0.096865	0.066764	-0.184594	0.096865	0.066764
A9H	ND	ND	ND	ND	ND	ND
A9I	0.141446	0.168962	0.0848	0.141446	0.168962	0.0848
A9K	-2.218018	0.563324	0.291309	-2.218018	0.563324	0.291309
A9L	-0.180059	0.150068	0.076292	-0.180059	0.150068	0.076292
A9M	-1.28883	0.430512	0.170479	-1.28883	0.430512	0.170479
A9N	1.003032	0.205308	0.147651	1.003032	0.205308	0.147651
A9P	-1.673289	0.344564	0.337409	-1.673289	0.344564	0.337409

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
A9Q	-1.149061	0.068907	0.155021	-1.149061	0.068907	0.155021
A9R	-2.19839	0.367233	0.251173	-2.19839	0.367233	0.251173
A9S	0.403174	0.224608	0.077807	0.403174	0.224608	0.077807
A9T	0.906075	0.328054	0.091164	0.906075	0.328054	0.091164
A9V	0.116547	0.038152	0.125508	0.116547	0.038152	0.125508
A9W	-1.422622	0.058916	0.263891	-1.422622	0.058916	0.263891
A9Y	-1.177207	0.310889	0.343219	-1.177207	0.310889	0.343219
A9*	Null	nan	nan	Null	nan	nan
V10A	0.478975	0.216778	0.071287	0.478975	0.216778	0.071287
V10C	ND	ND	ND	ND	ND	ND
V10D	-0.569728	0.362974	0.236961	-0.569728	0.362974	0.236961
V10E	-0.031127	0.052011	0.07948	-0.031127	0.052011	0.07948
V10F	-0.626348	0.265298	0.24832	-0.626348	0.265298	0.24832
V10G	-0.345622	0.094906	0.069719	-0.345622	0.094906	0.069719
V10H	-1.339311	0.645864	0.406927	-1.339311	0.645864	0.406927
V10I	-0.959336	0.187813	0.381375	-0.959336	0.187813	0.381375
V10K	-2.602692	0.381084	0.269559	-2.602692	0.381084	0.269559
V10L	-1.297352	0.301588	0.086965	-1.297352	0.301588	0.086965
V10M	-2.845304	1.289248	0.304053	-2.845304	1.289248	0.304053
V10N	-0.241096	0.118159	0.11277	-0.241096	0.118159	0.11277
V10P	0.34548	0.110329	0.095902	0.34548	0.110329	0.095902
V10Q	-1.832487	0.443114	0.198025	-1.832487	0.443114	0.198025
V10R	-2.217292	0.336805	0.169605	-2.217292	0.336805	0.169605
V10S	-0.203137	0.140978	0.048488	-0.203137	0.140978	0.048488
V10T	-0.926374	0.097807	0.246871	-0.926374	0.097807	0.246871
V10W	-1.582681	0.322656	0.235961	-1.582681	0.322656	0.235961
V10Y	-0.636287	0.421976	0.182306	-0.636287	0.421976	0.182306
V10*	ND	ND	ND	ND	ND	ND
D11A	0.318414	0.021914	0.135275	0.318414	0.021914	0.135275
D11C	-0.581506	0.282883	0.118533	-0.581506	0.282883	0.118533
D11E	0.269634	0.068533	0.053244	0.269634	0.068533	0.053244
D11F	-2.178431	0.285717	0.313724	-2.178431	0.285717	0.313724
D11G	0.211498	0.07788	0.046109	0.211498	0.07788	0.046109
D11H	-0.970961	0.271172	0.188365	-0.970961	0.271172	0.188365
D11I	ND	ND	ND	ND	ND	ND
D11K	0.837506	0.23658	0.063452	0.837506	0.23658	0.063452
D11L	-0.804088	0.094428	0.094366	-0.804088	0.094428	0.094366
D11M	-0.446691	0.050434	0.097279	-0.446691	0.050434	0.097279
D11N	-0.699912	0.053077	0.141804	-0.699912	0.053077	0.141804
D11P	-0.027216	0.078951	0.155961	-0.027216	0.078951	0.155961
D11Q	0.739642	0.040295	0.119775	0.739642	0.040295	0.119775
D11R	0.504602	0.189311	0.066074	0.504602	0.189311	0.066074
D11S	-0.22818	0.178366	0.081273	-0.22818	0.178366	0.081273
D11T	-0.028655	0.202869	0.102371	-0.028655	0.202869	0.102371
D11V	-0.556428	0.313019	0.127842	-0.556428	0.313019	0.127842
D11W	-0.988947	0.082266	0.105014	-0.988947	0.082266	0.105014
D11Y	-1.500944	0.175323	0.25574	-1.500944	0.175323	0.25574
D11*	Null	nan	nan	Null	nan	nan
R12A	-0.795197	0.323057	0.129756	-0.795197	0.323057	0.129756
R12C	0.314971	0.701727	0.338824	0.314971	0.701727	0.338824

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
R12D	0.172746	0.285918	0.096733	0.172746	0.285918	0.096733
R12E	0.593265	0.095759	0.085917	0.593265	0.095759	0.085917
R12F	-0.521906	0.224722	0.112031	-0.521906	0.224722	0.112031
R12G	-1.598468	0.214365	0.172479	-1.598468	0.214365	0.172479
R12H	Null	nan	nan	Null	nan	nan
R12I	-0.71642	0.489309	0.168855	-0.71642	0.489309	0.168855
R12K	0.911451	0.186208	0.134644	0.911451	0.186208	0.134644
R12L	0.46101	0.14072	0.101574	0.46101	0.14072	0.101574
R12M	0.518227	0.168345	0.082972	0.518227	0.168345	0.082972
R12N	0.491014	0.2082	0.124149	0.491014	0.2082	0.124149
R12P	-2.588745	1.005482	0.46565	-2.588745	1.005482	0.46565
R12Q	0.413412	0.088687	0.075474	0.413412	0.088687	0.075474
R12S	0.007539	0.084427	0.051758	0.007539	0.084427	0.051758
R12T	-0.653352	0.079648	0.062534	-0.653352	0.079648	0.062534
R12V	-0.001393	0.063022	0.072683	-0.001393	0.063022	0.072683
R12W	-0.07984	0.393917	0.104417	-0.07984	0.393917	0.104417
R12Y	0.149849	0.298427	0.171065	0.149849	0.298427	0.171065
R12*	Null	nan	nan	Null	nan	nan
V13A	-0.286673	0.453485	0.072367	-0.286673	0.453485	0.072367
V13C	-0.708609	0.134352	0.187444	-0.708609	0.134352	0.187444
V13D	-1.483453	0.140687	0.224413	-1.483453	0.140687	0.224413
V13E	-0.652633	0.126121	0.116234	-0.652633	0.126121	0.116234
V13F	-0.707558	0.040782	0.095948	-0.707558	0.040782	0.095948
V13G	-1.536397	0.465488	0.296449	-1.536397	0.465488	0.296449
V13H	-1.535606	0.280759	0.259818	-1.535606	0.280759	0.259818
V13I	0.018852	0.339353	0.144548	0.018852	0.339353	0.144548
V13K	-1.79452	0.708096	0.234409	-1.79452	0.708096	0.234409
V13L	-0.410226	0.064732	0.071829	-0.410226	0.064732	0.071829
V13M	-0.215035	0.229814	0.102651	-0.215035	0.229814	0.102651
V13N	-1.068206	0.211042	0.191657	-1.068206	0.211042	0.191657
V13P	-2.332183	0.043535	0.316494	-2.332183	0.043535	0.316494
V13Q	-1.167006	0.15141	0.205654	-1.167006	0.15141	0.205654
V13R	-1.402282	0.378748	0.224834	-1.402282	0.378748	0.224834
V13S	-0.416124	0.041978	0.114685	-0.416124	0.041978	0.114685
V13T	0.218408	0.114048	0.078966	0.218408	0.114048	0.078966
V13W	-0.911462	0.116718	0.124686	-0.911462	0.116718	0.124686
V13Y	-1.890299	0.233441	0.246746	-1.890299	0.233441	0.246746
V13*	Null	nan	nan	Null	nan	nan
I14A	0.376604	0.143324	0.089435	0.376604	0.143324	0.089435
I14C	0.445781	0.244458	0.153098	0.445781	0.244458	0.153098
I14D	Null	nan	nan	Null	nan	nan
I14E	Null	nan	nan	Null	nan	nan
I14F	Null	nan	nan	Null	nan	nan
I14G	-1.929402	0.282211	0.314878	-1.929402	0.282211	0.314878
I14H	Null	nan	nan	Null	nan	nan
I14K	ND	ND	ND	ND	ND	ND
I14L	1.092509	0.189243	0.070259	1.092509	0.189243	0.070259
I14M	-0.051279	0.351063	0.15049	-0.051279	0.351063	0.15049
I14N	Null	nan	nan	Null	nan	nan
I14P	-2.401558	0.076138	0.498568	-2.401558	0.076138	0.498568

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
I14Q	-3.37144	0.591933	0.077567	-3.37144	0.591933	0.077567
I14R	ND	ND	ND	ND	ND	ND
I14S	-1.833682	0.28113	0.236175	-1.833682	0.28113	0.236175
I14T	-0.837831	0.49151	0.208408	-0.837831	0.49151	0.208408
I14V	0.568391	0.168675	0.104564	0.568391	0.168675	0.104564
I14W	Null	nan	nan	Null	nan	nan
I14Y	Null	nan	nan	Null	nan	nan
I14*	Null	nan	nan	Null	nan	nan
G15A	-2.312781	0.386416	0.353418	-2.312781	0.386416	0.353418
G15C	Null	nan	nan	Null	nan	nan
G15D	Null	nan	nan	Null	nan	nan
G15E	-2.161786	0.251143	0.279907	-2.161786	0.251143	0.279907
G15F	ND	ND	ND	ND	ND	ND
G15H	Null	nan	nan	Null	nan	nan
G15I	Null	nan	nan	Null	nan	nan
G15K	Null	nan	nan	Null	nan	nan
G15L	-1.376205	0.174193	0.412731	-1.376205	0.174193	0.412731
G15M	Null	nan	nan	Null	nan	nan
G15N	Null	nan	nan	Null	nan	nan
G15P	-1.330027	0.178158	0.154713	-1.330027	0.178158	0.154713
G15Q	Null	nan	nan	Null	nan	nan
G15R	-1.7718	0.232846	0.526257	-1.7718	0.232846	0.526257
G15S	-2.556151	0.245625	0.41777	-2.556151	0.245625	0.41777
G15T	ND	ND	ND	ND	ND	ND
G15V	Null	nan	nan	Null	nan	nan
G15W	-2.138828	0.202429	0.560485	-2.138828	0.202429	0.560485
G15Y	Null	nan	nan	Null	nan	nan
G15*	Null	nan	nan	Null	nan	nan
M16A	0.416211	0.046354	0.090335	0.416211	0.046354	0.090335
M16C	0.164286	0.136389	0.075735	0.164286	0.136389	0.075735
M16D	-1.022921	0.143272	0.128131	-1.022921	0.143272	0.128131
M16E	-1.044876	0.154317	0.113263	-1.044876	0.154317	0.113263
M16F	-1.770745	0.257054	0.290081	-1.770745	0.257054	0.290081
M16G	-1.399253	0.055734	0.162046	-1.399253	0.055734	0.162046
M16H	-1.339961	0.270886	0.152303	-1.339961	0.270886	0.152303
M16I	ND	ND	ND	ND	ND	ND
M16K	-0.818132	0.130997	0.076409	-0.818132	0.130997	0.076409
M16L	-0.206775	0.064379	0.060221	-0.206775	0.064379	0.060221
M16N	-1.577818	0.15784	0.241005	-1.577818	0.15784	0.241005
M16P	ND	ND	ND	ND	ND	ND
M16Q	-0.345205	0.13556	0.048923	-0.345205	0.13556	0.048923
M16R	-0.077645	0.120757	0.049842	-0.077645	0.120757	0.049842
M16S	0.33604	0.109552	0.05056	0.33604	0.109552	0.05056
M16T	0.113482	0.154068	0.179096	0.113482	0.154068	0.179096
M16V	0.387904	0.154388	0.044553	0.387904	0.154388	0.044553
M16W	-0.304578	0.098788	0.073916	-0.304578	0.098788	0.073916
M16Y	-2.083463	0.529349	0.397027	-2.083463	0.529349	0.397027
M16*	-1.761302	0.629732	0.532787	-1.761302	0.629732	0.532787
E17A	-0.328916	0.647186	0.143826	-0.328916	0.647186	0.143826
E17C	-0.913174	0.228997	0.133819	-0.913174	0.228997	0.133819

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
E17D	0.026414	0.220889	0.125024	0.026414	0.220889	0.125024
E17F	-1.481166	0.423003	0.142628	-1.481166	0.423003	0.142628
E17G	0.091186	0.119136	0.065572	0.091186	0.119136	0.065572
E17H	-1.6484	0.095357	0.18713	-1.6484	0.095357	0.18713
E17I	Null	nan	nan	Null	nan	nan
E17K	-1.570776	0.3405	0.14159	-1.570776	0.3405	0.14159
E17L	-1.22373	0.070212	0.14163	-1.22373	0.070212	0.14163
E17M	-1.109718	0.534076	0.198776	-1.109718	0.534076	0.198776
E17N	-0.384086	0.043777	0.118606	-0.384086	0.043777	0.118606
E17P	-1.970535	0.210373	0.200475	-1.970535	0.210373	0.200475
E17Q	-1.414567	0.269621	0.151374	-1.414567	0.269621	0.151374
E17R	-1.878188	0.368895	0.192713	-1.878188	0.368895	0.192713
E17S	-0.477839	0.115026	0.046138	-0.477839	0.115026	0.046138
E17T	-1.650777	0.190188	0.203458	-1.650777	0.190188	0.203458
E17V	-1.880747	0.303201	0.203866	-1.880747	0.303201	0.203866
E17W	-2.053954	0.188754	0.218659	-2.053954	0.188754	0.218659
E17Y	-1.463397	0.318274	0.136113	-1.463397	0.318274	0.136113
E17*	Null	nan	nan	Null	nan	nan
N18A	-0.305781	0.449646	0.104075	-0.305781	0.449646	0.104075
N18C	-0.027512	0.424995	0.247092	-0.027512	0.424995	0.247092
N18D	ND	ND	ND	ND	ND	ND
N18E	-0.062783	0.024474	0.051692	-0.062783	0.024474	0.051692
N18F	0.439369	0.034984	0.17702	0.439369	0.034984	0.17702
N18G	-0.315947	0.119873	0.043162	-0.315947	0.119873	0.043162
N18H	Null	nan	nan	Null	nan	nan
N18I	Null	nan	nan	Null	nan	nan
N18K	-0.238715	0.306223	0.217307	-0.238715	0.306223	0.217307
N18L	0.408299	0.070163	0.099584	0.408299	0.070163	0.099584
N18M	-0.105091	0.176768	0.08205	-0.105091	0.176768	0.08205
N18P	Null	nan	nan	Null	nan	nan
N18Q	0.843206	0.45334	0.124276	0.843206	0.45334	0.124276
N18R	-0.267888	0.153631	0.124759	-0.267888	0.153631	0.124759
N18S	-0.053426	0.246324	0.162313	-0.053426	0.246324	0.162313
N18T	0.038048	0.287017	0.118136	0.038048	0.287017	0.118136
N18V	-0.968588	0.151916	0.090096	-0.968588	0.151916	0.090096
N18W	-0.28061	0.038561	0.069252	-0.28061	0.038561	0.069252
N18Y	Null	nan	nan	Null	nan	nan
N18*	Null	nan	nan	Null	nan	nan
A19C	0.345716	0.135495	0.093353	0.345716	0.135495	0.093353
A19D	0.267983	0.099377	0.135248	0.267983	0.099377	0.135248
A19E	-0.783896	0.068803	0.133958	-0.783896	0.068803	0.133958
A19F	0.537547	0.216537	0.055683	0.537547	0.216537	0.055683
A19G	0.30545	0.045818	0.074081	0.30545	0.045818	0.074081
A19H	0.223864	0.203507	0.11606	0.223864	0.203507	0.11606
A19I	0.170836	0.072881	0.096215	0.170836	0.072881	0.096215
A19K	-0.44047	0.039939	0.084187	-0.44047	0.039939	0.084187
A19L	-0.544193	0.103195	0.094423	-0.544193	0.103195	0.094423
A19M	-0.416844	0.144439	0.102628	-0.416844	0.144439	0.102628
A19N	0.384018	0.176496	0.078671	0.384018	0.176496	0.078671
A19P	Null	nan	nan	Null	nan	nan

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
A19Q	-0.072633	0.249359	0.134861	-0.072633	0.249359	0.134861
A19R	-0.906538	0.093999	0.099019	-0.906538	0.093999	0.099019
A19S	-0.131809	0.16508	0.096312	-0.131809	0.16508	0.096312
A19T	-0.196221	0.096964	0.178428	-0.196221	0.096964	0.178428
A19V	-1.003011	0.093678	0.107551	-1.003011	0.093678	0.107551
A19W	-0.024138	0.133061	0.057541	-0.024138	0.133061	0.057541
A19Y	0.488655	0.10238	0.113982	0.488655	0.10238	0.113982
A19*	Null	nan	nan	Null	nan	nan
M20A	-2.952453	0.455383	0.414994	-2.952453	0.455383	0.414994
M20C	-2.316384	0.104917	0.399624	-2.316384	0.104917	0.399624
M20D	Null	nan	nan	Null	nan	nan
M20E	-1.794705	0.340061	0.348184	-1.794705	0.340061	0.348184
M20F	-1.26574	0.697393	0.388693	-1.26574	0.697393	0.388693
M20G	-2.145402	0.563596	0.285249	-2.145402	0.563596	0.285249
M20H	Null	nan	nan	Null	nan	nan
M20I	Null	nan	nan	Null	nan	nan
M20K	Null	nan	nan	Null	nan	nan
M20L	-1.979247	0.213948	0.164297	-1.979247	0.213948	0.164297
M20N	-1.631221	0.82777	0.519933	-1.631221	0.82777	0.519933
M20P	-1.760191	0.824261	0.221361	-1.760191	0.824261	0.221361
M20Q	-1.191482	0.275039	0.473516	-1.191482	0.275039	0.473516
M20R	Null	nan	nan	Null	nan	nan
M20S	-2.244325	0.698233	0.25088	-2.244325	0.698233	0.25088
M20T	-2.007709	0.153448	0.368228	-2.007709	0.153448	0.368228
M20V	ND	ND	ND	ND	ND	ND
M20W	-1.454051	0.778336	0.43759	-1.454051	0.778336	0.43759
M20Y	ND	ND	ND	ND	ND	ND
M20*	Null	nan	nan	Null	nan	nan
P21A	-0.427802	0.296025	0.188561	-0.427802	0.296025	0.188561
P21C	0.511783	0.100556	0.130426	0.511783	0.100556	0.130426
P21D	ND	ND	ND	ND	ND	ND
P21E	-1.902267	0.154857	0.238557	-1.902267	0.154857	0.238557
P21F	-2.289288	0.605327	0.290554	-2.289288	0.605327	0.290554
P21G	-0.709654	0.151617	0.10484	-0.709654	0.151617	0.10484
P21H	-1.298227	0.251246	0.216328	-1.298227	0.251246	0.216328
P21I	-1.219042	0.239485	0.198152	-1.219042	0.239485	0.198152
P21K	-1.344451	0.295789	0.205662	-1.344451	0.295789	0.205662
P21L	-0.888313	0.304601	0.145473	-0.888313	0.304601	0.145473
P21M	-1.702244	0.392571	0.20963	-1.702244	0.392571	0.20963
P21N	-0.714316	0.413291	0.089046	-0.714316	0.413291	0.089046
P21Q	-0.51428	0.300292	0.189232	-0.51428	0.300292	0.189232
P21R	-1.312593	0.117944	0.197862	-1.312593	0.117944	0.197862
P21S	-0.03134	0.216023	0.077922	-0.03134	0.216023	0.077922
P21T	-0.957965	0.065934	0.121398	-0.957965	0.065934	0.121398
P21V	-0.46958	0.15194	0.094095	-0.46958	0.15194	0.094095
P21W	-3.406062	1.063193	0.56325	-3.406062	1.063193	0.56325
P21Y	-1.620441	0.690284	0.465029	-1.620441	0.690284	0.465029
P21*	Null	nan	nan	Null	nan	nan
W22A	Null	nan	nan	Null	nan	nan
W22C	Null	nan	nan	Null	nan	nan

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
W22D	Null	nan	nan	Null	nan	nan
W22E	Null	nan	nan	Null	nan	nan
W22F	0.074607	0.266427	0.326684	0.074607	0.266427	0.326684
W22G	Null	nan	nan	Null	nan	nan
W22H	-0.863454	0.107685	0.056627	-0.863454	0.107685	0.056627
W22I	Null	nan	nan	Null	nan	nan
W22K	Null	nan	nan	Null	nan	nan
W22L	ND	ND	ND	ND	ND	ND
W22M	Null	nan	nan	Null	nan	nan
W22N	Null	nan	nan	Null	nan	nan
W22P	Null	nan	nan	Null	nan	nan
W22Q	Null	nan	nan	Null	nan	nan
W22R	Null	nan	nan	Null	nan	nan
W22S	Null	nan	nan	Null	nan	nan
W22T	-1.453818	0.455808	0.577427	-1.453818	0.455808	0.577427
W22V	Null	nan	nan	Null	nan	nan
W22Y	Null	nan	nan	Null	nan	nan
W22*	Null	nan	nan	Null	nan	nan
N23A	0.40467	0.081462	0.121516	0.40467	0.081462	0.121516
N23C	0.127448	0.193512	0.250969	0.127448	0.193512	0.250969
N23D	Null	nan	nan	Null	nan	nan
N23E	-0.121633	0.145498	0.130539	-0.121633	0.145498	0.130539
N23F	-0.581096	0.342254	0.169329	-0.581096	0.342254	0.169329
N23G	0.152672	0.200105	0.036878	0.152672	0.200105	0.036878
N23H	-0.053407	0.210244	0.144104	-0.053407	0.210244	0.144104
N23I	Null	nan	nan	Null	nan	nan
N23K	-2.225531	0.853599	0.614486	-2.225531	0.853599	0.614486
N23L	-0.192579	0.044164	0.078926	-0.192579	0.044164	0.078926
N23M	0.119454	0.17694	0.191011	0.119454	0.17694	0.191011
N23P	-0.229267	0.060942	0.102062	-0.229267	0.060942	0.102062
N23Q	-0.028919	0.419589	0.150434	-0.028919	0.419589	0.150434
N23R	-0.115947	0.059166	0.101168	-0.115947	0.059166	0.101168
N23S	-0.344628	0.231749	0.164439	-0.344628	0.231749	0.164439
N23T	0.437661	0.080255	0.101564	0.437661	0.080255	0.101564
N23V	0.021819	0.135636	0.121079	0.021819	0.135636	0.121079
N23W	-0.341777	0.385049	0.146421	-0.341777	0.385049	0.146421
N23Y	Null	nan	nan	Null	nan	nan
N23*	Null	nan	nan	Null	nan	nan
L24A	0.496723	0.042753	0.155465	0.496723	0.042753	0.155465
L24C	-0.498789	0.426812	0.261029	-0.498789	0.426812	0.261029
L24D	0.436255	0.132051	0.152137	0.436255	0.132051	0.152137
L24E	0.241421	0.182859	0.084025	0.241421	0.182859	0.084025
L24F	-0.018711	0.234983	0.0791	-0.018711	0.234983	0.0791
L24G	0.917813	0.168818	0.077442	0.917813	0.168818	0.077442
L24H	ND	ND	ND	ND	ND	ND
L24I	0.453504	0.435146	0.102796	0.453504	0.435146	0.102796
L24K	-1.939271	0.233757	0.170486	-1.939271	0.233757	0.170486
L24M	-1.95974	0.006245	0.94526	-1.95974	0.006245	0.94526
L24N	0.116661	0.128919	0.188285	0.116661	0.128919	0.188285
L24P	Null	nan	nan	Null	nan	nan

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
L24Q	0.370577	0.384744	0.224196	0.370577	0.384744	0.224196
L24R	-1.844625	0.741494	0.207837	-1.844625	0.741494	0.207837
L24S	-0.725496	0.125245	0.143637	-0.725496	0.125245	0.143637
L24T	0.42817	0.109709	0.06643	0.42817	0.109709	0.06643
L24V	1.304508	0.277945	0.158621	1.304508	0.277945	0.158621
L24W	-0.857343	0.129796	0.072544	-0.857343	0.129796	0.072544
L24Y	-0.966791	0.313943	0.298337	-0.966791	0.313943	0.298337
L24*	Null	nan	nan	Null	nan	nan
P25A	-0.878856	0.144131	0.154196	-0.878856	0.144131	0.154196
P25C	-0.23344	0.077728	0.177836	-0.23344	0.077728	0.177836
P25D	-0.255347	0.047619	0.201071	-0.255347	0.047619	0.201071
P25E	-0.435161	0.132303	0.063518	-0.435161	0.132303	0.063518
P25F	0.516592	0.610048	0.114041	0.516592	0.610048	0.114041
P25G	-0.252092	0.061224	0.04793	-0.252092	0.061224	0.04793
P25H	-0.408289	0.292586	0.181165	-0.408289	0.292586	0.181165
P25I	0.350439	0.087691	0.199936	0.350439	0.087691	0.199936
P25K	-0.612932	0.075883	0.103435	-0.612932	0.075883	0.103435
P25L	-0.096139	0.061715	0.060595	-0.096139	0.061715	0.060595
P25M	0.080906	0.282491	0.069539	0.080906	0.282491	0.069539
P25N	0.122235	0.150586	0.094787	0.122235	0.150586	0.094787
P25Q	-0.857391	0.414966	0.195554	-0.857391	0.414966	0.195554
P25R	-0.282044	0.102807	0.10291	-0.282044	0.102807	0.10291
P25S	-0.038473	0.164995	0.071465	-0.038473	0.164995	0.071465
P25T	0.136078	0.131403	0.105966	0.136078	0.131403	0.105966
P25V	-0.564962	0.128648	0.045643	-0.564962	0.128648	0.045643
P25W	-0.249641	0.255095	0.062933	-0.249641	0.255095	0.062933
P25Y	0.185116	0.278576	0.188373	0.185116	0.278576	0.188373
P25*	Null	nan	nan	Null	nan	nan
A26C	-1.571587	0.328712	0.192469	-1.571587	0.328712	0.192469
A26D	ND	ND	ND	ND	ND	ND
A26E	-0.927874	0.419263	0.144605	-0.927874	0.419263	0.144605
A26F	-0.131288	0.168651	0.082831	-0.131288	0.168651	0.082831
A26G	-1.259458	0.083238	0.134515	-1.259458	0.083238	0.134515
A26H	-0.095878	0.163166	0.18108	-0.095878	0.163166	0.18108
A26I	-0.072548	0.113874	0.178493	-0.072548	0.113874	0.178493
A26K	-1.846991	0.497111	0.192401	-1.846991	0.497111	0.192401
A26L	-0.750856	0.217418	0.074583	-0.750856	0.217418	0.074583
A26M	-1.005042	0.176001	0.184443	-1.005042	0.176001	0.184443
A26N	-1.187332	0.388393	0.270842	-1.187332	0.388393	0.270842
A26P	-0.312186	0.093426	0.139418	-0.312186	0.093426	0.139418
A26Q	-0.605464	0.164121	0.198508	-0.605464	0.164121	0.198508
A26R	-1.236387	0.364731	0.194799	-1.236387	0.364731	0.194799
A26S	-1.098131	0.678746	0.118951	-1.098131	0.678746	0.118951
A26T	-1.000935	0.08504	0.172793	-1.000935	0.08504	0.172793
A26V	-1.29141	0.18305	0.123246	-1.29141	0.18305	0.123246
A26W	-1.085143	0.384518	0.153466	-1.085143	0.384518	0.153466
A26Y	0.085586	0.198518	0.243331	0.085586	0.198518	0.243331
A26*	-2.086997	0.117761	0.355434	-2.086997	0.117761	0.355434
D27A	Null	nan	nan	Null	nan	nan
D27C	Null	nan	nan	Null	nan	nan

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
D27E	-0.274413	0.142375	0.169612	-0.274413	0.142375	0.169612
D27F	Null	nan	nan	Null	nan	nan
D27G	-2.001598	0.324637	0.428258	-2.001598	0.324637	0.428258
D27H	Null	nan	nan	Null	nan	nan
D27I	Null	nan	nan	Null	nan	nan
D27K	Null	nan	nan	Null	nan	nan
D27L	Null	nan	nan	Null	nan	nan
D27M	Null	nan	nan	Null	nan	nan
D27N	Null	nan	nan	Null	nan	nan
D27P	Null	nan	nan	Null	nan	nan
D27Q	Null	nan	nan	Null	nan	nan
D27R	ND	ND	ND	ND	ND	ND
D27S	ND	ND	ND	ND	ND	ND
D27T	Null	nan	nan	Null	nan	nan
D27V	Null	nan	nan	Null	nan	nan
D27W	ND	ND	ND	ND	ND	ND
D27Y	Null	nan	nan	Null	nan	nan
D27*	-0.831948	0.061495	0.417552	-0.831948	0.061495	0.417552
L28A	-1.664887	0.274729	0.135294	-1.664887	0.274729	0.135294
L28C	ND	ND	ND	ND	ND	ND
L28D	Null	nan	nan	Null	nan	nan
L28E	-2.126523	0.268637	0.364319	-2.126523	0.268637	0.364319
L28F	0.326201	0.173329	0.076749	0.326201	0.173329	0.076749
L28G	-1.410193	0.491079	0.21511	-1.410193	0.491079	0.21511
L28H	Null	nan	nan	Null	nan	nan
L28I	Null	nan	nan	Null	nan	nan
L28K	-1.264724	0.309028	0.417853	-1.264724	0.309028	0.417853
L28M	-0.347377	0.183858	0.112752	-0.347377	0.183858	0.112752
L28N	-1.565162	0.799435	0.325581	-1.565162	0.799435	0.325581
L28P	-2.095162	0.259805	0.290252	-2.095162	0.259805	0.290252
L28Q	-2.696982	0.042751	0.341746	-2.696982	0.042751	0.341746
L28R	-2.408129	0.348599	0.343422	-2.408129	0.348599	0.343422
L28S	-1.614886	0.199514	0.287948	-1.614886	0.199514	0.287948
L28T	-1.889947	0.107359	0.187104	-1.889947	0.107359	0.187104
L28V	-2.207933	0.196483	0.217379	-2.207933	0.196483	0.217379
L28W	-1.967846	0.144236	0.292296	-1.967846	0.144236	0.292296
L28Y	0.2293	0.345707	0.066601	0.2293	0.345707	0.066601
L28*	ND	ND	ND	ND	ND	ND
A29C	-1.191341	0.266501	0.229722	-1.191341	0.266501	0.229722
A29D	Null	nan	nan	Null	nan	nan
A29E	-0.085666	0.166575	0.106446	-0.085666	0.166575	0.106446
A29F	-0.532149	0.236543	0.098511	-0.532149	0.236543	0.098511
A29G	-0.833975	0.065019	0.131868	-0.833975	0.065019	0.131868
A29H	-1.015092	0.081037	0.133243	-1.015092	0.081037	0.133243
A29I	-1.63115	0.274075	0.251386	-1.63115	0.274075	0.251386
A29K	-0.359333	0.148506	0.07013	-0.359333	0.148506	0.07013
A29L	-0.007351	0.132617	0.093461	-0.007351	0.132617	0.093461
A29M	-0.905594	0.368766	0.151408	-0.905594	0.368766	0.151408
A29N	-2.188096	0.412675	0.281644	-2.188096	0.412675	0.281644
A29P	-2.765418	0.190569	0.408673	-2.765418	0.190569	0.408673

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
A29Q	0.111734	0.181854	0.107904	0.111734	0.181854	0.107904
A29R	-0.611323	0.144011	0.104978	-0.611323	0.144011	0.104978
A29S	-0.321566	0.101669	0.108609	-0.321566	0.101669	0.108609
A29T	-0.398129	0.111986	0.100114	-0.398129	0.111986	0.100114
A29V	-0.855109	0.158222	0.107859	-0.855109	0.158222	0.107859
A29W	-0.316459	0.1629	0.124174	-0.316459	0.1629	0.124174
A29Y	-1.654693	0.586026	0.224053	-1.654693	0.586026	0.224053
A29*	Null	nan	nan	Null	nan	nan
W30A	0.078155	0.051413	0.122281	0.078155	0.051413	0.122281
W30C	ND	ND	ND	ND	ND	ND
W30D	-0.080607	0.098441	0.137786	-0.080607	0.098441	0.137786
W30E	0.484674	0.18724	0.101835	0.484674	0.18724	0.101835
W30F	0.935061	0.071351	0.096096	0.935061	0.071351	0.096096
W30G	0.443451	0.28156	0.106889	0.443451	0.28156	0.106889
W30H	0.887169	0.033887	0.130204	0.887169	0.033887	0.130204
W30I	-0.220035	0.077795	0.146307	-0.220035	0.077795	0.146307
W30K	0.868365	0.060574	0.089007	0.868365	0.060574	0.089007
W30L	0.707863	0.073076	0.066889	0.707863	0.073076	0.066889
W30M	0.928257	0.254354	0.102064	0.928257	0.254354	0.102064
W30N	0.463544	0.09413	0.097156	0.463544	0.09413	0.097156
W30P	-1.613784	0.136517	0.407458	-1.613784	0.136517	0.407458
W30Q	0.425701	0.150889	0.105486	0.425701	0.150889	0.105486
W30R	0.151376	0.08633	0.106493	0.151376	0.08633	0.106493
W30S	0.956178	0.207533	0.133782	0.956178	0.207533	0.133782
W30T	0.537258	0.096019	0.056521	0.537258	0.096019	0.056521
W30V	-0.405314	0.122772	0.098632	-0.405314	0.122772	0.098632
W30Y	0.706081	0.247766	0.08672	0.706081	0.247766	0.08672
W30*	Null	nan	nan	Null	nan	nan
F31A	-2.987076	1.324162	0.227321	-2.987076	1.324162	0.227321
F31C	-1.446568	0.322894	0.161984	-1.446568	0.322894	0.161984
F31D	Null	nan	nan	Null	nan	nan
F31E	-2.503525	0.62055	0.565209	-2.503525	0.62055	0.565209
F31G	-1.532062	0.027807	0.378169	-1.532062	0.027807	0.378169
F31H	ND	ND	ND	ND	ND	ND
F31I	-2.129959	0.300622	0.620896	-2.129959	0.300622	0.620896
F31K	Null	nan	nan	Null	nan	nan
F31L	-0.455298	0.186123	0.144905	-0.455298	0.186123	0.144905
F31M	-0.408941	0.164791	0.111605	-0.408941	0.164791	0.111605
F31N	Null	nan	nan	Null	nan	nan
F31P	-1.16903	0.52627	0.408165	-1.16903	0.52627	0.408165
F31Q	Null	nan	nan	Null	nan	nan
F31R	-2.466992	0.493189	0.30184	-2.466992	0.493189	0.30184
F31S	ND	ND	ND	ND	ND	ND
F31T	-2.696482	1.071249	0.368695	-2.696482	1.071249	0.368695
F31V	-0.450921	0.588961	0.073735	-0.450921	0.588961	0.073735
F31W	ND	ND	ND	ND	ND	ND
F31Y	-0.220766	0.475191	0.064998	-0.220766	0.475191	0.064998
F31*	Null	nan	nan	Null	nan	nan
K32A	-0.579102	0.171129	0.113336	-0.579102	0.171129	0.113336
K32C	-1.129088	0.096723	0.195013	-1.129088	0.096723	0.195013

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
K32D	Null	nan	nan	Null	nan	nan
K32E	-0.851386	0.117612	0.127962	-0.851386	0.117612	0.127962
K32F	-0.600714	0.180116	0.159963	-0.600714	0.180116	0.159963
K32G	-0.642965	0.074137	0.062167	-0.642965	0.074137	0.062167
K32H	-0.683171	0.632505	0.146427	-0.683171	0.632505	0.146427
K32I	-1.486358	0.388398	0.253302	-1.486358	0.388398	0.253302
K32L	-0.001218	0.154184	0.068894	-0.001218	0.154184	0.068894
K32M	-1.044189	0.193662	0.102658	-1.044189	0.193662	0.102658
K32N	Null	nan	nan	Null	nan	nan
K32P	ND	ND	ND	ND	ND	ND
K32Q	-0.611261	0.217402	0.175561	-0.611261	0.217402	0.175561
K32R	-0.556734	0.128541	0.076463	-0.556734	0.128541	0.076463
K32S	-0.275456	0.143988	0.107074	-0.275456	0.143988	0.107074
K32T	-0.819361	0.339446	0.089956	-0.819361	0.339446	0.089956
K32V	-0.310882	0.096436	0.098443	-0.310882	0.096436	0.098443
K32W	-1.324635	0.242909	0.172592	-1.324635	0.242909	0.172592
K32Y	-0.434718	0.029675	0.234742	-0.434718	0.029675	0.234742
K32*	Null	nan	nan	Null	nan	nan
R33A	0.116265	0.107393	0.053532	0.116265	0.107393	0.053532
R33C	Null	nan	nan	Null	nan	nan
R33D	0.180305	0.473208	0.161697	0.180305	0.473208	0.161697
R33E	0.298497	0.247897	0.051845	0.298497	0.247897	0.051845
R33F	0.187223	0.331118	0.149382	0.187223	0.331118	0.149382
R33G	-0.867165	0.096941	0.123607	-0.867165	0.096941	0.123607
R33H	ND	ND	ND	ND	ND	ND
R33I	0.32293	0.133949	0.137774	0.32293	0.133949	0.137774
R33K	0.160261	0.106066	0.113429	0.160261	0.106066	0.113429
R33L	-0.191206	0.217672	0.102816	-0.191206	0.217672	0.102816
R33M	0.183214	0.329993	0.109146	0.183214	0.329993	0.109146
R33N	0.109783	0.29855	0.16131	0.109783	0.29855	0.16131
R33P	-1.126476	0.419192	0.16513	-1.126476	0.419192	0.16513
R33Q	0.351843	0.11732	0.160298	0.351843	0.11732	0.160298
R33S	0.242787	0.170626	0.115551	0.242787	0.170626	0.115551
R33T	0.482882	0.103113	0.076813	0.482882	0.103113	0.076813
R33V	-0.073706	0.113611	0.073738	-0.073706	0.113611	0.073738
R33W	-0.442852	0.033617	0.098862	-0.442852	0.033617	0.098862
R33Y	0.288588	0.227725	0.123754	0.288588	0.227725	0.123754
R33*	Null	nan	nan	Null	nan	nan
N34A	-0.341949	0.126248	0.085619	-0.341949	0.126248	0.085619
N34C	0.019405	0.394869	0.113889	0.019405	0.394869	0.113889
N34D	Null	nan	nan	Null	nan	nan
N34E	-0.032796	0.171051	0.06639	-0.032796	0.171051	0.06639
N34F	-0.417802	0.356546	0.09672	-0.417802	0.356546	0.09672
N34G	0.333491	0.141647	0.063557	0.333491	0.141647	0.063557
N34H	0.149803	0.59553	0.212415	0.149803	0.59553	0.212415
N34I	Null	nan	nan	Null	nan	nan
N34K	0.354709	0.503894	0.181616	0.354709	0.503894	0.181616
N34L	-0.164276	0.130113	0.069167	-0.164276	0.130113	0.069167
N34M	-0.538833	0.299627	0.102728	-0.538833	0.299627	0.102728
N34P	Null	nan	nan	Null	nan	nan

