Stability analysis of the complete single mutant library of a protein domain

Chapter 4

Adapted from a manuscript coauthored with Ernest Lee and Stephen L. Mayo.

Abstract

Proper understanding and prediction of the fitness consequences upon sequence mutation is an outstanding challenge in protein engineering. Although mutational data traditionally has been difficult and expensive to acquire, recent advances in laboratory automation have enabled the thermodynamic evaluation of almost every single mutant in a small 56-residue protein. With a domain-level perspective, we explore mutational outcomes, distributions, positional sensitivity, and mutant amino acid tolerance. Benchmarking current stability prediction methods reveals unbalanced performance across different structural criteria, but a uniform capability to capture the trends of our unbiased dataset. The surprising neutrality of single mutations to the domain contrasts with the stark negative epistasis seen in small, heavily mutated datasets, especially in variants comprised solely of individually stabilizing mutations. The poor predictability of multiple mutants from single mutations indicates that the field must move beyond single global stability measurements in order to truly comprehend the mutational effects of proteins.

Introduction

Protein mutagenesis data has long provided insights into the forces responsible for protein stability and folding (1–5). The protein-engineering field flourished from the ability to test hypotheses by comparing the thermodynamic effect of single- and multiple-amino acid mutations against a wild-type sequence. Abstraction of these concepts into computationally tractable algorithms have pushed the field even further, allowing users to score near or distant mutant sequences (6–9). These tools have also made feasible the computational probing of the mechanisms surrounding domain mutational tolerance and evolvability (10–13).

Despite this success, the principles of stability engineering describe only the general trend of the effects of amino acid mutations. Results are often mixed when applied to any specific problem due to the number of possible secondary and tertiary environments found in proteins (1, 4). A typical solution to this issue of proper context is the acquisition of more and more data in your protein of interest. This was a daunting task in the past, as the efforts required to engineer and purify protein mutants scaled linearly with the number to be made. Recently, technological advancements in laboratory automation and next-generation sequencing (14–16) have lowered this barrier so that the construction effort is identical for orders of magnitude more variants. Here, we analyze thermodynamic data from almost every single-mutant of an entire protein domain obtained through a previous effort in laboratory automation. As experimental data of this magnitude will only become more common, we examine the general utility of large datasets, and explore the performance of today's scoring algorithms.

Rare in the literature due to its laborious nature, domain-level mutagenesis data can provide valuable insight on mutational distributions and average positional and amino acid effects in proteins. Early work on mutational effects was performed on globular proteins such as the globins, lac repressor, lysozyme, staphylococcal nuclease, and barnase. Numbers on the fraction of mutations experimentally found stabilizing or destabilizing varied with each report, suggesting that the distribution of stability effects was unique to each protein. Although a convincing universal distribution for globular proteins was recently reported, most of the evidence was computationally derived (17). Conclusions on positional and amino acid effects from the pioneering studies centered on the importance of core and surface patterning of polar and nonpolar amino acids and secondary structure propensities. But what dominates or decides the sensitivity of a position? And which amino acid is best tolerated by the protein of interest? Our domain-level perspective of single mutant data simultaneously verifies the nature of mutational distributions and illuminates topics on positional sensitivity and amino acid tolerance.

Structure-based stability prediction algorithms translate our general knowledge of non-covalent protein interactions into a context-sensitive output. The degree to which they succeed is entirely dependent upon the constitution of the test dataset. A long necessary independent analysis of prediction algorithms concluded that all of the tested methods did not perform as well as previously reported and frequently failed to capture details (18). The curious fact that the algorithms had previous success in predicting mutational effects was reconciled with their ability to recapitulate general trends in the independent test set. Unfortunately, the test dataset used by Potapov and colleagues was biased towards large-to-small amino acid mutations, a feature very common to datasets

extracted from the popular online aggregate database of stability data, Protherm (19). As most of the current algorithms were trained on datasets sampled from Protherm, their true capabilities may be more underwhelming than reported. Our comprehensive single-mutant library provides a truly unbiased dataset upon which to test the performance of current and future prediction algorithms.

Experimental data on the stability landscape of a protein elicits inquiry into additivity effects and the domain's mutational robustness. Recent work on these topics has put forth that proteins thermodynamically stabilized from wild type or under weak functional selection exhibit a threshold against deleterious mutations that, once exhausted, declines rapidly (11, 12). The larger than expected effect of detrimental mutations after the threshold defines the system as negatively epistatic. We weigh in on this theory with real datasets and, alternatively, determine what effects the combination of multiple *beneficial* mutations might bestow upon a protein. The literature contends that all manner of simple and complex effects can exist under this scenario (20). Does the knowledge of every favorable single mutation enable the engineering of hyper-stabilized proteins?

Using a streamlined laboratory automation method, we constructed 935 single-mutants of the small monomeric domain Gβ1, multiplying 55 of 56 positions by 17 of 19 possible mutant amino acids. Variants were purified and assayed for thermodynamic stability by observing tryptophan fluorescence in response to chemical denaturation. Rather than engage in the details of individual mutations, we chose to explore an array of topics important to protein engineering. We experimentally verify previous conceptions about the distribution of mutational effects for an entire protein domain, as well as

describe novel trends previously unexplored due to the lack of data. The unbiased nature of the dataset provided us a terrific avenue upon which to test popular stability prediction algorithms, as well as to guide the mutagenesis of multi-point mutants aimed to explore the epistatic effects of favorable mutations.