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
N34Q	0.546118	0.312706	0.089787	0.546118	0.312706	0.089787
N34R	0.109312	0.117446	0.078851	0.109312	0.117446	0.078851
N34S	0.078856	0.164841	0.101761	0.078856	0.164841	0.101761
N34T	-0.495738	0.258797	0.224115	-0.495738	0.258797	0.224115
N34V	-0.194464	0.148251	0.063157	-0.194464	0.148251	0.063157
N34W	-0.187578	0.142718	0.071514	-0.187578	0.142718	0.071514
N34Y	ND	ND	ND	ND	ND	ND
N34*	Null	nan	nan	Null	nan	nan
T35A	ND	ND	ND	ND	ND	ND
T35C	-0.708241	0.053135	0.178574	-0.708241	0.053135	0.178574
T35D	Null	nan	nan	Null	nan	nan
T35E	Null	nan	nan	Null	nan	nan
T35F	Null	nan	nan	Null	nan	nan
T35G	-0.706308	0.176002	0.124556	-0.706308	0.176002	0.124556
T35H	Null	nan	nan	Null	nan	nan
T35I	Null	nan	nan	Null	nan	nan
T35K	Null	nan	nan	Null	nan	nan
T35L	-2.641566	0.245972	0.279363	-2.641566	0.245972	0.279363
T35M	Null	nan	nan	Null	nan	nan
T35N	Null	nan	nan	Null	nan	nan
T35P	Null	nan	nan	Null	nan	nan
T35Q	Null	nan	nan	Null	nan	nan
T35R	-2.137438	0.193388	0.219522	-2.137438	0.193388	0.219522
T35S	-0.397348	0.486299	0.215718	-0.397348	0.486299	0.215718
T35V	-1.588366	0.547213	0.258468	-1.588366	0.547213	0.258468
T35W	Null	nan	nan	Null	nan	nan
T35Y	Null	nan	nan	Null	nan	nan
T35*	Null	nan	nan	Null	nan	nan
L36A	-1.262919	0.357953	0.219044	-1.262919	0.357953	0.219044
L36C	-1.07775	0.063861	0.160196	-1.07775	0.063861	0.160196
L36D	-0.285615	0.150051	0.110064	-0.285615	0.150051	0.110064
L36E	-0.209509	0.160342	0.058938	-0.209509	0.160342	0.058938
L36F	0.349927	0.109555	0.116812	0.349927	0.109555	0.116812
L36G	0.085059	0.085172	0.05434	0.085059	0.085172	0.05434
L36H	-0.120345	0.156705	0.135038	-0.120345	0.156705	0.135038
L36I	-0.055973	0.337193	0.143276	-0.055973	0.337193	0.143276
L36K	0.417831	0.093832	0.094142	0.417831	0.093832	0.094142
L36M	-0.106503	0.069766	0.106603	-0.106503	0.069766	0.106603
L36N	0.41145	0.02335	0.115041	0.41145	0.02335	0.115041
L36P	-2.46644	0.298496	0.266715	-2.46644	0.298496	0.266715
L36Q	0.385088	0.235541	0.146992	0.385088	0.235541	0.146992
L36R	0.188268	0.34108	0.075778	0.188268	0.34108	0.075778
L36S	-0.12961	0.28065	0.090998	-0.12961	0.28065	0.090998
L36T	-0.023961	0.083124	0.106688	-0.023961	0.083124	0.106688
L36V	-0.939099	0.054275	0.144993	-0.939099	0.054275	0.144993
L36W	-0.24282	0.105327	0.081036	-0.24282	0.105327	0.081036
L36Y	-0.012174	0.269969	0.099802	-0.012174	0.269969	0.099802
L36*	ND	ND	ND	ND	ND	ND
N37A	-0.245368	0.300699	0.062899	-0.245368	0.300699	0.062899
N37C	-0.663034	0.117711	0.153942	-0.663034	0.117711	0.153942

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
N37D	-0.025042	0.078056	0.137428	-0.025042	0.078056	0.137428
N37E	0.222869	0.159576	0.060092	0.222869	0.159576	0.060092
N37F	0.222095	0.239624	0.087537	0.222095	0.239624	0.087537
N37G	-0.244047	0.04969	0.053057	-0.244047	0.04969	0.053057
N37H	-0.160231	0.046852	0.100301	-0.160231	0.046852	0.100301
N37I	-0.330718	0.181195	0.10524	-0.330718	0.181195	0.10524
N37K	0.286752	0.048142	0.095373	0.286752	0.048142	0.095373
N37L	0.018227	0.080408	0.078587	0.018227	0.080408	0.078587
N37M	0.291724	0.40991	0.142371	0.291724	0.40991	0.142371
N37P	-0.731096	0.116603	0.079996	-0.731096	0.116603	0.079996
N37Q	0.03753	0.154154	0.091237	0.03753	0.154154	0.091237
N37R	0.130762	0.07305	0.033701	0.130762	0.07305	0.033701
N37S	0.041464	0.231181	0.107642	0.041464	0.231181	0.107642
N37T	-0.098031	0.231252	0.08141	-0.098031	0.231252	0.08141
N37V	0.135606	0.12195	0.079502	0.135606	0.12195	0.079502
N37W	0.103785	0.219981	0.107954	0.103785	0.219981	0.107954
N37Y	0.272424	0.09052	0.084744	0.272424	0.09052	0.084744
N37*	Null	nan	nan	Null	nan	nan
K38A	-0.028098	0.386311	0.100592	-0.028098	0.386311	0.100592
K38C	0.55229	0.0823	0.158633	0.55229	0.0823	0.158633
K38D	0.155953	0.070484	0.113384	0.155953	0.070484	0.113384
K38E	-0.050809	0.172648	0.12745	-0.050809	0.172648	0.12745
K38F	0.456405	0.175148	0.07528	0.456405	0.175148	0.07528
K38G	0.098357	0.226721	0.096928	0.098357	0.226721	0.096928
K38H	-0.171247	0.084308	0.040732	-0.171247	0.084308	0.040732
K38I	0.216307	0.118698	0.075924	0.216307	0.118698	0.075924
K38L	0.287453	0.068505	0.085708	0.287453	0.068505	0.085708
K38M	-0.023711	0.090109	0.194972	-0.023711	0.090109	0.194972
K38N	1.030419	0.191168	0.12758	1.030419	0.191168	0.12758
K38P	-1.777301	0.451629	0.266473	-1.777301	0.451629	0.266473
K38Q	-1.125653	0.189994	0.242056	-1.125653	0.189994	0.242056
K38R	0.266487	0.079194	0.09988	0.266487	0.079194	0.09988
K38S	0.409376	0.125635	0.065827	0.409376	0.125635	0.065827
K38T	0.457957	0.166856	0.078843	0.457957	0.166856	0.078843
K38V	0.140045	0.248012	0.075002	0.140045	0.248012	0.075002
K38W	-0.131523	0.187183	0.128423	-0.131523	0.187183	0.128423
K38Y	0.260847	0.286237	0.114772	0.260847	0.286237	0.114772
K38*	Null	nan	nan	Null	nan	nan
P39A	0.082372	0.149977	0.094154	0.082372	0.149977	0.094154
P39C	-0.139599	0.468341	0.095781	-0.139599	0.468341	0.095781
P39D	-0.736985	0.380762	0.245797	-0.736985	0.380762	0.245797
P39E	0.530026	0.158968	0.094346	0.530026	0.158968	0.094346
P39F	0.289087	0.234355	0.134185	0.289087	0.234355	0.134185
P39G	-0.203398	0.156845	0.073812	-0.203398	0.156845	0.073812
P39H	0.830231	0.205231	0.125331	0.830231	0.205231	0.125331
P39I	-0.259664	0.353987	0.114311	-0.259664	0.353987	0.114311
P39K	0.144647	0.104894	0.085337	0.144647	0.104894	0.085337
P39L	0.225135	0.288434	0.087992	0.225135	0.288434	0.087992
P39M	0.33667	0.031107	0.063052	0.33667	0.031107	0.063052
P39N	0.405075	0.123503	0.107448	0.405075	0.123503	0.107448

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
P39Q	-0.037447	0.03116	0.152157	-0.037447	0.03116	0.152157
P39R	-0.095591	0.20842	0.053189	-0.095591	0.20842	0.053189
P39S	0.024442	0.115922	0.089889	0.024442	0.115922	0.089889
P39T	0.640166	0.356766	0.089294	0.640166	0.356766	0.089294
P39V	0.344861	0.034472	0.081279	0.344861	0.034472	0.081279
P39W	0.135082	0.20271	0.089961	0.135082	0.20271	0.089961
P39Y	0.305947	0.281298	0.178341	0.305947	0.281298	0.178341
P39*	Null	nan	nan	Null	nan	nan
V40A	0.517594	0.08174	0.135126	0.517594	0.08174	0.135126
V40C	0.01787	0.376553	0.094831	0.01787	0.376553	0.094831
V40D	-2.7074	0.046463	0.569111	-2.7074	0.046463	0.569111
V40E	-1.10461	0.301997	0.231409	-1.10461	0.301997	0.231409
V40F	0.459582	0.15513	0.09408	0.459582	0.15513	0.09408
V40G	-0.050439	0.182281	0.043465	-0.050439	0.182281	0.043465
V40H	0.101972	0.100913	0.062124	0.101972	0.100913	0.062124
V40I	0.049057	0.344475	0.16615	0.049057	0.344475	0.16615
V40K	-2.459289	0.20735	0.395031	-2.459289	0.20735	0.395031
V40L	0.232409	0.062742	0.091512	0.232409	0.062742	0.091512
V40M	0.070021	0.353352	0.141629	0.070021	0.353352	0.141629
V40N	0.019185	0.165538	0.1435	0.019185	0.165538	0.1435
V40P	-2.014964	0.702033	0.499943	-2.014964	0.702033	0.499943
V40Q	ND	ND	ND	ND	ND	ND
V40R	-2.518108	0.386187	0.326798	-2.518108	0.386187	0.326798
V40S	0.266599	0.035784	0.089947	0.266599	0.035784	0.089947
V40T	0.542587	0.075873	0.108677	0.542587	0.075873	0.108677
V40W	-2.012247	0.471054	0.29267	-2.012247	0.471054	0.29267
V40Y	-1.842549	0.358514	0.332106	-1.842549	0.358514	0.332106
V40*	ND	ND	ND	ND	ND	ND
I41A	0.714484	0.019548	0.101286	0.714484	0.019548	0.101286
I41C	0.921558	0.012893	0.089076	0.921558	0.012893	0.089076
I41D	Null	nan	nan	Null	nan	nan
I41E	ND	ND	ND	ND	ND	ND
I41F	-0.091335	0.235015	0.084055	-0.091335	0.235015	0.084055
I41G	-1.123459	0.717119	0.174521	-1.123459	0.717119	0.174521
I41H	ND	ND	ND	ND	ND	ND
I41K	ND	ND	ND	ND	ND	ND
I41L	-0.089273	0.090417	0.060921	-0.089273	0.090417	0.060921
I41M	-0.447399	0.316449	0.063513	-0.447399	0.316449	0.063513
I41N	-2.232058	0.019251	0.198777	-2.232058	0.019251	0.198777
I41P	-1.831342	0.666337	0.228299	-1.831342	0.666337	0.228299
I41Q	-1.459682	0.52517	0.400591	-1.459682	0.52517	0.400591
I41R	-1.671415	0.330948	0.332794	-1.671415	0.330948	0.332794
I41S	-1.122081	0.15252	0.072034	-1.122081	0.15252	0.072034
I41T	0.179708	0.063986	0.082998	0.179708	0.063986	0.082998
I41V	0.730428	0.070202	0.073781	0.730428	0.070202	0.073781
I41W	-0.700809	0.72271	0.208452	-0.700809	0.72271	0.208452
I41Y	ND	ND	ND	ND	ND	ND
I41*	Null	nan	nan	Null	nan	nan
M42A	-0.34117	0.331984	0.110953	-0.34117	0.331984	0.110953
M42C	ND	ND	ND	ND	ND	ND

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
M42D	Null	nan	nan	Null	nan	nan
M42E	-1.027186	0.315943	0.126216	-1.027186	0.315943	0.126216
M42F	0.537107	0.264764	0.059187	0.537107	0.264764	0.059187
M42G	-0.672364	0.063199	0.071058	-0.672364	0.063199	0.071058
M42H	0.256883	0.273878	0.151554	0.256883	0.273878	0.151554
M42I	Null	nan	nan	Null	nan	nan
M42K	Null	nan	nan	Null	nan	nan
M42L	0.103523	0.185331	0.079953	0.103523	0.185331	0.079953
M42N	-1.470634	0.000567	0.379315	-1.470634	0.000567	0.379315
M42P	Null	nan	nan	Null	nan	nan
M42Q	0.364309	0.171945	0.120694	0.364309	0.171945	0.120694
M42R	-2.248877	0.00069	0.294124	-2.248877	0.00069	0.294124
M42S	0.163867	0.262007	0.115526	0.163867	0.262007	0.115526
M42T	-1.036864	0.431277	0.250432	-1.036864	0.431277	0.250432
M42V	-2.711417	0.871677	0.349212	-2.711417	0.871677	0.349212
M42W	-1.863172	0.332898	0.183971	-1.863172	0.332898	0.183971
M42Y	0.467547	0.346154	0.133034	0.467547	0.346154	0.133034
M42*	ND	ND	ND	ND	ND	ND
G43A	-1.711839	0.23372	0.453048	-1.711839	0.23372	0.453048
G43C	Null	nan	nan	Null	nan	nan
G43D	Null	nan	nan	Null	nan	nan
G43E	-1.103656	0.308623	0.358983	-1.103656	0.308623	0.358983
G43F	Null	nan	nan	Null	nan	nan
G43H	Null	nan	nan	Null	nan	nan
G43I	Null	nan	nan	Null	nan	nan
G43K	ND	ND	ND	ND	ND	ND
G43L	-1.319405	0.010135	0.380589	-1.319405	0.010135	0.380589
G43M	Null	nan	nan	Null	nan	nan
G43N	Null	nan	nan	Null	nan	nan
G43P	Null	nan	nan	Null	nan	nan
G43Q	Null	nan	nan	Null	nan	nan
G43R	-0.797322	0.111112	0.208134	-0.797322	0.111112	0.208134
G43S	-2.035288	0.612398	0.869731	-2.035288	0.612398	0.869731
G43T	Null	nan	nan	Null	nan	nan
G43V	-1.441452	0.018831	0.324966	-1.441452	0.018831	0.324966
G43W	ND	ND	ND	ND	ND	ND
G43Y	Null	nan	nan	Null	nan	nan
G43*	Null	nan	nan	Null	nan	nan
R44A	0.254575	0.073703	0.073528	0.254575	0.073703	0.073528
R44C	0.462548	0.140561	0.109483	0.462548	0.140561	0.109483
R44D	-1.168202	0.067047	0.240714	-1.168202	0.067047	0.240714
R44E	-0.899046	0.266334	0.089699	-0.899046	0.266334	0.089699
R44F	-0.474518	0.18006	0.09224	-0.474518	0.18006	0.09224
R44G	0.534574	0.068816	0.049057	0.534574	0.068816	0.049057
R44H	ND	ND	ND	ND	ND	ND
R44I	-0.497985	0.209001	0.164196	-0.497985	0.209001	0.164196
R44K	0.429508	0.092357	0.043068	0.429508	0.092357	0.043068
R44L	-1.08727	0.517648	0.088494	-1.08727	0.517648	0.088494
R44M	0.065684	0.110757	0.073811	0.065684	0.110757	0.073811
R44N	1.251744	0.032646	0.072964	1.251744	0.032646	0.072964

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
R44P	-2.69298	0.479227	0.236652	-2.69298	0.479227	0.236652
R44Q	-0.244016	0.122768	0.07771	-0.244016	0.122768	0.07771
R44S	0.613036	0.148214	0.079359	0.613036	0.148214	0.079359
R44T	0.098515	0.079045	0.04857	0.098515	0.079045	0.04857
R44V	-0.219787	0.02979	0.064675	-0.219787	0.02979	0.064675
R44W	-1.356479	0.32312	0.127036	-1.356479	0.32312	0.127036
R44Y	-0.366362	0.27952	0.103228	-0.366362	0.27952	0.103228
R44*	ND	ND	ND	ND	ND	ND
H45A	0.909803	0.153417	0.1106	0.909803	0.153417	0.1106
H45C	-0.736601	0.094094	0.143032	-0.736601	0.094094	0.143032
H45D	-0.586663	0.409928	0.134036	-0.586663	0.409928	0.134036
H45E	-0.985937	0.209659	0.101962	-0.985937	0.209659	0.101962
H45F	-0.84094	0.090575	0.1785	-0.84094	0.090575	0.1785
H45G	0.058733	0.076794	0.059763	0.058733	0.076794	0.059763
H45I	-1.534623	0.417052	0.222988	-1.534623	0.417052	0.222988
H45K	-1.048486	0.350708	0.088433	-1.048486	0.350708	0.088433
H45L	-0.886406	0.086457	0.071063	-0.886406	0.086457	0.071063
H45M	-0.515869	0.15513	0.129143	-0.515869	0.15513	0.129143
H45N	0.072031	0.107036	0.097642	0.072031	0.107036	0.097642
H45P	0.700128	0.090894	0.074121	0.700128	0.090894	0.074121
H45Q	-0.086933	0.120095	0.084626	-0.086933	0.120095	0.084626
H45R	-0.979458	0.237295	0.100435	-0.979458	0.237295	0.100435
H45S	1.142782	0.236523	0.086143	1.142782	0.236523	0.086143
H45T	0.891305	0.236035	0.05201	0.891305	0.236035	0.05201
H45V	-1.6551	0.112414	0.176452	-1.6551	0.112414	0.176452
H45W	-0.581609	0.049591	0.108088	-0.581609	0.049591	0.108088
H45Y	-0.912711	0.016873	0.098084	-0.912711	0.016873	0.098084
H45*	Null	nan	nan	Null	nan	nan
T46A	-1.498759	0.329796	0.157867	-1.498759	0.329796	0.157867
T46C	-0.207309	0.36487	0.119585	-0.207309	0.36487	0.119585
T46D	Null	nan	nan	Null	nan	nan
T46E	Null	nan	nan	Null	nan	nan
T46F	ND	ND	ND	ND	ND	ND
T46G	-1.082574	0.042898	0.082669	-1.082574	0.042898	0.082669
T46H	Null	nan	nan	Null	nan	nan
T46I	Null	nan	nan	Null	nan	nan
T46K	Null	nan	nan	Null	nan	nan
T46L	-2.675069	0.25354	0.2494	-2.675069	0.25354	0.2494
T46M	ND	ND	ND	ND	ND	ND
T46N	-1.230475	0.474903	0.284255	-1.230475	0.474903	0.284255
T46P	Null	nan	nan	Null	nan	nan
T46Q	Null	nan	nan	Null	nan	nan
T46R	Null	nan	nan	Null	nan	nan
T46S	0.596038	0.260684	0.097519	0.596038	0.260684	0.097519
T46V	-2.20074	0.152127	0.210443	-2.20074	0.152127	0.210443
T46W	Null	nan	nan	Null	nan	nan
T46Y	Null	nan	nan	Null	nan	nan
T46*	Null	nan	nan	Null	nan	nan
W47A	0.527781	0.156038	0.051667	0.527781	0.156038	0.051667
W47C	0.143894	0.151676	0.130879	0.143894	0.151676	0.130879

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
W47D	-0.688869	0.478484	0.305989	-0.688869	0.478484	0.305989
W47E	-0.430354	0.397347	0.088589	-0.430354	0.397347	0.088589
W47F	0.455255	0.066448	0.088453	0.455255	0.066448	0.088453
W47G	0.11355	0.074461	0.081385	0.11355	0.074461	0.081385
W47H	0.517042	0.027497	0.083642	0.517042	0.027497	0.083642
W47I	0.800232	0.662589	0.095929	0.800232	0.662589	0.095929
W47K	0.568519	0.073244	0.072384	0.568519	0.073244	0.072384
W47L	1.624326	0.123687	0.089273	1.624326	0.123687	0.089273
W47M	1.244016	0.112837	0.072584	1.244016	0.112837	0.072584
W47N	-0.27867	0.193834	0.065555	-0.27867	0.193834	0.065555
W47P	-1.355197	0.607104	0.347613	-1.355197	0.607104	0.347613
W47Q	0.857851	0.086919	0.075288	0.857851	0.086919	0.075288
W47R	0.444005	0.14411	0.076916	0.444005	0.14411	0.076916
W47S	0.14242	0.124331	0.035869	0.14242	0.124331	0.035869
W47T	-0.046736	0.087151	0.066172	-0.046736	0.087151	0.066172
W47V	1.008386	0.068563	0.083332	1.008386	0.068563	0.083332
W47Y	0.898841	0.420812	0.114912	0.898841	0.420812	0.114912
W47*	Null	nan	nan	Null	nan	nan
E48A	0.350102	0.146919	0.076983	0.350102	0.146919	0.076983
E48C	0.410355	0.101115	0.101016	0.410355	0.101115	0.101016
E48D	-0.141051	0.245368	0.118154	-0.141051	0.245368	0.118154
E48F	0.385005	0.169541	0.086368	0.385005	0.169541	0.086368
E48G	0.72587	0.108052	0.062308	0.72587	0.108052	0.062308
E48H	0.220753	0.4688	0.070935	0.220753	0.4688	0.070935
E48I	-0.137099	0.133425	0.130987	-0.137099	0.133425	0.130987
E48K	0.381664	0.093874	0.054573	0.381664	0.093874	0.054573
E48L	0.42652	0.023243	0.038578	0.42652	0.023243	0.038578
E48M	0.121475	0.061579	0.070813	0.121475	0.061579	0.070813
E48N	-0.057178	0.43561	0.115864	-0.057178	0.43561	0.115864
E48P	-0.341049	0.333831	0.132462	-0.341049	0.333831	0.132462
E48Q	0.255721	0.034489	0.065475	0.255721	0.034489	0.065475
E48R	0.567234	0.114099	0.036065	0.567234	0.114099	0.036065
E48S	0.390527	0.137578	0.07577	0.390527	0.137578	0.07577
E48T	-0.049218	0.117485	0.069782	-0.049218	0.117485	0.069782
E48V	0.443966	0.042718	0.077266	0.443966	0.042718	0.077266
E48W	0.728645	0.063606	0.06927	0.728645	0.063606	0.06927
E48Y	0.307329	0.215787	0.111685	0.307329	0.215787	0.111685
E48*	-1.244844	0.608521	0.44112	-1.244844	0.608521	0.44112
S49A	-0.059353	0.080859	0.083571	-0.059353	0.080859	0.083571
S49C	-0.478805	0.3672	0.137052	-0.478805	0.3672	0.137052
S49D	-0.399139	0.066696	0.073711	-0.399139	0.066696	0.073711
S49E	-0.705718	0.285213	0.138362	-0.705718	0.285213	0.138362
S49F	-0.512378	0.323965	0.187374	-0.512378	0.323965	0.187374
S49G	-0.561643	0.131197	0.084843	-0.561643	0.131197	0.084843
S49H	-0.032786	0.338434	0.149491	-0.032786	0.338434	0.149491
S49I	ND	ND	ND	ND	ND	ND
S49K	-1.521389	0.988884	0.35741	-1.521389	0.988884	0.35741
S49L	-1.709606	0.647755	0.125223	-1.709606	0.647755	0.125223
S49M	-2.358457	0.242493	0.135003	-2.358457	0.242493	0.135003
S49N	-1.124496	0.196823	0.142404	-1.124496	0.196823	0.142404

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
S49P	-1.340726	0.204511	0.158768	-1.340726	0.204511	0.158768
S49Q	-1.022245	0.364688	0.218208	-1.022245	0.364688	0.218208
S49R	-2.168028	0.053735	0.449871	-2.168028	0.053735	0.449871
S49T	0.09727	0.032312	0.04586	0.09727	0.032312	0.04586
S49V	-0.773287	0.187533	0.104594	-0.773287	0.187533	0.104594
S49W	-1.577235	0.651951	0.27972	-1.577235	0.651951	0.27972
S49Y	-1.365475	0.257916	0.185841	-1.365475	0.257916	0.185841
S49*	Null	nan	nan	Null	nan	nan
I50A	-0.518921	0.09823	0.146505	-0.518921	0.09823	0.146505
I50C	0.174627	0.395243	0.137145	0.174627	0.395243	0.137145
I50D	Null	nan	nan	Null	nan	nan
I50E	ND	ND	ND	ND	ND	ND
I50F	Null	nan	nan	Null	nan	nan
I50G	-2.396354	0.194663	0.170577	-2.396354	0.194663	0.170577
I50H	Null	nan	nan	Null	nan	nan
I50K	ND	ND	ND	ND	ND	ND
I50L	0.850197	0.029669	0.067993	0.850197	0.029669	0.067993
I50M	0.426879	0.131953	0.091538	0.426879	0.131953	0.091538
I50N	Null	nan	nan	Null	nan	nan
I50P	Null	nan	nan	Null	nan	nan
I50Q	-1.456447	0.512148	0.286838	-1.456447	0.512148	0.286838
I50R	-1.783224	0.441534	0.180994	-1.783224	0.441534	0.180994
I50S	-2.122367	0.235664	0.303448	-2.122367	0.235664	0.303448
I50T	-0.404154	0.190328	0.170935	-0.404154	0.190328	0.170935
I50V	-0.216834	0.114291	0.13479	-0.216834	0.114291	0.13479
I50W	ND	ND	ND	ND	ND	ND
I50Y	Null	nan	nan	Null	nan	nan
I50*	Null	nan	nan	Null	nan	nan
G51A	0.024584	0.090502	0.07502	0.024584	0.090502	0.07502
G51C	-0.343724	0.152584	0.103762	-0.343724	0.152584	0.103762
G51D	0.143452	0.013897	0.066573	0.143452	0.013897	0.066573
G51E	0.466452	0.201875	0.097036	0.466452	0.201875	0.097036
G51F	0.061267	0.152971	0.094905	0.061267	0.152971	0.094905
G51H	0.273707	0.185507	0.061678	0.273707	0.185507	0.061678
G51I	0.391271	0.188016	0.097788	0.391271	0.188016	0.097788
G51K	0.165881	0.126524	0.119014	0.165881	0.126524	0.119014
G51L	-0.394324	0.201209	0.079766	-0.394324	0.201209	0.079766
G51M	-0.030868	0.186229	0.160343	-0.030868	0.186229	0.160343
G51N	-0.02183	0.116693	0.097218	-0.02183	0.116693	0.097218
G51P	0.296334	0.155467	0.112536	0.296334	0.155467	0.112536
G51Q	0.069653	0.14499	0.131077	0.069653	0.14499	0.131077
G51R	-0.080215	0.182876	0.050037	-0.080215	0.182876	0.050037
G51S	-0.139203	0.094683	0.160173	-0.139203	0.094683	0.160173
G51T	0.551569	0.133695	0.053777	0.551569	0.133695	0.053777
G51V	0.221965	0.120465	0.063445	0.221965	0.120465	0.063445
G51W	0.116958	0.475353	0.125256	0.116958	0.475353	0.125256
G51Y	-0.047365	0.106696	0.072814	-0.047365	0.106696	0.072814
G51*	Null	nan	nan	Null	nan	nan
R52A	0.181166	0.162021	0.044884	0.181166	0.162021	0.044884
R52C	0.073159	0.270608	0.056053	0.073159	0.270608	0.056053

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
R52D	0.070969	0.425011	0.148077	0.070969	0.425011	0.148077
R52E	-0.155911	0.337268	0.110998	-0.155911	0.337268	0.110998
R52F	-0.306056	0.247892	0.10541	-0.306056	0.247892	0.10541
R52G	0.259092	0.167431	0.067795	0.259092	0.167431	0.067795
R52H	-0.071064	0.296703	0.126827	-0.071064	0.296703	0.126827
R52I	0.180446	0.124179	0.048044	0.180446	0.124179	0.048044
R52K	0.045266	0.0565	0.064632	0.045266	0.0565	0.064632
R52L	0.011365	0.046368	0.070269	0.011365	0.046368	0.070269
R52M	0.178203	0.119643	0.079051	0.178203	0.119643	0.079051
R52N	-0.072313	0.10664	0.106258	-0.072313	0.10664	0.106258
R52P	0.248333	0.075506	0.045941	0.248333	0.075506	0.045941
R52Q	0.005	0.250864	0.134494	0.005	0.250864	0.134494
R52S	0.085959	0.25023	0.069149	0.085959	0.25023	0.069149
R52T	0.197604	0.168432	0.094812	0.197604	0.168432	0.094812
R52V	-0.172778	0.073902	0.057367	-0.172778	0.073902	0.057367
R52W	0.007968	0.242201	0.076169	0.007968	0.242201	0.076169
R52Y	0.110182	0.192408	0.111974	0.110182	0.192408	0.111974
R52*	Null	nan	nan	Null	nan	nan
P53A	0.146931	0.40515	0.118575	0.146931	0.40515	0.118575
P53C	-0.360827	0.298207	0.139806	-0.360827	0.298207	0.139806
P53D	-0.085569	0.247308	0.089521	-0.085569	0.247308	0.089521
P53E	0.04288	0.321458	0.059388	0.04288	0.321458	0.059388
P53F	0.258687	0.21311	0.095473	0.258687	0.21311	0.095473
P53G	0.470714	0.139535	0.075921	0.470714	0.139535	0.075921
P53H	0.274614	0.061623	0.114048	0.274614	0.061623	0.114048
P53I	-0.284093	0.131824	0.110839	-0.284093	0.131824	0.110839
P53K	-0.079634	0.416501	0.138309	-0.079634	0.416501	0.138309
P53L	0.127964	0.163304	0.10858	0.127964	0.163304	0.10858
P53M	-0.309394	0.043978	0.101776	-0.309394	0.043978	0.101776
P53N	-0.024956	0.148472	0.133286	-0.024956	0.148472	0.133286
P53Q	0.212658	0.446752	0.142289	0.212658	0.446752	0.142289
P53R	0.098116	0.107799	0.072348	0.098116	0.107799	0.072348
P53S	0.105198	0.138047	0.058818	0.105198	0.138047	0.058818
P53T	0.159697	0.114704	0.096017	0.159697	0.114704	0.096017
P53V	-0.25533	0.147779	0.061175	-0.25533	0.147779	0.061175
P53W	0.251741	0.277009	0.049915	0.251741	0.277009	0.049915
P53Y	0.07755	0.207018	0.136562	0.07755	0.207018	0.136562
P53*	Null	nan	nan	Null	nan	nan
L54A	-1.773406	0.363918	0.225227	-1.773406	0.363918	0.225227
L54C	-1.394751	0.23836	0.19258	-1.394751	0.23836	0.19258
L54D	Null	nan	nan	Null	nan	nan
L54E	Null	nan	nan	Null	nan	nan
L54F	0.034086	0.470785	0.070794	0.034086	0.470785	0.070794
L54G	-1.879994	0.754667	0.285055	-1.879994	0.754667	0.285055
L54H	ND	ND	ND	ND	ND	ND
L54I	-0.398346	0.243015	0.037218	-0.398346	0.243015	0.037218
L54K	ND	ND	ND	ND	ND	ND
L54M	0.139649	0.204496	0.141653	0.139649	0.204496	0.141653
L54N	ND	ND	ND	ND	ND	ND
L54P	-1.831335	0.183441	0.179478	-1.831335	0.183441	0.179478

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
L54Q	Null	nan	nan	Null	nan	nan
L54R	-2.093569	0.839614	0.65422	-2.093569	0.839614	0.65422
L54S	-2.544355	0.042534	0.38874	-2.544355	0.042534	0.38874
L54T	-1.700499	0.164688	0.196039	-1.700499	0.164688	0.196039
L54V	-1.42078	0.297571	0.174191	-1.42078	0.297571	0.174191
L54W	Null	nan	nan	Null	nan	nan
L54Y	Null	nan	nan	Null	nan	nan
L54*	Null	nan	nan	Null	nan	nan
P55A	-0.041357	0.188313	0.058833	-0.041357	0.188313	0.058833
P55C	0.005945	0.402965	0.327338	0.005945	0.402965	0.327338
P55D	0.130679	0.554047	0.121706	0.130679	0.554047	0.121706
P55E	-0.014064	0.205188	0.071849	-0.014064	0.205188	0.071849
P55F	0.248759	0.610263	0.181803	0.248759	0.610263	0.181803
P55G	0.106084	0.022015	0.040359	0.106084	0.022015	0.040359
P55H	0.072853	0.107403	0.119339	0.072853	0.107403	0.119339
P55I	-0.044413	0.170478	0.066692	-0.044413	0.170478	0.066692
P55K	-0.135585	0.20488	0.076306	-0.135585	0.20488	0.076306
P55L	0.058366	0.104773	0.084931	0.058366	0.104773	0.084931
P55M	-0.080446	0.155007	0.076864	-0.080446	0.155007	0.076864
P55N	0.270078	0.209668	0.05591	0.270078	0.209668	0.05591
P55Q	-0.037262	0.133808	0.100719	-0.037262	0.133808	0.100719
P55R	-0.140673	0.112801	0.061385	-0.140673	0.112801	0.061385
P55S	0.014373	0.140552	0.05176	0.014373	0.140552	0.05176
P55T	0.172234	0.084581	0.080304	0.172234	0.084581	0.080304
P55V	-0.01438	0.241024	0.081623	-0.01438	0.241024	0.081623
P55W	0.017669	0.197169	0.084947	0.017669	0.197169	0.084947
P55Y	0.294714	0.356525	0.081774	0.294714	0.356525	0.081774
P55*	ND	ND	ND	ND	ND	ND
G56A	-0.286761	0.134306	0.155302	-0.286761	0.134306	0.155302
G56C	-0.224495	0.32941	0.219727	-0.224495	0.32941	0.219727
G56D	-0.123528	0.141796	0.076415	-0.123528	0.141796	0.076415
G56E	-0.080475	0.087334	0.084753	-0.080475	0.087334	0.084753
G56F	-0.175148	0.196549	0.206505	-0.175148	0.196549	0.206505
G56H	-0.09302	0.185455	0.130577	-0.09302	0.185455	0.130577
G56I	-0.68853	0.084131	0.094022	-0.68853	0.084131	0.094022
G56K	-0.069696	0.126098	0.157077	-0.069696	0.126098	0.157077
G56L	-0.337939	0.310991	0.15071	-0.337939	0.310991	0.15071
G56M	-0.126497	0.205372	0.111228	-0.126497	0.205372	0.111228
G56N	-0.366431	0.364321	0.124699	-0.366431	0.364321	0.124699
G56P	ND	ND	ND	ND	ND	ND
G56Q	-0.214707	0.14974	0.126411	-0.214707	0.14974	0.126411
G56R	-0.19479	0.127195	0.118293	-0.19479	0.127195	0.118293
G56S	-0.031379	0.290143	0.068616	-0.031379	0.290143	0.068616
G56T	-0.066243	0.133687	0.12112	-0.066243	0.133687	0.12112
G56V	-0.12423	0.157015	0.083152	-0.12423	0.157015	0.083152
G56W	-0.050036	0.517662	0.189992	-0.050036	0.517662	0.189992
G56Y	0.127947	0.372693	0.104984	0.127947	0.372693	0.104984
G56*	Null	nan	nan	Null	nan	nan
R57A	Null	nan	nan	Null	nan	nan
R57C	Null	nan	nan	Null	nan	nan

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
R57D	Null	nan	nan	Null	nan	nan
R57E	Null	nan	nan	Null	nan	nan
R57F	Null	nan	nan	Null	nan	nan
R57G	-2.11624	0.103718	0.301344	-2.11624	0.103718	0.301344
R57H	Null	nan	nan	Null	nan	nan
R57I	Null	nan	nan	Null	nan	nan
R57K	ND	ND	ND	ND	ND	ND
R57L	-2.823427	1.036024	0.35726	-2.823427	1.036024	0.35726
R57M	ND	ND	ND	ND	ND	ND
R57N	Null	nan	nan	Null	nan	nan
R57P	ND	ND	ND	ND	ND	ND
R57Q	Null	nan	nan	Null	nan	nan
R57S	Null	nan	nan	Null	nan	nan
R57T	ND	ND	ND	ND	ND	ND
R57V	-1.886144	0.122322	0.2173	-1.886144	0.122322	0.2173
R57W	ND	ND	ND	ND	ND	ND
R57Y	Null	nan	nan	Null	nan	nan
R57*	Null	nan	nan	Null	nan	nan
K58A	-0.07684	0.231146	0.074867	-0.07684	0.231146	0.074867
K58C	0.082915	0.318226	0.081426	0.082915	0.318226	0.081426
K58D	0.088541	0.143985	0.069193	0.088541	0.143985	0.069193
K58E	-0.18963	0.270173	0.067508	-0.18963	0.270173	0.067508
K58F	-0.142078	0.111129	0.072793	-0.142078	0.111129	0.072793
K58G	-0.21518	0.087197	0.043211	-0.21518	0.087197	0.043211
K58H	-0.173476	0.115058	0.152215	-0.173476	0.115058	0.152215
K58I	-0.117549	0.273982	0.171805	-0.117549	0.273982	0.171805
K58L	0.012172	0.262666	0.05667	0.012172	0.262666	0.05667
K58M	-0.131503	0.26993	0.104424	-0.131503	0.26993	0.104424
K58N	0.352655	0.083122	0.17623	0.352655	0.083122	0.17623
K58P	-0.035026	0.196525	0.093816	-0.035026	0.196525	0.093816
K58Q	0.146893	0.244965	0.056809	0.146893	0.244965	0.056809
K58R	-0.057165	0.047214	0.056122	-0.057165	0.047214	0.056122
K58S	-0.019697	0.140283	0.042304	-0.019697	0.140283	0.042304
K58T	-0.361345	0.195439	0.102948	-0.361345	0.195439	0.102948
K58V	0.104952	0.133037	0.059222	0.104952	0.133037	0.059222
K58W	-0.240607	0.15472	0.058939	-0.240607	0.15472	0.058939
K58Y	-0.288675	0.21956	0.07898	-0.288675	0.21956	0.07898
K58*	-1.695478	0.453331	0.272686	-1.695478	0.453331	0.272686
N59A	0.277445	0.16073	0.104155	0.277445	0.16073	0.104155
N59C	0.243218	0.245482	0.129416	0.243218	0.245482	0.129416
N59D	0.197571	0.286913	0.167098	0.197571	0.286913	0.167098
N59E	-0.037491	0.350497	0.068383	-0.037491	0.350497	0.068383
N59F	-0.297411	0.03498	0.079932	-0.297411	0.03498	0.079932
N59G	0.171928	0.103953	0.024411	0.171928	0.103953	0.024411
N59H	0.065186	0.293296	0.160001	0.065186	0.293296	0.160001
N59I	0.488956	0.132492	0.0806	0.488956	0.132492	0.0806
N59K	0.04433	0.050222	0.070936	0.04433	0.050222	0.070936
N59L	-0.606128	0.193807	0.07871	-0.606128	0.193807	0.07871
N59M	-0.323676	0.086147	0.137356	-0.323676	0.086147	0.137356
N59P	0.185201	0.163001	0.137647	0.185201	0.163001	0.137647

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
N59Q	-0.374398	0.50546	0.160608	-0.374398	0.50546	0.160608
N59R	-0.743417	0.12781	0.109966	-0.743417	0.12781	0.109966
N59S	0.145648	0.491455	0.081287	0.145648	0.491455	0.081287
N59T	0.056741	0.12219	0.092094	0.056741	0.12219	0.092094
N59V	0.400477	0.190103	0.094277	0.400477	0.190103	0.094277
N59W	-0.388496	0.10213	0.100976	-0.388496	0.10213	0.100976
N59Y	-0.587563	0.210807	0.212575	-0.587563	0.210807	0.212575
N59*	Null	nan	nan	Null	nan	nan
I60A	0.045252	0.080026	0.055758	0.045252	0.080026	0.055758
I60C	0.275707	0.276919	0.065178	0.275707	0.276919	0.065178
I60D	ND	ND	ND	ND	ND	ND
I60E	-1.248719	0.283197	0.330869	-1.248719	0.283197	0.330869
I60F	0.541465	0.204429	0.114456	0.541465	0.204429	0.114456
I60G	-1.306908	0.354226	0.065561	-1.306908	0.354226	0.065561
I60H	-0.117586	0.121117	0.113725	-0.117586	0.121117	0.113725
I60K	-1.328122	0.363103	0.160847	-1.328122	0.363103	0.160847
I60L	0.294041	0.164713	0.092089	0.294041	0.164713	0.092089
I60M	0.153736	0.2584	0.132691	0.153736	0.2584	0.132691
I60N	-0.967007	0.246511	0.128738	-0.967007	0.246511	0.128738
I60P	ND	ND	ND	ND	ND	ND
I60Q	-1.118766	0.288232	0.159697	-1.118766	0.288232	0.159697
I60R	-1.931627	0.167907	0.135578	-1.931627	0.167907	0.135578
I60S	-0.665508	0.084494	0.104593	-0.665508	0.084494	0.104593
I60T	-0.340452	0.568027	0.08019	-0.340452	0.568027	0.08019
I60V	0.123792	0.053139	0.049468	0.123792	0.053139	0.049468
I60W	-0.275319	0.164424	0.084361	-0.275319	0.164424	0.084361
I60Y	0.68977	0.148583	0.080234	0.68977	0.148583	0.080234
I60*	Null	nan	nan	Null	nan	nan
I61A	-0.846102	0.279064	0.180137	-0.846102	0.279064	0.180137
I61C	0.17709	0.158842	0.118033	0.17709	0.158842	0.118033
I61D	Null	nan	nan	Null	nan	nan
I61E	ND	ND	ND	ND	ND	ND
I61F	ND	ND	ND	ND	ND	ND
I61G	-0.823535	0.208461	0.066295	-0.823535	0.208461	0.066295
I61H	ND	ND	ND	ND	ND	ND
I61K	-2.029939	0.765536	0.232104	-2.029939	0.765536	0.232104
I61L	0.597883	0.061202	0.08526	0.597883	0.061202	0.08526
I61M	0.598767	0.150576	0.086518	0.598767	0.150576	0.086518
I61N	Null	nan	nan	Null	nan	nan
I61P	-1.558122	0.681672	0.284503	-1.558122	0.681672	0.284503
I61Q	-1.337254	0.134526	0.340929	-1.337254	0.134526	0.340929
I61R	-0.972285	0.069504	0.13249	-0.972285	0.069504	0.13249
I61S	-1.051935	0.366492	0.152807	-1.051935	0.366492	0.152807
I61T	-1.336322	0.768705	0.219069	-1.336322	0.768705	0.219069
I61V	-0.293164	0.189615	0.107235	-0.293164	0.189615	0.107235
I61W	-1.365534	0.932915	0.259912	-1.365534	0.932915	0.259912
I61Y	-0.772824	0.154595	0.104816	-0.772824	0.154595	0.104816
I61*	Null	nan	nan	Null	nan	nan
L62A	0.720198	0.16596	0.123016	0.720198	0.16596	0.123016
L62C	0.733007	0.140242	0.146432	0.733007	0.140242	0.146432

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
L62D	Null	nan	nan	Null	nan	nan
L62E	-1.779701	0.527284	0.321975	-1.779701	0.527284	0.321975
L62F	0.243962	0.243973	0.06615	0.243962	0.243973	0.06615
L62G	ND	ND	ND	ND	ND	ND
L62H	0.078033	0.254161	0.163746	0.078033	0.254161	0.163746
L62I	1.093305	0.304238	0.166421	1.093305	0.304238	0.166421
L62K	-1.183937	0.309408	0.159548	-1.183937	0.309408	0.159548
L62M	0.289015	0.323585	0.138182	0.289015	0.323585	0.138182
L62N	-1.225726	0.462243	0.202681	-1.225726	0.462243	0.202681
L62P	-1.011073	0.563065	0.324575	-1.011073	0.563065	0.324575
L62Q	-0.439716	0.025637	0.066653	-0.439716	0.025637	0.066653
L62R	ND	ND	ND	ND	ND	ND
L62S	-0.525074	0.232337	0.083932	-0.525074	0.232337	0.083932
L62T	0.893335	0.13677	0.097058	0.893335	0.13677	0.097058
L62V	0.897067	0.077084	0.094457	0.897067	0.077084	0.094457
L62W	-0.90407	0.301967	0.114141	-0.90407	0.301967	0.114141
L62Y	-0.935905	0.330075	0.164861	-0.935905	0.330075	0.164861
L62*	Null	nan	nan	Null	nan	nan
S63A	0.343146	0.063912	0.079763	0.343146	0.063912	0.079763
S63C	0.469623	0.120262	0.102589	0.469623	0.120262	0.102589
S63D	-2.302626	0.824393	0.36751	-2.302626	0.824393	0.36751
S63E	-1.339032	0.427406	0.248423	-1.339032	0.427406	0.248423
S63F	-0.950162	0.120379	0.100607	-0.950162	0.120379	0.100607
S63G	0.418656	0.060517	0.070089	0.418656	0.060517	0.070089
S63H	-0.334001	0.177027	0.090792	-0.334001	0.177027	0.090792
S63I	-0.374839	0.143093	0.136442	-0.374839	0.143093	0.136442
S63K	-0.660505	0.197621	0.095598	-0.660505	0.197621	0.095598
S63L	-0.404197	0.126519	0.055088	-0.404197	0.126519	0.055088
S63M	-0.018971	0.057368	0.059383	-0.018971	0.057368	0.059383
S63N	-1.610536	0.160132	0.377192	-1.610536	0.160132	0.377192
S63P	-2.354638	0.327945	0.400337	-2.354638	0.327945	0.400337
S63Q	-0.203879	0.260706	0.077373	-0.203879	0.260706	0.077373
S63R	-0.169477	0.112645	0.041846	-0.169477	0.112645	0.041846
S63T	0.315829	0.137925	0.033666	0.315829	0.137925	0.033666
S63V	0.169458	0.033691	0.071507	0.169458	0.033691	0.071507
S63W	-0.219484	0.144338	0.077306	-0.219484	0.144338	0.077306
S63Y	-0.809006	0.470881	0.201848	-0.809006	0.470881	0.201848
S63*	Null	nan	nan	Null	nan	nan
S64A	0.321919	0.413449	0.067204	0.321919	0.413449	0.067204
S64C	0.357211	0.273844	0.058881	0.357211	0.273844	0.058881
S64D	-0.228134	0.2644	0.133346	-0.228134	0.2644	0.133346
S64E	0.129586	0.106021	0.072643	0.129586	0.106021	0.072643
S64F	0.391777	0.098607	0.075905	0.391777	0.098607	0.075905
S64G	0.432119	0.153587	0.048563	0.432119	0.153587	0.048563
S64H	-0.44359	0.367741	0.220305	-0.44359	0.367741	0.220305
S64I	0.22508	0.123451	0.147387	0.22508	0.123451	0.147387
S64K	0.097306	0.079031	0.059591	0.097306	0.079031	0.059591
S64L	0.328203	0.242201	0.040068	0.328203	0.242201	0.040068
S64M	0.371718	0.080653	0.060773	0.371718	0.080653	0.060773
S64N	0.22839	0.232105	0.149799	0.22839	0.232105	0.149799

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
S64P	-1.211363	0.188869	0.179743	-1.211363	0.188869	0.179743
S64Q	0.60023	0.08606	0.088245	0.60023	0.08606	0.088245
S64R	0.173775	0.072737	0.063207	0.173775	0.072737	0.063207
S64T	-0.234072	0.14146	0.127756	-0.234072	0.14146	0.127756
S64V	0.06341	0.175764	0.088838	0.06341	0.175764	0.088838
S64W	-0.13772	0.065363	0.089701	-0.13772	0.065363	0.089701
S64Y	-0.121718	0.313368	0.134506	-0.121718	0.313368	0.134506
S64*	ND	ND	ND	ND	ND	ND
Q65A	-0.029661	0.137779	0.094555	-0.029661	0.137779	0.094555
Q65C	0.254606	0.150437	0.09087	0.254606	0.150437	0.09087
Q65D	0.095082	0.077357	0.094727	0.095082	0.077357	0.094727
Q65E	0.064395	0.501558	0.053377	0.064395	0.501558	0.053377
Q65F	-0.00329	0.177797	0.099195	-0.00329	0.177797	0.099195
Q65G	-0.760865	0.164382	0.076142	-0.760865	0.164382	0.076142
Q65H	0.114743	0.137487	0.038612	0.114743	0.137487	0.038612
Q65I	0.255965	0.128146	0.090299	0.255965	0.128146	0.090299
Q65K	-0.207959	0.242251	0.063494	-0.207959	0.242251	0.063494
Q65L	0.034867	0.183711	0.070979	0.034867	0.183711	0.070979
Q65M	-0.340335	0.263921	0.087803	-0.340335	0.263921	0.087803
Q65N	-0.229923	0.211188	0.102735	-0.229923	0.211188	0.102735
Q65P	-0.742416	0.208653	0.122837	-0.742416	0.208653	0.122837
Q65R	0.131263	0.13616	0.056653	0.131263	0.13616	0.056653
Q65S	0.119396	0.032832	0.065412	0.119396	0.032832	0.065412
Q65T	0.119002	0.088482	0.087271	0.119002	0.088482	0.087271
Q65V	0.306627	0.090604	0.067089	0.306627	0.090604	0.067089
Q65W	0.323663	0.12567	0.075285	0.323663	0.12567	0.075285
Q65Y	0.262422	0.058391	0.130166	0.262422	0.058391	0.130166
Q65*	Null	nan	nan	Null	nan	nan
P66A	-0.106615	0.227556	0.143563	-0.106615	0.227556	0.143563
P66C	ND	ND	ND	ND	ND	ND
P66D	-0.955554	0.093629	0.210335	-0.955554	0.093629	0.210335
P66E	-0.042597	0.309909	0.064836	-0.042597	0.309909	0.064836
P66F	0.67837	0.049446	0.155348	0.67837	0.049446	0.155348
P66G	0.004905	0.161943	0.041612	0.004905	0.161943	0.041612
P66H	0.259802	0.323162	0.353838	0.259802	0.323162	0.353838
P66I	ND	ND	ND	ND	ND	ND
P66K	-0.006722	0.071584	0.037867	-0.006722	0.071584	0.037867
P66L	0.036091	0.088632	0.080762	0.036091	0.088632	0.080762
P66M	-0.052038	0.345809	0.054524	-0.052038	0.345809	0.054524
P66N	0.054995	0.333119	0.164121	0.054995	0.333119	0.164121
P66Q	Null	nan	nan	Null	nan	nan
P66R	0.173597	0.104077	0.04554	0.173597	0.104077	0.04554
P66S	0.2928	0.215031	0.091216	0.2928	0.215031	0.091216
P66T	-0.1582	0.211681	0.12905	-0.1582	0.211681	0.12905
P66V	0.131562	0.134443	0.089072	0.131562	0.134443	0.089072
P66W	0.085126	0.102235	0.081262	0.085126	0.102235	0.081262
P66Y	0.208406	0.277628	0.270416	0.208406	0.277628	0.270416
P66*	Null	nan	nan	Null	nan	nan
G67A	0.289482	0.40142	0.167644	0.289482	0.40142	0.167644
G67C	0.313571	0.105375	0.169886	0.313571	0.105375	0.169886

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
G67D	0.36453	0.016746	0.12981	0.36453	0.016746	0.12981
G67E	0.566803	0.029243	0.056125	0.566803	0.029243	0.056125
G67F	-0.294592	0.329938	0.205778	-0.294592	0.329938	0.205778
G67H	-0.39488	0.275136	0.220657	-0.39488	0.275136	0.220657
G67I	-0.13397	0.325507	0.168704	-0.13397	0.325507	0.168704
G67K	-0.129636	0.045668	0.04699	-0.129636	0.045668	0.04699
G67L	0.307753	0.282328	0.085335	0.307753	0.282328	0.085335
G67M	0.43732	0.135523	0.113859	0.43732	0.135523	0.113859
G67N	0.016221	0.104499	0.05911	0.016221	0.104499	0.05911
G67P	0.526881	0.095073	0.160126	0.526881	0.095073	0.160126
G67Q	ND	ND	ND	ND	ND	ND
G67R	-0.639784	0.042313	0.09056	-0.639784	0.042313	0.09056
G67S	0.249805	0.241299	0.062343	0.249805	0.241299	0.062343
G67T	0.597699	0.242757	0.079595	0.597699	0.242757	0.079595
G67V	0.359742	0.111783	0.072498	0.359742	0.111783	0.072498
G67W	-0.587144	0.092254	0.185899	-0.587144	0.092254	0.185899
G67Y	-0.073925	0.136331	0.192673	-0.073925	0.136331	0.192673
G67*	Null	nan	nan	Null	nan	nan
T68A	0.251632	0.086077	0.058052	0.251632	0.086077	0.058052
T68C	0.410023	0.172323	0.0725	0.410023	0.172323	0.0725
T68D	0.389492	0.197449	0.139419	0.389492	0.197449	0.139419
T68E	0.186146	0.092964	0.041077	0.186146	0.092964	0.041077
T68F	0.308897	0.203629	0.103707	0.308897	0.203629	0.103707
T68G	0.017208	0.187532	0.046581	0.017208	0.187532	0.046581
T68H	0.428918	0.238216	0.103083	0.428918	0.238216	0.103083
T68I	-0.022946	0.126923	0.145077	-0.022946	0.126923	0.145077
T68K	0.025066	0.213135	0.054985	0.025066	0.213135	0.054985
T68L	0.085305	0.148896	0.057208	0.085305	0.148896	0.057208
T68M	-0.859495	0.080905	0.290938	-0.859495	0.080905	0.290938
T68N	-0.088752	0.08998	0.054239	-0.088752	0.08998	0.054239
T68P	0.292875	0.401716	0.106051	0.292875	0.401716	0.106051
T68Q	0.359781	0.176468	0.103448	0.359781	0.176468	0.103448
T68R	-0.111425	0.092511	0.076945	-0.111425	0.092511	0.076945
T68S	0.237291	0.309419	0.107779	0.237291	0.309419	0.107779
T68V	-0.013517	0.209991	0.083071	-0.013517	0.209991	0.083071
T68W	-0.040302	0.352098	0.134851	-0.040302	0.352098	0.134851
T68Y	0.106313	0.557212	0.085041	0.106313	0.557212	0.085041
T68*	ND	ND	ND	ND	ND	ND
D69A	0.056951	0.3885	0.089132	0.056951	0.3885	0.089132
D69C	0.413451	0.268865	0.234623	0.413451	0.268865	0.234623
D69E	0.030705	0.107167	0.110301	0.030705	0.107167	0.110301
D69F	-0.029086	0.341599	0.105464	-0.029086	0.341599	0.105464
D69G	0.159578	0.267028	0.052863	0.159578	0.267028	0.052863
D69H	-0.117972	0.062112	0.074547	-0.117972	0.062112	0.074547
D69I	-0.115516	0.195265	0.117258	-0.115516	0.195265	0.117258
D69K	-0.171851	0.056175	0.07117	-0.171851	0.056175	0.07117
D69L	0.059475	0.197062	0.072374	0.059475	0.197062	0.072374
D69M	-0.281545	0.279328	0.116551	-0.281545	0.279328	0.116551
D69N	-0.069301	0.081877	0.056377	-0.069301	0.081877	0.056377
D69P	0.128878	0.274084	0.068528	0.128878	0.274084	0.068528

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
D69Q	-0.271471	0.167243	0.122607	-0.271471	0.167243	0.122607
D69R	0.073202	0.213739	0.083592	0.073202	0.213739	0.083592
D69S	0.141603	0.230058	0.091267	0.141603	0.230058	0.091267
D69T	-0.165906	0.077819	0.066497	-0.165906	0.077819	0.066497
D69V	0.015235	0.219939	0.086915	0.015235	0.219939	0.086915
D69W	0.31123	0.119206	0.063588	0.31123	0.119206	0.063588
D69Y	0.130133	0.290814	0.1672	0.130133	0.290814	0.1672
D69*	Null	nan	nan	Null	nan	nan
D70A	-0.261442	0.139567	0.095183	-0.261442	0.139567	0.095183
D70C	0.16018	0.197032	0.1361	0.16018	0.197032	0.1361
D70E	-0.000502	0.272344	0.075335	-0.000502	0.272344	0.075335
D70F	0.031474	0.012941	0.089884	0.031474	0.012941	0.089884
D70G	-0.077922	0.049734	0.044182	-0.077922	0.049734	0.044182
D70H	0.260689	0.134026	0.098646	0.260689	0.134026	0.098646
D70I	-0.088705	0.236664	0.109258	-0.088705	0.236664	0.109258
D70K	-0.002274	0.13473	0.066137	-0.002274	0.13473	0.066137
D70L	0.132306	0.019804	0.056804	0.132306	0.019804	0.056804
D70M	0.213184	0.3176	0.097475	0.213184	0.3176	0.097475
D70N	-0.306798	0.213825	0.056306	-0.306798	0.213825	0.056306
D70P	0.151729	0.329171	0.086872	0.151729	0.329171	0.086872
D70Q	0.257253	0.133588	0.104952	0.257253	0.133588	0.104952
D70R	0.077052	0.114539	0.072787	0.077052	0.114539	0.072787
D70S	0.03565	0.051189	0.054918	0.03565	0.051189	0.054918
D70T	0.15029	0.140454	0.078123	0.15029	0.140454	0.078123
D70V	-0.050864	0.073788	0.076935	-0.050864	0.073788	0.076935
D70W	-0.069317	0.230347	0.081312	-0.069317	0.230347	0.081312
D70Y	-0.004851	0.059962	0.109071	-0.004851	0.059962	0.109071
D70*	Null	nan	nan	Null	nan	nan
R71A	-0.048141	0.139845	0.091552	-0.048141	0.139845	0.091552
R71C	Null	nan	nan	Null	nan	nan
R71D	0.169613	0.256285	0.123263	0.169613	0.256285	0.123263
R71E	0.098249	0.049388	0.08422	0.098249	0.049388	0.08422
R71F	-0.030571	0.139972	0.080637	-0.030571	0.139972	0.080637
R71G	0.180493	0.162098	0.11506	0.180493	0.162098	0.11506
R71H	ND	ND	ND	ND	ND	ND
R71I	0.051003	0.419676	0.078509	0.051003	0.419676	0.078509
R71K	-0.14523	0.136921	0.059723	-0.14523	0.136921	0.059723
R71L	-0.126923	0.104152	0.052102	-0.126923	0.104152	0.052102
R71M	-0.134023	0.203306	0.115418	-0.134023	0.203306	0.115418
R71N	0.169087	0.27457	0.0976	0.169087	0.27457	0.0976
R71P	-0.033324	0.222569	0.073654	-0.033324	0.222569	0.073654
R71Q	-0.063793	0.290838	0.13139	-0.063793	0.290838	0.13139
R71S	0.191545	0.028149	0.073414	0.191545	0.028149	0.073414
R71T	0.124479	0.26693	0.066483	0.124479	0.26693	0.066483
R71V	-0.210517	0.273576	0.0684	-0.210517	0.273576	0.0684
R71W	-0.043721	0.102623	0.114598	-0.043721	0.102623	0.114598
R71Y	-0.088199	0.242332	0.06201	-0.088199	0.242332	0.06201
R71*	Null	nan	nan	Null	nan	nan
V72A	0.428257	0.264792	0.110853	0.428257	0.264792	0.110853
V72C	0.23161	0.110116	0.16005	0.23161	0.110116	0.16005

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
V72D	0.317953	0.180529	0.104368	0.317953	0.180529	0.104368
V72E	-0.15911	0.218791	0.105765	-0.15911	0.218791	0.105765
V72F	0.606767	0.188281	0.102414	0.606767	0.188281	0.102414
V72G	0.363638	0.087247	0.047417	0.363638	0.087247	0.047417
V72H	0.753307	0.078338	0.152208	0.753307	0.078338	0.152208
V72I	0.339909	0.353137	0.122743	0.339909	0.353137	0.122743
V72K	-0.365863	0.129965	0.063364	-0.365863	0.129965	0.063364
V72L	0.11862	0.168874	0.083415	0.11862	0.168874	0.083415
V72M	0.260065	0.171332	0.094815	0.260065	0.171332	0.094815
V72N	0.276257	0.153826	0.115903	0.276257	0.153826	0.115903
V72P	0.083092	0.38369	0.130927	0.083092	0.38369	0.130927
V72Q	0.250648	0.128789	0.118131	0.250648	0.128789	0.118131
V72R	-0.055098	0.136562	0.081809	-0.055098	0.136562	0.081809
V72S	0.21024	0.018619	0.1024	0.21024	0.018619	0.1024
V72T	0.253136	0.380701	0.15924	0.253136	0.380701	0.15924
V72W	0.429263	0.074477	0.175012	0.429263	0.074477	0.175012
V72Y	0.714289	0.23767	0.164565	0.714289	0.23767	0.164565
V72*	Null	nan	nan	Null	nan	nan
T73A	-0.283718	0.270711	0.132132	-0.283718	0.270711	0.132132
T73C	ND	ND	ND	ND	ND	ND
T73D	-0.553583	0.560648	0.212438	-0.553583	0.560648	0.212438
T73E	0.090139	0.243602	0.146089	0.090139	0.243602	0.146089
T73F	-0.412579	0.332141	0.201463	-0.412579	0.332141	0.201463
T73G	-0.010764	0.142712	0.088093	-0.010764	0.142712	0.088093
T73H	0.038617	0.156386	0.144744	0.038617	0.156386	0.144744
T73I	-0.183654	0.361141	0.175775	-0.183654	0.361141	0.175775
T73K	0.222663	0.267403	0.100281	0.222663	0.267403	0.100281
T73L	0.213858	0.02812	0.096825	0.213858	0.02812	0.096825
T73M	Null	nan	nan	Null	nan	nan
T73N	-0.300868	0.223055	0.107199	-0.300868	0.223055	0.107199
T73P	0.207021	0.203049	0.123242	0.207021	0.203049	0.123242
T73Q	0.011603	0.598945	0.190411	0.011603	0.598945	0.190411
T73R	0.050727	0.143581	0.099161	0.050727	0.143581	0.099161
T73S	0.075696	0.074237	0.141951	0.075696	0.074237	0.141951
T73V	-0.035227	0.103093	0.072949	-0.035227	0.103093	0.072949
T73W	0.177734	0.168414	0.091793	0.177734	0.168414	0.091793
T73Y	0.10171	0.452989	0.221869	0.10171	0.452989	0.221869
T73*	Null	nan	nan	Null	nan	nan
W74A	0.116528	0.302593	0.150217	0.116528	0.302593	0.150217
W74C	Null	nan	nan	Null	nan	nan
W74D	-0.555187	0.35817	0.230824	-0.555187	0.35817	0.230824
W74E	0.12651	0.169144	0.066287	0.12651	0.169144	0.066287
W74F	ND	ND	ND	ND	ND	ND
W74G	-0.150133	0.110248	0.075792	-0.150133	0.110248	0.075792
W74H	0.235382	0.307601	0.350617	0.235382	0.307601	0.350617
W74I	0.014934	0.270888	0.191215	0.014934	0.270888	0.191215
W74K	-0.215493	0.371728	0.149339	-0.215493	0.371728	0.149339
W74L	0.248772	0.333782	0.108766	0.248772	0.333782	0.108766
W74M	0.213134	0.237987	0.125841	0.213134	0.237987	0.125841
W74N	0.056966	0.315235	0.098918	0.056966	0.315235	0.098918

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
W74P	-0.336734	0.1042	0.165519	-0.336734	0.1042	0.165519
W74Q	0.070452	0.278045	0.119377	0.070452	0.278045	0.119377
W74R	-0.189971	0.144908	0.065604	-0.189971	0.144908	0.065604
W74S	0.104742	0.104514	0.0303	0.104742	0.104514	0.0303
W74T	0.197754	0.063435	0.081055	0.197754	0.063435	0.081055
W74V	0.279203	0.049848	0.094388	0.279203	0.049848	0.094388
W74Y	0.296996	0.040313	0.196423	0.296996	0.040313	0.196423
W74*	Null	nan	nan	Null	nan	nan
V75A	-0.45141	0.642398	0.232469	-0.45141	0.642398	0.232469
V75C	-0.254562	0.116209	0.081485	-0.254562	0.116209	0.081485
V75D	-0.386486	0.106685	0.144396	-0.386486	0.106685	0.144396
V75E	-0.082631	0.025222	0.168828	-0.082631	0.025222	0.168828
V75F	-0.06149	0.080894	0.054689	-0.06149	0.080894	0.054689
V75G	-0.268696	0.02702	0.065044	-0.268696	0.02702	0.065044
V75H	-0.084485	0.454756	0.124237	-0.084485	0.454756	0.124237
V75I	0.566569	0.058027	0.060363	0.566569	0.058027	0.060363
V75K	0.109546	0.144128	0.072305	0.109546	0.144128	0.072305
V75L	0.315335	0.112082	0.059294	0.315335	0.112082	0.059294
V75M	0.072696	0.172823	0.153027	0.072696	0.172823	0.153027
V75N	-0.214207	0.330851	0.11517	-0.214207	0.330851	0.11517
V75P	ND	ND	ND	ND	ND	ND
V75Q	0.105906	0.101104	0.073359	0.105906	0.101104	0.073359
V75R	0.129617	0.126183	0.054858	0.129617	0.126183	0.054858
V75S	-0.079553	0.149763	0.081612	-0.079553	0.149763	0.081612
V75T	0.040529	0.210862	0.084847	0.040529	0.210862	0.084847
V75W	0.247242	0.11136	0.071244	0.247242	0.11136	0.071244
V75Y	-0.131643	0.236908	0.078001	-0.131643	0.236908	0.078001
V75*	ND	ND	ND	ND	ND	ND
K76A	-0.056764	0.062421	0.054223	-0.056764	0.062421	0.054223
K76C	-0.128702	0.200593	0.091287	-0.128702	0.200593	0.091287
K76D	0.199265	0.399132	0.121146	0.199265	0.399132	0.121146
K76E	0.584681	0.236852	0.092682	0.584681	0.236852	0.092682
K76F	0.275717	0.242203	0.118113	0.275717	0.242203	0.118113
K76G	-0.174878	0.110661	0.063437	-0.174878	0.110661	0.063437
K76H	-0.510772	0.010472	0.161458	-0.510772	0.010472	0.161458
K76I	0.438528	0.299675	0.162147	0.438528	0.299675	0.162147
K76L	0.122128	0.337916	0.087581	0.122128	0.337916	0.087581
K76M	0.024262	0.152853	0.127061	0.024262	0.152853	0.127061
K76N	-0.170053	0.202666	0.122656	-0.170053	0.202666	0.122656
K76P	-0.273575	0.324262	0.105238	-0.273575	0.324262	0.105238
K76Q	-0.012069	0.342151	0.114231	-0.012069	0.342151	0.114231
K76R	-0.111644	0.220224	0.095282	-0.111644	0.220224	0.095282
K76S	-0.247507	0.197946	0.05273	-0.247507	0.197946	0.05273
K76T	-0.181255	0.384807	0.061624	-0.181255	0.384807	0.061624
K76V	0.251792	0.127647	0.059924	0.251792	0.127647	0.059924
K76W	0.098263	0.244195	0.098597	0.098263	0.244195	0.098597
K76Y	0.077203	0.16649	0.10201	0.077203	0.16649	0.10201
K76*	Null	nan	nan	Null	nan	nan
S77A	0.148609	0.000199	0.15288	0.148609	0.000199	0.15288
S77C	-0.160459	0.121309	0.118185	-0.160459	0.121309	0.118185

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
S77D	0.469379	0.126706	0.061725	0.469379	0.126706	0.061725
S77E	0.603844	0.107439	0.072532	0.603844	0.107439	0.072532
S77F	0.698144	0.264044	0.117786	0.698144	0.264044	0.117786
S77G	-0.185688	0.265061	0.077282	-0.185688	0.265061	0.077282
S77H	0.111931	0.144116	0.101685	0.111931	0.144116	0.101685
S77I	0.533847	0.183271	0.101853	0.533847	0.183271	0.101853
S77K	0.488524	0.111413	0.118842	0.488524	0.111413	0.118842
S77L	0.769964	0.527988	0.142826	0.769964	0.527988	0.142826
S77M	0.495228	0.0437	0.06718	0.495228	0.0437	0.06718
S77N	0.138541	0.065817	0.032988	0.138541	0.065817	0.032988
S77P	-1.680038	0.144468	0.341116	-1.680038	0.144468	0.341116
S77Q	0.408394	0.125683	0.119501	0.408394	0.125683	0.119501
S77R	0.474825	0.109412	0.083253	0.474825	0.109412	0.083253
S77T	0.608325	0.169146	0.103417	0.608325	0.169146	0.103417
S77V	0.731656	0.067679	0.079075	0.731656	0.067679	0.079075
S77W	0.43589	0.261778	0.061234	0.43589	0.261778	0.061234
S77Y	0.619035	0.088172	0.097962	0.619035	0.088172	0.097962
S77*	Null	nan	nan	Null	nan	nan
V78A	0.378043	0.176553	0.1203	0.378043	0.176553	0.1203
V78C	Null	nan	nan	Null	nan	nan
V78D	0.439085	0.355883	0.112247	0.439085	0.355883	0.112247
V78E	0.799959	0.116036	0.070692	0.799959	0.116036	0.070692
V78F	0.13785	0.302592	0.160354	0.13785	0.302592	0.160354
V78G	0.486742	0.085522	0.043022	0.486742	0.085522	0.043022
V78H	0.443611	0.206341	0.091961	0.443611	0.206341	0.091961
V78I	-0.489585	0.324338	0.088563	-0.489585	0.324338	0.088563
V78K	0.214516	0.085782	0.096974	0.214516	0.085782	0.096974
V78L	-0.400233	0.108842	0.083023	-0.400233	0.108842	0.083023
V78M	-0.011928	0.316492	0.111454	-0.011928	0.316492	0.111454
V78N	0.493073	0.10156	0.077361	0.493073	0.10156	0.077361
V78P	0.557716	0.152697	0.097346	0.557716	0.152697	0.097346
V78Q	0.244415	0.320636	0.080213	0.244415	0.320636	0.080213
V78R	0.412726	0.218261	0.056101	0.412726	0.218261	0.056101
V78S	0.33183	0.161825	0.08649	0.33183	0.161825	0.08649
V78T	0.17432	0.064047	0.093283	0.17432	0.064047	0.093283
V78W	0.654776	0.26926	0.110838	0.654776	0.26926	0.110838
V78Y	0.580902	0.305451	0.15695	0.580902	0.305451	0.15695
V78*	ND	ND	ND	ND	ND	ND
D79A	-0.252265	0.056712	0.076526	-0.252265	0.056712	0.076526
D79C	ND	ND	ND	ND	ND	ND
D79E	-0.253149	0.130727	0.136988	-0.253149	0.130727	0.136988
D79F	-0.256681	0.257203	0.098025	-0.256681	0.257203	0.098025
D79G	-0.220534	0.054244	0.082045	-0.220534	0.054244	0.082045
D79H	0.054197	0.180994	0.11399	0.054197	0.180994	0.11399
D79I	-0.075332	0.217221	0.144045	-0.075332	0.217221	0.144045
D79K	-0.37535	0.176038	0.110232	-0.37535	0.176038	0.110232
D79L	-0.361664	0.153362	0.096949	-0.361664	0.153362	0.096949
D79M	-0.013112	0.068953	0.092405	-0.013112	0.068953	0.092405
D79N	-0.0383	0.277273	0.077744	-0.0383	0.277273	0.077744
D79P	-0.053468	0.126661	0.068177	-0.053468	0.126661	0.068177

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
D79Q	-0.284641	0.178435	0.082396	-0.284641	0.178435	0.082396
D79R	-0.229605	0.059573	0.048831	-0.229605	0.059573	0.048831
D79S	-0.157553	0.198958	0.057872	-0.157553	0.198958	0.057872
D79T	0.14487	0.228901	0.09854	0.14487	0.228901	0.09854
D79V	-0.244249	0.022647	0.095169	-0.244249	0.022647	0.095169
D79W	-0.281891	0.341424	0.105225	-0.281891	0.341424	0.105225
D79Y	-0.385442	0.175112	0.125421	-0.385442	0.175112	0.125421
D79*	Null	nan	nan	Null	nan	nan
E80A	-0.437612	0.049587	0.088465	-0.437612	0.049587	0.088465
E80C	0.020903	0.185821	0.130969	0.020903	0.185821	0.130969
E80D	Null	nan	nan	Null	nan	nan
E80F	0.226115	0.019553	0.120835	0.226115	0.019553	0.120835
E80G	0.018864	0.137036	0.063557	0.018864	0.137036	0.063557
E80H	0.574017	0.591753	0.160964	0.574017	0.591753	0.160964
E80I	-0.085599	0.155313	0.109103	-0.085599	0.155313	0.109103
E80K	-0.341938	0.075441	0.054734	-0.341938	0.075441	0.054734
E80L	-0.209658	0.247888	0.059627	-0.209658	0.247888	0.059627
E80M	-0.149722	0.201585	0.067666	-0.149722	0.201585	0.067666
E80N	-0.222594	0.311402	0.172304	-0.222594	0.311402	0.172304
E80P	-0.187195	0.22912	0.085948	-0.187195	0.22912	0.085948
E80Q	-0.223494	0.07925	0.112811	-0.223494	0.07925	0.112811
E80R	-0.306736	0.147029	0.053434	-0.306736	0.147029	0.053434
E80S	-0.263483	0.040496	0.086701	-0.263483	0.040496	0.086701
E80T	0.204346	0.413346	0.084261	0.204346	0.413346	0.084261
E80V	0.047934	0.121514	0.056103	0.047934	0.121514	0.056103
E80W	-0.38734	0.191743	0.142729	-0.38734	0.191743	0.142729
E80Y	0.173802	0.522804	0.254107	0.173802	0.522804	0.254107
E80*	-1.074671	0.570999	0.261782	-1.074671	0.570999	0.261782
A81C	0.492112	0.025832	0.070703	0.492112	0.025832	0.070703
A81D	Null	nan	nan	Null	nan	nan
A81E	Null	nan	nan	Null	nan	nan
A81F	-0.554736	0.172635	0.06911	-0.554736	0.172635	0.06911
A81G	-0.175629	0.156616	0.076424	-0.175629	0.156616	0.076424
A81H	-1.81191	0.079934	0.309424	-1.81191	0.079934	0.309424
A81I	0.301475	0.328652	0.0786	0.301475	0.328652	0.0786
A81K	-0.546124	0.089039	0.335845	-0.546124	0.089039	0.335845
A81L	0.256431	0.365936	0.051213	0.256431	0.365936	0.051213
A81M	0.303926	0.328293	0.059352	0.303926	0.328293	0.059352
A81N	Null	nan	nan	Null	nan	nan
A81P	ND	ND	ND	ND	ND	ND
A81Q	ND	ND	ND	ND	ND	ND
A81R	ND	ND	ND	ND	ND	ND
A81S	-0.313907	0.2961	0.110211	-0.313907	0.2961	0.110211
A81T	0.53925	0.437402	0.09125	0.53925	0.437402	0.09125
A81V	0.577924	0.08344	0.065789	0.577924	0.08344	0.065789
A81W	-0.77523	0.184301	0.259072	-0.77523	0.184301	0.259072
A81Y	-0.975587	0.24761	0.172741	-0.975587	0.24761	0.172741
A81*	Null	nan	nan	Null	nan	nan
I82A	0.323538	0.220993	0.050392	0.323538	0.220993	0.050392
I82C	-0.031707	0.554421	0.128448	-0.031707	0.554421	0.128448

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
I82D	-0.503405	0.583292	0.085161	-0.503405	0.583292	0.085161
I82E	0.045166	0.125511	0.034396	0.045166	0.125511	0.034396
I82F	Null	nan	nan	Null	nan	nan
I82G	-0.272681	0.220576	0.053409	-0.272681	0.220576	0.053409
I82H	-0.11238	0.162135	0.113092	-0.11238	0.162135	0.113092
I82K	0.107562	0.187356	0.038948	0.107562	0.187356	0.038948
I82L	-0.200089	0.240863	0.026376	-0.200089	0.240863	0.026376
I82M	-0.057511	0.156842	0.052798	-0.057511	0.156842	0.052798
I82N	Null	nan	nan	Null	nan	nan
I82P	-0.56621	0.625815	0.144809	-0.56621	0.625815	0.144809
I82Q	-0.062365	0.184694	0.110971	-0.062365	0.184694	0.110971
I82R	0.035195	0.070135	0.036145	0.035195	0.070135	0.036145
I82S	0.360475	0.328921	0.074244	0.360475	0.328921	0.074244
I82T	0.094635	0.212893	0.071926	0.094635	0.212893	0.071926
I82V	0.093987	0.081965	0.061658	0.093987	0.081965	0.061658
I82W	0.227648	0.215865	0.078551	0.227648	0.215865	0.078551
I82Y	ND	ND	ND	ND	ND	ND
I82*	ND	ND	ND	ND	ND	ND
A83C	0.128998	0.128857	0.117767	0.128998	0.128857	0.117767
A83D	-0.193334	0.235817	0.088794	-0.193334	0.235817	0.088794
A83E	-0.074961	0.55091	0.057666	-0.074961	0.55091	0.057666
A83F	0.221441	0.171852	0.076228	0.221441	0.171852	0.076228
A83G	-0.077439	0.047811	0.043207	-0.077439	0.047811	0.043207
A83H	-0.098197	0.363211	0.102803	-0.098197	0.363211	0.102803
A83I	0.186708	0.389744	0.098224	0.186708	0.389744	0.098224
A83K	-0.065337	0.227597	0.055943	-0.065337	0.227597	0.055943
A83L	0.052106	0.192194	0.054445	0.052106	0.192194	0.054445
A83M	0.144401	0.068145	0.080659	0.144401	0.068145	0.080659
A83N	0.116688	0.117838	0.068282	0.116688	0.117838	0.068282
A83P	0.203429	0.492902	0.185138	0.203429	0.492902	0.185138
A83Q	-0.055019	0.08699	0.072286	-0.055019	0.08699	0.072286
A83R	-0.011263	0.308249	0.057833	-0.011263	0.308249	0.057833
A83S	0.022451	0.248012	0.135651	0.022451	0.248012	0.135651
A83T	0.13074	0.329813	0.111916	0.13074	0.329813	0.111916
A83V	0.097211	0.004661	0.067894	0.097211	0.004661	0.067894
A83W	0.296204	0.174432	0.085022	0.296204	0.174432	0.085022
A83Y	-0.15773	0.143415	0.101611	-0.15773	0.143415	0.101611
A83*	Null	nan	nan	Null	nan	nan
A84C	Null	nan	nan	Null	nan	nan
A84D	-0.118712	0.022949	0.085551	-0.118712	0.022949	0.085551
A84E	-0.456791	0.324411	0.062058	-0.456791	0.324411	0.062058
A84F	0.466733	0.012323	0.316032	0.466733	0.012323	0.316032
A84G	-0.330522	0.311131	0.056103	-0.330522	0.311131	0.056103
A84H	0.296255	0.262694	0.108495	0.296255	0.262694	0.108495
A84I	0.036611	0.225338	0.078139	0.036611	0.225338	0.078139
A84K	-0.252406	0.320904	0.052363	-0.252406	0.320904	0.052363
A84L	0.023606	0.104099	0.040057	0.023606	0.104099	0.040057
A84M	-0.194272	0.268777	0.048621	-0.194272	0.268777	0.048621
A84N	0.141116	0.425166	0.0617	0.141116	0.425166	0.0617
A84P	-0.170548	0.18801	0.038209	-0.170548	0.18801	0.038209