Results and discussion

The mutational distribution of Gβ1

The overall distribution of ΔΔG effects in the Gβ1 domain is shown in Figure 4-1. As a significant fraction of mutations lead to severely destabilized or insoluble proteins, it is not a normal distribution. Thus, we report an overall median ΔΔG of -0.25 kcal/mol with an interquartile range of 1.88 kcal/mol. If neutral mutations are those with a ΔΔG of ±0.5 kcal/mol then the fractions of positive, neutral, and negative mutations are 16%, 41%, and 43%, respectively. While technically these values indicate that most mutations are destabilizing, 57% of mutations (positive + neutral) would have at worst almost no effect on protein stability. Roughly 12% of mutations tested could not be accurately measured by our stability assay ("unf" in the mutational distribution) and likely represent evolutionary dead ends.

This data illustrates that across the entire domain, the native sequence is robust to point mutations. This topic has seen much debate in the literature with evidence in favor for and against mutational robustness. Discrepancies likely arise (beyond those due to measuring different proteins) from varying stringencies in functional selection and by incorporating only the residues allowed through amber codon suppression. Our stability data likely represents the upper limits of robustness, as layering an activity requirement

upon sequences will lower the yield of neutral and positive mutational outcomes. Omitting unbiased tryptophan and cysteine incorporation, although necessary for the fidelity of the stability assay, also likely skews the fractional outcomes toward mutational tolerance. Finally, because the Gβ1 domain is small (56 residues), its ratio of surface-to-buried positions fosters a tolerant distribution that may not be the case for larger proteins with larger cores. Previous measures of a mutation's functional inactivation probability across a protein domain are divergent, with values from 5% (21) to 34% (10). Again, differences in protein identity, selection stringency, and incorporated residues are likely the answer to these inconsistencies. In Gβ1's case, having 88% of single mutations available for mutation presents an enormous amount of "safe" potential evolutionary trajectories for stabilization. This finding is somewhat surprising considering the relatively low thermodynamic threshold of Gβ1 (~ 4–5 kcal/mol) in comparison to other proteins (up to 15 kcal/mol). Of course, the interactions between subsequent mutations will ultimately decide the feasibility of any particular path, and will likely shrink the potential complexity (22).

Although it appears intuitive that the mutational distributions of any two unrelated monomeric proteins will differ, this was found not to be the case in a recent computational study (17). Using the FoldX algorithm, the authors computed mutational distributions for a large panel of proteins and show that they all follow a similar asymmetric distribution. While future experimental efforts will be able to support or refute the full finding, we can currently verify the presence of the universal distribution in our dataset and confirm the FoldX algorithm's suitability for this type of study. We fitted our dataset (excluding those mutations labeled unfolded) to the bi-Gaussian and

individual core/surface distributions described in (17). In addition, we compare the derived means and variances to those from an identical dataset produced by FoldX3 (Table 4-1). Both sets of data fit the Gaussian equations well and exhibit the overall universal trend of sharper, stabilizing surface distributions with wider, destabilizing core distributions (Figure 4-2). In fact, the exquisite agreements between the calculated and experimental standard deviations provide noteworthy support for a Gaussian description of the surface and core mutational distributions. A two-population t-test to determine the similarity between the calculated and experimental individual Gaussian means finds the core distributions statistically identical ($\alpha = 0.025$, p = 0.30) and the surface distributions different ($\alpha = 0.025$, p = $9.2x10^{-13}$). This discrepancy can be explained by the historically poor treatment of surface energetics in stability algorithms along with small margin of error due to the tighter distribution. As shown in the literature (17, 18, 23) and later in our analysis, it appears that algorithms like FoldX3 can fail to capture specific details but still produce the correct experimental trend.

Positional sensitivity and mutant amino acid tolerance

Efficient visualization of each individual mutant in the library is accomplished through the use of a heat map (Figure 4-3). This perspective makes it abundantly clear that sequence position, not identity of the incorporated mutant amino acid, dominates mutational effects. This phenomenon is supported by the mutational sensitivity of wild-type non-polar amino acids that contribute to the hydrophobic core of the protein. Box plots of the stability effects separated by RESCLASS (Figure 4-4), an algorithm that uses the geometry from a crystal structure to designate core, boundary, or surface positions,

illustrates this behavior extremely well. The 75th percentile of the core mutation distribution sits below the inter-quartile range of the other categories, illustrating the intolerance of core positions to mutation. Only amino acid mutations to proline or glycine serve as exceptions to this result as they are generally deleterious regardless of the position due to their unique phi psi distributions.

Although we know that random core mutations are deleterious to protein stability, what quantitatively determines positional sensitivity? We approached this question by using supervised classification on a large number of attributes to train a linear regression model to predict the average $\Delta\Delta G$ of each position in the Gβ1 domain. The best model gives a correlation coefficient of 0.83 and includes weights from a measure of the hydrophobicity of the wild-type amino acid and RESCLASS categories. However, the major contributor to the model was occluded surface packing value (OSP), which alone gives a correlation coefficient of 0.78 (Figure 4-5). As a metric for protein packing, OSP is routinely used to analyze structural datasets and protein folding predictions (24). That its found to be the chief determinant of domain-wide positional sensitivity is not unrealistic, as the result extends previous work on core mutations (25, 26) and makes intuitive sense: heavily occluded amino acids would be less likely to accommodate disruptions in their packing environments. The major outliers to this correlation are Gly41 and Tyr45, whose average $\Delta\Delta G$ values are greatly destabilized in comparison to their OSP. Position 41's sensitivity can be explained by its proximity to Trp43, the reporter amino acid for the stability assay, which will severely limit the allowed mutation types (small or flexible). And while position 45 is located on an outer beta strand, it is only one of two hydrophobic amino acids responsible (the other being Trp43) for

shielding an edge of the protein's core from solvent. Overall, OSP does a superb job in identifying the most sensitive positions to mutation and should be a part of the protein engineer's toolbox.