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
A84Q	-0.2587	0.280542	0.141147	-0.2587	0.280542	0.141147
A84R	-0.261641	0.15513	0.061391	-0.261641	0.15513	0.061391
A84S	0.063535	0.139821	0.087096	0.063535	0.139821	0.087096
A84T	0.122183	0.115946	0.10208	0.122183	0.115946	0.10208
A84V	0.013736	0.26313	0.063951	0.013736	0.26313	0.063951
A84W	ND	ND	ND	ND	ND	ND
A84Y	0.204087	0.497859	0.106255	0.204087	0.497859	0.106255
A84*	Null	nan	nan	Null	nan	nan
C85A	0.152698	0.046364	0.060605	0.152698	0.046364	0.060605
C85D	-0.20608	0.429499	0.073107	-0.20608	0.429499	0.073107
C85E	-0.264261	0.362238	0.042961	-0.264261	0.362238	0.042961
C85F	0.923853	0.309261	0.286558	0.923853	0.309261	0.286558
C85G	-0.01923	0.178935	0.04013	-0.01923	0.178935	0.04013
C85H	0.873249	0.176566	0.108943	0.873249	0.176566	0.108943
C85I	0.794269	0.217945	0.087139	0.794269	0.217945	0.087139
C85K	-0.119028	0.378167	0.055964	-0.119028	0.378167	0.055964
C85L	1.008159	0.089142	0.10165	1.008159	0.089142	0.10165
C85M	0.811862	0.059986	0.084539	0.811862	0.059986	0.084539
C85N	0.532947	0.199622	0.086025	0.532947	0.199622	0.086025
C85P	0.228106	0.25904	0.053789	0.228106	0.25904	0.053789
C85Q	0.404165	0.032079	0.094451	0.404165	0.032079	0.094451
C85R	-0.07404	0.216731	0.064646	-0.07404	0.216731	0.064646
C85S	-0.010201	0.077041	0.043418	-0.010201	0.077041	0.043418
C85T	-0.003783	0.161568	0.033742	-0.003783	0.161568	0.033742
C85V	0.251387	0.080515	0.057368	0.251387	0.080515	0.057368
C85W	0.454431	0.079646	0.059012	0.454431	0.079646	0.059012
C85Y	0.735476	0.506399	0.145458	0.735476	0.506399	0.145458
C85*	Null	nan	nan	Null	nan	nan
G86A	0.072589	0.362641	0.071508	0.072589	0.362641	0.071508
G86C	0.016911	0.307844	0.084039	0.016911	0.307844	0.084039
G86D	-0.073905	0.221473	0.11699	-0.073905	0.221473	0.11699
G86E	0.06828	0.372609	0.085698	0.06828	0.372609	0.085698
G86F	0.122735	0.305387	0.070465	0.122735	0.305387	0.070465
G86H	0.099081	0.14045	0.077782	0.099081	0.14045	0.077782
G86I	0.446522	0.239254	0.062006	0.446522	0.239254	0.062006
G86K	0.191441	0.20454	0.062164	0.191441	0.20454	0.062164
G86L	0.072605	0.048273	0.044829	0.072605	0.048273	0.044829
G86M	0.122501	0.433426	0.043668	0.122501	0.433426	0.043668
G86N	0.07474	0.168869	0.043402	0.07474	0.168869	0.043402
G86P	0.74206	0.1465	0.056867	0.74206	0.1465	0.056867
G86Q	0.193568	0.177573	0.077495	0.193568	0.177573	0.077495
G86R	0.260309	0.073082	0.018556	0.260309	0.073082	0.018556
G86S	0.231534	0.02966	0.057172	0.231534	0.02966	0.057172
G86T	0.380829	0.025321	0.056565	0.380829	0.025321	0.056565
G86V	0.351535	0.217313	0.060227	0.351535	0.217313	0.060227
G86W	-0.081452	0.238812	0.198267	-0.081452	0.238812	0.198267
G86Y	-0.152425	0.370165	0.065007	-0.152425	0.370165	0.065007
G86*	Null	nan	nan	Null	nan	nan
D87A	-0.095741	0.212889	0.192739	-0.095741	0.212889	0.192739
D87C	-0.454146	0.108473	0.110237	-0.454146	0.108473	0.110237

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
D87E	-0.011093	0.100003	0.046009	-0.011093	0.100003	0.046009
D87F	0.285017	0.158032	0.048651	0.285017	0.158032	0.048651
D87G	0.146667	0.149491	0.051386	0.146667	0.149491	0.051386
D87H	0.212643	0.430761	0.079428	0.212643	0.430761	0.079428
D87I	0.245531	0.1221	0.113639	0.245531	0.1221	0.113639
D87K	0.130005	0.16121	0.054193	0.130005	0.16121	0.054193
D87L	0.081368	0.057293	0.04013	0.081368	0.057293	0.04013
D87M	0.202803	0.173841	0.048754	0.202803	0.173841	0.048754
D87N	0.098705	0.122149	0.092128	0.098705	0.122149	0.092128
D87P	0.302212	0.065626	0.056806	0.302212	0.065626	0.056806
D87Q	0.187511	0.099998	0.122422	0.187511	0.099998	0.122422
D87R	-0.022868	0.495527	0.066195	-0.022868	0.495527	0.066195
D87S	0.129465	0.423832	0.035804	0.129465	0.423832	0.035804
D87T	0.074707	0.107115	0.059961	0.074707	0.107115	0.059961
D87V	-0.042843	0.263422	0.070907	-0.042843	0.263422	0.070907
D87W	0.069997	0.552712	0.075081	0.069997	0.552712	0.075081
D87Y	-0.049726	0.075402	0.085095	-0.049726	0.075402	0.085095
D87*	Null	nan	nan	Null	nan	nan
V88A	0.226678	0.318568	0.115924	0.226678	0.318568	0.115924
V88C	-0.053714	0.346035	0.099743	-0.053714	0.346035	0.099743
V88D	0.355887	0.018585	0.085495	0.355887	0.018585	0.085495
V88E	0.023735	0.115137	0.104933	0.023735	0.115137	0.104933
V88F	0.518621	0.311254	0.059719	0.518621	0.311254	0.059719
V88G	0.240584	0.039236	0.056597	0.240584	0.039236	0.056597
V88H	0.548148	0.237454	0.124277	0.548148	0.237454	0.124277
V88I	0.420329	0.34465	0.110918	0.420329	0.34465	0.110918
V88K	0.168018	0.090094	0.037329	0.168018	0.090094	0.037329
V88L	0.348722	0.190549	0.052666	0.348722	0.190549	0.052666
V88M	0.19727	0.132067	0.092348	0.19727	0.132067	0.092348
V88N	-0.237833	0.163582	0.058337	-0.237833	0.163582	0.058337
V88P	0.415195	0.424659	0.124431	0.415195	0.424659	0.124431
V88Q	0.085217	0.463181	0.080994	0.085217	0.463181	0.080994
V88R	0.091787	0.564684	0.05754	0.091787	0.564684	0.05754
V88S	0.306046	0.192308	0.059293	0.306046	0.192308	0.059293
V88T	0.263339	0.049067	0.088563	0.263339	0.049067	0.088563
V88W	0.382103	0.136473	0.047045	0.382103	0.136473	0.047045
V88Y	ND	ND	ND	ND	ND	ND
V88*	Null	nan	nan	Null	nan	nan
P89A	0.117905	0.233173	0.232813	0.117905	0.233173	0.232813
P89C	0.392345	0.376802	0.232737	0.392345	0.376802	0.232737
P89D	-0.703087	0.634716	0.168652	-0.703087	0.634716	0.168652
P89E	-0.060435	0.438581	0.139195	-0.060435	0.438581	0.139195
P89F	0.067248	0.32084	0.247569	0.067248	0.32084	0.247569
P89G	0.050576	0.376342	0.097568	0.050576	0.376342	0.097568
P89H	0.00881	0.054633	0.08315	0.00881	0.054633	0.08315
P89I	-0.244788	0.043757	0.139717	-0.244788	0.043757	0.139717
P89K	-0.040804	0.537871	0.103845	-0.040804	0.537871	0.103845
P89L	0.031493	0.141412	0.069911	0.031493	0.141412	0.069911
P89M	ND	ND	ND	ND	ND	ND
P89N	-0.261352	0.300107	0.052621	-0.261352	0.300107	0.052621

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
P89Q	0.003358	0.267976	0.08747	0.003358	0.267976	0.08747
P89R	-0.040169	0.227416	0.118214	-0.040169	0.227416	0.118214
P89S	-0.051504	0.319634	0.089476	-0.051504	0.319634	0.089476
P89T	ND	ND	ND	ND	ND	ND
P89V	0.0731	0.265568	0.111443	0.0731	0.265568	0.111443
P89W	0.052078	0.55669	0.163221	0.052078	0.55669	0.163221
P89Y	-0.072105	0.145266	0.072219	-0.072105	0.145266	0.072219
P89*	Null	nan	nan	Null	nan	nan
E90A	-0.136828	0.568164	0.084104	-0.136828	0.568164	0.084104
E90C	ND	ND	ND	ND	ND	ND
E90D	0.053967	0.137975	0.117857	0.053967	0.137975	0.117857
E90F	ND	ND	ND	ND	ND	ND
E90G	-0.314472	0.079356	0.06128	-0.314472	0.079356	0.06128
E90H	0.006964	0.15177	0.105786	0.006964	0.15177	0.105786
E90I	ND	ND	ND	ND	ND	ND
E90K	-0.858225	0.086091	0.169583	-0.858225	0.086091	0.169583
E90L	-0.858598	0.21347	0.126837	-0.858598	0.21347	0.126837
E90M	-0.689134	0.185545	0.077089	-0.689134	0.185545	0.077089
E90N	-0.123669	0.232773	0.07458	-0.123669	0.232773	0.07458
E90P	Null	nan	nan	Null	nan	nan
E90Q	-0.325896	0.246683	0.202304	-0.325896	0.246683	0.202304
E90R	-0.493224	0.316103	0.088269	-0.493224	0.316103	0.088269
E90S	0.169941	0.118365	0.100916	0.169941	0.118365	0.100916
E90T	-0.590796	0.038639	0.115885	-0.590796	0.038639	0.115885
E90V	-0.493375	0.107659	0.078494	-0.493375	0.107659	0.078494
E90W	-0.373371	0.215966	0.102214	-0.373371	0.215966	0.102214
E90Y	-0.334474	0.070168	0.181154	-0.334474	0.070168	0.181154
E90*	Null	nan	nan	Null	nan	nan
I91A	1.042866	0.15741	0.07165	1.042866	0.15741	0.07165
I91C	0.646227	0.101826	0.161898	0.646227	0.101826	0.161898
I91D	-2.014601	0.783812	0.253129	-2.014601	0.783812	0.253129
I91E	-1.016723	0.14572	0.093971	-1.016723	0.14572	0.093971
I91F	Null	nan	nan	Null	nan	nan
I91G	1.03477	0.04214	0.053504	1.03477	0.04214	0.053504
I91H	-0.339373	0.371851	0.174167	-0.339373	0.371851	0.174167
I91K	0.072374	0.238949	0.050555	0.072374	0.238949	0.050555
I91L	0.638111	0.162533	0.098616	0.638111	0.162533	0.098616
I91M	0.595238	0.076912	0.089597	0.595238	0.076912	0.089597
I91N	-0.092879	0.027018	0.121598	-0.092879	0.027018	0.121598
I91P	0.227489	0.110361	0.084553	0.227489	0.110361	0.084553
I91Q	-0.0227	0.212912	0.167541	-0.0227	0.212912	0.167541
I91R	-0.185289	0.56381	0.057924	-0.185289	0.56381	0.057924
I91S	0.778548	0.404149	0.087832	0.778548	0.404149	0.087832
I91T	0.571104	0.256661	0.069263	0.571104	0.256661	0.069263
I91V	0.632837	0.637068	0.124855	0.632837	0.637068	0.124855
I91W	0.368498	0.305729	0.118843	0.368498	0.305729	0.118843
I91Y	0.270208	0.382994	0.104634	0.270208	0.382994	0.104634
I91*	Null	nan	nan	Null	nan	nan
M92A	0.549621	0.151307	0.065676	0.549621	0.151307	0.065676
M92C	0.632375	0.286578	0.079107	0.632375	0.286578	0.079107

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
M92D	-1.723517	0.655359	0.372251	-1.723517	0.655359	0.372251
M92E	-1.829311	0.25768	0.220215	-1.829311	0.25768	0.220215
M92F	0.874195	0.430258	0.100236	0.874195	0.430258	0.100236
M92G	-0.190773	0.209847	0.071835	-0.190773	0.209847	0.071835
M92H	0.493699	0.211883	0.079144	0.493699	0.211883	0.079144
M92I	0.912665	0.04592	0.084383	0.912665	0.04592	0.084383
M92K	-1.008181	0.566081	0.244331	-1.008181	0.566081	0.244331
M92L	0.64567	0.160422	0.086685	0.64567	0.160422	0.086685
M92N	0.393175	0.047591	0.064774	0.393175	0.047591	0.064774
M92P	-0.480891	0.140324	0.182623	-0.480891	0.140324	0.182623
M92Q	0.341911	0.197233	0.07605	0.341911	0.197233	0.07605
M92R	-1.423954	0.202179	0.183316	-1.423954	0.202179	0.183316
M92S	0.456167	0.368599	0.053845	0.456167	0.368599	0.053845
M92T	0.826594	0.117582	0.189599	0.826594	0.117582	0.189599
M92V	0.571769	0.309201	0.09258	0.571769	0.309201	0.09258
M92W	0.343996	0.050751	0.051318	0.343996	0.050751	0.051318
M92Y	0.711877	0.314193	0.053312	0.711877	0.314193	0.053312
M92*	Null	nan	nan	Null	nan	nan
V93A	0.474191	0.194419	0.040418	0.474191	0.194419	0.040418
V93C	0.138963	0.206482	0.072673	0.138963	0.206482	0.072673
V93D	Null	nan	nan	Null	nan	nan
V93E	Null	nan	nan	Null	nan	nan
V93F	-0.659841	0.369339	0.225433	-0.659841	0.369339	0.225433
V93G	-0.206141	0.055228	0.052727	-0.206141	0.055228	0.052727
V93H	Null	nan	nan	Null	nan	nan
V93I	-0.124037	0.17316	0.059555	-0.124037	0.17316	0.059555
V93K	ND	ND	ND	ND	ND	ND
V93L	-0.095671	0.113802	0.058401	-0.095671	0.113802	0.058401
V93M	-0.776964	0.05595	0.144107	-0.776964	0.05595	0.144107
V93N	-1.349158	0.254365	0.085647	-1.349158	0.254365	0.085647
V93P	-1.110513	0.590634	0.338693	-1.110513	0.590634	0.338693
V93Q	Null	nan	nan	Null	nan	nan
V93R	-1.412942	0.367618	0.335409	-1.412942	0.367618	0.335409
V93S	-0.340304	0.057785	0.065824	-0.340304	0.057785	0.065824
V93T	0.024087	0.116779	0.069916	0.024087	0.116779	0.069916
V93W	ND	ND	ND	ND	ND	ND
V93Y	Null	nan	nan	Null	nan	nan
V93*	Null	nan	nan	Null	nan	nan
I94A	-0.820451	0.574034	0.116653	-0.820451	0.574034	0.116653
I94C	-0.387931	0.248752	0.15621	-0.387931	0.248752	0.15621
I94D	Null	nan	nan	Null	nan	nan
I94E	Null	nan	nan	Null	nan	nan
I94F	Null	nan	nan	Null	nan	nan
I94G	-1.735347	0.200052	0.114994	-1.735347	0.200052	0.114994
I94H	Null	nan	nan	Null	nan	nan
I94K	ND	ND	ND	ND	ND	ND
I94L	0.512004	0.134942	0.048609	0.512004	0.134942	0.048609
I94M	-0.392177	0.163052	0.094247	-0.392177	0.163052	0.094247
I94N	-2.110788	0.131863	0.202729	-2.110788	0.131863	0.202729
I94P	-1.82809	0.514441	0.161376	-1.82809	0.514441	0.161376

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
I94Q	-0.338092	0.236822	0.101361	-0.338092	0.236822	0.101361
I94R	Null	nan	nan	Null	nan	nan
I94S	-0.790964	0.461205	0.092237	-0.790964	0.461205	0.092237
I94T	-0.57566	0.08697	0.064552	-0.57566	0.08697	0.064552
I94V	-0.142149	0.195344	0.066396	-0.142149	0.195344	0.066396
I94W	Null	nan	nan	Null	nan	nan
I94Y	Null	nan	nan	Null	nan	nan
I94*	Null	nan	nan	Null	nan	nan
G95A	Null	nan	nan	Null	nan	nan
G95C	Null	nan	nan	Null	nan	nan
G95D	Null	nan	nan	Null	nan	nan
G95E	Null	nan	nan	Null	nan	nan
G95F	Null	nan	nan	Null	nan	nan
G95H	Null	nan	nan	Null	nan	nan
G95I	Null	nan	nan	Null	nan	nan
G95K	ND	ND	ND	ND	ND	ND
G95L	-1.321424	0.042369	0.24161	-1.321424	0.042369	0.24161
G95M	Null	nan	nan	Null	nan	nan
G95N	Null	nan	nan	Null	nan	nan
G95P	-1.779887	0.186882	0.408402	-1.779887	0.186882	0.408402
G95Q	Null	nan	nan	Null	nan	nan
G95R	Null	nan	nan	Null	nan	nan
G95S	Null	nan	nan	Null	nan	nan
G95T	Null	nan	nan	Null	nan	nan
G95V	Null	nan	nan	Null	nan	nan
G95W	ND	ND	ND	ND	ND	ND
G95Y	Null	nan	nan	Null	nan	nan
G95*	Null	nan	nan	Null	nan	nan
G96A	Null	nan	nan	Null	nan	nan
G96C	Null	nan	nan	Null	nan	nan
G96D	Null	nan	nan	Null	nan	nan
G96E	ND	ND	ND	ND	ND	ND
G96F	Null	nan	nan	Null	nan	nan
G96H	Null	nan	nan	Null	nan	nan
G96I	Null	nan	nan	Null	nan	nan
G96K	ND	ND	ND	ND	ND	ND
G96L	-0.347066	0.634165	0.277068	-0.347066	0.634165	0.277068
G96M	ND	ND	ND	ND	ND	ND
G96N	Null	nan	nan	Null	nan	nan
G96P	Null	nan	nan	Null	nan	nan
G96Q	Null	nan	nan	Null	nan	nan
G96R	-1.767201	0.937387	0.270797	-1.767201	0.937387	0.270797
G96S	Null	nan	nan	Null	nan	nan
G96T	Null	nan	nan	Null	nan	nan
G96V	Null	nan	nan	Null	nan	nan
G96W	ND	ND	ND	ND	ND	ND
G96Y	Null	nan	nan	Null	nan	nan
G96*	Null	nan	nan	Null	nan	nan
G97A	0.073802	0.305079	0.099426	0.073802	0.305079	0.099426
G97C	-0.905031	0.476187	0.19891	-0.905031	0.476187	0.19891

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
G97D	-0.650518	0.320358	0.143876	-0.650518	0.320358	0.143876
G97E	-0.424593	0.391934	0.061846	-0.424593	0.391934	0.061846
G97F	-1.211768	0.380177	0.106747	-1.211768	0.380177	0.106747
G97H	-1.541859	0.372481	0.095433	-1.541859	0.372481	0.095433
G97I	Null	nan	nan	Null	nan	nan
G97K	-0.950801	0.217876	0.060576	-0.950801	0.217876	0.060576
G97L	-1.384659	0.180988	0.153375	-1.384659	0.180988	0.153375
G97M	-0.626265	0.155157	0.106355	-0.626265	0.155157	0.106355
G97N	-0.427785	0.267297	0.069831	-0.427785	0.267297	0.069831
G97P	-2.455005	0.357641	0.306112	-2.455005	0.357641	0.306112
G97Q	-0.021675	0.140179	0.049704	-0.021675	0.140179	0.049704
G97R	-0.340717	0.352542	0.074374	-0.340717	0.352542	0.074374
G97S	-0.390731	0.164425	0.067899	-0.390731	0.164425	0.067899
G97T	-0.437134	0.156443	0.080883	-0.437134	0.156443	0.080883
G97V	-2.008201	0.203885	0.191863	-2.008201	0.203885	0.191863
G97W	-2.214095	0.839955	0.142479	-2.214095	0.839955	0.142479
G97Y	-1.496712	0.483698	0.146015	-1.496712	0.483698	0.146015
G97*	Null	nan	nan	Null	nan	nan
R98A	0.045951	0.128812	0.053661	0.045951	0.128812	0.053661
R98C	Null	nan	nan	Null	nan	nan
R98D	-1.043681	0.260432	0.220089	-1.043681	0.260432	0.220089
R98E	-1.659886	0.196616	0.185991	-1.659886	0.196616	0.185991
R98F	0.779584	0.136086	0.087455	0.779584	0.136086	0.087455
R98G	0.505287	0.230329	0.080155	0.505287	0.230329	0.080155
R98H	0.858887	0.070818	0.131232	0.858887	0.070818	0.131232
R98I	-1.551889	0.398493	0.208284	-1.551889	0.398493	0.208284
R98K	-0.113436	0.353078	0.07871	-0.113436	0.353078	0.07871
R98L	-1.246332	0.492268	0.078614	-1.246332	0.492268	0.078614
R98M	-0.901989	0.678736	0.288727	-0.901989	0.678736	0.288727
R98N	-0.595799	0.415167	0.067357	-0.595799	0.415167	0.067357
R98P	-2.137198	0.243606	0.171377	-2.137198	0.243606	0.171377
R98Q	-1.358556	0.217649	0.164007	-1.358556	0.217649	0.164007
R98S	-1.028567	0.224612	0.054981	-1.028567	0.224612	0.054981
R98T	-0.59207	0.100428	0.110397	-0.59207	0.100428	0.110397
R98V	-1.340475	0.0247	0.300414	-1.340475	0.0247	0.300414
R98W	0.852667	0.11822	0.066314	0.852667	0.11822	0.066314
R98Y	0.920043	0.171635	0.07696	0.920043	0.171635	0.07696
R98*	Null	nan	nan	Null	nan	nan
V99A	0.194465	0.308086	0.075416	0.194465	0.308086	0.075416
V99C	0.398699	0.276294	0.095611	0.398699	0.276294	0.095611
V99D	Null	nan	nan	Null	nan	nan
V99E	Null	nan	nan	Null	nan	nan
V99F	ND	ND	ND	ND	ND	ND
V99G	0.372453	0.018293	0.061269	0.372453	0.018293	0.061269
V99H	-1.412935	0.283579	0.404264	-1.412935	0.283579	0.404264
V99I	0.592402	0.379316	0.040576	0.592402	0.379316	0.040576
V99K	ND	ND	ND	ND	ND	ND
V99L	0.348107	0.12977	0.028684	0.348107	0.12977	0.028684
V99M	0.258413	0.116003	0.084865	0.258413	0.116003	0.084865
V99N	0.093385	0.082154	0.110867	0.093385	0.082154	0.110867

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
V99P	-1.427301	0.468	0.158427	-1.427301	0.468	0.158427
V99Q	-1.575005	0.47304	0.312308	-1.575005	0.47304	0.312308
V99R	-0.849371	0.658467	0.292127	-0.849371	0.658467	0.292127
V99S	0.173059	0.122032	0.049648	0.173059	0.122032	0.049648
V99T	-0.750074	0.609827	0.12371	-0.750074	0.609827	0.12371
V99W	Null	nan	nan	Null	nan	nan
V99Y	Null	nan	nan	Null	nan	nan
V99*	Null	nan	nan	Null	nan	nan
Y100A	-0.319616	0.145222	0.091985	-0.319616	0.145222	0.091985
Y100C	-0.037659	0.304571	0.161274	-0.037659	0.304571	0.161274
Y100D	-1.281924	0.456635	0.272296	-1.281924	0.456635	0.272296
Y100E	ND	ND	ND	ND	ND	ND
Y100F	ND	ND	ND	ND	ND	ND
Y100G	-1.965319	0.530037	0.170025	-1.965319	0.530037	0.170025
Y100H	ND	ND	ND	ND	ND	ND
Y100I	ND	ND	ND	ND	ND	ND
Y100K	ND	ND	ND	ND	ND	ND
Y100L	-0.984365	0.302047	0.128275	-0.984365	0.302047	0.128275
Y100M	-0.829519	0.198074	0.131517	-0.829519	0.198074	0.131517
Y100N	-1.070512	0.074967	0.131224	-1.070512	0.074967	0.131224
Y100P	-0.526692	0.50471	0.353055	-0.526692	0.50471	0.353055
Y100Q	-1.432712	0.287854	0.164832	-1.432712	0.287854	0.164832
Y100R	ND	ND	ND	ND	ND	ND
Y100S	-0.603093	0.154502	0.118445	-0.603093	0.154502	0.118445
Y100T	-1.920862	0.352624	0.405054	-1.920862	0.352624	0.405054
Y100V	-1.389177	0.108894	0.200217	-1.389177	0.108894	0.200217
Y100W	0.291978	0.192792	0.070766	0.291978	0.192792	0.070766
Y100*	Null	nan	nan	Null	nan	nan
E101A	-0.037563	0.408996	0.048965	-0.037563	0.408996	0.048965
E101C	-0.00994	0.434243	0.08004	-0.00994	0.434243	0.08004
E101D	-0.295015	0.206339	0.069369	-0.295015	0.206339	0.069369
E101F	0.114699	0.134174	0.06214	0.114699	0.134174	0.06214
E101G	0.272774	0.221407	0.037682	0.272774	0.221407	0.037682
E101H	0.130153	0.171502	0.060593	0.130153	0.171502	0.060593
E101I	-0.129595	0.274468	0.054516	-0.129595	0.274468	0.054516
E101K	0.097136	0.252737	0.066681	0.097136	0.252737	0.066681
E101L	0.002649	0.162326	0.061538	0.002649	0.162326	0.061538
E101M	0.164558	0.195963	0.042418	0.164558	0.195963	0.042418
E101N	-0.151446	0.346042	0.056945	-0.151446	0.346042	0.056945
E101P	-1.951765	0.488085	0.149214	-1.951765	0.488085	0.149214
E101Q	0.042053	0.257852	0.07472	0.042053	0.257852	0.07472
E101R	0.054559	0.111579	0.031258	0.054559	0.111579	0.031258
E101S	-0.051353	0.031734	0.047886	-0.051353	0.031734	0.047886
E101T	-0.176015	0.022249	0.047957	-0.176015	0.022249	0.047957
E101V	-0.11707	0.162289	0.058725	-0.11707	0.162289	0.058725
E101W	0.295671	0.187938	0.050039	0.295671	0.187938	0.050039
E101Y	-0.048394	0.601069	0.050241	-0.048394	0.601069	0.050241
E101*	Null	nan	nan	Null	nan	nan
Q102A	0.24836	0.059913	0.058312	0.24836	0.059913	0.058312
Q102C	0.218148	0.57172	0.178188	0.218148	0.57172	0.178188

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
Q102D	-0.057573	0.392498	0.073416	-0.057573	0.392498	0.073416
Q102E	0.321358	0.029568	0.046559	0.321358	0.029568	0.046559
Q102F	0.762628	0.151916	0.069189	0.762628	0.151916	0.069189
Q102G	0.122871	0.180306	0.049774	0.122871	0.180306	0.049774
Q102H	0.499828	0.42216	0.083761	0.499828	0.42216	0.083761
Q102I	0.670161	0.045796	0.053494	0.670161	0.045796	0.053494
Q102K	0.008247	0.206629	0.057026	0.008247	0.206629	0.057026
Q102L	0.618421	0.148012	0.038483	0.618421	0.148012	0.038483
Q102M	0.206822	0.228613	0.076991	0.206822	0.228613	0.076991
Q102N	0.493389	0.169878	0.071903	0.493389	0.169878	0.071903
Q102P	-0.174412	0.110198	0.053766	-0.174412	0.110198	0.053766
Q102R	0.202813	0.276315	0.060308	0.202813	0.276315	0.060308
Q102S	0.270965	0.120615	0.041492	0.270965	0.120615	0.041492
Q102T	0.505889	0.049195	0.061433	0.505889	0.049195	0.061433
Q102V	0.50888	0.02508	0.054129	0.50888	0.02508	0.054129
Q102W	0.805234	0.119495	0.101728	0.805234	0.119495	0.101728
Q102Y	0.618991	0.316399	0.087576	0.618991	0.316399	0.087576
Q102*	Null	nan	nan	Null	nan	nan
F103A	0.830451	0.19327	0.067673	0.830451	0.19327	0.067673
F103C	ND	ND	ND	ND	ND	ND
F103D	ND	ND	ND	ND	ND	ND
F103E	-1.933638	0.757237	0.387782	-1.933638	0.757237	0.387782
F103G	-0.098842	0.043449	0.105423	-0.098842	0.043449	0.105423
F103H	-0.084574	0.15103	0.118226	-0.084574	0.15103	0.118226
F103I	ND	ND	ND	ND	ND	ND
F103K	-1.131921	0.542354	0.295089	-1.131921	0.542354	0.295089
F103L	-0.53322	0.355769	0.049659	-0.53322	0.355769	0.049659
F103M	-0.495985	0.09872	0.049094	-0.495985	0.09872	0.049094
F103N	-0.481889	0.159139	0.199763	-0.481889	0.159139	0.199763
F103P	-0.86133	0.192971	0.245934	-0.86133	0.192971	0.245934
F103Q	ND	ND	ND	ND	ND	ND
F103R	-1.209504	0.743359	0.43181	-1.209504	0.743359	0.43181
F103S	0.581643	0.216679	0.058151	0.581643	0.216679	0.058151
F103T	0.192945	0.137	0.065443	0.192945	0.137	0.065443
F103V	0.121685	0.198691	0.036505	0.121685	0.198691	0.036505
F103W	-0.082976	0.417631	0.095579	-0.082976	0.417631	0.095579
F103Y	-1.637003	0.070963	0.218831	-1.637003	0.070963	0.218831
F103*	Null	nan	nan	Null	nan	nan
L104A	ND	ND	ND	ND	ND	ND
L104C	Null	nan	nan	Null	nan	nan
L104D	Null	nan	nan	Null	nan	nan
L104E	ND	ND	ND	ND	ND	ND
L104F	Null	nan	nan	Null	nan	nan
L104G	-0.741285	0.17179	0.189329	-0.741285	0.17179	0.189329
L104H	Null	nan	nan	Null	nan	nan
L104I	ND	ND	ND	ND	ND	ND
L104K	ND	ND	ND	ND	ND	ND
L104M	Null	nan	nan	Null	nan	nan
L104N	Null	nan	nan	Null	nan	nan
L104P	Null	nan	nan	Null	nan	nan

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
L104Q	ND	ND	ND	ND	ND	ND
L104R	-1.211744	0.187436	0.193124	-1.211744	0.187436	0.193124
L104S	Null	nan	nan	Null	nan	nan
L104T	Null	nan	nan	Null	nan	nan
L104V	-1.445437	0.347741	0.456899	-1.445437	0.347741	0.456899
L104W	Null	nan	nan	Null	nan	nan
L104Y	Null	nan	nan	Null	nan	nan
L104*	Null	nan	nan	Null	nan	nan
P105A	0.11294	0.329581	0.056281	0.11294	0.329581	0.056281
P105C	0.180494	0.4823	0.072117	0.180494	0.4823	0.072117
P105D	-0.260746	0.274652	0.046446	-0.260746	0.274652	0.046446
P105E	0.080946	0.237903	0.035066	0.080946	0.237903	0.035066
P105F	0.169347	0.336609	0.089422	0.169347	0.336609	0.089422
P105G	0.047684	0.245483	0.041908	0.047684	0.245483	0.041908
P105H	-0.094274	0.146922	0.094268	-0.094274	0.146922	0.094268
P105I	-0.148519	0.306671	0.096546	-0.148519	0.306671	0.096546
P105K	0.089054	0.125425	0.026634	0.089054	0.125425	0.026634
P105L	0.162087	0.089377	0.035236	0.162087	0.089377	0.035236
P105M	0.21106	0.078675	0.071874	0.21106	0.078675	0.071874
P105N	0.04877	0.065581	0.046866	0.04877	0.065581	0.046866
P105Q	0.092351	0.311284	0.091387	0.092351	0.311284	0.091387
P105R	0.175127	0.209524	0.032955	0.175127	0.209524	0.032955
P105S	0.041318	0.22136	0.064911	0.041318	0.22136	0.064911
P105T	0.193681	0.003289	0.057882	0.193681	0.003289	0.057882
P105V	-0.053565	0.193278	0.041256	-0.053565	0.193278	0.041256
P105W	0.008445	0.187172	0.060092	0.008445	0.187172	0.060092
P105Y	0.136341	0.156876	0.122003	0.136341	0.156876	0.122003
P105*	ND	ND	ND	ND	ND	ND
K106A	0.101772	0.441285	0.105044	0.101772	0.441285	0.105044
K106C	ND	ND	ND	ND	ND	ND
K106D	-0.119334	0.266511	0.099904	-0.119334	0.266511	0.099904
K106E	0.173552	0.062962	0.074665	0.173552	0.062962	0.074665
K106F	ND	ND	ND	ND	ND	ND
K106G	0.198685	0.157713	0.036108	0.198685	0.157713	0.036108
K106H	0.488947	0.245321	0.095827	0.488947	0.245321	0.095827
K106I	0.163069	0.123842	0.092849	0.163069	0.123842	0.092849
K106L	-0.274516	0.282779	0.090328	-0.274516	0.282779	0.090328
K106M	-0.460485	0.287503	0.063403	-0.460485	0.287503	0.063403
K106N	0.371563	0.047195	0.122906	0.371563	0.047195	0.122906
K106P	-0.143067	0.363737	0.058755	-0.143067	0.363737	0.058755
K106Q	0.631843	0.417328	0.089062	0.631843	0.417328	0.089062
K106R	0.190714	0.217424	0.042006	0.190714	0.217424	0.042006
K106S	0.051835	0.195987	0.059562	0.051835	0.195987	0.059562
K106T	0.156872	0.05714	0.070335	0.156872	0.05714	0.070335
K106V	0.220433	0.074314	0.049415	0.220433	0.074314	0.049415
K106W	0.253378	0.150207	0.099617	0.253378	0.150207	0.099617
K106Y	0.387209	0.364598	0.160134	0.387209	0.364598	0.160134
K106*	Null	nan	nan	Null	nan	nan
A107C	-0.025658	0.061177	0.186599	-0.025658	0.061177	0.186599
A107D	Null	nan	nan	Null	nan	nan

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
A107E	Null	nan	nan	Null	nan	nan
A107F	Null	nan	nan	Null	nan	nan
A107G	-0.281922	0.355609	0.088262	-0.281922	0.355609	0.088262
A107H	Null	nan	nan	Null	nan	nan
A107I	0.863368	0.147451	0.157905	0.863368	0.147451	0.157905
A107K	Null	nan	nan	Null	nan	nan
A107L	-0.636853	0.151195	0.120092	-0.636853	0.151195	0.120092
A107M	ND	ND	ND	ND	ND	ND
A107N	Null	nan	nan	Null	nan	nan
A107P	Null	nan	nan	Null	nan	nan
A107Q	Null	nan	nan	Null	nan	nan
A107R	Null	nan	nan	Null	nan	nan
A107S	-0.004634	0.141113	0.079155	-0.004634	0.141113	0.079155
A107T	0.372023	0.301911	0.057682	0.372023	0.301911	0.057682
A107V	ND	ND	ND	ND	ND	ND
A107W	Null	nan	nan	Null	nan	nan
A107Y	Null	nan	nan	Null	nan	nan
A107*	Null	nan	nan	Null	nan	nan
Q108A	-0.047264	0.098259	0.05366	-0.047264	0.098259	0.05366
Q108C	-0.043604	0.016128	0.101961	-0.043604	0.016128	0.101961
Q108D	ND	ND	ND	ND	ND	ND
Q108E	-0.016009	0.271719	0.094786	-0.016009	0.271719	0.094786
Q108F	-0.006214	0.12144	0.07283	-0.006214	0.12144	0.07283
Q108G	0.113101	0.219512	0.026893	0.113101	0.219512	0.026893
Q108H	0.151955	0.175444	0.069091	0.151955	0.175444	0.069091
Q108I	-0.388836	0.526452	0.069785	-0.388836	0.526452	0.069785
Q108K	0.200665	0.203734	0.062239	0.200665	0.203734	0.062239
Q108L	0.006425	0.081649	0.052398	0.006425	0.081649	0.052398
Q108M	-0.143545	0.138096	0.086654	-0.143545	0.138096	0.086654
Q108N	0.130532	0.258417	0.073625	0.130532	0.258417	0.073625
Q108P	-1.211687	0.248485	0.118421	-1.211687	0.248485	0.118421
Q108R	0.039241	0.26222	0.059504	0.039241	0.26222	0.059504
Q108S	0.080863	0.310881	0.033293	0.080863	0.310881	0.033293
Q108T	0.180746	0.156225	0.068846	0.180746	0.156225	0.068846
Q108V	-0.097902	0.08961	0.062056	-0.097902	0.08961	0.062056
Q108W	-0.016944	0.280593	0.051644	-0.016944	0.280593	0.051644
Q108Y	-0.028479	0.403216	0.123261	-0.028479	0.403216	0.123261
Q108*	ND	ND	ND	ND	ND	ND
K109A	0.343604	0.275457	0.091495	0.343604	0.275457	0.091495
K109C	0.264489	0.438066	0.132757	0.264489	0.438066	0.132757
K109D	-1.825645	0.463618	0.336676	-1.825645	0.463618	0.336676
K109E	-0.204653	0.511448	0.132465	-0.204653	0.511448	0.132465
K109F	0.209619	0.18615	0.077067	0.209619	0.18615	0.077067
K109G	0.165315	0.094417	0.036856	0.165315	0.094417	0.036856
K109H	0.465864	0.402686	0.078435	0.465864	0.402686	0.078435
K109I	0.021728	0.461847	0.092591	0.021728	0.461847	0.092591
K109L	0.278378	0.032849	0.038264	0.278378	0.032849	0.038264
K109M	-0.113166	0.134019	0.086175	-0.113166	0.134019	0.086175
K109N	0.331119	0.221206	0.020366	0.331119	0.221206	0.020366
K109P	-1.624668	0.272381	0.253079	-1.624668	0.272381	0.253079