Complementary to the analysis on positional sensitivity is determining which amino acid scan of the domain best captures the average ΔΔG for each position. Actual experimental data on the system can help alleviate the issues with complex environments that protein packing alone predicted poorly. Whereas alanine mutagenesis is most often used to derive functional hotspots, it is unclear which amino acid can best forecast overall destabilizing, neutral, or stabilizing sites. The result of both a ΔΔG deviation method and a linear ranking method (see Methods) show serine as the highest-ranking amino acid, with a mixture of methionine, threonine, and glutamine rounding out the top four (Figure 4-6). All of these amino acids are non-charged, polar, and fairly amphiphilic in nature, making them reasonable choices for an amino acid stability scan.

If the structure of the protein of interest is available, then adjusting the scan by RESCLASS would likely lead to higher prediction accuracy. In our study, methionine and alanine dominated the core rankings, while threonine, serine, and glutamine topped the rankings for boundary and surface positions. These results again show a preference for uncharged amphiphilic amino acids, along with moderate "like dissolves like" tendencies for core and boundary/surface predictions. Alanine ends up performing respectably well across the core and boundary segments of the protein, although it is a decidedly poor indicator of positional sensitivity on the surface, where the majority of other amino acids outcompete it. In total, the two deviation calculations give similar

results, unifying and strengthening the evidence for small amphiphillic residues such as serine as the first choice for stability scanning mutagenesis.

Instead of averaging by position, we now average by mutant amino acid and ask, which amino acid is best or least tolerated by the domain, and why? Because reasonable solutions to this query require a comprehensive number of mutations per position, our dataset is uniquely positioned to explore this topic. The worst amino acid for general incorporation is proline, followed by glycine (Figure 4-7). This isn't surprising, as the special amino acids are well known to be debilitating to protein stability. Aspartic acid is the third worst incorporated amino acid, most likely due to its highly acidic nature. It also contains the smallest amount of nonpolar atoms in comparison to the other charged amino acids, strengthening its relative charge and snowballing its destabilizing nature. At the other end of the spectrum are the hydrophobic amino acids, and in particular, the large aromatics tyrosine and phenylalanine. Why are these tolerated so well on a domain that, due to its small size, features a much larger surface-to-core ratio than most other proteins?

Close observation of the data in Figure 4-7 shows that among the functionally identical amino acid pairs (D/E and N/Q), the residue carrying an extra methylene was tolerated better across the protein. This, coupled with a high surface-to-core ratio when compared to the average protein, suggested that the Gβ1 domain may be unique in its accommodation of hydrophobic mutations, perhaps in an effort to bury more hydrophobic surface area. Since average mutant amino acid data is recapitulated well by Popmusic2 (Figure 4-8) we investigated our hypothesis by calculating the systematic scan of four other proteins, all larger than Gβ1, with the Popmusic2 web server. The proteins selected

isolated the effects size, secondary structure composition, and packing density might have on amino acid tolerance (Table 4-2). Amazingly, all four had very similar average mutant amino acid rankings, essentially duplicating the experimental Gβ1 results (Figure 4-8). The tolerance to large hydrophobic amino acids across a domain appears to be a general feature of soluble globular proteins.

Native proteins feature very modest amounts of solvent-exposed hydrophobic residues due to the possibility of alternative folded states that better bury the nonpolar surface area. Yet, the very presence of modest amounts of surface hydrophobic residues indicates that the physical mechanism underlying this behavior has some buffer preventing disastrous aggregated outcomes. Single incorporations of nonpolar residues are not likely to alter the native conformation, and as seen by the data, can stabilize the fold. Previous experimental work on staphylococcal nuclease supported the notion that this "reverse hydrophobic effect" is almost nonexistent across single mutations (27). Interestingly, they note that fully exposed positions better tolerate aromatic incorporation than partially buried sites, arguing that mutation sites are still susceptible to steric clashes and packing effects despite being close to the protein surface. We found similar results when the average stability effects of the top four tolerated amino acids (Phe, Tyr, Leu, Ile) in boundary and surface sites were broken down into two populations of packing density. However, when the data was broken down into quartiles, there is a bump in average stability in the partially exposed quartile, providing evidence for preferential packing between the incorporated hydrophobic amino acid and the nonpolar atoms of other native residues near the surface (Table 4-3). In sum, considering the hydrophobic mutability of protein cores (28, 29), the support for partially exposed hydrophobic

clusters, and that the native conformation is unlikely to change because of a single mutation on plastic protein surfaces, non-polar residues offer the best chance at making neutral or stabilizing interactions across a protein domain.

Stability prediction algorithm performance

Three popular prediction algorithms, Popmusic2, FoldX3, and Rosetta, were used to calculate the stability change of the 935 mutations in our domain mutagenesis dataset. Popmusic2 is a reduced-representation statistical energy function trained to recapitulate a large experimental dataset from the Protherm database. FoldX3 is similarly trained, but uses an empirically derived energy function mixed with weighted statistical terms. Rosetta mixes statistical potentials with an all-atom physical potential, and was trained to recover native sequence composition for protein design. Three versions of Rosetta are used, each with increasing amounts of backbone flexibility. The specific details and parameters used for each algorithm are described in the methods. Unfolded mutations for which only approximate data is available were filtered, leaving 825 mutations. Unrealistic predicted energies from the FoldX3 and Rosetta calculations prompted further filtering by removing mutations with abnormally high van der Waals clash or repulsive energies, respectively. Algorithm performance was evaluated by correlation coefficients (Table 4-4) and fraction correct % (Table 4-5). In addition, these metrics are reported for the datasets broken down by volume change, RESCLASS, and polarity change to assess performance by mutation type.