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
K109Q	-0.240799	0.389965	0.275687	-0.240799	0.389965	0.275687
K109R	0.255028	0.428573	0.067104	0.255028	0.428573	0.067104
K109S	0.07877	0.084608	0.068357	0.07877	0.084608	0.068357
K109T	0.270891	0.203143	0.036992	0.270891	0.203143	0.036992
K109V	0.066794	0.448566	0.045717	0.066794	0.448566	0.045717
K109W	0.346198	0.279092	0.114293	0.346198	0.279092	0.114293
K109Y	0.356448	0.156092	0.108266	0.356448	0.156092	0.108266
K109*	Null	nan	nan	Null	nan	nan
L110A	-0.704337	0.278985	0.094081	-0.704337	0.278985	0.094081
L110C	-0.404459	0.000789	0.12524	-0.404459	0.000789	0.12524
L110D	Null	nan	nan	Null	nan	nan
L110E	ND	ND	ND	ND	ND	ND
L110F	ND	ND	ND	ND	ND	ND
L110G	-1.971494	0.481129	0.258251	-1.971494	0.481129	0.258251
L110H	Null	nan	nan	Null	nan	nan
L110I	0.203503	0.40174	0.102628	0.203503	0.40174	0.102628
L110K	-1.198374	0.284298	0.230412	-1.198374	0.284298	0.230412
L110M	ND	ND	ND	ND	ND	ND
L110N	-1.734008	0.251267	0.183394	-1.734008	0.251267	0.183394
L110P	Null	nan	nan	Null	nan	nan
L110Q	Null	nan	nan	Null	nan	nan
L110R	Null	nan	nan	Null	nan	nan
L110S	-1.033849	0.237471	0.165196	-1.033849	0.237471	0.165196
L110T	-0.493152	0.270103	0.094554	-0.493152	0.270103	0.094554
L110V	-0.432457	0.382993	0.050471	-0.432457	0.382993	0.050471
L110W	Null	nan	nan	Null	nan	nan
L110Y	ND	ND	ND	ND	ND	ND
L110*	Null	nan	nan	Null	nan	nan
Y111A	-0.678335	0.230512	0.123185	-0.678335	0.230512	0.123185
Y111C	0.675761	0.398223	0.199023	0.675761	0.398223	0.199023
Y111D	Null	nan	nan	Null	nan	nan
Y111E	ND	ND	ND	ND	ND	ND
Y111F	0.294315	0.155844	0.154779	0.294315	0.155844	0.154779
Y111G	-1.167126	0.643523	0.262277	-1.167126	0.643523	0.262277
Y111H	-0.066718	0.177266	0.115749	-0.066718	0.177266	0.115749
Y111I	0.313182	0.174483	0.074184	0.313182	0.174483	0.074184
Y111K	-1.102216	0.422732	0.244556	-1.102216	0.422732	0.244556
Y111L	0.050031	0.19359	0.057847	0.050031	0.19359	0.057847
Y111M	0.142467	0.123381	0.077071	0.142467	0.123381	0.077071
Y111N	-0.772766	0.329533	0.215385	-0.772766	0.329533	0.215385
Y111P	Null	nan	nan	Null	nan	nan
Y111Q	-1.215312	0.323682	0.324673	-1.215312	0.323682	0.324673
Y111R	-2.784406	0.671725	0.476246	-2.784406	0.671725	0.476246
Y111S	-1.91984	0.6458	0.284638	-1.91984	0.6458	0.284638
Y111T	-0.253557	0.288305	0.154737	-0.253557	0.288305	0.154737
Y111V	0.286936	0.176612	0.040094	0.286936	0.176612	0.040094
Y111W	-0.373203	0.045193	0.120722	-0.373203	0.045193	0.120722
Y111*	Null	nan	nan	Null	nan	nan
L112A	-0.323968	0.16219	0.100749	-0.323968	0.16219	0.100749
L112C	0.142108	0.225263	0.091421	0.142108	0.225263	0.091421

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
L112D	Null	nan	nan	Null	nan	nan
L112E	-2.536713	0.171563	0.385586	-2.536713	0.171563	0.385586
L112F	0.28835	0.37912	0.037956	0.28835	0.37912	0.037956
L112G	-1.838527	0.767749	0.267522	-1.838527	0.767749	0.267522
L112H	0.095997	0.070339	0.067035	0.095997	0.070339	0.067035
L112I	0.11139	0.41365	0.057286	0.11139	0.41365	0.057286
L112K	-0.554019	0.39929	0.195107	-0.554019	0.39929	0.195107
L112M	0.067801	0.344593	0.043589	0.067801	0.344593	0.043589
L112N	-1.161534	0.29774	0.233234	-1.161534	0.29774	0.233234
L112P	ND	ND	ND	ND	ND	ND
L112Q	-0.349849	0.030868	0.149207	-0.349849	0.030868	0.149207
L112R	-1.28957	0.68985	0.209413	-1.28957	0.68985	0.209413
L112S	ND	ND	ND	ND	ND	ND
L112T	-0.129759	0.069258	0.055823	-0.129759	0.069258	0.055823
L112V	0.032628	0.122179	0.033588	0.032628	0.122179	0.033588
L112W	-0.07923	0.021102	0.081315	-0.07923	0.021102	0.081315
L112Y	-0.009895	0.109849	0.129582	-0.009895	0.109849	0.129582
L112*	Null	nan	nan	Null	nan	nan
T113A	-0.090473	0.20696	0.146781	-0.090473	0.20696	0.146781
T113C	-0.098167	0.197998	0.08927	-0.098167	0.197998	0.08927
T113D	ND	ND	ND	ND	ND	ND
T113E	Null	nan	nan	Null	nan	nan
T113F	ND	ND	ND	ND	ND	ND
T113G	-2.274265	0.472064	0.21224	-2.274265	0.472064	0.21224
T113H	Null	nan	nan	Null	nan	nan
T113I	Null	nan	nan	Null	nan	nan
T113K	Null	nan	nan	Null	nan	nan
T113L	-2.396973	1.020055	0.529554	-2.396973	1.020055	0.529554
T113M	Null	nan	nan	Null	nan	nan
T113N	0.091266	0.081785	0.079169	0.091266	0.081785	0.079169
T113P	-1.910519	0.87669	0.451941	-1.910519	0.87669	0.451941
T113Q	Null	nan	nan	Null	nan	nan
T113R	Null	nan	nan	Null	nan	nan
T113S	-0.178177	0.284401	0.067809	-0.178177	0.284401	0.067809
T113V	0.660791	0.379726	0.056385	0.660791	0.379726	0.056385
T113W	Null	nan	nan	Null	nan	nan
T113Y	Null	nan	nan	Null	nan	nan
T113*	Null	nan	nan	Null	nan	nan
H114A	0.752878	0.158397	0.073672	0.752878	0.158397	0.073672
H114C	0.553881	0.104167	0.091485	0.553881	0.104167	0.091485
H114D	0.214903	0.058703	0.075254	0.214903	0.058703	0.075254
H114E	0.682791	0.184367	0.052238	0.682791	0.184367	0.052238
H114F	0.571446	0.033327	0.055357	0.571446	0.033327	0.055357
H114G	0.552782	0.051314	0.029384	0.552782	0.051314	0.029384
H114I	0.542188	0.11186	0.068073	0.542188	0.11186	0.068073
H114K	0.876897	0.069709	0.061486	0.876897	0.069709	0.061486
H114L	0.697249	0.055962	0.066379	0.697249	0.055962	0.066379
H114M	0.625473	0.104464	0.053152	0.625473	0.104464	0.053152
H114N	0.801632	0.284862	0.083471	0.801632	0.284862	0.083471
H114P	ND	ND	ND	ND	ND	ND

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
H114Q	0.767374	0.155916	0.047074	0.767374	0.155916	0.047074
H114R	0.757887	0.103011	0.072031	0.757887	0.103011	0.072031
H114S	0.64187	0.249187	0.070219	0.64187	0.249187	0.070219
H114T	0.574653	0.188616	0.070721	0.574653	0.188616	0.070721
H114V	0.83668	0.077784	0.05694	0.83668	0.077784	0.05694
H114W	0.267853	0.196994	0.055928	0.267853	0.196994	0.055928
H114Y	0.261282	0.192362	0.074329	0.261282	0.192362	0.074329
H114*	Null	nan	nan	Null	nan	nan
I115A	0.200245	0.192498	0.074414	0.200245	0.192498	0.074414
I115C	0.070533	0.389811	0.107156	0.070533	0.389811	0.107156
I115D	Null	nan	nan	Null	nan	nan
I115E	-0.651869	0.340606	0.402131	-0.651869	0.340606	0.402131
I115F	0.106946	0.16758	0.083245	0.106946	0.16758	0.083245
I115G	-1.109682	0.378897	0.166184	-1.109682	0.378897	0.166184
I115H	-0.112486	0.083386	0.085478	-0.112486	0.083386	0.085478
I115K	ND	ND	ND	ND	ND	ND
I115L	ND	ND	ND	ND	ND	ND
I115M	0.028839	0.165055	0.076826	0.028839	0.165055	0.076826
I115N	0.029993	0.326048	0.112705	0.029993	0.326048	0.112705
I115P	ND	ND	ND	ND	ND	ND
I115Q	-0.498872	0.116662	0.127	-0.498872	0.116662	0.127
I115R	-2.156712	0.544961	0.272145	-2.156712	0.544961	0.272145
I115S	0.512317	0.042536	0.068688	0.512317	0.042536	0.068688
I115T	0.197815	0.132237	0.041898	0.197815	0.132237	0.041898
I115V	-0.163126	0.271746	0.050934	-0.163126	0.271746	0.050934
I115W	-0.511826	0.063861	0.096445	-0.511826	0.063861	0.096445
I115Y	-0.399368	0.068317	0.107746	-0.399368	0.068317	0.107746
I115*	Null	nan	nan	Null	nan	nan
D116A	0.484183	0.132607	0.119938	0.484183	0.132607	0.119938
D116C	0.778708	0.278309	0.061746	0.778708	0.278309	0.061746
D116E	0.464419	0.619057	0.070685	0.464419	0.619057	0.070685
D116F	0.758472	0.050699	0.06964	0.758472	0.050699	0.06964
D116G	0.741586	0.133205	0.058693	0.741586	0.133205	0.058693
D116H	1.108138	0.357593	0.089995	1.108138	0.357593	0.089995
D116I	0.987545	0.424015	0.106917	0.987545	0.424015	0.106917
D116K	0.7125	0.054599	0.078543	0.7125	0.054599	0.078543
D116L	0.873884	0.123092	0.049582	0.873884	0.123092	0.049582
D116M	1.071348	0.068511	0.072773	1.071348	0.068511	0.072773
D116N	0.632694	0.66503	0.104908	0.632694	0.66503	0.104908
D116P	-0.103387	0.371885	0.052459	-0.103387	0.371885	0.052459
D116Q	0.614454	0.123061	0.041781	0.614454	0.123061	0.041781
D116R	0.850926	0.119946	0.092166	0.850926	0.119946	0.092166
D116S	0.487143	0.094065	0.062036	0.487143	0.094065	0.062036
D116T	0.609079	0.074769	0.056969	0.609079	0.074769	0.056969
D116V	0.738585	0.272631	0.075276	0.738585	0.272631	0.075276
D116W	0.694421	0.052927	0.060054	0.694421	0.052927	0.060054
D116Y	1.076413	0.111828	0.074516	1.076413	0.111828	0.074516
D116*	Null	nan	nan	Null	nan	nan
A117C	ND	ND	ND	ND	ND	ND
A117D	0.066634	0.398387	0.164653	0.066634	0.398387	0.164653

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
A117E	-0.009263	0.138502	0.022282	-0.009263	0.138502	0.022282
A117F	-1.991846	0.548117	0.241723	-1.991846	0.548117	0.241723
A117G	-0.157207	0.181988	0.034726	-0.157207	0.181988	0.034726
A117H	-0.481996	0.098277	0.095459	-0.481996	0.098277	0.095459
A117I	0.156843	0.442218	0.12599	0.156843	0.442218	0.12599
A117K	-0.622607	0.448448	0.075942	-0.622607	0.448448	0.075942
A117L	-1.237202	0.294691	0.075117	-1.237202	0.294691	0.075117
A117M	-0.205113	0.56414	0.087818	-0.205113	0.56414	0.087818
A117N	-0.425998	0.151795	0.059631	-0.425998	0.151795	0.059631
A117P	-1.102744	0.398709	0.131385	-1.102744	0.398709	0.131385
A117Q	0.15911	0.398464	0.190124	0.15911	0.398464	0.190124
A117R	0.014769	0.293656	0.064602	0.014769	0.293656	0.064602
A117S	-0.033138	0.093203	0.073413	-0.033138	0.093203	0.073413
A117T	-0.42704	0.252281	0.116232	-0.42704	0.252281	0.116232
A117V	0.068142	0.259483	0.054195	0.068142	0.259483	0.054195
A117W	-1.323466	0.271351	0.242801	-1.323466	0.271351	0.242801
A117Y	-1.737945	0.381922	0.19351	-1.737945	0.381922	0.19351
A117*	Null	nan	nan	Null	nan	nan
E118A	ND	ND	ND	ND	ND	ND
E118C	0.339288	0.216668	0.148547	0.339288	0.216668	0.148547
E118D	-0.144873	0.256828	0.138045	-0.144873	0.256828	0.138045
E118F	0.452061	0.709867	0.073688	0.452061	0.709867	0.073688
E118G	0.503075	0.030307	0.070956	0.503075	0.030307	0.070956
E118H	0.694741	0.11437	0.132591	0.694741	0.11437	0.132591
E118I	0.288126	0.203987	0.046599	0.288126	0.203987	0.046599
E118K	0.597721	0.066054	0.063651	0.597721	0.066054	0.063651
E118L	0.386037	0.290305	0.063255	0.386037	0.290305	0.063255
E118M	0.335531	0.265839	0.058095	0.335531	0.265839	0.058095
E118N	0.473976	0.187923	0.082355	0.473976	0.187923	0.082355
E118P	0.479923	0.039722	0.063797	0.479923	0.039722	0.063797
E118Q	0.511814	0.131611	0.161386	0.511814	0.131611	0.161386
E118R	0.640651	0.107622	0.134152	0.640651	0.107622	0.134152
E118S	0.552822	0.215606	0.078169	0.552822	0.215606	0.078169
E118T	0.591392	0.052829	0.084201	0.591392	0.052829	0.084201
E118V	0.355807	0.162427	0.031614	0.355807	0.162427	0.031614
E118W	0.096541	0.049531	0.10839	0.096541	0.049531	0.10839
E118Y	0.233005	0.14214	0.216652	0.233005	0.14214	0.216652
E118*	Null	nan	nan	Null	nan	nan
V119A	-0.378568	0.369914	0.154189	-0.378568	0.369914	0.154189
V119C	0.539217	0.004422	0.220067	0.539217	0.004422	0.220067
V119D	0.13997	0.137259	0.1025	0.13997	0.137259	0.1025
V119E	0.661759	0.161765	0.05064	0.661759	0.161765	0.05064
V119F	-2.102305	0.51005	0.203615	-2.102305	0.51005	0.203615
V119G	-0.506404	0.28559	0.081084	-0.506404	0.28559	0.081084
V119H	-0.586667	0.305036	0.116047	-0.586667	0.305036	0.116047
V119I	0.00503	0.075527	0.062304	0.00503	0.075527	0.062304
V119K	-0.476245	0.550583	0.059938	-0.476245	0.550583	0.059938
V119L	0.012207	0.378572	0.052763	0.012207	0.378572	0.052763
V119M	-0.015788	0.411363	0.093588	-0.015788	0.411363	0.093588
V119N	0.697375	0.205608	0.046877	0.697375	0.205608	0.046877

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
V119P	0.003158	0.350924	0.04562	0.003158	0.350924	0.04562
V119Q	0.586285	0.330883	0.067581	0.586285	0.330883	0.067581
V119R	0.633277	0.063602	0.059767	0.633277	0.063602	0.059767
V119S	0.267161	0.163644	0.04934	0.267161	0.163644	0.04934
V119T	0.680148	0.140721	0.05016	0.680148	0.140721	0.05016
V119W	-0.400638	0.544861	0.070984	-0.400638	0.544861	0.070984
V119Y	ND	ND	ND	ND	ND	ND
V119*	Null	nan	nan	Null	nan	nan
E120A	-0.507564	0.644991	0.075547	-0.507564	0.644991	0.075547
E120C	0.221023	0.289757	0.14453	0.221023	0.289757	0.14453
E120D	0.698358	0.008322	0.096075	0.698358	0.008322	0.096075
E120F	0.505713	0.405896	0.096115	0.505713	0.405896	0.096115
E120G	0.425309	0.131581	0.024597	0.425309	0.131581	0.024597
E120H	0.323124	0.064293	0.08984	0.323124	0.064293	0.08984
E120I	-0.017603	0.181501	0.062728	-0.017603	0.181501	0.062728
E120K	0.616798	0.031712	0.060466	0.616798	0.031712	0.060466
E120L	0.355917	0.120488	0.05144	0.355917	0.120488	0.05144
E120M	0.36425	0.1311	0.054382	0.36425	0.1311	0.054382
E120N	0.494283	0.183703	0.065475	0.494283	0.183703	0.065475
E120P	-0.959085	0.305179	0.103572	-0.959085	0.305179	0.103572
E120Q	0.37373	0.425624	0.055714	0.37373	0.425624	0.055714
E120R	0.817379	0.151716	0.065521	0.817379	0.151716	0.065521
E120S	-0.110669	0.348652	0.04312	-0.110669	0.348652	0.04312
E120T	0.408812	0.019561	0.042148	0.408812	0.019561	0.042148
E120V	-0.186507	0.147848	0.072262	-0.186507	0.147848	0.072262
E120W	0.378529	0.178502	0.064091	0.378529	0.178502	0.064091
E120Y	0.395101	0.323009	0.068554	0.395101	0.323009	0.068554
E120*	Null	nan	nan	Null	nan	nan
G121A	-1.714597	0.363004	0.108403	-1.714597	0.363004	0.108403
G121C	Null	nan	nan	Null	nan	nan
G121D	Null	nan	nan	Null	nan	nan
G121E	-2.492295	0.121528	0.242003	-2.492295	0.121528	0.242003
G121F	ND	ND	ND	ND	ND	ND
G121H	-1.739515	0.998313	0.205976	-1.739515	0.998313	0.205976
G121I	ND	ND	ND	ND	ND	ND
G121K	-2.620836	0.493072	0.221041	-2.620836	0.493072	0.221041
G121L	-2.714727	0.598492	0.255776	-2.714727	0.598492	0.255776
G121M	ND	ND	ND	ND	ND	ND
G121N	-2.627773	0.257411	0.165555	-2.627773	0.257411	0.165555
G121P	-2.400494	0.121897	0.183965	-2.400494	0.121897	0.183965
G121Q	-2.485075	0.005433	0.2473	-2.485075	0.005433	0.2473
G121R	-1.941241	0.698864	0.262792	-1.941241	0.698864	0.262792
G121S	-1.851268	0.228216	0.201569	-1.851268	0.228216	0.201569
G121T	-2.629815	0.497872	0.308259	-2.629815	0.497872	0.308259
G121V	-1.676576	0.574708	0.106257	-1.676576	0.574708	0.106257
G121W	-2.224785	0.108306	0.14654	-2.224785	0.108306	0.14654
G121Y	-0.800921	0.324184	0.318796	-0.800921	0.324184	0.318796
G121*	-1.50774	0.246258	0.354123	-1.50774	0.246258	0.354123
D122A	-2.23356	0.302381	0.234363	-2.23356	0.302381	0.234363
D122C	-1.891397	0.576145	0.298913	-1.891397	0.576145	0.298913

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
D122E	-1.225294	0.475604	0.11713	-1.225294	0.475604	0.11713
D122F	-1.864995	0.892082	0.392078	-1.864995	0.892082	0.392078
D122G	-1.739832	0.038634	0.13314	-1.739832	0.038634	0.13314
D122H	-2.036106	0.566934	0.523315	-2.036106	0.566934	0.523315
D122I	-2.138532	0.375284	0.40166	-2.138532	0.375284	0.40166
D122K	-1.920541	0.471862	0.303575	-1.920541	0.471862	0.303575
D122L	-2.041232	0.108978	0.212633	-2.041232	0.108978	0.212633
D122M	-1.74919	0.122547	0.220998	-1.74919	0.122547	0.220998
D122N	-2.888468	0.41796	0.378943	-2.888468	0.41796	0.378943
D122P	-1.822968	0.432423	0.338301	-1.822968	0.432423	0.338301
D122Q	-1.460761	0.119641	0.122683	-1.460761	0.119641	0.122683
D122R	-2.087126	0.176821	0.163676	-2.087126	0.176821	0.163676
D122S	-1.114428	0.197385	0.140572	-1.114428	0.197385	0.140572
D122T	-1.127998	0.616359	0.102656	-1.127998	0.616359	0.102656
D122V	-1.808115	0.704714	0.224778	-1.808115	0.704714	0.224778
D122W	-1.624567	0.134646	0.162199	-1.624567	0.134646	0.162199
D122Y	Null	nan	nan	Null	nan	nan
D122*	Null	nan	nan	Null	nan	nan
T123A	0.837164	0.328674	0.052539	0.837164	0.328674	0.052539
T123C	-0.355863	0.242764	0.152797	-0.355863	0.242764	0.152797
T123D	ND	ND	ND	ND	ND	ND
T123E	-0.938396	0.521356	0.109313	-0.938396	0.521356	0.109313
T123F	-2.3739	0.542349	0.099497	-2.3739	0.542349	0.099497
T123G	-1.639723	0.053325	0.127081	-1.639723	0.053325	0.127081
T123H	-0.774483	0.173818	0.176619	-0.774483	0.173818	0.176619
T123I	-0.896894	0.062438	0.410273	-0.896894	0.062438	0.410273
T123K	0.166962	0.100063	0.061639	0.166962	0.100063	0.061639
T123L	-1.484167	0.444303	0.117617	-1.484167	0.444303	0.117617
T123M	-1.093815	0.057164	0.113796	-1.093815	0.057164	0.113796
T123N	Null	nan	nan	Null	nan	nan
T123P	-2.281484	0.514311	0.237234	-2.281484	0.514311	0.237234
T123Q	-1.295795	0.148197	0.116653	-1.295795	0.148197	0.116653
T123R	0.173403	0.173494	0.039736	0.173403	0.173494	0.039736
T123S	0.017754	0.194005	0.043849	0.017754	0.194005	0.043849
T123V	-0.09384	0.519831	0.059942	-0.09384	0.519831	0.059942
T123W	-1.425613	0.034273	0.239551	-1.425613	0.034273	0.239551
T123Y	-1.58159	0.610233	0.297217	-1.58159	0.610233	0.297217
T123*	ND	ND	ND	ND	ND	ND
H124A	0.402668	0.220142	0.083158	0.402668	0.220142	0.083158
H124C	0.680855	0.082589	0.089604	0.680855	0.082589	0.089604
H124D	ND	ND	ND	ND	ND	ND
H124E	-0.396933	0.572885	0.045548	-0.396933	0.572885	0.045548
H124F	-0.985956	0.100863	0.078621	-0.985956	0.100863	0.078621
H124G	0.504463	0.036603	0.058559	0.504463	0.036603	0.058559
H124I	0.139483	0.22862	0.060672	0.139483	0.22862	0.060672
H124K	0.394024	0.293029	0.053359	0.394024	0.293029	0.053359
H124L	-0.286423	0.208812	0.050882	-0.286423	0.208812	0.050882
H124M	-0.020112	0.152227	0.034084	-0.020112	0.152227	0.034084
H124N	0.112803	0.112885	0.111709	0.112803	0.112885	0.111709
H124P	0.308936	0.136466	0.048693	0.308936	0.136466	0.048693

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
H124Q	0.747764	0.384369	0.092521	0.747764	0.384369	0.092521
H124R	0.32763	0.071457	0.067088	0.32763	0.071457	0.067088
H124S	0.348911	0.177877	0.053416	0.348911	0.177877	0.053416
H124T	0.371156	0.206788	0.045221	0.371156	0.206788	0.045221
H124V	0.399637	0.314061	0.074792	0.399637	0.314061	0.074792
H124W	-1.210644	0.363694	0.109909	-1.210644	0.363694	0.109909
H124Y	-0.724331	0.418987	0.090462	-0.724331	0.418987	0.090462
H124*	Null	nan	nan	Null	nan	nan
F125A	-0.461068	0.313063	0.078084	-0.461068	0.313063	0.078084
F125C	0.027918	0.13269	0.098559	0.027918	0.13269	0.098559
F125D	Null	nan	nan	Null	nan	nan
F125E	Null	nan	nan	Null	nan	nan
F125G	0.271108	0.153411	0.056406	0.271108	0.153411	0.056406
F125H	-0.864865	0.152626	0.139962	-0.864865	0.152626	0.139962
F125I	0.111972	0.193341	0.104912	0.111972	0.193341	0.104912
F125K	ND	ND	ND	ND	ND	ND
F125L	0.503303	0.13467	0.043919	0.503303	0.13467	0.043919
F125M	0.570293	0.224163	0.069031	0.570293	0.224163	0.069031
F125N	ND	ND	ND	ND	ND	ND
F125P	-1.107984	0.381277	0.116539	-1.107984	0.381277	0.116539
F125Q	Null	nan	nan	Null	nan	nan
F125R	ND	ND	ND	ND	ND	ND
F125S	-1.308428	0.686643	0.250867	-1.308428	0.686643	0.250867
F125T	-1.700126	0.316583	0.263186	-1.700126	0.316583	0.263186
F125V	-1.691484	0.259629	0.137704	-1.691484	0.259629	0.137704
F125W	0.256745	0.255638	0.106513	0.256745	0.255638	0.106513
F125Y	-1.118942	0.326409	0.303884	-1.118942	0.326409	0.303884
F125*	Null	nan	nan	Null	nan	nan
P126A	0.461057	0.286677	0.087181	0.461057	0.286677	0.087181
P126C	0.568197	0.464207	0.152941	0.568197	0.464207	0.152941
P126D	0.725907	0.190644	0.098181	0.725907	0.190644	0.098181
P126E	0.109279	0.305835	0.065921	0.109279	0.305835	0.065921
P126F	-1.08082	0.280545	0.13886	-1.08082	0.280545	0.13886
P126G	-0.046363	0.360772	0.067417	-0.046363	0.360772	0.067417
P126H	0.349883	0.14499	0.069395	0.349883	0.14499	0.069395
P126I	0.61421	0.419548	0.131238	0.61421	0.419548	0.131238
P126K	0.341087	0.056411	0.073396	0.341087	0.056411	0.073396
P126L	-0.442534	0.595176	0.063799	-0.442534	0.595176	0.063799
P126M	0.513261	0.244759	0.077603	0.513261	0.244759	0.077603
P126N	0.411927	0.075434	0.051615	0.411927	0.075434	0.051615
P126Q	0.826298	0.119027	0.221199	0.826298	0.119027	0.221199
P126R	0.246912	0.097022	0.064942	0.246912	0.097022	0.064942
P126S	0.240777	0.405035	0.085771	0.240777	0.405035	0.085771
P126T	0.545161	0.148313	0.135027	0.545161	0.148313	0.135027
P126V	0.34372	0.195995	0.070182	0.34372	0.195995	0.070182
P126W	0.271414	0.212386	0.047515	0.271414	0.212386	0.047515
P126Y	0.346727	0.109878	0.102003	0.346727	0.109878	0.102003
P126*	Null	nan	nan	Null	nan	nan
D127A	0.366394	0.44423	0.067965	0.366394	0.44423	0.067965
D127C	0.462619	0.339017	0.088511	0.462619	0.339017	0.088511

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
D127E	0.252224	0.180537	0.04616	0.252224	0.180537	0.04616
D127F	0.520381	0.277972	0.072529	0.520381	0.277972	0.072529
D127G	0.395018	0.146003	0.043901	0.395018	0.146003	0.043901
D127H	0.504506	0.501162	0.046659	0.504506	0.501162	0.046659
D127I	0.483241	0.102384	0.055258	0.483241	0.102384	0.055258
D127K	0.558819	0.248215	0.052288	0.558819	0.248215	0.052288
D127L	0.423389	0.058337	0.045174	0.423389	0.058337	0.045174
D127M	0.35952	0.104443	0.065381	0.35952	0.104443	0.065381
D127N	0.211358	0.101389	0.052258	0.211358	0.101389	0.052258
D127P	0.345836	0.162953	0.035847	0.345836	0.162953	0.035847
D127Q	0.398391	0.384421	0.099599	0.398391	0.384421	0.099599
D127R	0.668671	0.184194	0.057997	0.668671	0.184194	0.057997
D127S	0.39748	0.443363	0.044648	0.39748	0.443363	0.044648
D127T	0.428696	0.225427	0.058699	0.428696	0.225427	0.058699
D127V	0.388104	0.138402	0.044364	0.388104	0.138402	0.044364
D127W	0.325492	0.285558	0.113091	0.325492	0.285558	0.113091
D127Y	0.664905	0.108542	0.081083	0.664905	0.108542	0.081083
D127*	Null	nan	nan	Null	nan	nan
Y128A	-0.730421	0.415532	0.069828	-0.730421	0.415532	0.069828
Y128C	Null	nan	nan	Null	nan	nan
Y128D	-0.505498	0.188079	0.129649	-0.505498	0.188079	0.129649
Y128E	-0.21808	0.308304	0.070778	-0.21808	0.308304	0.070778
Y128F	-0.635038	0.111324	0.1913	-0.635038	0.111324	0.1913
Y128G	-0.203549	0.031361	0.056306	-0.203549	0.031361	0.056306
Y128H	ND	ND	ND	ND	ND	ND
Y128I	0.168379	0.2793	0.11771	0.168379	0.2793	0.11771
Y128K	0.391635	0.054884	0.045322	0.391635	0.054884	0.045322
Y128L	-0.138107	0.050706	0.025107	-0.138107	0.050706	0.025107
Y128M	0.098751	0.262187	0.076538	0.098751	0.262187	0.076538
Y128N	-0.397743	0.190325	0.06983	-0.397743	0.190325	0.06983
Y128P	-0.06975	0.105303	0.057372	-0.06975	0.105303	0.057372
Y128Q	-0.581948	0.605409	0.120765	-0.581948	0.605409	0.120765
Y128R	0.22353	0.281927	0.043549	0.22353	0.281927	0.043549
Y128S	-0.073854	0.108435	0.052149	-0.073854	0.108435	0.052149
Y128T	-0.143411	0.16791	0.043075	-0.143411	0.16791	0.043075
Y128V	-0.182221	0.389797	0.060515	-0.182221	0.389797	0.060515
Y128W	0.515006	0.245654	0.055469	0.515006	0.245654	0.055469
Y128*	Null	nan	nan	Null	nan	nan
E129A	0.144107	0.223966	0.066572	0.144107	0.223966	0.066572
E129C	0.168662	0.440757	0.061538	0.168662	0.440757	0.061538
E129D	ND	ND	ND	ND	ND	ND
E129F	0.346372	0.185691	0.04942	0.346372	0.185691	0.04942
E129G	0.404553	0.456395	0.092621	0.404553	0.456395	0.092621
E129H	0.488774	0.05106	0.064971	0.488774	0.05106	0.064971
E129I	0.348527	0.212759	0.139554	0.348527	0.212759	0.139554
E129K	0.4242	0.396996	0.112964	0.4242	0.396996	0.112964
E129L	0.100326	0.162062	0.04066	0.100326	0.162062	0.04066
E129M	0.337935	0.030222	0.08298	0.337935	0.030222	0.08298
E129N	0.078711	0.220837	0.080731	0.078711	0.220837	0.080731
E129P	0.302174	0.034597	0.04163	0.302174	0.034597	0.04163

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
E129Q	0.413498	0.234057	0.096138	0.413498	0.234057	0.096138
E129R	0.242537	0.260332	0.075945	0.242537	0.260332	0.075945
E129S	0.159344	0.164516	0.063942	0.159344	0.164516	0.063942
E129T	0.121251	0.322569	0.056758	0.121251	0.322569	0.056758
E129V	0.195081	0.369854	0.074999	0.195081	0.369854	0.074999
E129W	0.412673	0.072438	0.058061	0.412673	0.072438	0.058061
E129Y	ND	ND	ND	ND	ND	ND
E129*	Null	nan	nan	Null	nan	nan
P130A	0.320942	0.152908	0.044904	0.320942	0.152908	0.044904
P130C	0.608971	0.355322	0.156277	0.608971	0.355322	0.156277
P130D	0.284295	0.165642	0.06312	0.284295	0.165642	0.06312
P130E	0.354869	0.135799	0.029157	0.354869	0.135799	0.029157
P130F	0.257956	0.090964	0.069034	0.257956	0.090964	0.069034
P130G	0.360202	0.037876	0.033687	0.360202	0.037876	0.033687
P130H	0.506519	0.031562	0.094148	0.506519	0.031562	0.094148
P130I	0.408819	0.145662	0.065455	0.408819	0.145662	0.065455
P130K	0.346637	0.057583	0.049445	0.346637	0.057583	0.049445
P130L	0.461062	0.068861	0.062856	0.461062	0.068861	0.062856
P130M	0.420749	0.144002	0.032673	0.420749	0.144002	0.032673
P130N	0.227777	0.152473	0.042308	0.227777	0.152473	0.042308
P130Q	0.418277	0.033393	0.066944	0.418277	0.033393	0.066944
P130R	0.511217	0.028279	0.037866	0.511217	0.028279	0.037866
P130S	0.523357	0.120695	0.053705	0.523357	0.120695	0.053705
P130T	0.416213	0.129032	0.046982	0.416213	0.129032	0.046982
P130V	0.410545	0.031289	0.032273	0.410545	0.031289	0.032273
P130W	0.434155	0.064736	0.046488	0.434155	0.064736	0.046488
P130Y	-0.079493	0.223553	0.10479	-0.079493	0.223553	0.10479
P130*	ND	ND	ND	ND	ND	ND
D131A	-0.619088	0.089103	0.106689	-0.619088	0.089103	0.106689
D131C	Null	nan	nan	Null	nan	nan
D131E	-0.745685	0.099545	0.133482	-0.745685	0.099545	0.133482
D131F	0.142934	0.185108	0.333436	0.142934	0.185108	0.333436
D131G	-0.6837	0.184314	0.052827	-0.6837	0.184314	0.052827
D131H	ND	ND	ND	ND	ND	ND
D131I	-0.25763	0.072769	0.103933	-0.25763	0.072769	0.103933
D131K	-0.473316	0.381337	0.055471	-0.473316	0.381337	0.055471
D131L	-0.121112	0.433247	0.065842	-0.121112	0.433247	0.065842
D131M	-0.666639	0.077403	0.101693	-0.666639	0.077403	0.101693
D131N	-0.284212	0.107749	0.062496	-0.284212	0.107749	0.062496
D131P	0.057002	0.10902	0.049796	0.057002	0.10902	0.049796
D131Q	-0.731785	0.088571	0.154815	-0.731785	0.088571	0.154815
D131R	-0.653402	0.109997	0.091518	-0.653402	0.109997	0.091518
D131S	-0.610667	0.155466	0.101197	-0.610667	0.155466	0.101197
D131T	-0.418531	0.555471	0.079659	-0.418531	0.555471	0.079659
D131V	-0.446957	0.234024	0.09834	-0.446957	0.234024	0.09834
D131W	0.072284	0.297987	0.093884	0.072284	0.297987	0.093884
D131Y	0.270566	0.349301	0.144328	0.270566	0.349301	0.144328
D131*	Null	nan	nan	Null	nan	nan
D132A	0.041891	0.211039	0.037777	0.041891	0.211039	0.037777
D132C	0.026612	0.04765	0.069746	0.026612	0.04765	0.069746

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
D132E	-0.359178	0.032964	0.061398	-0.359178	0.032964	0.061398
D132F	0.249367	0.081197	0.087351	0.249367	0.081197	0.087351
D132G	0.147133	0.154121	0.036942	0.147133	0.154121	0.036942
D132H	-0.01157	0.159331	0.078689	-0.01157	0.159331	0.078689
D132I	0.506605	0.177348	0.062793	0.506605	0.177348	0.062793
D132K	0.339027	0.105332	0.050946	0.339027	0.105332	0.050946
D132L	0.236783	0.038844	0.033508	0.236783	0.038844	0.033508
D132M	-0.138821	0.246071	0.062082	-0.138821	0.246071	0.062082
D132N	0.334271	0.227747	0.1055	0.334271	0.227747	0.1055
D132P	0.31596	0.115898	0.045307	0.31596	0.115898	0.045307
D132Q	0.051588	0.2471	0.074952	0.051588	0.2471	0.074952
D132R	0.423338	0.110867	0.021233	0.423338	0.110867	0.021233
D132S	0.31292	0.250575	0.059095	0.31292	0.250575	0.059095
D132T	0.453163	0.085675	0.039974	0.453163	0.085675	0.039974
D132V	0.191831	0.197639	0.052036	0.191831	0.197639	0.052036
D132W	0.116022	0.196055	0.062799	0.116022	0.196055	0.062799
D132Y	0.494586	0.274939	0.135208	0.494586	0.274939	0.135208
D132*	Null	nan	nan	Null	nan	nan
W133A	ND	ND	ND	ND	ND	ND
W133C	ND	ND	ND	ND	ND	ND
W133D	Null	nan	nan	Null	nan	nan
W133E	ND	ND	ND	ND	ND	ND
W133F	0.310643	0.221436	0.048966	0.310643	0.221436	0.048966
W133G	ND	ND	ND	ND	ND	ND
W133H	-0.23198	0.29072	0.129711	-0.23198	0.29072	0.129711
W133I	-0.51326	0.344969	0.220438	-0.51326	0.344969	0.220438
W133K	-1.810236	0.240585	0.309322	-1.810236	0.240585	0.309322
W133L	0.348752	0.09128	0.054917	0.348752	0.09128	0.054917
W133M	0.153889	0.119954	0.046372	0.153889	0.119954	0.046372
W133N	-0.181241	0.5766	0.336014	-0.181241	0.5766	0.336014
W133P	-1.243831	0.365274	0.23208	-1.243831	0.365274	0.23208
W133Q	ND	ND	ND	ND	ND	ND
W133R	Null	nan	nan	Null	nan	nan
W133S	ND	ND	ND	ND	ND	ND
W133T	-2.186069	0.720911	0.134047	-2.186069	0.720911	0.134047
W133V	-1.040612	0.21683	0.241703	-1.040612	0.21683	0.241703
W133Y	0.262824	0.003009	0.057106	0.262824	0.003009	0.057106
W133*	Null	nan	nan	Null	nan	nan
E134A	0.026751	0.022437	0.046948	0.026751	0.022437	0.046948
E134C	0.036286	0.334771	0.103109	0.036286	0.334771	0.103109
E134D	-0.938782	0.260976	0.25208	-0.938782	0.260976	0.25208
E134F	0.165284	0.187634	0.038138	0.165284	0.187634	0.038138
E134G	0.312458	0.141188	0.032509	0.312458	0.141188	0.032509
E134H	0.158011	0.205728	0.044947	0.158011	0.205728	0.044947
E134I	0.135736	0.409518	0.11804	0.135736	0.409518	0.11804
E134K	0.237286	0.31387	0.055416	0.237286	0.31387	0.055416
E134L	0.231322	0.136897	0.048074	0.231322	0.136897	0.048074
E134M	0.159909	0.445552	0.079471	0.159909	0.445552	0.079471
E134N	0.050013	0.122892	0.041814	0.050013	0.122892	0.041814
E134P	0.315138	0.05679	0.043638	0.315138	0.05679	0.043638

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
E134Q	0.225213	0.366472	0.066273	0.225213	0.366472	0.066273
E134R	0.193508	0.047536	0.04629	0.193508	0.047536	0.04629
E134S	-0.035811	0.070271	0.050107	-0.035811	0.070271	0.050107
E134T	0.045623	0.15607	0.051282	0.045623	0.15607	0.051282
E134V	0.238819	0.375491	0.039096	0.238819	0.375491	0.039096
E134W	0.021687	0.156518	0.066749	0.021687	0.156518	0.066749
E134Y	0.11015	0.277426	0.091992	0.11015	0.277426	0.091992
E134*	Null	nan	nan	Null	nan	nan
S135A	-0.518261	0.241157	0.094036	-0.518261	0.241157	0.094036
S135C	-0.152476	0.090216	0.130466	-0.152476	0.090216	0.130466
S135D	ND	ND	ND	ND	ND	ND
S135E	0.306421	0.135421	0.078771	0.306421	0.135421	0.078771
S135F	0.432139	0.628791	0.109309	0.432139	0.628791	0.109309
S135G	0.167931	0.183019	0.04808	0.167931	0.183019	0.04808
S135H	0.017266	0.087052	0.090606	0.017266	0.087052	0.090606
S135I	0.605633	0.060965	0.111673	0.605633	0.060965	0.111673
S135K	0.647132	0.042574	0.058281	0.647132	0.042574	0.058281
S135L	0.468314	0.465989	0.055929	0.468314	0.465989	0.055929
S135M	0.047777	0.317608	0.043686	0.047777	0.317608	0.043686
S135N	0.763744	0.04043	0.0807	0.763744	0.04043	0.0807
S135P	0.297891	0.270098	0.071404	0.297891	0.270098	0.071404
S135Q	0.611214	0.187083	0.047107	0.611214	0.187083	0.047107
S135R	0.815634	0.058769	0.050292	0.815634	0.058769	0.050292
S135T	0.551627	0.16896	0.076675	0.551627	0.16896	0.076675
S135V	-0.132661	0.058905	0.065046	-0.132661	0.058905	0.065046
S135W	-0.157965	0.304565	0.126419	-0.157965	0.304565	0.126419
S135Y	0.077474	0.04719	0.066945	0.077474	0.04719	0.066945
S135*	Null	nan	nan	Null	nan	nan
V136A	0.269021	0.515666	0.086594	0.269021	0.515666	0.086594
V136C	ND	ND	ND	ND	ND	ND
V136D	0.150272	0.45594	0.07346	0.150272	0.45594	0.07346
V136E	-0.021389	0.231195	0.103968	-0.021389	0.231195	0.103968
V136F	0.182274	0.289177	0.069843	0.182274	0.289177	0.069843
V136G	0.18539	0.199161	0.035431	0.18539	0.199161	0.035431
V136H	-0.063373	0.025336	0.109763	-0.063373	0.025336	0.109763
V136I	0.1407	0.100641	0.147697	0.1407	0.100641	0.147697
V136K	0.111079	0.126789	0.046505	0.111079	0.126789	0.046505
V136L	0.030018	0.121955	0.047798	0.030018	0.121955	0.047798
V136M	0.150013	0.257894	0.125692	0.150013	0.257894	0.125692
V136N	0.171946	0.560831	0.053966	0.171946	0.560831	0.053966
V136P	0.128583	0.375388	0.05564	0.128583	0.375388	0.05564
V136Q	0.03553	0.295014	0.126197	0.03553	0.295014	0.126197
V136R	0.160228	0.096446	0.042304	0.160228	0.096446	0.042304
V136S	0.068232	0.219228	0.053223	0.068232	0.219228	0.053223
V136T	0.028512	0.119661	0.057231	0.028512	0.119661	0.057231
V136W	-0.008375	0.224602	0.094649	-0.008375	0.224602	0.094649
V136Y	0.034021	0.584109	0.145549	0.034021	0.584109	0.145549
V136*	Null	nan	nan	Null	nan	nan
F137A	0.183931	0.161918	0.051576	0.183931	0.161918	0.051576
F137C	0.732775	0.278232	0.069596	0.732775	0.278232	0.069596

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
F137D	0.143587	0.137763	0.088527	0.143587	0.137763	0.088527
F137E	0.198879	0.337294	0.039509	0.198879	0.337294	0.039509
F137G	-0.068313	0.091251	0.067827	-0.068313	0.091251	0.067827
F137H	0.079297	0.045836	0.068966	0.079297	0.045836	0.068966
F137I	0.89629	0.236178	0.12408	0.89629	0.236178	0.12408
F137K	0.159171	0.149351	0.051278	0.159171	0.149351	0.051278
F137L	0.324463	0.140824	0.052027	0.324463	0.140824	0.052027
F137M	0.254593	0.543288	0.054506	0.254593	0.543288	0.054506
F137N	0.375245	0.054299	0.052083	0.375245	0.054299	0.052083
F137P	-0.560717	0.209955	0.18247	-0.560717	0.209955	0.18247
F137Q	0.160713	0.358697	0.077254	0.160713	0.358697	0.077254
F137R	0.27833	0.091854	0.062242	0.27833	0.091854	0.062242
F137S	0.466858	0.255555	0.056755	0.466858	0.255555	0.056755
F137T	0.310917	0.017998	0.060028	0.310917	0.017998	0.060028
F137V	0.410176	0.270472	0.055184	0.410176	0.270472	0.055184
F137W	-0.155526	0.170369	0.057702	-0.155526	0.170369	0.057702
F137Y	-0.685483	0.163156	0.282939	-0.685483	0.163156	0.282939
F137*	Null	nan	nan	Null	nan	nan
S138A	-0.274474	0.239942	0.09829	-0.274474	0.239942	0.09829
S138C	ND	ND	ND	ND	ND	ND
S138D	0.048502	0.177799	0.114726	0.048502	0.177799	0.114726
S138E	0.484972	0.134953	0.07593	0.484972	0.134953	0.07593
S138F	-0.549933	0.123744	0.098747	-0.549933	0.123744	0.098747
S138G	-0.33486	0.044908	0.050573	-0.33486	0.044908	0.050573
S138H	ND	ND	ND	ND	ND	ND
S138I	0.527549	0.55896	0.131547	0.527549	0.55896	0.131547
S138K	0.295705	0.11862	0.044601	0.295705	0.11862	0.044601
S138L	0.312166	0.266738	0.035477	0.312166	0.266738	0.035477
S138M	0.039816	0.476034	0.050337	0.039816	0.476034	0.050337
S138N	Null	nan	nan	Null	nan	nan
S138P	-0.17995	0.121609	0.087195	-0.17995	0.121609	0.087195
S138Q	0.564012	0.038074	0.062443	0.564012	0.038074	0.062443
S138R	0.485231	0.115483	0.035504	0.485231	0.115483	0.035504
S138T	0.241575	0.587474	0.060102	0.241575	0.587474	0.060102
S138V	0.093895	0.051815	0.031923	0.093895	0.051815	0.031923
S138W	-0.131131	0.153925	0.053558	-0.131131	0.153925	0.053558
S138Y	-0.52196	0.152442	0.104062	-0.52196	0.152442	0.104062
S138*	Null	nan	nan	Null	nan	nan
E139A	-0.144449	0.192699	0.066357	-0.144449	0.192699	0.066357
E139C	-0.198498	0.084966	0.065107	-0.198498	0.084966	0.065107
E139D	-0.163093	0.024664	0.1281	-0.163093	0.024664	0.1281
E139F	-0.019624	0.116745	0.106699	-0.019624	0.116745	0.106699
E139G	-0.127806	0.101102	0.062403	-0.127806	0.101102	0.062403
E139H	0.006733	0.180216	0.047518	0.006733	0.180216	0.047518
E139I	-0.038586	0.057345	0.08326	-0.038586	0.057345	0.08326
E139K	-0.017878	0.246919	0.054826	-0.017878	0.246919	0.054826
E139L	0.040732	0.188154	0.034082	0.040732	0.188154	0.034082
E139M	-0.05921	0.200411	0.044389	-0.05921	0.200411	0.044389
E139N	-0.072791	0.040526	0.065223	-0.072791	0.040526	0.065223
E139P	-0.402935	0.230928	0.127018	-0.402935	0.230928	0.127018

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
E139Q	-0.09311	0.199921	0.072478	-0.09311	0.199921	0.072478
E139R	-0.08808	0.149129	0.0372	-0.08808	0.149129	0.0372
E139S	-0.10486	0.123518	0.043862	-0.10486	0.123518	0.043862
E139T	0.095007	0.295254	0.04944	0.095007	0.295254	0.04944
E139V	-0.130126	0.026589	0.045133	-0.130126	0.026589	0.045133
E139W	-0.204022	0.259889	0.047559	-0.204022	0.259889	0.047559
E139Y	-0.200806	0.309018	0.08371	-0.200806	0.309018	0.08371
E139*	Null	nan	nan	Null	nan	nan
F140A	0.412717	0.391767	0.073651	0.412717	0.391767	0.073651
F140C	ND	ND	ND	ND	ND	ND
F140D	0.426685	0.071439	0.076449	0.426685	0.071439	0.076449
F140E	0.481784	0.471096	0.101941	0.481784	0.471096	0.101941
F140G	0.459605	0.120341	0.059866	0.459605	0.120341	0.059866
F140H	-0.053215	0.254537	0.083172	-0.053215	0.254537	0.083172
F140I	0.575706	0.256151	0.072634	0.575706	0.256151	0.072634
F140K	0.410686	0.58974	0.091691	0.410686	0.58974	0.091691
F140L	0.593901	0.149878	0.062862	0.593901	0.149878	0.062862
F140M	0.396933	0.191155	0.07866	0.396933	0.191155	0.07866
F140N	0.553848	0.204373	0.06699	0.553848	0.204373	0.06699
F140P	0.466947	0.102189	0.046236	0.466947	0.102189	0.046236
F140Q	0.319445	0.155516	0.062722	0.319445	0.155516	0.062722
F140R	0.410888	0.110342	0.075989	0.410888	0.110342	0.075989
F140S	0.472018	0.243734	0.086529	0.472018	0.243734	0.086529
F140T	0.587246	0.091628	0.060958	0.587246	0.091628	0.060958
F140V	0.55445	0.518933	0.053398	0.55445	0.518933	0.053398
F140W	0.067125	0.001235	0.122337	0.067125	0.001235	0.122337
F140Y	ND	ND	ND	ND	ND	ND
F140*	Null	nan	nan	Null	nan	nan
H141A	-0.205733	0.082756	0.052952	-0.205733	0.082756	0.052952
H141C	-0.232552	0.237338	0.15369	-0.232552	0.237338	0.15369
H141D	-0.98817	0.760019	0.228608	-0.98817	0.760019	0.228608
H141E	-0.167685	0.14802	0.044739	-0.167685	0.14802	0.044739
H141F	-0.102172	0.389089	0.044264	-0.102172	0.389089	0.044264
H141G	0.028026	0.053099	0.024324	0.028026	0.053099	0.024324
H141I	0.3366	0.040214	0.077104	0.3366	0.040214	0.077104
H141K	-0.244242	0.171251	0.051135	-0.244242	0.171251	0.051135
H141L	-0.15948	0.241158	0.051616	-0.15948	0.241158	0.051616
H141M	-0.273142	0.184138	0.061461	-0.273142	0.184138	0.061461
H141N	-0.22364	0.234855	0.072171	-0.22364	0.234855	0.072171
H141P	-0.789793	0.124862	0.08023	-0.789793	0.124862	0.08023
H141Q	-0.274218	0.534541	0.059207	-0.274218	0.534541	0.059207
H141R	-0.322051	0.265722	0.061074	-0.322051	0.265722	0.061074
H141S	-0.068722	0.157665	0.038904	-0.068722	0.157665	0.038904
H141T	0.079802	0.071367	0.06278	0.079802	0.071367	0.06278
H141V	0.085666	0.059142	0.048755	0.085666	0.059142	0.048755
H141W	-0.174053	0.348486	0.086427	-0.174053	0.348486	0.086427
H141Y	-0.082553	0.215168	0.163231	-0.082553	0.215168	0.163231
H141*	Null	nan	nan	Null	nan	nan
D142A	0.467817	0.142685	0.038218	0.467817	0.142685	0.038218
D142C	-0.053111	0.131797	0.182387	-0.053111	0.131797	0.182387

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
D142E	0.110562	0.445318	0.038905	0.110562	0.445318	0.038905
D142F	0.267592	0.084337	0.053915	0.267592	0.084337	0.053915
D142G	0.215601	0.09489	0.044395	0.215601	0.09489	0.044395
D142H	0.296699	0.124971	0.051438	0.296699	0.124971	0.051438
D142I	-0.035127	0.344842	0.050297	-0.035127	0.344842	0.050297
D142K	0.25413	0.126665	0.039656	0.25413	0.126665	0.039656
D142L	0.314035	0.041308	0.041411	0.314035	0.041308	0.041411
D142M	0.172685	0.038088	0.053455	0.172685	0.038088	0.053455
D142N	0.082707	0.104618	0.049355	0.082707	0.104618	0.049355
D142P	0.296885	0.251685	0.049195	0.296885	0.251685	0.049195
D142Q	0.0611	0.291865	0.050473	0.0611	0.291865	0.050473
D142R	0.222027	0.020288	0.037319	0.222027	0.020288	0.037319
D142S	0.168009	0.204183	0.040375	0.168009	0.204183	0.040375
D142T	0.082826	0.024972	0.053618	0.082826	0.024972	0.053618
D142V	0.267437	0.068593	0.045559	0.267437	0.068593	0.045559
D142W	0.242873	0.100075	0.041915	0.242873	0.100075	0.041915
D142Y	0.145168	0.09801	0.066264	0.145168	0.09801	0.066264
D142*	-1.266889	0.132738	0.423865	-1.266889	0.132738	0.423865
A143C	ND	ND	ND	ND	ND	ND
A143D	ND	ND	ND	ND	ND	ND
A143E	-0.050607	0.263428	0.07221	-0.050607	0.263428	0.07221
A143F	0.097293	0.18798	0.044281	0.097293	0.18798	0.044281
A143G	-0.116927	0.039907	0.042823	-0.116927	0.039907	0.042823
A143H	0.016363	0.151457	0.05501	0.016363	0.151457	0.05501
A143I	-0.120142	0.185799	0.064359	-0.120142	0.185799	0.064359
A143K	0.090805	0.146233	0.062898	0.090805	0.146233	0.062898
A143L	-0.019067	0.121968	0.041052	-0.019067	0.121968	0.041052
A143M	0.163028	0.166838	0.042906	0.163028	0.166838	0.042906
A143N	-0.135601	0.146265	0.040716	-0.135601	0.146265	0.040716
A143P	-0.402503	0.086003	0.039475	-0.402503	0.086003	0.039475
A143Q	-0.169968	0.071983	0.078273	-0.169968	0.071983	0.078273
A143R	-0.015307	0.084811	0.05195	-0.015307	0.084811	0.05195
A143S	-0.044616	0.042947	0.062736	-0.044616	0.042947	0.062736
A143T	-0.28648	0.319632	0.028678	-0.28648	0.319632	0.028678
A143V	-0.08729	0.134707	0.037104	-0.08729	0.134707	0.037104
A143W	-0.18052	0.159887	0.06129	-0.18052	0.159887	0.06129
A143Y	ND	ND	ND	ND	ND	ND
A143*	Null	nan	nan	Null	nan	nan
D144A	-0.242671	0.030048	0.080492	-0.242671	0.030048	0.080492
D144C	0.307657	0.187752	0.107516	0.307657	0.187752	0.107516
D144E	-0.762415	0.0223	0.227151	-0.762415	0.0223	0.227151
D144F	-0.359709	0.352218	0.109684	-0.359709	0.352218	0.109684
D144G	0.254019	0.123239	0.0536	0.254019	0.123239	0.0536
D144H	-0.124113	0.264102	0.106319	-0.124113	0.264102	0.106319
D144I	-0.112417	0.466297	0.12687	-0.112417	0.466297	0.12687
D144K	-0.527556	0.024286	0.091842	-0.527556	0.024286	0.091842
D144L	-0.326384	0.026218	0.049531	-0.326384	0.026218	0.049531
D144M	-0.323297	0.105408	0.04808	-0.323297	0.105408	0.04808
D144N	-0.201943	0.239119	0.056303	-0.201943	0.239119	0.056303
D144P	ND	ND	ND	ND	ND	ND

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
D144Q	-0.28108	0.212747	0.105917	-0.28108	0.212747	0.105917
D144R	-0.705434	0.265204	0.0648	-0.705434	0.265204	0.0648
D144S	0.059419	0.207369	0.077093	0.059419	0.207369	0.077093
D144T	-0.263549	0.064368	0.071349	-0.263549	0.064368	0.071349
D144V	-0.220122	0.030661	0.034463	-0.220122	0.030661	0.034463
D144W	-0.078082	0.099988	0.09652	-0.078082	0.099988	0.09652
D144Y	-0.699293	0.248685	0.173948	-0.699293	0.248685	0.173948
D144*	Null	nan	nan	Null	nan	nan
A145C	0.030769	0.325972	0.090443	0.030769	0.325972	0.090443
A145D	0.149044	0.116095	0.073253	0.149044	0.116095	0.073253
A145E	-0.016689	0.224944	0.053047	-0.016689	0.224944	0.053047
A145F	0.17212	0.154896	0.128833	0.17212	0.154896	0.128833
A145G	0.145605	0.045413	0.031312	0.145605	0.045413	0.031312
A145H	0.168866	0.195991	0.076047	0.168866	0.195991	0.076047
A145I	0.003684	0.333792	0.095698	0.003684	0.333792	0.095698
A145K	0.000777	0.131311	0.044858	0.000777	0.131311	0.044858
A145L	-0.035231	0.209077	0.03932	-0.035231	0.209077	0.03932
A145M	-0.028177	0.074116	0.049446	-0.028177	0.074116	0.049446
A145N	0.191849	0.106057	0.0442	0.191849	0.106057	0.0442
A145P	-0.237063	0.060179	0.051154	-0.237063	0.060179	0.051154
A145Q	-0.095273	0.306725	0.080749	-0.095273	0.306725	0.080749
A145R	-0.081533	0.165296	0.034647	-0.081533	0.165296	0.034647
A145S	0.024799	0.352717	0.052862	0.024799	0.352717	0.052862
A145T	-0.096519	0.216749	0.093268	-0.096519	0.216749	0.093268
A145V	-0.067227	0.205682	0.07839	-0.067227	0.205682	0.07839
A145W	-0.018651	0.224671	0.046492	-0.018651	0.224671	0.046492
A145Y	-0.063637	0.168916	0.076246	-0.063637	0.168916	0.076246
A145*	ND	ND	ND	ND	ND	ND
Q146A	-0.116415	0.065762	0.055492	-0.116415	0.065762	0.055492
Q146C	0.182309	0.094085	0.070519	0.182309	0.094085	0.070519
Q146D	0.100084	0.068156	0.073632	0.100084	0.068156	0.073632
Q146E	-0.06405	0.376004	0.051626	-0.06405	0.376004	0.051626
Q146F	0.24463	0.184551	0.122462	0.24463	0.184551	0.122462
Q146G	0.054449	0.063055	0.025737	0.054449	0.063055	0.025737
Q146H	-0.002834	0.369378	0.070265	-0.002834	0.369378	0.070265
Q146I	0.051907	0.192737	0.05241	0.051907	0.192737	0.05241
Q146K	-0.16507	0.352449	0.065337	-0.16507	0.352449	0.065337
Q146L	0.003459	0.194503	0.054113	0.003459	0.194503	0.054113
Q146M	-0.020912	0.305874	0.080753	-0.020912	0.305874	0.080753
Q146N	-0.015916	0.352095	0.054678	-0.015916	0.352095	0.054678
Q146P	-0.387225	0.155389	0.054413	-0.387225	0.155389	0.054413
Q146R	-0.293407	0.036378	0.026842	-0.293407	0.036378	0.026842
Q146S	0.008933	0.105087	0.035857	0.008933	0.105087	0.035857
Q146T	-0.072503	0.093064	0.042088	-0.072503	0.093064	0.042088
Q146V	-0.075601	0.147319	0.02063	-0.075601	0.147319	0.02063
Q146W	-0.09878	0.348714	0.0384	-0.09878	0.348714	0.0384
Q146Y	0.070414	0.110818	0.059447	0.070414	0.110818	0.059447
Q146*	Null	nan	nan	Null	nan	nan
N147A	0.362302	0.010331	0.053054	0.362302	0.010331	0.053054
N147C	-0.319619	0.394602	0.111393	-0.319619	0.394602	0.111393

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
N147D	-1.542414	0.479481	0.344026	-1.542414	0.479481	0.344026
N147E	-0.267787	0.254362	0.048427	-0.267787	0.254362	0.048427
N147F	-2.092428	0.300037	0.201843	-2.092428	0.300037	0.201843
N147G	-0.673095	0.193022	0.078198	-0.673095	0.193022	0.078198
N147H	-0.370629	0.005489	0.109465	-0.370629	0.005489	0.109465
N147I	ND	ND	ND	ND	ND	ND
N147K	-1.071151	0.327477	0.083141	-1.071151	0.327477	0.083141
N147L	-1.06595	0.052865	0.071948	-1.06595	0.052865	0.071948
N147M	-0.69564	0.145046	0.10401	-0.69564	0.145046	0.10401
N147P	-1.394062	0.081268	0.130811	-1.394062	0.081268	0.130811
N147Q	-0.194216	0.506887	0.046013	-0.194216	0.506887	0.046013
N147R	-0.76284	0.114056	0.051197	-0.76284	0.114056	0.051197
N147S	0.083135	0.081714	0.041808	0.083135	0.081714	0.041808
N147T	-0.291015	0.320751	0.036491	-0.291015	0.320751	0.036491
N147V	-0.77805	0.291987	0.056693	-0.77805	0.291987	0.056693
N147W	-1.667268	0.243572	0.110756	-1.667268	0.243572	0.110756
N147Y	Null	nan	nan	Null	nan	nan
N147*	ND	ND	ND	ND	ND	ND
S148A	-2.08381	0.2708	0.276575	-2.08381	0.2708	0.276575
S148C	-2.167903	0.992664	0.256353	-2.167903	0.992664	0.256353
S148D	-1.202779	0.680211	0.199255	-1.202779	0.680211	0.199255
S148E	-1.45747	0.206864	0.23987	-1.45747	0.206864	0.23987
S148F	-1.852031	0.15515	0.119062	-1.852031	0.15515	0.119062
S148G	-2.001315	0.149355	0.082205	-2.001315	0.149355	0.082205
S148H	-2.375586	0.147841	0.223728	-2.375586	0.147841	0.223728
S148I	-1.468013	0.631048	0.214776	-1.468013	0.631048	0.214776
S148K	-2.187706	0.413592	0.184086	-2.187706	0.413592	0.184086
S148L	-1.947342	0.068652	0.116437	-1.947342	0.068652	0.116437
S148M	-2.094591	0.207552	0.262767	-2.094591	0.207552	0.262767
S148N	ND	ND	ND	ND	ND	ND
S148P	-2.1691	0.041721	0.198319	-2.1691	0.041721	0.198319
S148Q	-2.32085	0.195565	0.219946	-2.32085	0.195565	0.219946
S148R	-1.909303	0.544045	0.155758	-1.909303	0.544045	0.155758
S148T	0.278095	0.222404	0.052073	0.278095	0.222404	0.052073
S148V	-2.211719	0.562044	0.152712	-2.211719	0.562044	0.152712
S148W	-1.382132	0.482602	0.184278	-1.382132	0.482602	0.184278
S148Y	-2.034147	0.580535	0.199885	-2.034147	0.580535	0.199885
S148*	Null	nan	nan	Null	nan	nan
H149A	-0.835001	0.67795	0.093743	-0.835001	0.67795	0.093743
H149C	-0.388411	0.526455	0.138455	-0.388411	0.526455	0.138455
H149D	ND	ND	ND	ND	ND	ND
H149E	-0.558509	0.364217	0.128547	-0.558509	0.364217	0.128547
H149F	-1.944795	0.220166	0.162736	-1.944795	0.220166	0.162736
H149G	-0.7382	0.249189	0.066063	-0.7382	0.249189	0.066063
H149I	-1.19016	0.122756	0.141504	-1.19016	0.122756	0.141504
H149K	-0.55951	0.130392	0.083982	-0.55951	0.130392	0.083982
H149L	-1.969912	0.372265	0.171764	-1.969912	0.372265	0.171764
H149M	-2.019863	0.362532	0.196439	-2.019863	0.362532	0.196439
H149N	-0.372278	0.498741	0.14823	-0.372278	0.498741	0.14823
H149P	0.419037	0.064269	0.051897	0.419037	0.064269	0.051897

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
H149Q	-1.174762	0.072336	0.141429	-1.174762	0.072336	0.141429
H149R	-0.121052	0.248619	0.063304	-0.121052	0.248619	0.063304
H149S	-1.26645	0.248397	0.110144	-1.26645	0.248397	0.110144
H149T	-0.098602	0.241633	0.061332	-0.098602	0.241633	0.061332
H149V	0.299314	0.192545	0.059161	0.299314	0.192545	0.059161
H149W	-2.1575	0.72935	0.298712	-2.1575	0.72935	0.298712
H149Y	Null	nan	nan	Null	nan	nan
H149*	Null	nan	nan	Null	nan	nan
S150A	-0.025687	0.286279	0.066689	-0.025687	0.286279	0.066689
S150C	-0.252332	0.069897	0.213384	-0.252332	0.069897	0.213384
S150D	0.191367	0.360347	0.094662	0.191367	0.360347	0.094662
S150E	-0.183802	0.232528	0.061361	-0.183802	0.232528	0.061361
S150F	-0.154342	0.21201	0.062185	-0.154342	0.21201	0.062185
S150G	-0.158072	0.110879	0.042099	-0.158072	0.110879	0.042099
S150H	-0.219694	0.253691	0.078269	-0.219694	0.253691	0.078269
S150I	ND	ND	ND	ND	ND	ND
S150K	-0.016071	0.400503	0.048464	-0.016071	0.400503	0.048464
S150L	-0.568229	0.078265	0.05555	-0.568229	0.078265	0.05555
S150M	-0.2787	0.161802	0.111362	-0.2787	0.161802	0.111362
S150N	-0.561747	0.364129	0.120594	-0.561747	0.364129	0.120594
S150P	-0.409305	0.099134	0.067137	-0.409305	0.099134	0.067137
S150Q	-0.885544	0.369819	0.127996	-0.885544	0.369819	0.127996
S150R	0.168259	0.347399	0.052984	0.168259	0.347399	0.052984
S150T	-0.379773	0.0626	0.058278	-0.379773	0.0626	0.058278
S150V	-0.573182	0.13519	0.049748	-0.573182	0.13519	0.049748
S150W	0.22529	0.175529	0.076206	0.22529	0.175529	0.076206
S150Y	-0.242341	0.252606	0.12335	-0.242341	0.252606	0.12335
S150*	Null	nan	nan	Null	nan	nan
Y151A	0.453413	0.085998	0.061427	0.453413	0.085998	0.061427
Y151C	0.515757	0.223588	0.092441	0.515757	0.223588	0.092441
Y151D	Null	nan	nan	Null	nan	nan
Y151E	-0.682554	0.223727	0.096204	-0.682554	0.223727	0.096204
Y151F	-0.027597	0.200293	0.063059	-0.027597	0.200293	0.063059
Y151G	0.100526	0.158047	0.070526	0.100526	0.158047	0.070526
Y151H	0.271976	0.119711	0.036417	0.271976	0.119711	0.036417
Y151I	0.121344	0.408966	0.069214	0.121344	0.408966	0.069214
Y151K	0.191635	0.154808	0.066226	0.191635	0.154808	0.066226
Y151L	-0.017294	0.18269	0.041456	-0.017294	0.18269	0.041456
Y151M	-0.101382	0.097261	0.058138	-0.101382	0.097261	0.058138
Y151N	0.132954	0.274422	0.083012	0.132954	0.274422	0.083012
Y151P	-0.872159	0.365826	0.141246	-0.872159	0.365826	0.141246
Y151Q	0.583552	0.215473	0.082381	0.583552	0.215473	0.082381
Y151R	0.1829	0.184835	0.038521	0.1829	0.184835	0.038521
Y151S	0.202214	0.033187	0.030418	0.202214	0.033187	0.030418
Y151T	0.384129	0.012255	0.045657	0.384129	0.012255	0.045657
Y151V	0.552957	0.230954	0.034759	0.552957	0.230954	0.034759
Y151W	-0.168368	0.073923	0.087807	-0.168368	0.073923	0.087807
Y151*	Null	nan	nan	Null	nan	nan
C152A	0.101074	0.10354	0.04752	0.101074	0.10354	0.04752
C152D	0.279227	0.282789	0.084735	0.279227	0.282789	0.084735

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
C152E	0.116394	0.117002	0.034483	0.116394	0.117002	0.034483
C152F	0.074314	0.445906	0.096177	0.074314	0.445906	0.096177
C152G	0.417903	0.114998	0.055492	0.417903	0.114998	0.055492
C152H	0.495131	0.19275	0.112589	0.495131	0.19275	0.112589
C152I	0.401517	0.055359	0.089774	0.401517	0.055359	0.089774
C152K	0.513714	0.187451	0.073045	0.513714	0.187451	0.073045
C152L	0.331649	0.240165	0.054965	0.331649	0.240165	0.054965
C152M	0.603859	0.218139	0.082787	0.603859	0.218139	0.082787
C152N	0.57544	0.153536	0.095213	0.57544	0.153536	0.095213
C152P	ND	ND	ND	ND	ND	ND
C152Q	0.37433	0.202486	0.080109	0.37433	0.202486	0.080109
C152R	0.411386	0.146001	0.055639	0.411386	0.146001	0.055639
C152S	0.733119	0.647613	0.067441	0.733119	0.647613	0.067441
C152T	0.625859	0.437414	0.058969	0.625859	0.437414	0.058969
C152V	0.066722	0.0048	0.062039	0.066722	0.0048	0.062039
C152W	0.433486	0.123034	0.09494	0.433486	0.123034	0.09494
C152Y	ND	ND	ND	ND	ND	ND
C152*	Null	nan	nan	Null	nan	nan
F153A	-0.337208	0.09709	0.131007	-0.337208	0.09709	0.131007
F153C	0.306838	0.343525	0.119788	0.306838	0.343525	0.119788
F153D	ND	ND	ND	ND	ND	ND
F153E	-0.541517	0.09861	0.08234	-0.541517	0.09861	0.08234
F153G	-0.99418	0.427345	0.246267	-0.99418	0.427345	0.246267
F153H	-0.375667	0.258464	0.113332	-0.375667	0.258464	0.113332
F153I	0.169574	0.354426	0.096533	0.169574	0.354426	0.096533
F153K	-0.195575	0.170034	0.054643	-0.195575	0.170034	0.054643
F153L	0.4361	0.286262	0.047746	0.4361	0.286262	0.047746
F153M	0.27813	0.088337	0.08986	0.27813	0.088337	0.08986
F153N	-0.252195	0.009024	0.048187	-0.252195	0.009024	0.048187
F153P	-1.726927	0.08694	0.194103	-1.726927	0.08694	0.194103
F153Q	-0.71294	0.228405	0.106155	-0.71294	0.228405	0.106155
F153R	-0.530213	0.346152	0.163709	-0.530213	0.346152	0.163709
F153S	-0.25741	0.055733	0.08485	-0.25741	0.055733	0.08485
F153T	0.107895	0.045865	0.071245	0.107895	0.045865	0.071245
F153V	-0.168301	0.068649	0.057268	-0.168301	0.068649	0.057268
F153W	-0.51396	0.358965	0.147375	-0.51396	0.358965	0.147375
F153Y	-0.429285	0.368183	0.129983	-0.429285	0.368183	0.129983
F153*	Null	nan	nan	Null	nan	nan
E154A	0.842312	0.414856	0.075637	0.842312	0.414856	0.075637
E154C	0.703864	0.576379	0.061974	0.703864	0.576379	0.061974
E154D	0.602054	0.166829	0.077153	0.602054	0.166829	0.077153
E154F	0.532958	0.507638	0.075528	0.532958	0.507638	0.075528
E154G	0.549209	0.017413	0.038253	0.549209	0.017413	0.038253
E154H	0.529285	0.041401	0.090207	0.529285	0.041401	0.090207
E154I	0.600984	0.056172	0.063116	0.600984	0.056172	0.063116
E154K	1.047673	0.207543	0.12325	1.047673	0.207543	0.12325
E154L	0.758277	0.083654	0.052496	0.758277	0.083654	0.052496
E154M	1.015258	0.177128	0.099266	1.015258	0.177128	0.099266
E154N	0.658014	0.070477	0.048473	0.658014	0.070477	0.048473
E154P	-1.421343	0.076805	0.234983	-1.421343	0.076805	0.234983

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
E154Q	0.584605	0.212548	0.075847	0.584605	0.212548	0.075847
E154R	0.808566	0.229761	0.054243	0.808566	0.229761	0.054243
E154S	0.729324	0.079303	0.040453	0.729324	0.079303	0.040453
E154T	0.753507	0.211873	0.05259	0.753507	0.211873	0.05259
E154V	1.119649	0.109459	0.079798	1.119649	0.109459	0.079798
E154W	0.61901	0.243298	0.073548	0.61901	0.243298	0.073548
E154Y	0.661189	0.316149	0.088452	0.661189	0.316149	0.088452
E154*	Null	nan	nan	Null	nan	nan
I155A	-0.229846	0.281662	0.086591	-0.229846	0.281662	0.086591
I155C	-0.043806	0.345769	0.09967	-0.043806	0.345769	0.09967
I155D	ND	ND	ND	ND	ND	ND
I155E	-0.071538	0.189189	0.067356	-0.071538	0.189189	0.067356
I155F	-0.105201	0.03414	0.08989	-0.105201	0.03414	0.08989
I155G	-0.645301	0.15168	0.082036	-0.645301	0.15168	0.082036
I155H	-0.155965	0.269494	0.084812	-0.155965	0.269494	0.084812
I155K	-0.083325	0.477838	0.045035	-0.083325	0.477838	0.045035
I155L	-0.106669	0.046736	0.079121	-0.106669	0.046736	0.079121
I155M	-0.009669	0.113589	0.082858	-0.009669	0.113589	0.082858
I155N	-0.648878	0.102012	0.052741	-0.648878	0.102012	0.052741
I155P	ND	ND	ND	ND	ND	ND
I155Q	-0.172609	0.098686	0.075025	-0.172609	0.098686	0.075025
I155R	0.102921	0.159822	0.054377	0.102921	0.159822	0.054377
I155S	-0.177263	0.101929	0.078529	-0.177263	0.101929	0.078529
I155T	-0.140911	0.418981	0.077215	-0.140911	0.418981	0.077215
I155V	-0.148475	0.117733	0.058684	-0.148475	0.117733	0.058684
I155W	0.293523	0.093716	0.094105	0.293523	0.093716	0.094105
I155Y	-0.140694	0.300686	0.089146	-0.140694	0.300686	0.089146
I155*	Null	nan	nan	Null	nan	nan
L156A	0.231095	0.057154	0.098439	0.231095	0.057154	0.098439
L156C	-0.913153	0.554717	0.153025	-0.913153	0.554717	0.153025
L156D	-0.333723	0.419439	0.145673	-0.333723	0.419439	0.145673
L156E	-0.368285	0.359232	0.07954	-0.368285	0.359232	0.07954
L156F	0.781829	0.193908	0.06317	0.781829	0.193908	0.06317
L156G	-0.077748	0.022302	0.103407	-0.077748	0.022302	0.103407
L156H	0.334435	0.042529	0.071852	0.334435	0.042529	0.071852
L156I	0.113507	0.365406	0.106393	0.113507	0.365406	0.106393
L156K	0.65405	0.228508	0.069746	0.65405	0.228508	0.069746
L156M	0.72834	0.043556	0.053424	0.72834	0.043556	0.053424
L156N	-0.175884	0.224332	0.059886	-0.175884	0.224332	0.059886
L156P	-1.800905	0.189435	0.290615	-1.800905	0.189435	0.290615
L156Q	0.447571	0.415392	0.159669	0.447571	0.415392	0.159669
L156R	0.642856	0.068963	0.063288	0.642856	0.068963	0.063288
L156S	0.599233	0.170974	0.049029	0.599233	0.170974	0.049029
L156T	ND	ND	ND	ND	ND	ND
L156V	-0.536009	0.066009	0.098362	-0.536009	0.066009	0.098362
L156W	0.396068	0.291407	0.046114	0.396068	0.291407	0.046114
L156Y	1.037471	0.086083	0.062928	1.037471	0.086083	0.062928
L156*	Null	nan	nan	Null	nan	nan
E157A	0.071887	0.129014	0.046327	0.071887	0.129014	0.046327
E157C	-0.165696	0.165664	0.142039	-0.165696	0.165664	0.142039

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
E157D	0.047007	0.225789	0.100346	0.047007	0.225789	0.100346
E157F	-0.357111	0.20521	0.046461	-0.357111	0.20521	0.046461
E157G	0.243426	0.167559	0.046394	0.243426	0.167559	0.046394
E157H	-0.247393	0.034638	0.074433	-0.247393	0.034638	0.074433
E157I	-0.327962	0.320657	0.109284	-0.327962	0.320657	0.109284
E157K	0.118457	0.057136	0.075352	0.118457	0.057136	0.075352
E157L	-0.269044	0.089	0.07116	-0.269044	0.089	0.07116
E157M	-0.125885	0.191334	0.070505	-0.125885	0.191334	0.070505
E157N	-0.073781	0.134792	0.048194	-0.073781	0.134792	0.048194
E157P	-1.19412	0.457708	0.259685	-1.19412	0.457708	0.259685
E157Q	0.206644	0.31462	0.111591	0.206644	0.31462	0.111591
E157R	0.095197	0.133488	0.048808	0.095197	0.133488	0.048808
E157S	0.119244	0.074381	0.044021	0.119244	0.074381	0.044021
E157T	-0.109943	0.178324	0.026713	-0.109943	0.178324	0.026713
E157V	-0.212105	0.260062	0.053703	-0.212105	0.260062	0.053703
E157W	-0.167407	0.527299	0.116825	-0.167407	0.527299	0.116825
E157Y	-0.27642	0.347972	0.053455	-0.27642	0.347972	0.053455
E157*	Null	nan	nan	Null	nan	nan
R158A	-0.146942	0.26873	0.047617	-0.146942	0.26873	0.047617
R158C	-0.069513	0.275487	0.084992	-0.069513	0.275487	0.084992
R158D	-0.322753	0.199619	0.108675	-0.322753	0.199619	0.108675
R158E	-0.163414	0.29593	0.038361	-0.163414	0.29593	0.038361
R158F	-0.186788	0.336992	0.074103	-0.186788	0.336992	0.074103
R158G	0.088791	0.151379	0.054499	0.088791	0.151379	0.054499
R158H	-0.032367	0.15649	0.09506	-0.032367	0.15649	0.09506
R158I	-0.298066	0.38141	0.046333	-0.298066	0.38141	0.046333
R158K	0.04307	0.110082	0.036897	0.04307	0.110082	0.036897
R158L	ND	ND	ND	ND	ND	ND
R158M	-0.044343	0.272875	0.057339	-0.044343	0.272875	0.057339
R158N	0.152121	0.294588	0.074965	0.152121	0.294588	0.074965
R158P	-0.24025	0.597181	0.053229	-0.24025	0.597181	0.053229
R158Q	ND	ND	ND	ND	ND	ND
R158S	0.050277	0.164083	0.080875	0.050277	0.164083	0.080875
R158T	-0.022997	0.021505	0.076427	-0.022997	0.021505	0.076427
R158V	-0.012189	0.102621	0.036349	-0.012189	0.102621	0.036349
R158W	ND	ND	ND	ND	ND	ND
R158Y	ND	ND	ND	ND	ND	ND
R158*	-1.51552	0.43876	0.247804	-1.51552	0.43876	0.247804
R159A	0.132409	0.236644	0.097972	0.132409	0.236644	0.097972
R159C	0.382074	0.045914	0.165103	0.382074	0.045914	0.165103
R159D	0.406127	0.076776	0.080014	0.406127	0.076776	0.080014
R159E	0.473272	0.125346	0.047458	0.473272	0.125346	0.047458
R159F	0.370461	0.032443	0.081212	0.370461	0.032443	0.081212
R159G	0.369054	0.41151	0.037446	0.369054	0.41151	0.037446
R159H	0.323235	0.249904	0.079968	0.323235	0.249904	0.079968
R159I	0.39535	0.074437	0.105797	0.39535	0.074437	0.105797
R159K	0.150092	0.029852	0.05683	0.150092	0.029852	0.05683
R159L	0.381551	0.065934	0.077351	0.381551	0.065934	0.077351
R159M	0.430365	0.061064	0.047027	0.430365	0.061064	0.047027
R159N	0.258661	0.13327	0.076065	0.258661	0.13327	0.076065