When asked to recapitulate energetic details of the full dataset the hybrid energy functions perform quite poorly, as the purely statistical Popmusic2 method led the pack

with a correlation coefficient of 0.56 (Table 4-4). After filtering mutations with large clashes, FoldX3 shows improved performance while the flexible backbone Rosetta methods achieve the best overall correlations to the dataset. Although full backbone minimization reduces the number of outliers due to repulsive clashes, it is outperformed by constrained minimization, even against our unbiased dataset (Chapter 3). The notion that too much backbone freedom may simultaneously hurt and help structure prediction of a mixed dataset (30) is upheld by our results. When asked to recover the fraction of positive, neutral, and negative mutations in the data (Table 4-5), all of the algorithms perform almost equally. This result speaks to the utility each method has in predicting the correct trend in large datasets (17, 30). That they all do so equally well is both reassuring to users in the field and frustrating to developers looking for avenues of improvement.

Mutations that remove volume (-VolΔ) are better predicted than those that add volume (+VolΔ) across all algorithms and both tables. The closer a mutant protein's conformation is to wild type, the better each prediction algorithm performs, as most are capable of only torsion preferences or rotameric flips. The methods that do introduce backbone flexibility perform better, but can be restricted by limited sampling of correct conformations (30). The preference for large-to-small mutations then implies limited structural rearrangements across the domain for this mutation type, a conclusion supported by work on T4 lysozyme (31). The overwhelming number of alanine and glycine mutants (-VolΔ) in the Protherm database may also partially explain the affinity Popmusic2 and FoldX3 have for this mutation type. A surprising detail is the continued advantage constrained minimization exhibits over unconstrained minimization, even

across the small-to-large mutations (r = 0.56 vs. r = 0.52) that are expected to introduce sizeable backbone rearrangements.

As a large determinant of mutational sensitivity, tertiary structure can be expected to play a role in algorithm performance. The overwhelming number of destabilizing mutations in the core inflates the fraction of mutations easily predicted (Table 4-5), while mutations closer to the surface are binned at > 50% accuracy. However, knowledge of a mutant's ΔΔG direction is no guarantee of correlation coefficient accuracy, as shown by the poor performance of core mutants in Table 4-4. The fact that destabilizing variants exist across a larger energetic range than more benign mutations likely promotes this inaccuracy in predicting buried positions.

Breaking down the data by polarity changes highlights the underlying principles that govern the prediction algorithms. The effective van der Waals potential in FoldX3 and Rosetta give these methods an advantage in predicting the core packing effects of nonpolar-to-nonpolar mutations (Table 4-4). Likewise, Rosetta's suboptimal treatment of buried electrostatics (nonpolar-to-polar, r = 0.34) and nonpolar exposure (polar-to-nonpolar, r = 0.42) is ameliorated by allowing backbone flexibility. No particular method excels in polar-to-polar mutations, likely a result of the lack of explicit solvent from any of the calculations. Differences between algorithm frameworks may be best embodied by comparing the core (nonpolar-to-nonpolar/polar) and surface (polar-to-nonpolar/polar) prediction accuracies of FoldX3 and Popmusic2. Despite very similar training sets, each method's competency lies in inverse structural environments: Core mutations come easier to FoldX3 due to its effective treatment of sterics, while surface mutations are

better captured by Popmusic2 because the statistical nature of its potential can implicitly capture complex multi-body effects.

The differences seen in Table 4-4 between the algorithms are smoothed when the stringency in prediction accuracy is lowered, as in Table 4-5. Only the poor performance in polar-to-polar mutations by FoldX3 and Rosetta is effectively reproduced from the previous metric. Popmusic2 performs admirably in predicting the fraction correct in all polarity change categories. This jack-of-all-trades quality likely stems from the fact that each statistical term is weighted by solvent accessibility, allowing it to grossly fractionate between debilitating core mutations and neutral surface mutations.

Any recommendation on the prediction algorithm of choice must be tempered by the type of question being asked. Queries concerning the specific and accurate stability of particular single-mutants would probably be best estimated by the constrained backbone minimization Rosetta protocol. However, attention should be paid to filter unreasonable repulsive energies and to consider that accuracy can drop with polar-to-polar mutations, buried positions, and mutations that add volume. If computational power is limited or the number of mutations is greater than 10^3, Popmusic2 is significantly faster and is effective in predicting the trends in the data (Figure 4-8). The actual magnitude of each calculation should be viewed skeptically, unless the mutation involves only polar residues, in which case Popmusic2 performs better than any iteration of Rosetta. FoldX3 serves as the middle ground between the other methods in terms of both speed and accuracy. While observations from mutational trend studies in which FoldX was used are likely to be duplicated by other algorithms, specific values should be taken lightly, especially those involving polar surface mutations.

Additivity of multiple mutations

The wild-type sequence of the Gβ1 domain is very tolerant to single mutations, as evidenced by the mutational outcome percentages and heat map distribution reported in the first section. Mutability trials, where single mutations are added until exhausting some threshold, reemphasize the observed leniency surface positions have over those in the core. These calculations assume each mutation is completely additive, an unrealistic assumption for proteins, although they do provide reference points for the mutational load of a perfectly additive system. If the threshold is 4.5 kcal/mol (the $\Delta G(H_2O)$ of wild-type Gβ1 is 4.04 ± 0.4 kcal/mol), then an average of 6.28 ± 4.9 random mutations over 1000 trajectories are needed to break the protein. Predictably, the required number of mutations increases as one progresses through the RESCLASS categories from 2.07 ± 1.1 mutations (core), to 10.27 ± 8.9 mutations (boundary), to a maximum of 17.46 ± 15.0 mutations on the surface. Given the stated evidence, one might assume that making multiple mutations on the surface of Gβ1 would be more successful than mutating a similar number of positions in the core.