(Continued on next page)

Mutation	-Lon			+Lon		
	Selection coefficient	Standard deviation	Standard error	Selection coefficient	Standard deviation	Standard error
R159P	0.490356	0.083603	0.035952	0.490356	0.083603	0.035952
R159Q	Null	nan	nan	Null	nan	nan
R159S	0.366392	0.415275	0.126562	0.366392	0.415275	0.126562
R159T	0.22844	0.292083	0.076299	0.22844	0.292083	0.076299
R159V	0.185262	0.283357	0.039961	0.185262	0.283357	0.039961
R159W	Null	nan	nan	Null	nan	nan
R159Y	0.273215	0.615767	0.060937	0.273215	0.615767	0.060937
R159*	0.063761	0.505331	0.110904	0.063761	0.505331	0.110904

Table 5.6: *In vitro* DHFR reactions

variant	run	DHF concentration ([DHF]) and DHFR velocity (V)							
		[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
E154V	1	39.30	6.66	38.68	6.33	38.09	6.65	37.48	6.49
		36.93	6.53	36.30	6.94	35.72	6.75	35.11	6.81
		34.52	6.77	33.93	6.29	33.33	6.37	32.73	6.97
		32.14	6.97	31.56	6.41	30.99	6.49	30.39	6.98
		29.83	6.24	29.28	6.11	28.70	5.69	28.13	6.57
		27.58	6.20	27.04	6.70	26.51	6.46	25.97	5.63
		25.46	5.29	24.93	5.59	24.39	5.49	23.91	5.70
		23.40	5.22	22.90	6.07	22.40	5.48	21.92	5.12
		21.46	5.34	20.98	4.65	20.53	5.07	20.07	5.22
		19.63	4.98	19.18	5.14	18.74	4.99	18.32	5.00
		17.91	4.81	17.48	4.67	17.09	4.51	16.68	4.62
		16.28	4.21	15.91	4.37	15.52	4.06	15.14	3.86
		14.76	4.25	14.43	3.65	14.06	3.91	13.71	3.97
		13.39	3.71	13.05	3.39	12.73	3.45	12.41	3.74
		12.11	3.64	11.80	3.22	11.50	3.43	11.21	2.85
		E154V	2	10.92	3.07	10.63	3.08	10.35	2.90
38.65	11.71			38.14	12.15	37.63	9.84	37.14	11.90
36.63	11.73			36.13	10.96	35.65	10.79	35.10	12.29
34.61	11.68			34.07	11.46	33.56	12.33	33.05	12.13
32.54	11.86			32.03	11.83	31.52	11.73	31.02	12.61
30.52	11.05			30.02	11.70	29.52	11.69	29.02	12.09
28.54	10.99			28.05	12.38	27.55	10.89	27.05	11.24
26.60	10.51			26.14	9.72	25.65	9.79	25.20	10.89
24.74	10.20			24.27	10.24	23.82	10.79	23.39	10.59
22.92	9.85			22.48	9.51	22.03	10.65	21.61	9.94
21.19	10.30			20.76	9.02	20.34	10.64	19.90	8.11
19.48	9.32			19.06	8.79	18.65	9.56	18.25	9.15
17.84	9.29			17.45	9.16	17.06	8.96	16.67	9.13
16.30	7.46			15.93	8.23	15.57	7.46	15.20	8.17
14.85	7.68			14.49	8.44	14.15	7.01	13.81	8.11
13.46	7.59			13.14	6.95	12.82	7.33	12.50	7.32
12.20	6.62	11.89	7.39	11.58	6.13	11.28	7.14		

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
E154V	3	10.99	6.54	10.70	6.64	10.43	6.36	10.15	5.66
		9.88	5.93						
		41.77	11.03	41.31	12.61	40.85	12.15	40.36	11.26
		39.88	10.95	39.42	11.46	38.95	10.81	38.48	11.33
		38.00	11.33	37.49	12.18	37.03	10.00	36.52	11.77
		36.04	10.54	35.58	11.71	35.09	10.65	34.63	12.02
		34.18	11.88	33.71	10.76	33.25	11.50	32.81	9.71
		32.34	11.94	31.89	8.38	31.42	10.98	30.97	10.23
		30.52	9.99	30.08	9.12	29.62	10.84	29.18	10.71
		28.75	9.76	28.32	9.91	27.86	10.23	27.43	10.07
		27.01	9.69	26.58	9.12	26.18	10.14	25.75	8.35
		25.34	9.39	24.95	9.55	24.53	8.22	24.14	9.51
		23.75	9.93	23.35	8.55	22.96	8.80	22.58	8.79
		22.20	8.82	21.82	8.39	21.45	8.36	21.06	8.76
		20.73	9.07	20.35	7.93	20.01	7.37	19.65	8.52
		19.30	7.62	18.97	7.74	18.63	7.73	18.30	7.70
		17.98	7.03	17.64	7.81	17.31	6.95	17.02	6.90
		16.68	7.40	16.38	5.96	16.09	7.02	15.78	7.21
		15.49	6.97	15.19	6.82	14.91	5.91	14.62	6.32
		14.36	5.89	14.07	6.51	13.80	6.11	13.55	5.23
13.29	6.22	13.03	5.92	12.78	6.23	12.52	4.49		
12.28	5.54	12.04	6.20	11.80	5.07	11.57	5.09		
11.34	5.24	11.11	5.10	10.89	4.40	10.68	4.81		
10.46	4.84	10.24	4.70	10.04	4.94	9.83	5.30		
9.63	4.33	9.43	4.04	9.24	4.34	9.06	4.43		
8.87	4.15								
I41A	1	31.18	4.29	30.26	4.16	29.39	3.80	28.48	4.06
		27.52	4.11	26.55	4.42	25.62	4.11	24.64	4.60
		23.68	4.22	22.75	4.18	21.83	4.03	20.93	3.90
		20.04	3.93	19.17	3.86	18.32	3.73	17.48	3.64
		16.66	3.52	15.87	3.47	15.11	3.38	14.36	3.20
		13.64	3.08	12.94	3.05	12.27	2.96	11.61	2.88
		10.99	2.74	10.38	2.71	9.81	2.54	9.24	2.37
		8.72	2.30	8.21	2.15	7.73	2.11	7.28	1.93
6.84	1.86	6.42	1.85						
I41A	2	33.25	4.32	32.29	4.75	31.32	4.45	30.35	4.12
		29.39	4.48	28.39	4.83	27.36	4.58	26.36	4.53
		25.38	4.31	24.42	4.19	23.48	4.28	22.57	4.16
		21.67	4.09	20.78	3.99	19.91	3.93	19.07	3.78
		18.24	3.67	17.44	3.63	16.66	3.52	15.89	3.42
		15.17	3.29	14.46	3.19	13.77	3.12	13.11	3.04
		12.47	2.90	11.85	2.79	11.27	2.68	10.70	2.57
		10.15	2.46	9.63	2.33	9.13	2.26	8.65	2.11
8.20	1.96								
I41A	3	28.72	7.01	27.30	6.53	25.91	6.59	24.49	6.67
		23.09	6.29	21.71	6.65	20.32	6.29	18.99	6.18
		17.68	6.14	16.40	5.91	15.16	5.66	13.97	5.38
		12.85	5.06	11.80	4.81	10.80	4.52	9.88	4.11
		9.02	3.91	8.21	3.70	7.44	3.44	6.73	3.06
6.10	2.87	5.52	2.57	4.99	2.32				
M42F	1	29.92	6.28	28.56	5.97	27.27	5.67	26.03	5.52

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
M42F	2	24.85	5.27	23.71	5.16	22.59	4.85	21.53	4.79
		20.50	4.76	19.49	4.54	18.51	4.37	17.56	4.26
		16.63	4.11	15.73	3.95	14.86	3.82	14.02	3.71
		13.20	3.71	12.41	3.46	11.65	3.42	10.91	3.24
		10.21	3.18	9.52	3.03	8.85	2.97	8.22	2.87
		7.60	2.76	7.02	2.63	6.46	2.45	5.93	2.34
		47.32	7.19	45.76	6.94	44.18	7.17	42.62	7.22
		41.01	7.20	39.45	7.00	37.87	7.19	36.32	6.84
		34.80	6.83	33.28	6.70	31.82	6.66	30.38	6.60
		28.98	6.39	27.60	6.08	26.28	5.90	24.98	5.92
		23.73	5.58	22.52	5.41	21.32	5.28	20.17	5.15
		19.06	4.91	17.99	4.87	16.97	4.58	15.99	4.42
		15.05	4.17	14.15	3.88	13.29	3.80	12.48	3.63
		11.70	3.44	10.97	3.26	10.27	3.06	9.61	2.95
M42F	3	56.41	8.79	54.43	8.74	52.48	8.76	50.46	9.19
		48.37	9.43	46.28	9.33	44.20	9.25	42.16	9.01
		40.14	8.81	38.17	8.57	36.21	8.60	34.31	8.28
		32.44	8.15	30.62	7.90	28.87	7.63	27.20	7.30
		25.60	6.99	24.08	6.65	22.62	6.29	21.24	5.93
		19.96	5.69	18.73	5.36	17.56	4.98	16.47	4.75
		15.46	4.34	14.50	4.15	13.61	3.85	12.77	3.53
		12.00	3.29	11.27	3.12	10.60	2.82	9.97	2.75
		9.38	2.56						
		26.76	3.88	25.46	3.95	24.14	3.97	22.79	4.10
L62V	1	21.39	4.23	19.99	4.03	18.64	3.99	17.35	3.61
		16.14	3.64	14.89	3.67	13.64	3.78	12.39	3.62
		11.17	3.56	10.00	3.42	8.88	3.24	7.81	3.06
		6.82	2.84	5.90	2.60				
		46.86	3.15	46.42	3.13	45.98	2.91	45.53	3.00
L62V	2	45.09	2.81	44.66	3.05	44.26	2.84	43.83	3.10
		43.34	3.21	42.85	2.89	42.42	3.46	41.97	2.78
		41.54	2.80	41.09	3.65	40.61	3.35	40.09	3.18
		39.59	3.60	39.09	3.51	38.59	3.48	38.08	4.03
		37.56	2.89	37.08	3.35	36.53	3.42	36.05	3.36
		35.55	3.26	35.02	3.49	34.52	3.28	34.02	3.45
		33.51	3.68	32.99	3.49	32.48	3.59	31.99	3.36
		31.47	3.17	30.98	3.47	30.48	3.14	29.97	3.37
		29.47	3.09	28.98	3.03	28.51	3.40	28.03	2.89
		27.54	3.35	27.07	3.29	26.60	3.28	26.15	2.78
		25.69	3.53	25.23	2.82	24.77	3.10	24.34	2.96
		23.89	2.71	23.45	2.91	23.03	2.76	22.61	3.00
		22.19	2.76	21.77	2.82	21.36	2.69	20.95	2.52
		20.56	2.91	20.18	2.65	19.79	2.44	19.41	2.50
		19.04	2.55	18.67	2.16	18.32	2.21	17.94	2.36
		17.60	2.57	17.25	2.19	16.92	2.18	16.58	2.19
		16.26	2.13	15.93	2.27	15.61	2.08	15.31	2.04
		15.01	1.96	14.70	1.87	14.40	2.01	14.12	1.91
13.85	1.83	13.57	1.72	13.30	1.66	13.03	1.70		
12.78	1.64	12.53	1.70	12.28	1.68	12.05	1.67		
11.80	1.69	11.59	1.53	11.36	1.31	11.14	1.32		
10.93	1.36								

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
L62V	3	40.36	3.69	39.86	3.57	39.29	3.69	38.73	3.93
		38.21	3.76	37.63	3.93	37.11	3.70	36.54	4.02
		35.98	3.61	35.44	3.35	34.92	3.44	34.34	3.67
		33.78	3.75	33.23	3.84	32.65	3.71	32.09	3.59
		31.52	3.59	30.99	3.59	30.43	3.68	29.90	3.29
		29.35	3.60	28.82	3.39	28.28	3.66	27.74	3.43
		27.24	3.50	26.72	3.42	26.21	3.44	25.70	3.11
		25.19	3.25	24.69	3.23	24.21	3.23	23.71	3.24
		23.22	3.16	22.77	3.21	22.30	2.98	21.83	2.89
		21.39	2.88	20.93	2.89	20.50	2.83	20.06	2.97
		19.62	2.90	19.19	2.74	18.79	2.76	18.37	2.48
		17.96	2.66	17.58	2.56	17.18	2.51	16.79	2.60
		16.42	2.42	16.04	2.45	15.68	2.44	15.32	2.22
		14.97	2.49	14.62	2.35	14.28	2.14	13.94	2.18
		13.62	2.24	13.29	2.27	12.98	2.05	12.66	2.02
		12.36	2.12	12.06	2.06	11.76	2.00	11.48	1.84
11.20	1.83	10.92	2.00	10.64	1.76	10.39	1.77		
10.12	1.76	9.86	1.84						
L62V	4	38.87	3.86	37.54	3.92	36.22	3.86	34.91	3.80
		33.56	3.98	32.21	3.99	30.86	4.06	29.51	4.02
		28.18	3.88	26.87	3.76	25.59	3.75	24.35	3.68
		23.13	3.60	21.94	3.56	20.75	3.37	19.61	3.32
		18.49	3.29	17.40	3.15	16.35	3.11	15.34	2.95
		14.37	2.76	13.44	2.69	12.55	2.53	11.69	2.44
10.89	2.27								
C85L	1	59.29	7.59	57.71	8.20	56.04	8.39	54.38	7.71
		52.79	7.92	51.15	7.96	49.55	7.76	48.00	7.49
		46.39	7.89	44.80	7.91	43.16	7.98	41.57	7.82
		39.99	7.70	38.42	7.55	36.91	7.36	35.42	7.16
		33.96	7.09	32.55	6.82	31.18	6.59	29.83	6.48
		28.55	6.08	27.30	5.94	26.09	5.77	24.93	5.48
		23.82	5.35	22.75	5.10	21.71	4.88	20.74	4.73
		19.78	4.46	18.87	4.30	18.00	4.23		
C85L	2	41.01	5.82	39.75	6.03	38.52	6.06	37.22	6.29
		36.04	5.56	34.91	5.67	33.76	5.29	32.63	5.55
		31.47	5.52	30.32	5.75	29.13	5.87	27.90	6.02
		26.69	5.92	25.50	5.70	24.36	5.36	23.26	5.42
		22.19	5.30	21.14	5.25	20.10	4.91	19.10	4.84
		18.12	4.74	17.17	4.49	16.25	4.42	15.36	4.21
		14.51	4.00	13.69	4.08	12.90	3.78	12.13	3.65
11.41	3.48	10.72	3.25						
C85L	3	33.48	5.09	32.45	4.71	31.40	4.91	30.35	5.17
		29.29	5.19	28.32	4.77	27.30	5.22	26.20	5.48
		25.06	5.54	23.95	5.57	22.82	5.41	21.76	5.09
		20.74	4.83	19.76	4.88	18.75	5.11	17.71	5.01
		16.68	4.88	15.67	4.84	14.71	4.51	13.78	4.48
		12.89	4.24	12.02	4.19	11.20	3.82	10.43	3.62
9.68	3.56								
R98Y	1	54.35	4.41	53.42	4.84	52.45	4.15	51.52	4.68
		50.56	4.63	49.60	4.92	48.56	5.18	47.54	4.93
		46.53	4.80	45.54	4.97	44.55	5.10	43.54	4.85

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
R98Y	2	42.53	5.04	41.51	4.98	40.48	4.84	39.48	4.73
		38.52	4.62	37.56	4.56	36.60	4.63	35.67	4.56
		34.74	4.66	33.82	4.45	32.91	4.38	32.02	4.28
		31.14	4.23	30.28	4.16	29.43	4.17	28.60	4.00
		27.78	3.96	26.99	3.91	26.21	3.76	25.45	3.62
		24.70	3.65	23.96	3.56	23.25	3.58	22.55	3.42
		21.86	3.28	21.19	3.18	20.54	3.26	19.91	3.06
		19.30	2.98	18.70	2.83	18.11	2.83	17.54	2.76
		16.98	2.66	16.45	2.63	15.93	2.51	15.42	2.43
		14.93	2.46	14.45	2.25	13.99	2.23	13.53	2.13
		13.10	2.14	12.68	1.97	12.27	1.99	11.87	1.92
		11.50	1.83	11.13	1.80	10.78	1.66	10.45	1.65
		40.40	4.14	39.63	3.75	38.85	3.75	38.07	3.79
		37.27	3.95	36.45	4.22	35.62	3.83	34.79	4.03
		33.96	4.16	33.08	4.17	32.22	4.23	31.39	4.07
		30.55	4.06	29.72	3.95	28.90	4.00	28.08	3.97
		27.25	4.03	26.46	3.78	25.65	3.80	24.86	3.88
		24.08	3.96	23.32	3.68	22.57	3.80	21.84	3.69
		21.11	3.43	20.40	3.38	19.71	3.32	19.03	3.29
		18.38	3.08	17.72	3.21	17.10	2.96	16.48	3.01
15.90	2.82	15.31	2.73	14.75	2.84	14.22	2.69		
13.68	2.68	13.17	2.46	12.67	2.28	12.20	2.24		
11.74	2.15	11.28	2.21	10.85	2.16	10.43	2.03		
10.02	1.90								
R98Y	3	29.22	5.40	28.13	5.35	27.06	5.10	26.03	4.99
		24.98	5.16	23.90	5.19	22.83	5.39	21.74	5.32
		20.67	5.25	19.61	5.16	18.54	5.01	17.52	4.82
		16.55	4.79	15.58	4.63	14.63	4.60	13.72	4.37
		12.82	4.38	11.96	4.07	11.13	3.92	10.32	3.76
		9.56	3.60	8.83	3.44	8.14	3.27	7.49	3.15
		6.87	3.02	6.28	2.85	5.73	2.71		
		26.40	8.58	25.78	8.11	25.13	8.46	24.50	7.95
L156Y	1	23.86	8.05	23.22	7.64	22.54	8.78	21.88	8.03
		21.24	8.21	20.57	8.77	19.91	8.62	19.28	8.57
		18.63	8.44	17.95	8.28	17.27	8.96	16.61	8.07
		15.94	9.07	15.25	8.50	14.58	9.07	13.93	8.38
		13.27	8.36	12.62	8.31	11.99	8.19	11.38	8.13
		10.76	7.70	10.15	7.63	9.55	7.66	8.96	7.56
		8.40	7.39	7.85	7.17	7.31	6.93	6.79	6.33
		6.29	6.61	5.81	6.15				
L156Y	2	34.51	7.92	33.91	8.21	33.32	7.38	32.72	8.14
		32.12	7.92	31.51	8.16	30.86	8.46	30.21	7.78
		29.57	8.36	28.94	8.93	28.28	8.90	27.59	8.46
		26.92	8.17	26.28	8.92	25.63	8.64	24.97	8.08
		24.31	8.53	23.64	9.10	22.97	8.93	22.30	8.45
		21.64	8.59	20.98	8.86	20.32	8.09	19.69	8.48
		19.03	8.36	18.41	8.39	17.78	7.75	17.15	8.55
		16.54	7.43	15.95	7.64	15.35	7.17	14.77	7.36
		14.22	7.42	13.65	6.89	13.11	7.21	12.57	6.90
		12.04	6.77	11.53	6.83	11.03	6.15	10.54	6.19
		10.06	6.14	9.58	6.13	9.10	6.02	8.66	5.51

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
L156Y	3	8.20	5.61	7.78	5.61				
		47.11	7.40	45.94	7.46	44.77	7.71	43.57	7.67
		42.38	8.00	41.16	7.73	39.96	7.60	38.78	7.07
		37.62	7.53	36.43	7.59	35.28	7.58	34.12	7.25
		32.97	7.13	31.85	7.06	30.74	6.80	29.67	6.81
		28.60	6.67	27.56	6.54	26.52	6.35	25.50	6.33
		24.53	6.28	23.56	6.24	22.61	5.87	21.69	6.06
		20.78	5.55	19.89	5.47	19.05	5.53	18.21	5.29
		17.41	4.96	16.63	4.92	15.87	4.70	15.15	4.38
		14.45	4.47						
Q102L	1	30.23	2.67	29.66	2.81	29.05	2.95	28.43	3.12
		27.81	3.08	27.16	2.90	26.53	3.11	25.88	3.09
		25.24	3.04	24.58	3.03	23.92	3.20	23.26	3.16
		22.60	3.15	21.94	3.14	21.28	2.95	20.63	3.09
		19.99	3.07	19.36	3.02	18.73	2.85	18.12	2.88
		17.52	2.83	16.92	2.78	16.33	2.82	15.77	2.69
		15.19	2.53	14.64	2.56	14.11	2.59	13.57	2.54
		13.06	2.43	12.55	2.33	12.06	2.31	11.58	2.17
		11.11	2.09	10.66	2.15	10.21	2.03	9.79	2.02
		9.36	2.01	8.96	1.97	8.56	1.90	8.18	1.67
Q102L	2	7.82	1.66						
		47.95	2.95	47.36	2.93	46.73	2.79	46.15	2.78
		45.54	2.70	44.94	2.94	44.34	2.84	43.72	2.70
		43.13	3.04	42.50	2.89	41.89	3.08	41.28	2.95
		40.66	3.10	40.03	2.87	39.41	2.79	38.81	2.74
		38.19	2.90	37.57	2.87	36.97	2.92	36.38	2.80
		35.75	2.67	35.17	3.01	34.58	2.67	34.00	2.73
		33.42	2.76	32.85	2.62	32.28	2.71	31.72	2.58
		31.18	2.69	30.64	2.53	30.11	2.48	29.57	2.41
		29.06	2.33	28.55	2.42	28.04	2.37	27.56	2.39
		27.06	2.29	26.58	2.34	26.10	2.24	25.64	2.33
		25.18	2.31	24.73	2.04	24.28	2.12	23.85	2.04
		23.43	1.94	22.99	2.05	22.59	2.08	22.17	2.06
		21.77	1.88	21.38	1.96	20.99	1.89	20.62	1.84
		20.24	1.65	19.87	1.78	19.52	1.69	19.16	1.68
		18.82	1.67	18.48	1.68	18.15	1.53	17.82	1.53
		17.51	1.50	17.20	1.34	16.90	1.36	16.60	1.35
		16.31	1.56	16.03	1.39	15.74	1.28	15.47	1.29
		15.21	1.24	14.94	1.25	14.68	1.20	14.42	1.26
		14.17	1.18	13.90	1.25	13.66	1.29	13.42	1.18
13.18	1.24	12.94	1.11	12.73	0.91	12.56	0.73		
12.42	0.49	12.33	0.40	12.25	0.38	12.18	0.29		
12.12	0.38	12.03	0.51						
Q102L	3	36.53	3.23	35.89	2.77	35.27	3.08	34.61	3.32
		33.95	3.09	33.31	2.92	32.67	2.92	32.01	3.17
		31.35	3.00	30.71	3.09	30.04	3.08	29.40	3.25
		28.72	2.96	28.07	3.29	27.41	3.19	26.73	3.15
		26.06	3.10	25.40	3.25	24.73	3.11	24.09	3.14
		23.44	3.12	22.81	3.01	22.18	3.07	21.57	2.97
		20.95	3.00	20.36	2.84	19.77	2.85	19.19	2.71
		18.61	2.65	18.07	2.47	17.53	2.57	17.00	2.48

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})		
H45S	1	16.48	2.56	15.97	2.51	15.47	2.41	14.99	2.35		
		14.52	2.17	14.06	2.25	13.62	2.07	13.18	2.10		
		12.76	1.86	12.35	1.98	11.95	1.87	11.56	1.74		
		11.19	1.84	10.82	1.67	10.46	1.88	10.11	1.69		
		44.83	5.37	43.64	5.46	42.42	5.47	41.16	5.79		
		39.88	5.72	38.64	5.65	37.40	5.81	36.18	5.46		
		34.99	5.20	33.85	5.28	32.70	5.27	31.56	5.03		
		30.46	5.01	29.37	4.93	28.32	4.80	27.26	4.59		
		26.23	4.47	25.25	4.39	24.28	4.39	23.34	4.30		
		22.43	4.12	21.55	4.00	20.69	3.88	19.85	3.63		
		19.05	3.70	18.26	3.57	17.51	3.49	16.76	3.32		
H45S	2	16.04	3.15	15.34	3.09	14.67	3.03	14.02	2.89		
		13.39	2.73	12.79	2.70	12.20	2.51	11.65	2.49		
		42.30	5.46	41.10	5.49	39.92	5.40	38.73	5.24		
		37.55	5.17	36.36	5.74	35.09	5.93	33.82	5.77		
		32.58	5.60	31.38	5.47	30.20	5.16	29.04	5.20		
		27.91	5.01	26.85	4.71	25.87	4.34	24.94	4.32		
		23.92	4.97	22.76	5.62	21.62	4.54	20.67	4.27		
		19.74	4.07	18.85	3.89	17.96	3.97	17.12	3.91		
		16.30	3.67	15.52	3.62	14.77	3.37	14.03	3.33		
		13.31	3.17	12.62	2.96	11.97	3.01	11.33	2.73		
		H45S	3	10.72	2.68						
35.27	3.86			34.53	3.01	33.81	3.59	32.98	3.73		
32.13	3.77			31.26	4.05	30.41	3.76	29.58	3.77		
28.78	3.65			27.97	3.67	27.15	3.67	26.35	3.61		
25.57	3.59			24.78	3.49	24.00	3.45	23.23	3.55		
22.45	3.64			21.70	3.39	20.94	3.32	20.19	3.36		
19.46	3.20			18.73	3.08	18.03	3.19	17.34	3.12		
16.66	2.98			16.00	2.94	15.34	2.92	14.72	2.77		
14.10	2.72			13.52	2.68	12.93	2.63	12.38	2.51		
11.84	2.34			11.32	2.31	10.81	2.25	10.33	2.12		
I91G	1			9.86	2.12	9.41	2.07	8.98	2.11	8.57	1.84
		48.03	10.70	46.51	11.79	44.91	11.86	43.36	11.19		
		41.86	10.89	40.37	10.94	38.91	10.90	37.50	10.38		
		36.09	10.59	34.66	10.65	33.25	9.80	31.86	10.07		
		30.52	9.96	29.21	9.55	27.93	9.12	26.68	9.33		
		25.47	8.75	24.29	8.41	23.15	8.29	22.03	7.93		
		20.94	7.90	19.89	7.90	18.87	7.40	17.88	7.06		
		16.94	6.61	16.04	6.52	15.17	6.23	14.34	6.15		
		13.53	5.90	12.78	5.63	12.05	5.17	11.36	4.77		
		10.71	4.57	10.09	4.51	9.50	4.23	8.94	3.89		
		I91G	2	41.85	10.50	40.49	9.94	39.11	9.57	37.73	10.94
36.31	10.50			34.91	10.83	33.43	11.10	31.92	10.95		
30.46	10.83			29.03	10.62	27.64	9.83	26.26	9.85		
24.94	9.75			23.66	9.70	22.40	9.16	21.20	8.90		
20.01	8.71			18.86	8.20	17.75	7.86	16.69	7.70		
15.64	7.34			14.68	7.00	13.73	7.25	12.84	6.37		
11.98	6.19			11.17	6.00	10.39	5.72	9.65	5.39		
8.95	4.98										
I91G	3			34.18	7.41	33.20	6.75	32.25	7.28	31.28	7.07
				30.33	6.88	29.47	6.07	28.69	5.57	27.93	5.91

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
Q102W	1	27.15	5.88	26.35	6.18	25.50	6.39	24.64	6.41
		23.79	6.06	22.99	5.69	22.23	5.70	21.50	5.31
		20.76	5.60	20.04	5.42	19.31	5.11	18.59	5.20
		17.90	5.04	17.23	4.79	16.58	4.62	15.96	4.53
		15.38	4.20	14.80	4.03	14.27	4.03	13.74	3.75
		13.24	3.59	12.76	3.59	12.30	3.35	11.85	3.22
		11.43	3.11	11.01	2.93	10.62	2.81		
		43.40	2.23	42.98	2.21	42.57	2.48	42.15	1.94
		41.75	2.07	41.37	2.15	40.93	2.05	40.54	2.14
		40.14	2.35	39.71	1.91	39.30	2.33	38.90	1.85
		38.50	2.29	38.09	2.25	37.69	1.93	37.27	2.19
		36.86	2.28	36.47	1.91	36.07	1.94	35.68	2.03
		35.30	2.19	34.90	2.07	34.52	1.94	34.14	2.19
		33.78	2.00	33.40	2.26	33.03	1.84	32.66	1.55
		32.31	2.06	31.95	1.79	31.60	1.68	31.24	1.89
		30.89	1.65	30.58	2.02	30.21	1.82	29.90	1.93
		29.56	1.35	29.25	1.82	28.92	1.40	28.60	1.67
		28.31	1.80	28.02	1.45	27.74	1.40	27.49	1.56
		27.24	0.97	27.03	0.94	26.82	1.01	26.65	0.87
		26.49	0.79	26.34	0.81	26.16	0.81	25.95	1.17
		25.70	1.28	25.42	1.60	25.11	1.67	24.80	1.69
		24.47	1.79	24.11	1.89	23.77	1.92	23.41	1.79
		23.07	1.61	22.72	1.71	22.39	1.80	22.07	1.73
		21.77	1.69	21.46	1.59	21.17	1.56	20.88	1.47
		20.61	1.48	20.34	1.41	20.09	1.36	19.83	1.21
		19.60	1.31	19.36	1.33	19.14	1.04	18.94	0.90
		18.77	0.69	18.62	0.79	18.55	0.09	18.58	-0.44
		18.67	-0.48	18.73	0.12	18.63	0.89	18.41	1.24
		18.15	1.47	17.88	1.46	17.59	1.58	17.29	1.56
		16.98	1.49	16.70	1.31	16.44	1.38	16.20	1.21
16.00	1.08	15.80	0.95	15.62	0.84	15.47	0.73		
15.31	0.87	15.17	0.90	15.03	0.64	14.87	0.88		
14.73	0.70	14.57	0.85	14.42	0.73	14.25	0.73		
14.07	1.07	13.89	1.02	13.72	0.91	13.53	0.96		
13.35	1.03	13.17	0.87	13.00	0.94	12.83	0.89		
Q102W	2	32.26	2.87	31.76	2.59	31.25	2.86	30.75	2.70
		30.23	2.97	29.70	2.58	29.17	2.67	28.67	2.60
		28.14	2.74	27.62	2.87	27.08	2.80	26.56	2.77
		26.02	2.63	25.50	2.57	24.98	2.59	24.46	2.55
		23.96	2.57	23.45	2.60	22.96	2.59	22.47	2.52
		21.98	2.48	21.51	2.56	21.03	2.41	20.57	2.60
		20.10	2.23	19.66	2.44	19.22	2.40	18.78	2.23
		18.34	2.24	17.93	2.12	17.52	2.15	17.12	2.21
		16.72	2.06	16.34	2.01	15.96	1.97	15.58	1.89
		15.22	1.92	14.86	1.86	14.49	1.73	14.14	1.87
		13.79	1.88	13.46	1.88	13.11	1.77	12.79	1.70
		12.47	1.69	12.15	1.49	11.88	1.43	11.65	0.91
Q102W	3	11.47	0.86	11.30	1.01	11.09	1.30	10.81	1.50
		31.07	2.73	30.56	2.64	30.09	2.28	29.64	2.36
		29.23	2.34	28.76	2.54	28.27	2.63	27.77	2.54
		27.29	2.66	26.77	2.64	26.26	2.63	25.76	2.63

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
I91A	1	25.24	2.55	24.74	2.78	24.24	2.65	23.73	2.63
		23.23	2.47	22.72	2.67	22.22	2.74	21.72	2.65
		21.24	2.47	20.74	2.53	20.27	2.51	19.80	2.52
		19.33	2.47	18.89	2.36	18.42	2.34	17.97	2.36
		17.54	2.34	17.10	2.19	16.67	2.25	16.24	2.27
		15.83	2.14	15.43	1.93	15.02	2.12	14.64	2.17
		14.25	2.11	13.87	1.97	13.50	1.92	13.14	1.88
		12.78	2.04	12.43	1.78	12.08	1.76	11.75	1.74
		11.43	1.76	11.10	1.59	10.80	1.59	10.50	1.44
		10.21	1.52	9.92	1.51				
		43.34	8.52	42.21	8.39	41.04	8.43	39.86	8.75
		38.66	9.06	37.44	9.29	36.21	8.89	35.00	9.14
		33.78	8.89	32.56	8.77	31.37	8.75	30.18	8.70
		29.02	8.63	27.87	8.43	26.76	8.13	25.66	8.09
		24.60	7.78	23.54	7.60	22.51	7.20	21.49	7.58
		20.49	7.36	19.51	7.15	18.56	7.00	17.63	6.82
16.73	6.41	15.87	6.58	15.02	6.27	14.21	5.79		
13.43	5.66	12.68	5.28	11.96	5.10	11.28	5.02		
10.63	4.76	10.01	4.41	9.41	4.41	8.85	4.12		
I91A	2	37.21	9.04	35.97	8.95	34.69	9.25	33.42	9.49
		32.11	9.61	30.82	9.50	29.51	9.59	28.23	9.48
		26.96	9.38	25.69	9.20	24.45	9.17	23.24	9.24
		22.03	8.59	20.87	8.70	19.73	8.19	18.60	8.11
		17.50	7.87	16.44	7.70	15.43	7.36	14.44	7.17
		13.47	6.79	12.56	6.73	11.68	6.14	10.83	6.00
		10.02	6.00	9.25	5.59				
I91A	3	34.02	8.79	32.88	8.26	31.75	8.48	30.60	8.37
		29.46	8.55	28.30	8.57	27.10	9.20	25.87	9.20
		24.62	9.28	23.40	8.96	22.17	8.87	20.97	8.97
		19.80	8.55	18.66	8.27	17.58	8.12	16.51	7.89
		15.48	7.33	14.49	7.11	13.55	6.85	12.65	6.49
11.79	6.08	10.98	5.92	10.19	5.68				
T113V	1	27.50	3.63	26.68	3.87	25.88	3.41	25.14	3.22
		24.43	3.18	23.85	2.17	23.16	3.95	22.46	2.64
		21.86	2.73	21.30	2.59	20.73	2.58	20.16	2.45
		19.63	2.35	19.10	2.29	18.61	2.28	18.12	2.22
		17.64	2.16	17.18	1.99	16.75	1.91	16.35	1.85
		15.94	1.78	15.55	1.79	15.17	1.68	14.81	1.58
		14.45	1.59	14.12	1.55	13.78	1.45	13.45	1.40
		13.13	1.33	12.83	1.30	12.53	1.25	12.24	1.26
		11.97	1.27	11.70	1.24	11.43	1.24	11.18	1.15
		10.94	1.10	10.71	1.01	10.51	1.06	10.29	1.01
10.06	0.96	9.83	1.13	9.59	1.04	9.39	0.90		
9.19	0.80								
T113V	2	44.40	7.14	42.78	7.72	41.22	6.71	39.74	6.77
		38.34	6.14	37.00	5.97	35.73	5.78	34.50	5.49
		33.31	5.21	32.20	4.91	31.16	4.53	30.15	4.64
		29.16	4.44	28.22	4.11	27.31	4.04	26.42	4.00
		25.60	3.57	24.82	3.37	24.05	3.17	23.33	3.25
		22.63	3.05	21.96	3.11	21.31	2.97	20.70	2.82
		20.11	2.53	19.54	2.58	19.00	2.43	18.47	2.28

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})		
T113V	3	17.98	2.22	17.49	2.23	17.02	2.14	16.56	1.99		
		16.10	2.01	15.68	2.02	15.26	1.93	14.85	1.89		
		14.45	1.85	14.07	1.76	13.69	1.57	13.32	1.55		
		12.97	1.68	12.63	1.54	12.29	1.50	11.97	1.52		
		11.65	1.45	11.34	1.30	11.05	1.31	10.76	1.26		
		10.47	1.18	10.20	1.28	9.94	1.10	9.69	1.19		
		9.44	0.98	9.20	1.05	8.97	1.04	8.74	0.99		
		8.52	1.10	8.31	0.87	8.10	0.84	7.91	0.84		
		7.71	0.89	7.54	0.71	7.35	0.90	7.17	0.88		
		7.00	0.82	6.81	0.75	6.66	0.71	6.49	0.65		
		6.35	0.58	6.19	0.67						
		49.13	5.48	47.96	5.78	46.80	4.94	45.73	4.67		
		44.75	4.22	43.79	4.64	42.73	4.80	41.72	4.34		
		40.82	3.82	39.95	4.00	39.06	3.86	38.18	3.77		
		37.40	3.48	36.62	3.59	35.85	3.64	35.10	3.29		
		34.39	3.26	33.69	3.00	33.01	3.02	32.38	3.06		
		31.75	2.82	31.16	2.77	30.56	2.69	29.97	2.47		
		29.40	2.47	28.84	2.74	28.29	2.53	27.76	2.34		
		27.24	2.41	26.74	2.37	26.23	2.18	25.73	2.09		
		25.26	2.13	24.78	2.07	24.31	1.96	23.86	2.10		
		23.42	1.94	22.98	2.00	22.55	1.92	22.13	1.96		
		21.72	1.81	21.31	1.69	20.92	1.80	20.53	1.67		
		20.14	1.71	19.79	1.64	19.41	1.57	19.07	1.46		
		18.71	1.55	18.36	1.60	18.03	1.66	17.71	1.47		
		17.39	1.46	17.08	1.43	16.78	1.40	16.49	1.33		
		16.18	1.32	15.90	1.23	15.62	1.26	15.34	1.27		
		15.07	1.07	14.80	1.17	14.53	1.05	14.26	1.17		
14.02	1.16	13.76	1.14	13.52	1.02	13.27	1.10				
13.04	1.12	12.81	1.05	12.59	0.97	12.37	0.96				
12.15	1.09	11.94	1.00	11.74	0.84	11.54	0.87				
11.36	0.81	11.16	0.76	10.99	0.79	10.81	0.87				
10.64	0.78	10.50	0.59	10.39	0.28	10.34	0.16				
10.35	-0.15										
T113V	4	34.84	5.33	33.72	5.20	32.59	5.27	31.47	4.81		
		30.46	4.43	29.52	4.27	28.61	4.02	27.78	3.56		
		27.00	3.53	26.23	3.53	25.46	3.40	24.74	3.06		
		24.07	2.96	23.43	2.94	22.81	2.71	22.20	2.67		
		21.61	2.72	21.03	2.72	20.47	2.50	19.91	2.58		
		19.37	2.45	18.84	2.46	18.33	2.16	17.83	2.29		
		17.32	2.30	16.84	2.05	16.37	2.10	15.91	2.03		
		15.46	2.10	15.03	1.97	14.59	1.84	14.19	1.85		
		13.77	1.83	13.37	1.79	13.00	1.75	12.63	1.58		
		12.27	1.67	11.92	1.54	11.58	1.50	11.26	1.42		
		10.95	1.51	10.63	1.40	10.33	1.35	10.04	1.27		
		9.75	1.34	9.48	1.13	9.22	1.25	8.96	1.24		
		8.71	1.06	8.47	0.99	8.23	1.24	8.00	1.00		
		7.78	0.83	7.58	1.00	7.38	0.87	7.17	0.99		
		6.98	0.88	6.81	0.78	6.63	0.71	6.47	0.72		
		W30F	1	34.48	9.96	33.14	9.97	31.79	10.05	30.42	9.85
				29.10	9.61	27.79	9.71	26.50	9.27	25.22	9.47
23.97	9.26			22.75	8.93	21.58	8.55	20.43	8.26		