Data from previous efforts in individually designing the core and surface of Gβ1, coupled with our comprehensive dataset, allows us to examine the relative performance of our design procedures as well as the additivity of mutations in different regions of the protein. The core mutant dataset is a compilation of libraries designed from different sources of structural diversity where only hydrophobic amino acids were allowed (23). Each mutant is 2–5 fold away from wild type, above the 2.1 random mutation reference mark, yet more than 80% of the dataset is thermodynamically neutral or better than wild

type (Figure 4-8). This result endorses the all-atom two-body energy function (similar to Rosetta) we used to predict these sequences for modeling hydrophobic core interactions. Rosetta's advantage in accurately predicting nonpolar-to-nonpolar single mutations (Table 4-4) supports this finding. The surface mutant dataset is a single library of mutants, designed using the same energy function as the core study, which aimed to improve overall stability through mutations to the β-sheet surface of the protein (unpublished results). No mutant carrying more than 1 mutation in this library was stabilized from wild type, and a mild inverse relationship exists between stability and the number of mutations (Figure 4-8). In addition, the observed number of mutations isn't remotely close to the random mutation reference point for this segment of the protein, reemphasizing the reduced capacity of physics-based energy functions to capture surface interactions.

The discrepancy in algorithm performance in different protein environments could also be attributed to the nature of additivity in each environment. When the core mutant data is plotted against the sum of individual $\Delta\Delta G$ stability values from our single mutant dataset, the linear trend line ($r = 0.86$) is jilted above the perfect additivity line ($y = x$), signifying that the mutants are destabilized compared to what the simple sum would predict (Figure 4-8). As the majority of mutations are stabilizing, one explanation may be that there is some limiting level of local stabilization that a protein core can reach before interactions elsewhere in the protein become more important globally (20). Conjecture aside, this plot serves as a reference to an identical chart featuring the surface mutant data (Figure 4-8). Here the trend line dramatically intercepts the perfect additivity line, demonstrating much more pronounced non-additivity than in the core mutants. The

particular number of mutations is also very important, as the slope of the trend line radically changes after introducing 3 mutations to the wild-type background, mirroring the largely simple additivity found in double-mutants from another large data study (14). Why is the additivity of multiple mutations so different between the structural environments?

One biophysical explanation would be that the capacity of surface residues to change conformation, along with the greater number of potential interaction partners (solvent molecules) as compared to the situation in the core, allows for the differences seen in mutational non-additivity. Interactions not modeled by traditional protein design software, such as buried waters and extensive hydrogen bond networks along the surface, could be so important to the enthalpy of the protein that they overcame any gain in entropy from joining the bulk solvent. Finally, a parallel can be made between our observations and recent work in linking fitness robustness and epistasis. Bershtein and coworkers describe a quantitative inverse correlation between higher tolerance to mutations and the level of negative epistasis (12). Enzymes (β-lactamase) tested under low fitness (ampicillin) levels exhibited higher degrees of negative epistasis; that is, the effect of mutations after exhausting some threshold level was greater than when under higher levels of fitness stringency. Similarly, surface positions, seemingly tolerant of most single mutations, display markedly stronger non-additivity than core positions. This finding extends the robustness-epistasis theory from describing global, random mutations to capturing particular tertiary-structure effects of proteins. Unfortunately, the ~ 100 multiple mutants in this combined study represent an incredibly small slice of the number

of potential two-, three-, four-, and fivefold mutants possible, and therefore this parallel may only exist for the variants examined here.

The surface mutations in the previous additivity study didn't include any greatly stabilizing single mutations, an element that may have had an effect on the observed epistasis. In fact, the Bershtein theory is reported to represent combined *deleterious* mutational effects. What is the result of combining multiple mutations that are all individually significantly *stabilizing*? Drawing from our comprehensive dataset, we constructed three variants (Table 4-6) by first selecting mutations that each stabilized the Gβ1 domain by more than 1 kcal/mol. This led to 33 individual mutations over 16 positions, a great majority of which introduced large hydrophobic residues. We selected one mutation per position with an eye towards limiting the number of incorporated non-polar residues and combined them into a single variant (16-fold). The mutations were then visualized on the 1PGA structure and funneled down to 8 mutations, removing clustered, interacting residues (8-fold). Lastly, 2 more mutations were screened out due to their mutation of special residues, which tend to have entropic effects on protein stability (6-fold).

The three variants were expressed, purified, and assayed for stability in the same way as the entire single mutant dataset. The 16-fold mutant had no soluble expression as determined by the criteria set in Chapter 3. The 6 and 8-fold mutants expressed normally yet both displayed neutral ΔΔG stability values, despite featuring individual mutations that, when summed, should stabilize the protein by more than 8 and 10 kcal/mol, respectively (Table 4-6). Rationally combining multiple efficacious single mutations from our stability map proved not to be a successful avenue for protein stabilization.

Upon investigating the results, all but two of the mutations in the 16-fold mutant occurred in the boundary or surface of the Gβ1 domain, and more than half of them involved a polar-to-nonpolar mutation. Introducing that many hydrophobic residues to the protein surface is thought to be dubious for aggregation reasons, unlike single nonpolar incorporations that can be stabilizing (1, 2, 32, 33). Anticipating the 16-fold mutant's issues, the 6- and 8-fold mutants were designed more conservatively, avoiding the insertion of potentially destabilizing interactions and evening the ratio of polar and nonpolar mutations. Mutated sites were selected for their three-dimensional distance from each other, thereby promoting conditions for perfect additivity (20, 34). The subsequent severe non-additivity encountered by the variants suggests very large unfolded state effects are at play. Whether these effects are due to the tertiary location, residue type, number of the mutations, or some combination of these attributes is unknown. Future work on these variants should attempt to definitely explain the source of non-additivity by tracing their potential mutational paths through sequence space.

Structure-based protein engineering and design seeks to modify the properties of proteins through the calculation of folded state energetics. Our analysis of the additivity of mutations in the Gβ1 domain demonstrates the difficulty in identifying distant, stabilizing sequences without explicit consideration to the unfolded state. Despite a wild-type sequence experimentally determined to be accommodating of most single mutations, multi-fold variants became increasingly harder to predict, especially if the mutations occurred on the surface. Modern techniques minimize this deficiency in structure-based design by testing libraries of engineered sequences to spread the risk in selecting a mispredicted variant. Although this can be effective (23, 35, 36), novel, efficient, and

effective methods for the unfolded state, among other absent concerns and approximations, will help to usher in truly robust protein design. Concurrently, efforts to capture experimental data on more than just the global stability of mutant proteins deserve attention. For example, multiplexing thermodynamic stability with solubility, expression level, and proteolysis resistance would produce quality high-density datasets that will deepen our understanding of both *in vitro* and *in vivo* stability.

Conclusions

Our aggregate analysis of stability data on every single mutant in Gβ1 provided the first experimental look at the mutational distribution of a protein domain. The rather tolerant nature of the protein, especially to hydrophobic residues, illustrates the plasticity of non-core residues and the heavy desire to bury nonpolar surface area. Atomic packing density linearly correlated with positional sensitivity, and scanning with serine, not alanine, served as the best experimental indicator of positional hotspots. The unbiased nature of the dataset provided an even playing field upon which to test popular stability-prediction algorithms. Although Rosetta was the clear performance leader, every method could satisfactorily recapitulate the general trends of the data. Upon examining the additivity of previous design efforts, we learned that non-additivity was prevalent throughout the protein, but especially on the surface. Attempts to utilize the data for rational stability engineering failed in the face of tremendous non-additivity.

Large mutagenesis datasets will only become more common with the maturation of automation technologies and next-generation sequencing. This incoming avalanche of data will provide breadth in complementing traditional in-depth analysis of important

systems. However, the volume of data current cutting-edge technologies can produce (10^5 variants, or all double mutants in a 50 aa protein) still pales in comparison to the enormity of potential sequence space. This is especially troubling considering that ours and another study show that double mutants are fairly additive (14), in contrast to what we see with variants three or more mutations away from wild type. Computational solutions exist that can traverse the ocean of potential sequences, but require high-density datasets from which to train new methods to properly evaluate proteins distant from the starting sequence.

Materials and methods

Dataset

Thermodynamic stability data from 935 single mutants of the β1 domain of wild-type Streptococcal protein G (Gβ1) was generated as described in Chapter 3 of this thesis. Briefly, each gene was constructed through laboratory automation and sequence verified. Every protein was expressed, purified, and its chemical melting point (C_m) determined by measuring tryptophan fluorescence in response to a 24-point GdmCl gradient. The $C_{m\ Mutant}$ value was used to calculate ΔΔG from the following equation:

$$\Delta\Delta G = \ \overline{m} * (C_{m\ Mutant} - C_{m\ WT})$$

where $\overline{m}$ is the average of the wild-type and mutant m-values (37), a parameter obtained from the linear extrapolation method (38) for determining changes in free energy. Using this equation, stabilizing mutations held positive values, while destabilizing mutations held negative values.

ΔΔG distribution fitting

The experimental ΔΔG data was binned into a histogram with 0.5 kcal/mol intervals. In order to match previous work, all stability data was multiplied by -1, making positive values represent destabilizing mutations and vice versa. We used the following Gaussian model to fit the individual core and surface distributions, with F the percent fraction and x the ΔΔG values (17):

$$F_s(x) = \frac{100}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

The bi-Gaussian fit model was given as a superposition of two Gaussians with different means and variances, and P_1 the fraction of the first Gaussian:

$$F_{bi}(x) = 100\left\{\frac{P_1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right] + \frac{1-P_1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right]\right\}$$

If the fit is good, the mean and variance of the first Gaussian should correspond with the surface distribution, while the mean and variance of the second Gaussian should correspond with the core distribution. A two-sided t-test was conducted using the Mathematica HypothesisTesting package (Wolfram Research) in order to determine agreements between the means of two given Gaussian distributions (assuming equal and unknown variances). P-values were generated with a 95% confidence interval ($\alpha = 0.025$).

Amino acid scanning analysis

Single mutant data of 19 amino acids across 55 of 56 positions in the Gβ1 domain allowed us to identify the best amino acid for experimental stability scanning. We experimented with two ways to rank the deviations between the ΔΔG for a particular

mutation and the average ΔΔG at each position in the protein. Let us define $\delta\Delta\Delta G_{i,a}$ as the deviation of the stability $\Delta\Delta G_i$ for amino acid mutant i from the average stability at a single residue position a where $1 \leq i \leq 19$ and $1 \leq a \leq 56$. This is given by the following equation:

$$\delta\Delta\Delta G_{i,a} = \left| \Delta\Delta G_{i,a} - \frac{\sum_{k=1}^{19} \Delta\Delta G_{k,a}}{19} \right|$$

In the first ranking system, we used a weighting directly proportional to the actual $\delta\Delta\Delta G_{i,a}$ value calculated. We derived $\delta\Delta\Delta G_{i,a}$ for every single possible mutant and then summed the results for each amino acid across all residue positions. The scoring function for each individual amino acid i in this weighted deviation method is given by:

$$score_i = \sum_{k=1}^{56} \delta\Delta\Delta G_{i,k}$$

As the weighted deviation method can be biased by the large stability changes common to core positions versus those on the surface, an alternative deviation method was developed. In this ranking system, we sorted $\delta\Delta\Delta G_{i,a}$ for all amino acids i from smallest to largest, at residue position a. The ranked position of a given mutation i at residue a is defined as $R(\delta\Delta\Delta G_{i,a})$. For a particular residue a, we would assign a score of 1 to the first-ranking single mutation, a score of 2 to the second-ranking single mutation, and so on through the last amino acid. Once completed, we then summed the individual scores $R(\delta\Delta\Delta G_{i,a})$ for each amino acid across all positions to obtain an aggregate score for that particular amino acid. This score for a given amino acid i in this ranked deviation method is given by:

$$score_i = \sum_{k=1}^{56} R(\delta\Delta\Delta G_{i,k})$$

As the actual scores in each method aren't physically relevant, they are normalized with respect to the score determined for alanine incorporation.