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
W30F	2	19.34	7.98	18.29	7.48	17.28	7.23	16.31	6.84
		15.39	6.64	14.51	6.23	13.67	5.95	12.89	5.60
		12.15	5.47	11.43	5.11	10.76	4.87	10.12	4.59
		9.52	4.34	8.96	3.99	8.43	3.91	7.93	3.62
		7.46	3.36						
		36.32	9.89	35.57	10.12	34.81	10.14	34.09	9.60
		33.38	9.17	32.68	9.62	31.92	9.79	31.14	9.98
		30.36	9.70	29.58	10.59	28.81	9.85	28.07	9.90
		27.33	10.06	26.56	9.59	25.81	9.79	25.08	9.77
		24.34	9.62	23.62	9.74	22.91	9.45	22.21	9.16
		21.55	9.22	20.89	8.63	20.24	8.22	19.63	8.35
		19.02	8.20	18.43	7.42	17.86	7.58	17.31	7.16
		16.76	6.88	16.24	6.87	15.74	6.90	15.24	6.53
		14.75	6.39	14.28	5.79	13.84	5.84	13.40	5.69
W30F	3	12.98	5.85	12.57	5.59	12.18	5.09	11.80	5.22
		11.43	4.38	11.06	4.67	10.71	4.33	10.38	4.46
		41.52	10.83	40.03	10.97	38.57	10.71	37.14	10.56
		35.72	10.59	34.30	10.46	32.90	10.05	31.51	10.09
		30.16	9.83	28.83	9.70	27.54	9.37	26.29	9.07
		25.07	8.94	23.89	8.74	22.74	8.38	21.62	8.17
		20.56	7.66	19.53	7.46	18.55	7.26	17.59	6.91
		16.68	6.59	15.80	6.24	14.97	5.98	14.17	5.82
		13.40	5.54	12.67	5.22	11.97	5.06	11.31	4.88
		10.67	4.57	10.06	4.50	9.48	4.15	8.92	3.97
		8.40	3.81	7.90	3.78				
		42.50	8.24	41.86	8.48	41.21	9.13	40.56	8.47
		39.86	9.09	39.16	9.23	38.47	9.03	37.80	8.91
		37.13	9.18	36.46	9.77	35.77	8.78	35.10	9.11
34.44	8.58	33.79	8.72	33.13	8.73	32.48	8.91		
31.83	8.77	31.18	8.75	30.55	8.59	29.93	8.00		
29.30	7.86	28.69	7.86	28.08	7.88	27.47	8.07		
26.89	7.93	26.31	7.45	25.73	7.64	25.17	7.29		
24.62	7.34	24.08	6.94	23.54	7.14	23.03	6.53		
22.53	6.53	22.02	6.86	21.53	6.43	21.05	6.31		
20.59	5.62	20.13	5.83	19.69	5.63	19.26	5.43		
18.84	5.72	18.43	5.37	18.03	5.24	17.64	5.41		
17.26	4.80	16.88	5.10	16.52	4.68	16.15	4.91		
15.79	4.87	15.44	4.85	15.08	4.86	14.74	4.24		
14.40	4.55								
D116I	1	31.62	9.07	30.81	8.76	29.98	9.76	29.16	9.39
		28.37	8.51	27.57	9.17	26.78	9.27	26.02	8.55
		25.27	8.02	24.54	8.44	23.76	9.03	23.01	8.80
		22.26	8.76	21.55	7.80	20.84	7.76	20.16	8.01
		19.43	8.08	18.70	8.45	17.96	8.24	17.23	8.34
		16.50	7.80	15.82	7.42	15.12	7.57	14.44	7.61
D116I	2	13.76	7.55	13.08	7.76	12.42	7.41	11.78	6.96
		55.35	8.85	54.64	8.42	53.95	9.00	53.22	8.01
		52.49	7.78	51.78	8.34	51.04	8.37	50.30	8.10
		49.58	8.42	48.81	7.92	48.05	8.41	47.30	8.31
		46.55	8.51	45.81	8.45	45.04	8.92	44.27	9.48
		43.51	9.21	42.78	9.08	42.01	8.58	41.27	8.14

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
D116I	3	40.52	8.73	39.78	8.64	39.07	8.51	38.34	7.82
		37.62	8.11	36.90	8.31	36.19	8.15	35.49	7.94
		34.77	8.69	34.08	7.96	33.40	7.53	32.72	7.79
		32.05	8.16	31.38	7.33	30.73	7.17	30.07	7.20
		29.43	7.63	28.79	6.86	28.16	7.06	27.55	7.47
		26.93	6.63	26.33	6.42	25.74	6.87	25.15	6.92
		24.58	6.83	24.00	6.15	23.44	6.34	22.88	6.16
		22.33	6.33	21.81	5.97	21.28	5.84	20.77	5.73
		20.25	5.71	19.76	5.57	19.27	5.35	18.79	5.58
		18.33	5.20	17.86	5.22	17.41	4.76	16.96	5.23
		16.53	4.61	16.10	4.90	15.69	4.78	15.28	4.43
		14.89	4.33	14.49	4.19	14.11	4.30	13.74	4.49
		13.39	4.23	13.02	4.02	12.68	3.75	12.33	3.70
		12.00	3.78	11.67	3.70				
		33.48	7.49	32.84	8.37	32.16	8.11	31.50	7.71
		30.82	7.97	30.15	8.29	29.49	7.06	28.83	7.23
		28.15	7.57	27.45	7.58	26.81	7.56	26.15	7.66
		25.48	7.79	24.81	7.82	24.14	7.70	23.50	6.79
		22.86	7.48	22.25	7.12	21.62	7.06	20.99	7.70
		20.36	7.77	19.72	7.10	19.12	7.00	18.48	8.04
17.87	6.61	17.25	7.04	16.64	7.03	16.05	6.81		
15.46	6.56	14.88	6.31	14.28	6.49	13.74	6.07		
13.17	6.08	12.64	6.10	12.11	5.63	11.59	5.97		
D116I	4	27.23	6.50	26.67	6.41	26.11	6.59	25.54	6.30
		24.96	6.72	24.36	6.44	23.77	6.89	23.18	6.50
		22.61	6.95	22.02	6.73	21.42	6.69	20.85	7.05
		20.23	6.87	19.66	7.21	19.05	7.27	18.43	6.58
		17.83	6.52	17.23	6.71	16.62	7.09	16.02	6.24
		15.43	6.90	14.83	7.18	14.24	7.10	13.65	6.89
		13.09	6.48	12.51	6.51	11.94	5.92		
M42Y	1	45.61	14.87	43.80	14.93	41.88	15.74	40.05	14.03
		38.22	14.88	36.35	15.03	34.48	14.85	32.68	14.37
		30.93	13.61	29.23	13.59	27.56	13.29	25.92	12.76
		24.33	12.30	22.82	11.92	21.37	11.40	19.99	10.61
		18.69	10.02	17.47	9.53	16.31	8.95	15.22	8.44
		14.18	7.90	13.22	7.56	12.31	6.97	11.45	6.59
		10.65	6.26	9.90	5.67	9.20	5.48	8.54	5.00
		7.94	4.73						
M42Y	2	45.52	16.09	43.60	15.92	41.65	16.62	39.70	15.89
		37.83	15.32	35.99	14.95	34.19	14.62	32.38	15.38
		30.55	15.06	28.75	14.52	27.01	14.35	25.30	14.18
		23.63	13.62	22.01	13.21	20.47	12.75	18.97	12.13
		17.54	11.48	16.17	11.05	14.87	10.48	13.64	10.11
		12.47	9.39	11.37	8.85	10.34	8.20	9.38	7.62
		8.49	7.34						
M42Y	3	45.84	18.14	43.69	17.24	41.63	16.68	39.49	18.17
		37.41	17.31	35.34	17.27	33.30	16.85	31.29	16.97
		29.26	16.74	27.30	16.34	25.33	16.03	23.44	15.04
		21.65	15.15	19.90	14.36	18.24	13.59	16.65	12.66
		15.15	11.99	13.76	11.32	12.44	10.60	11.19	10.08
		10.04	9.47	8.96	8.60	7.96	7.76	7.04	7.47

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
I41V	1	6.21	6.66						
		26.19	4.24	25.36	4.49	24.47	4.72	23.57	4.70
		22.68	4.42	21.81	4.67	20.93	4.75	20.03	4.65
		19.14	4.70	18.27	4.51	17.42	4.26	16.59	4.21
		15.77	4.40	14.94	4.31	14.12	4.24	13.31	4.28
		12.52	4.00	11.74	3.99	10.99	3.80	10.28	3.71
I41V	2	9.57	3.64	8.90	3.44	8.25	3.25	7.64	3.15
		46.98	4.26	46.19	3.71	45.42	4.23	44.60	4.24
		43.83	3.99	43.03	4.03	42.27	3.97	41.53	4.00
		40.78	3.92	40.04	3.83	39.30	3.93	38.58	3.37
		37.86	3.67	37.16	3.83	36.46	3.66	35.78	3.39
		35.10	3.64	34.45	3.33	33.79	3.40	33.15	3.25
		32.50	3.19	31.90	3.13	31.27	3.16	30.64	3.12
		30.04	3.25	29.44	3.06	28.87	2.70	28.35	2.74
		27.85	2.43	27.41	2.22	27.00	2.15	26.50	2.40
		26.03	2.52	25.56	2.25	25.06	1.90	24.64	2.22
		24.39	0.79	24.18	1.04	23.91	2.34	23.49	2.85
		22.99	2.64	22.49	2.31	22.04	2.15	21.61	2.01
		21.16	2.65	20.76	2.38	20.25	2.85	19.79	2.06
		19.36	2.26	18.95	2.16	18.59	1.60	18.27	1.79
17.92	1.79	17.57	1.73	17.25	1.73	16.93	1.60		
16.62	1.69	16.31	1.62	16.03	1.42	15.76	1.36		
15.49	1.33	15.23	1.32	14.98	1.23	14.74	1.33		
14.50	1.20	14.25	1.38	13.97	1.53	13.69	1.54		
13.40	1.45	13.12	1.38	12.87	1.26	12.63	1.18		
12.41	1.18	12.18	1.14						
I41V	3	31.48	7.49	30.09	7.17	28.62	7.92	27.01	8.92
		25.31	7.18	23.78	8.03	22.19	8.66	20.51	8.26
		18.88	8.08	17.33	8.35	15.70	7.83	14.14	7.69
		12.59	7.42	11.16	6.84	9.83	6.63	8.58	6.50
		7.36	5.89	6.26	5.48	5.25	4.83	4.36	4.23
L24V	1	31.95	11.46	30.57	11.68	29.16	11.92	27.81	11.31
		26.48	11.02	25.16	11.07	23.83	10.85	22.55	10.98
		21.20	11.19	19.85	11.30	18.56	10.43	17.32	10.28
		16.11	9.58	14.98	9.19	13.92	8.60	12.92	8.12
		11.97	7.72	11.06	7.51	10.18	7.01	9.33	7.00
		8.54	6.39	7.80	5.89	7.13	5.39	6.51	5.17
		5.93	4.65	5.39	4.31	4.89	3.86	4.44	3.55
		4.03	3.18	3.65	3.12	3.30	2.85		
L24V	2	42.22	10.54	40.98	10.63	39.77	10.16	38.57	10.75
		37.41	10.15	36.22	10.55	35.04	10.52	33.87	9.97
		32.69	10.27	31.52	10.31	30.39	9.97	29.28	9.34
		28.21	9.19	27.16	9.02	26.13	8.82	25.12	8.55
		24.14	8.33	23.19	7.97	22.28	7.92	21.37	7.64
		20.50	7.36	19.68	7.09	18.86	6.77	18.10	6.42
		17.35	6.08	16.63	6.17	15.94	5.80	15.28	5.64
		14.64	5.31	14.02	5.31	13.44	5.05		
L24V	3	43.61	11.14	42.29	10.93	40.94	11.62	39.56	11.40
		38.14	11.48	36.75	11.74	35.35	11.55	33.99	11.29
		32.62	11.30	31.29	11.18	29.96	10.83	28.68	10.51
		27.43	10.39	26.21	10.15	25.02	9.68	23.86	9.45

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
L24V	4	22.73	9.24	21.63	9.20	20.57	8.96	19.52	8.68
		18.52	8.26	17.55	8.29	16.60	7.74	15.69	7.50
		14.81	6.99	13.97	6.97	13.16	6.59	12.39	6.37
		11.66	6.16						
		39.44	13.09	37.87	13.25	36.28	13.69	34.64	13.72
		33.02	13.30	31.40	13.44	29.79	13.17	28.18	13.50
		26.57	13.23	24.96	13.43	23.40	13.20	21.85	12.68
		20.35	12.45	18.88	11.82	17.45	11.75	16.08	11.20
L62A	1	14.77	10.82	13.51	10.28	12.29	9.90	11.15	9.21
		25.08	3.34	24.57	3.27	24.05	3.37	23.52	3.47
		22.98	3.71	22.46	3.56	21.94	3.30	21.43	3.53
		20.92	3.15	20.43	3.25	19.93	3.34	19.43	3.21
		18.94	3.31	18.46	3.20	17.98	3.07	17.51	3.22
		17.04	3.08	16.58	3.13	16.10	3.01	15.66	2.97
		15.22	2.88	14.78	2.87	14.36	2.68	13.93	2.70
		13.53	2.66	13.13	2.78	12.73	2.46	12.34	2.62
		11.97	2.48	11.60	2.35	11.24	2.36	10.90	2.25
		10.56	2.17	10.23	2.07	9.91	2.11	9.60	1.98
L62A	2	9.30	1.95						
		47.98	2.00	47.69	2.04	47.43	1.92	47.15	1.89
		46.87	1.83	46.58	1.77	46.30	1.95	46.06	2.10
		45.76	2.25	45.47	1.85	45.17	1.94	44.91	1.60
		44.63	2.05	44.32	1.91	44.03	1.80	43.78	2.09
		43.53	1.84	43.24	1.84	42.97	1.84	42.70	1.46
		42.41	1.74	42.15	1.71	41.87	1.65	41.58	1.61
		41.31	1.89	41.05	2.02	40.76	1.72	40.49	1.68
		40.21	1.94	39.96	1.98	39.69	2.10	39.42	1.73
		39.15	1.39	38.91	1.82	38.63	1.56	38.40	1.72
		38.14	1.60	37.90	1.42	37.64	1.53	37.39	1.65
		37.14	1.54	36.91	1.45	36.65	1.66	36.41	1.62
		36.19	0.99	36.01	1.57	35.76	1.30	35.52	1.87
		35.27	1.62	35.08	1.12	34.84	1.18	34.63	1.26
		34.44	1.08	34.27	1.34	34.06	1.30	33.91	1.18
		33.74	1.26	33.59	1.14	33.39	1.06	33.21	1.19
		33.03	1.19	32.82	1.59	32.57	1.60	32.36	1.65
		32.09	1.67	31.84	1.52	31.59	1.46	31.35	1.62
		31.10	1.56	30.91	1.12	30.75	0.74	30.71	-0.42
		30.78	-0.15	30.81	0.16	30.71	0.74	30.58	1.23
		30.44	0.81	30.30	0.82	30.13	1.16	29.94	0.91
		29.80	1.14	29.63	1.25	29.44	1.21	29.24	1.19
		29.04	1.38	28.82	1.12	28.60	1.63	28.37	1.50
		28.15	1.56	27.95	1.38	27.74	1.27	27.54	1.23
		27.33	1.40	27.11	1.95	26.86	1.28	26.63	1.40
		26.46	1.52	26.22	1.67	26.00	1.57	25.81	1.37
		25.62	1.36	25.42	1.40	25.24	1.26	25.08	1.56
		24.90	1.02	24.74	0.87	24.62	0.76	24.46	0.98
24.31	0.68	24.20	0.89	24.02	1.37	23.87	0.32		
23.79	0.80	23.59	0.98	23.45	1.25	23.29	0.80		
23.19	0.55	23.06	0.47	22.96	1.39	22.87	1.70		
22.68	0.82	22.56	0.12	22.50	1.81	22.31	0.79		
22.21	1.07	22.10	0.36	22.02	0.85	21.90	0.39		

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
L62A	3	21.82	1.92	21.61	0.68	21.48	0.54	21.46	1.85
		21.30	-0.32	21.23	0.51	21.04	1.43	20.95	-0.28
		20.85	0.26	20.71	0.78	20.68	0.83	20.55	1.03
		20.37	1.08	20.26	0.94	20.13	1.43	19.96	0.57
		19.84	0.59	19.71	1.14	19.60	0.74	19.47	0.52
		19.37	1.10	19.26	0.34	19.13	0.93	19.02	0.90
		18.90	0.93	18.79	1.03	18.71	0.66	18.58	0.72
		18.46	0.80	18.36	0.68	18.23	0.74	18.10	0.65
		17.99	0.77	17.87	0.80	17.76	0.83	17.63	0.71
		17.53	0.80	17.42	0.67	17.30	0.78	17.20	0.86
		17.08	0.70	16.98	0.80	16.87	0.57	16.76	0.50
		16.66	0.70	16.54	0.77	16.43	0.69	16.32	0.66
		16.22	0.77	16.12	0.64	16.01	0.62	15.92	0.71
		15.82	0.58	15.72	0.40	15.62	0.69	15.53	0.60
		15.42	0.62	15.34	0.56	15.24	0.65	15.15	0.49
		15.05	0.53	14.96	0.67	14.87	0.61	14.79	0.45
		14.71	0.60	14.61	0.69	14.55	0.50	14.46	0.56
		14.36	0.75	14.28	0.57	14.18	0.57	14.10	0.58
		14.01	0.44	13.93	0.70	13.84	0.73	13.75	0.72
		13.65	0.48	13.56	0.58	13.46	0.48	13.37	0.66
		13.28	0.50	13.19	0.55	13.11	0.50	13.02	0.56
		12.93	0.50	12.86	0.52	12.77	0.72	12.70	0.51
		12.62	0.49	12.55	0.59	12.47	0.53		
		40.03	3.67	39.50	3.59	38.94	3.57	38.41	3.60
		37.85	3.60	37.30	3.67	36.73	3.75	36.18	3.88
		35.63	3.69	35.07	3.68	34.53	3.71	33.98	3.49
		33.43	3.69	32.89	3.50	32.35	3.56	31.82	3.48
		31.30	3.45	30.78	3.38	30.27	3.34	29.75	3.34
		29.26	3.19	28.76	3.14	28.28	3.06	27.80	3.08
		27.33	3.02	26.87	3.08	26.41	2.89	25.96	2.85
		25.52	2.86	25.09	2.89	24.66	2.80	24.24	2.72
		23.82	2.71	23.41	2.71	23.00	2.66	22.61	2.60
		22.22	2.54	21.83	2.48	21.45	2.34	21.09	2.36
20.72	2.47	20.36	2.32	20.01	2.28	19.67	2.17		
19.35	2.13	19.04	1.95	18.74	1.91	18.45	1.86		
18.17	1.86	17.89	1.74	17.62	1.75	17.37	1.67		
17.11	1.74	16.86	1.71	16.60	1.76	16.35	1.65		
16.09	1.69	15.83	1.62	15.58	1.74	15.32	1.66		
15.08	1.69	14.83	1.62	14.59	1.59	14.36	1.52		
14.14	1.51	13.92	1.48	13.70	1.47				
W47V	1	26.81	3.24	26.47	3.27	26.12	3.14	25.77	3.15
		25.42	3.20	25.06	3.45	24.71	3.45	24.35	3.35
		23.99	3.27	23.63	3.45	23.27	3.38	22.90	3.34
		22.55	3.33	22.18	3.34	21.83	3.13	21.47	3.31
		21.12	3.40	20.77	3.33	20.42	3.38	20.08	3.07
		19.73	3.26	19.39	3.11	19.06	3.20	18.73	3.10
		18.40	3.10	18.08	2.98	17.77	2.90	17.45	3.00
		17.14	2.81	16.83	2.76	16.52	2.75	16.23	2.66
		15.94	2.75	15.65	2.57	15.36	2.52	15.09	2.45
		14.82	2.39	14.55	2.50	14.28	2.41	14.03	2.33
		13.77	2.37	13.52	2.27	13.27	2.36	13.03	2.29

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
W47V	2	12.79	2.30	12.55	2.25	12.31	2.15	12.08	2.05
		11.84	2.29	11.62	2.01	11.38	2.17	11.15	2.12
		10.93	2.06	10.71	1.92	10.49	2.08	10.27	2.03
		10.05	2.07	9.83	1.92	9.61	2.07	9.41	2.00
		48.67	3.50	47.98	3.33	47.29	3.49	46.60	3.39
		45.90	3.66	45.19	3.58	44.50	3.29	43.81	3.35
		43.11	3.36	42.43	3.31	41.77	3.23	41.11	3.13
		40.47	3.05	39.83	3.25	39.18	3.04	38.57	3.03
		37.96	2.77	37.39	2.92	36.80	2.80	36.24	2.87
		35.66	2.83	35.10	2.76	34.56	2.64	34.01	2.68
		33.46	2.80	32.90	2.84	32.33	2.78	31.77	2.77
		31.22	2.72	30.66	2.67	30.12	2.68	29.58	2.77
		29.06	2.63	28.55	2.55	28.04	2.49	27.55	2.45
		27.06	2.35	26.59	2.22	26.12	2.25	25.74	1.59
		25.35	2.31	24.88	2.22	24.40	2.32	23.96	2.11
		23.54	1.98	23.17	1.65	22.88	1.10	22.70	0.92
		22.48	1.19	22.23	1.27	21.97	1.36	21.69	1.51
		21.39	1.64	21.06	1.64	20.73	1.63	20.41	1.48
		20.09	1.58	19.78	1.56	19.47	1.57	19.16	1.53
		18.86	1.46	18.56	1.47	18.27	1.45	17.99	1.35
		17.72	1.27	17.46	1.23	17.21	1.26	16.96	1.20
		16.72	1.16	16.48	1.16	16.25	1.14	16.01	1.25
		15.77	1.20	15.52	1.17	15.28	1.13	15.04	1.14
		14.80	1.12	14.57	1.14	14.35	1.11	14.12	1.15
		13.88	1.12	13.66	1.13	13.45	1.00	13.23	1.04
		13.01	1.00	12.81	0.83	12.62	0.92	12.44	0.90
12.26	0.81	12.07	0.90	11.90	0.88	11.70	0.95		
11.54	0.78	11.37	0.89	11.17	1.05	11.01	0.92		
10.83	0.62	10.64	0.86	10.48	0.66	10.33	0.52		
10.18	0.88	10.00	0.63	9.88	0.68				
W47V	3	39.87	3.97	39.48	3.85	39.08	4.06	38.68	3.85
		38.27	4.00	37.87	3.78	37.45	4.31	37.07	4.02
		36.67	3.91	36.28	3.94	35.90	3.86	35.52	2.97
		35.17	3.65	34.79	4.18	34.38	3.53	34.01	3.73
		33.67	3.17	33.31	3.58	32.95	3.79	32.56	3.91
		32.19	3.60	31.84	3.53	31.48	3.47	31.13	3.48
		30.80	3.24	30.46	3.28	30.13	3.21	29.80	3.27
		29.47	3.00	29.15	3.26	28.84	2.91	28.53	3.09
		28.23	2.97	27.92	3.08	27.62	2.93	27.32	2.99
		27.03	3.02	26.74	2.73	26.45	2.91	26.16	2.97
		25.87	2.89	25.58	2.82	25.29	2.66	25.01	2.74
		24.72	2.75	24.44	2.88	24.16	2.75	23.88	2.78
		23.59	2.74	23.31	2.86	23.03	2.63	22.76	2.69
		22.48	2.44	22.22	2.55	21.94	2.56	21.67	2.62
		21.41	2.67	21.14	2.52	20.90	2.58	20.64	2.42
		20.40	2.30	20.18	2.07	19.97	1.99	19.78	1.78
		19.62	1.37	19.49	1.27	19.37	1.28	19.23	1.44
		19.06	1.86	18.86	2.24	18.62	2.38	18.37	2.39
		18.12	2.41	17.89	2.04	17.69	1.68	17.53	1.42
		17.39	1.53	17.23	1.62	17.05	1.76	16.87	1.70
		16.69	1.70	16.52	1.61	16.37	1.43	16.21	1.56

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
W47L	1	16.05	1.59	15.89	1.63	15.74	1.59	15.57	1.64
		15.41	1.56	15.25	1.55				
		45.09	2.65	44.57	2.45	44.06	2.53	43.55	2.30
		43.05	2.14	42.55	2.07	42.05	1.74	41.52	2.25
		41.00	2.36	40.43	2.37	39.82	2.13	39.31	2.14
		38.78	2.24	38.24	2.44	37.67	2.34	37.14	2.12
		36.58	2.26	36.05	2.28	35.50	2.13	34.96	2.25
		34.42	2.35	33.85	2.30	33.31	2.34	32.76	2.31
		32.20	2.41	31.64	2.43	31.08	2.27	30.51	2.49
		29.95	2.37	29.37	2.49	28.84	2.32	28.27	2.21
		27.72	2.35	27.16	2.21	26.61	2.29	26.08	2.24
		25.53	2.20	25.01	2.25	24.48	2.14	23.96	2.15
		23.45	2.27	22.96	2.10	22.45	1.99	21.96	2.09
		21.48	1.97	21.01	1.92	20.55	1.88	20.09	1.92
		19.63	1.83	19.20	1.84	18.75	1.71	18.33	1.76
		17.91	1.82	17.51	1.73	17.10	1.73	16.69	1.50
		16.31	1.67	15.93	1.53	15.55	1.59	15.18	1.62
		14.83	1.57	14.48	1.48	14.13	1.50	13.80	1.29
		13.47	1.33	13.15	1.27	12.82	1.29	12.51	1.36
		12.21	1.29	11.91	1.27	11.62	1.17	11.34	1.22
W47L	2	11.06	1.17						
		44.80	2.42	44.20	2.84	43.61	2.33	43.06	2.24
		42.57	2.10	42.05	2.44	41.51	2.32	40.99	2.29
		40.46	2.06	39.95	2.14	39.44	2.07	38.94	2.30
		38.43	2.20	37.93	2.32	37.38	2.15	36.84	2.18
		36.33	2.28	35.76	2.23	35.24	2.26	34.73	2.29
		34.20	2.39	33.68	2.19	33.16	2.14	32.68	2.11
		32.15	2.19	31.62	2.20	31.11	2.05	30.60	2.04
		30.09	2.36	29.58	2.19	29.07	2.00	28.58	2.06
		28.07	2.07	27.60	2.00	27.11	2.12	26.63	2.02
		26.16	2.00	25.68	2.07	25.22	1.96	24.77	1.70
		24.31	1.89	23.86	1.97	23.42	1.89	22.98	1.84
		22.56	1.77	22.12	1.81	21.72	1.79	21.30	1.67
		20.88	1.65	20.49	1.56	20.10	1.62	19.70	1.49
		19.32	1.72	18.94	1.64	18.58	1.72	18.21	1.47
		17.84	1.46	17.49	1.41	17.14	1.53	16.81	1.45
		16.47	1.35	16.15	1.41	15.82	1.38	15.50	1.28
		15.17	1.36	14.87	1.24	14.57	1.23	14.27	1.26
		13.99	1.27	13.71	1.13	13.43	1.16	13.14	1.18
		12.86	1.13	12.61	1.16	12.34	1.12	12.09	1.04
11.84	1.10	11.59	0.97	11.35	1.02	11.11	0.96		
W47L	3	10.88	1.06						
		25.50	2.35	24.86	2.33	24.19	2.48	23.53	2.36
		22.91	2.13	22.30	2.36	21.63	2.41	20.98	2.47
		20.32	2.42	19.66	2.50	18.95	2.79	18.19	2.85
		17.44	2.93	16.67	2.86	15.96	2.54	15.24	2.55
		14.52	2.78	13.78	2.57	13.10	2.35	12.44	2.29
		11.80	2.45	11.18	2.29	10.56	2.20	9.97	2.10
		9.40	2.02	8.85	2.03	8.31	1.89	7.80	1.83
7.31	1.76	6.83	1.66						
W47L	4	27.30	3.28	26.41	3.26	25.51	3.32	24.61	3.32

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
W47L	5	23.72	3.19	22.87	3.13	22.06	2.89	21.20	3.37
		20.29	3.40	19.37	3.41	18.44	3.57	17.48	3.59
		16.53	3.39	15.63	3.22	14.77	3.22	13.94	3.14
		13.11	3.04	12.31	2.88	11.56	2.72	10.83	2.68
		10.14	2.53	9.47	2.49	8.84	2.24	8.24	2.18
		7.67	2.06						
		36.43	3.85	34.71	3.81	33.01	3.88	31.23	3.71
		29.42	4.12	27.67	3.76	26.03	3.64	24.43	3.46
		22.85	3.26	21.42	3.08	20.09	2.80	18.87	2.77
		17.69	2.59	16.53	2.35	15.48	2.42	14.47	2.13
H114V	1	13.49	1.96	12.60	2.00	11.74	1.82	10.93	1.83
		10.18	1.59	9.44	1.44	8.80	1.51	8.15	1.34
		27.35	5.75	26.59	6.03	25.83	6.10	24.97	6.78
		24.11	6.36	23.35	5.92	22.59	5.87	21.86	5.76
		21.11	6.09	20.37	5.69	19.62	6.13	18.81	6.23
		18.00	6.31	17.19	6.26	16.37	6.16	15.57	6.13
		14.77	6.29	13.98	5.97	13.21	5.71	12.47	5.68
		11.73	5.90	10.97	5.92	10.25	5.21	9.56	5.26
		8.89	5.02	8.25	4.92	7.64	4.69	7.03	4.82
		6.47	4.52	5.92	4.13				
H114V	2	38.79	8.33	36.92	8.40	35.11	8.27	33.32	8.34
		31.58	7.86	29.88	7.94	28.17	7.78	26.54	7.65
		24.94	7.49	23.35	7.30	21.82	7.01	20.33	6.73
		18.90	6.61	17.52	6.44	16.17	6.20	14.87	5.99
		13.62	5.71	12.44	5.23	11.32	4.98	10.27	4.81
H114V	3	47.55	6.22	46.78	6.00	46.01	6.16	45.24	6.31
		44.46	6.27	43.66	6.05	42.88	5.84	42.08	6.17
		41.29	6.13	40.52	5.93	39.73	5.91	38.96	6.07
		38.18	6.00	37.39	6.10	36.62	5.94	35.86	5.77
		35.11	5.92	34.36	5.44	33.62	5.53	32.90	5.81
		32.18	5.51	31.46	5.57	30.77	5.31	30.06	5.25
		29.38	5.09	28.72	5.14	28.06	5.09	27.42	4.86
		26.78	5.06	26.15	4.72	25.54	4.75	24.93	4.52
		24.34	4.49	23.77	4.17	23.21	4.30	22.64	4.31
		22.09	3.88	21.56	4.00	21.03	4.04	20.52	3.98
		20.01	3.76	19.53	3.86	19.05	3.94	18.58	3.54
		18.13	3.58	17.68	3.35	17.24	3.16	16.81	3.20
		16.40	3.20	15.99	2.99	15.59	3.13	15.21	3.16
		14.83	2.89	14.46	2.91	14.09	2.88	13.74	2.70
		13.40	2.62	13.06	2.50	12.73	2.59	12.41	2.39
H114V	4	12.10	2.35	11.80	2.38				
		33.46	6.62	32.64	6.28	31.81	6.38	30.97	6.54
		30.14	6.75	29.31	6.71	28.45	6.66	27.63	6.26
		26.78	6.17	25.95	6.23	25.14	6.60	24.32	5.89
		23.51	6.22	22.70	6.12	21.88	5.96	21.10	5.87
		20.30	6.24	19.53	6.06	18.77	5.86	18.01	5.69
		17.27	5.85	16.54	5.40	15.83	5.61	15.11	5.19
		14.43	5.49	13.76	5.22	13.10	4.96	12.46	4.91
		11.83	4.98	11.23	4.48	10.63	4.62	10.06	4.33
		9.50	4.16	8.96	4.10	8.43	4.00		
WT	1	32.27	5.25	31.21	5.61	30.12	5.89	29.01	5.73

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
WT	2	27.91	5.65	26.85	5.57	25.77	5.54	24.71	5.66
		23.63	5.60	22.58	5.38	21.56	5.36	20.57	5.16
		19.59	5.10	18.64	4.94	17.67	4.95	16.74	4.78
		15.84	4.69	14.94	4.56	14.07	4.52	13.22	4.43
		12.39	4.19	11.59	4.09	10.84	3.95	10.10	3.73
		9.34	3.70	8.65	3.19	7.98	3.75	7.36	3.32
		24.05	4.55	23.20	4.16	22.35	4.30	21.50	4.15
		20.66	4.08	19.84	4.22	19.02	4.15	18.19	4.26
		17.38	4.20	16.61	3.90	15.85	3.92	15.13	3.62
		14.42	3.56	13.74	3.57	13.10	3.31	12.47	3.24
		11.76	3.12	11.17	2.52	10.60	2.96	10.01	2.99
		9.47	2.80	8.95	2.45	8.43	2.58	7.97	2.15
		7.53	2.10	7.06	1.91	6.64	2.44	6.25	1.51
		5.89	1.74	5.50	1.75	5.18	1.69	4.89	0.96
WT	3	4.58	1.62	4.30	1.02	4.03	1.12		
		23.81	5.99	22.63	6.00	21.42	5.96	20.24	5.99
		19.06	6.05	17.90	5.73	16.76	5.80	15.66	5.49
		14.57	5.59	13.50	5.30	12.46	5.15	11.47	4.90
		10.51	4.78	9.57	4.62	8.68	4.50	7.84	4.22
		7.04	3.88	6.29	3.66	5.58	3.44	4.94	3.24
		4.34	2.91	3.79	2.61	3.30	2.34	2.87	2.09
		2.48	1.87	2.14	1.54	1.85	1.40	1.60	1.18
D116M	1	26.11	8.65	25.36	8.01	24.63	8.04	23.90	8.56
		23.16	8.89	22.42	8.96	21.70	8.34	20.98	8.00
		20.28	7.57	19.58	8.23	18.89	7.75	18.19	7.80
		17.50	7.67	16.79	7.89	16.07	8.15	15.36	7.96
		14.65	7.82	13.94	7.68	13.25	7.75	12.58	7.76
		11.92	7.65	11.26	7.51	10.60	7.96		
D116M	2	39.59	9.43	38.82	8.98	38.04	9.21	37.25	9.23
		36.45	8.93	35.66	9.02	34.91	9.34	34.11	8.77
		33.36	8.83	32.57	8.59	31.76	8.41	30.98	8.90
		30.16	8.94	29.36	8.48	28.55	8.86	27.77	9.44
		26.97	8.62	26.14	9.33	25.35	9.11	24.53	9.31
		23.75	9.45	22.93	8.42	22.15	9.15	21.35	8.86
		20.56	9.28	19.76	9.08	18.97	9.25	18.21	8.50
		17.48	8.22	16.72	8.36	15.99	8.42	15.28	8.53
		14.57	8.01	13.89	7.97	13.20	7.81	12.54	7.45
		11.91	6.95	11.27	6.96	10.65	7.00	10.05	7.04
D116M	3	50.24	7.07	49.64	7.73	49.01	7.18	48.37	7.29
		47.74	7.85	47.08	7.47	46.37	7.71	45.67	8.70
		44.97	7.41	44.24	8.35	43.56	8.18	42.87	8.42
		42.18	7.52	41.50	7.82	40.83	7.54	40.14	7.74
		39.46	8.25	38.78	7.80	38.12	7.99	37.45	7.40
		36.79	7.45	36.15	8.02	35.48	7.24	34.81	7.72
		34.17	7.61	33.49	7.17	32.86	7.25	32.24	7.86
		31.60	6.92	30.94	7.11	30.32	6.84	29.70	6.60
		29.10	6.96	28.47	7.24	27.89	6.95	27.28	6.99
		26.71	6.71	26.13	6.65	25.54	6.52	24.98	6.47
		24.41	6.33	23.86	6.32	23.30	6.79	22.74	6.25
		22.22	6.03	21.68	5.84	21.14	5.59	20.65	6.12
20.13	5.71	19.63	5.73	19.14	5.64	18.66	5.36		

(Continued on next page)

variant	run	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})	[DHF] (μM)	V (s^{-1})
D116M	4	18.19	5.06	17.71	5.25	17.27	4.94	16.80	5.06
		16.37	5.02	15.93	5.23	15.51	4.85	15.08	4.74
		14.68	4.95	14.28	4.52	13.88	3.99	13.52	4.20
		13.14	4.08	12.78	4.14	12.42	3.79	12.06	3.87
		11.73	4.03	11.39	3.85	11.06	3.74	10.74	3.67
		33.11	7.80	32.47	7.88	31.81	7.76	31.15	7.62
		30.49	7.44	29.86	7.58	29.16	7.60	28.50	7.69
		27.80	7.97	27.14	7.78	26.49	7.63	25.81	7.91
		25.17	7.11	24.48	7.33	23.80	7.91	23.15	7.24
		22.49	7.59	21.83	7.46	21.17	7.11	20.53	7.52
		19.89	7.18	19.26	7.39	18.63	7.04	18.02	6.84
		17.40	6.75	16.81	7.00	16.21	6.61	15.64	6.80
		15.06	6.60	14.48	6.19	13.92	6.37	13.38	6.11
		12.84	6.03	12.31	6.46	11.80	5.89	11.28	5.60
W30M	1	10.78	6.05	10.29	5.57				
		36.93	10.18	35.60	10.28	32.04	10.18	30.67	10.73
		29.35	9.47	28.17	8.69	27.07	8.44	25.99	8.08
		24.92	7.97	23.89	8.09	22.87	7.89	21.83	7.67
		20.83	7.78	19.84	7.36	18.90	7.18	17.98	6.71
		17.10	6.69	16.26	6.32	15.44	6.31	14.67	5.98
		13.93	5.72	13.22	5.24	12.55	5.04	11.92	4.58
		11.33	4.27	10.77	4.10	10.24	3.87	9.73	3.69
		9.25	3.61	8.81	3.29	8.38	3.13	7.98	2.91
		7.60	2.74	7.25	2.60	6.91	2.51		
W30M	2	40.40	10.45	39.05	10.31	37.71	10.51	36.37	10.18
		35.05	10.22	33.72	10.22	32.39	10.27	31.08	10.04
		29.79	9.89	28.53	9.68	27.28	9.60	26.06	9.09
		24.87	9.11	23.70	8.89	22.56	8.47	21.47	8.20
		20.39	7.96	19.37	8.10	18.36	7.59	17.39	7.28
		16.46	7.03	15.57	6.82	14.71	6.57	13.89	6.29
		13.08	5.93	12.33	5.72	11.60	5.43	10.91	5.13

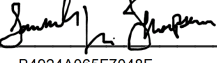
References

- [1] K. A. Reynolds, R. N. McLaughlin, and R. Ranganathan. Hot spots for allosteric regulation on protein surfaces. *Cell*, 147(7):1564–75, 2011.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

B4924A065F7048F... Author Signature

6/9/2020
Date