Prediction algorithms

The webserver for Popmusic version 2.1, located at http://babylone.ulb.ac.be/popmusic, was used by performing a "Systematic" command on the wild-type crystal structure of Gβ1 (1PGA).

The latest release of FoldX (version 3.0, beta 5) was retrieved from http://foldx.crg.es. The crystal structure of Gβ1 (1PGA) was prepared by using the "RepairPDB" command to perform Asn, Gln, and His flips, alleviate small Van der Waals' clashes, and optimize wild-type rotamer packing. Every mutation in the dataset was constructed through the "BuildModel" command, and the difference in energy between the WT reference and the corresponding mutant was averaged over five trials.

The latest release of Rosetta (version 3.3) was retrieved from http://www.rosettacommons.org. The ddg_monomer application was used to generate single mutant stability data from a pre-minimized version of the crystal structure of Gβ1 (1PGA). We explicitly followed the available online documentation in order to prepare all necessary input files. Option sets described in the documentation pertain to the various Rosetta iterations tested in this paper (no bb min: low-resolution protocol; cst bb min: high-resolution protocol; full bb min: high-resolution protocol with an empty distance restraints file).

The performance of each algorithm was evaluated by Pearson's correlation coefficient and the fraction correct, defined as the number of correctly categorized

mutants (stabilizing (>= 0.5 kcal/mol), neutral (< 0.5 kcal/mol and > -0.5 kcal/mol), and destabilizing (<= -0.5 kcal/mol)) divided by the total number of mutations in the set.

Mutability determination

In order to provide an upper bound on the mutability of the protein we assumed perfect additivity of the ΔΔG values for single mutants. We defined n_{crit} as the critical number of mutations needed to exceed the threshold ΔΔG of 4.5 kcal/mol. For a series of N mutations, the net stability of such a domain is:

$$\Delta\Delta G_{net} = \sum_{k=1}^{N} \Delta\Delta G_k$$

The probability of stabilizing/destabilizing a protein domain with a random mutation was examined by making each single mutant equally probable. We ran a simulation in Mathematica for 1000 or 10000 trials with 200 random mutations in each trial. This many random mutations were made in order to have a high likelihood that we would reach the unfolding threshold before this value. We found the critical number of mutations n_{crit} by checking the net stability after each successive mutation. When $\Delta\Delta G_{net,k} > 4.5$ kcal/mol, then $n_{crit} = k$.

References

1. Bowie JU, Reidhaar-Olson JF, Lim WA & Sauer RT (1990) Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* 247(4948):1306–1310.

2. Rennell D, Bouvier SE, Hardy LW & Poteete AR (1991) Systematic mutation of bacteriophage T4 lysozyme. *Journal of Molecular Biology* 222(1):67–88.

3. Matthews BW (1993) Structural and genetic analysis of protein stability. *Annual review of biochemistry* 62:139–160.

4. Fersht AR & Serrano L (1993) Principles of Protein Stability Derived from Protein Engineering Experiments. *Current Opinion in Structural Biology* 3(1):75–83.

5. Markiewicz P, Kleina LG, Cruz C, Ehret S & Miller JH (1994) Genetic studies of the lac repressor. XIV. Analysis of 4000 altered Escherichia coli lac repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence. *Journal of Molecular Biology* 240(5):421–433.

6. Dahiyat BI & Mayo SL (1997) De novo protein design: fully automated sequence selection. *Science* 278(5335):82–87.

7. Guerois R, Nielsen JE & Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of Molecular Biology* 320(2):369–387.

8. Kuhlman B, et al. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302(5649):1364–1368.

9. Dehouck Y, et al. (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* 25(19):2537–2543.

10. Guo HH, Choe J & Loeb LA (2004) Protein tolerance to random amino acid change. *Proceedings of the National Academy of Sciences of the United States of America* 101(25):9205–9210.

11. Bloom JD, et al. (2005) Thermodynamic prediction of protein neutrality. *Proceedings of the National Academy of Sciences of the United States of America* 102(3):606–611.

12. Bershtein S, Segal M, Bekerman R, Tokuriki N & Tawfik DS (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444(7121):929–932.

13. Tokuriki N & Tawfik DS (2009) Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology* 19(5):596–604.

14. Fowler DM, et al. (2010) High-resolution mapping of protein sequence-function relationships. *Nature methods* 7(9):741–746.

15. Hietpas RT, Jensen JD & Bolon DNA (2011) Experimental illumination of a fitness landscape. *Proceedings of the National Academy of Sciences of the United States of America* 108(19):7896–7901.

16. Araya CL & Fowler DM (2011) Deep mutational scanning: assessing protein function on a massive scale. *Trends in Biotechnology* 29(9):435–442.

17. Tokuriki N, Stricher F, Schymkowitz J, Serrano L & Tawfik DS (2007) The stability effects of protein mutations appear to be universally distributed. *Journal of Molecular Biology* 369(5):1318–1332.

18. Potapov V, Cohen M & Schreiber G (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein engineering, design & selection* 22(9):553–560.

19. Kumar MD, et al. (2006) ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Research* 34(Database issue):D204–206.

20. Wells JA (1990) Additivity of mutational effects in proteins. *Biochemistry* 29(37):8509–8517.

21. Axe DD, Foster NW & Fersht AR (1998) A search for single substitutions that eliminate enzymatic function in a bacterial ribonuclease. *Biochemistry* 37(20):7157–7166.

22. Weinreich DM, Delaney NF, Depristo MA & Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312(5770):111–114.

23. Allen BD, Nisthal A & Mayo SL (2010) Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proceedings of the National Academy of Sciences of the United States of America* 107(46):19838–19843.

24. Fleming PJ & Richards FM (2000) Protein packing: dependence on protein size, secondary structure and amino acid composition. *Journal of Molecular Biology* 299(2):487–498.

25. Serrano L, Kellis JT, Jr., Cann P, Matouschek A & Fersht AR (1992) The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *Journal of Molecular Biology* 224(3):783–804.

26. Jackson SE, Moracci M, elMasry N, Johnson CM & Fersht AR (1993) Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* 32(42):11259–11269.

27. Schwehm JM, Kristyanne ES, Biggers CC & Stites WE (1998) Stability effects of increasing the hydrophobicity of solvent-exposed side chains in staphylococcal nuclease. *Biochemistry* 37(19):6939–6948.

28. Axe DD, Foster NW & Fersht AR (1996) Active barnase variants with completely random hydrophobic cores. *Proceedings of the National Academy of Sciences of the United States of America* 93(11):5590–5594.

29. Gassner NC, Baase WA & Matthews BW (1996) A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proceedings of the National Academy of Sciences of the United States of America* 93(22):12155–12158.

30. Kellogg EH, Leaver-Fay A & Baker D (2011) Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* 79(3):830–838.

31. Baase WA, Liu L, Tronrud DE & Matthews BW (2010) Lessons from the lysozyme of phage T4. *Protein Science* 19(4):631–641.

32. Cordes MH & Sauer RT (1999) Tolerance of a protein to multiple polar-to-hydrophobic surface substitutions. *Protein Science* 8(2):318–325.

33. Taverna DM & Goldstein RA (2002) Why are proteins so robust to site mutations? *Journal of Molecular Biology* 315(3):479–484.

34. Skinner MM & Terwilliger TC (1996) Potential use of additivity of mutational effects in simplifying protein engineering. *Proceedings of the National Academy of Sciences of the United States of America* 93(20):10753–10757.

35. Chica RA, Moore MM, Allen BD & Mayo SL (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proceedings of the National Academy of Sciences of the United States of America* 107(47):20257–20262.

36. Guntas G, Purbeck C & Kuhlman B (2010) Engineering a protein-protein interface using a computationally designed library. *Proceedings of the National Academy of Sciences of the United States of America* 107(45):19296–19301.

37. Myers JK, Pace CN & Scholtz JM (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Science* 4(10):2138–2148.

38.   Santoro MM & Bolen DW (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27(21):8063–8068.
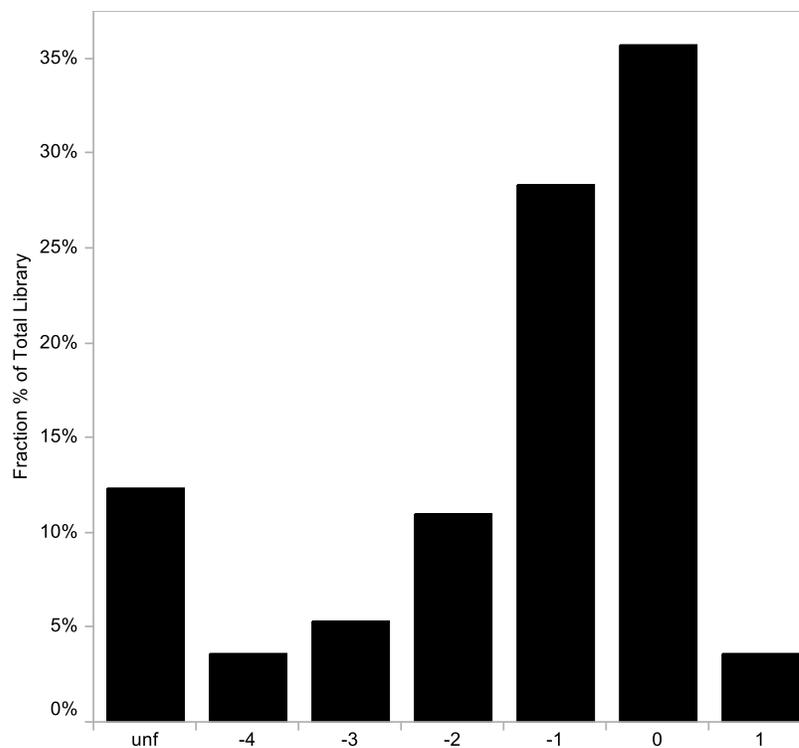
Figure 4-1: Single mutant stability distribution for the Gβ1 domain. As an example, if ΔΔG stability data is represented by χ, then the zero bin holds data with values $0 \geq \chi > 1$. The "unf" bin holds mutant stability data that could not be determined, and is likely insoluble or in an alternative conformation.

Type	Experimental data					FoldX3 data				
	μ_1	σ_1	μ_2	σ_2	R^2	μ_1	σ_1	μ_2	σ_2	R^2
Surface[a]	-0.13	0.65	-	-	0.996	-0.53	0.68	-	-	0.995
Core[a]	-	-	0.28	1.23	0.932	-	-	-0.17	1.25	0.901
All	-0.13	0.53	1.01	1.43	0.999	-0.55	0.64	1.02	1.75	0.997

[a] Surface and core determination done as described in the methods

Table 4-1: Gaussian fitting parameters for the mutational distributions

(a) Experimental core (red) and surface (blue) distributions

(b) Experimental bi-gaussian distribution

(c) FoldX core (red) and surface (blue) distributions

(d) FoldX bi-gaussian distribution
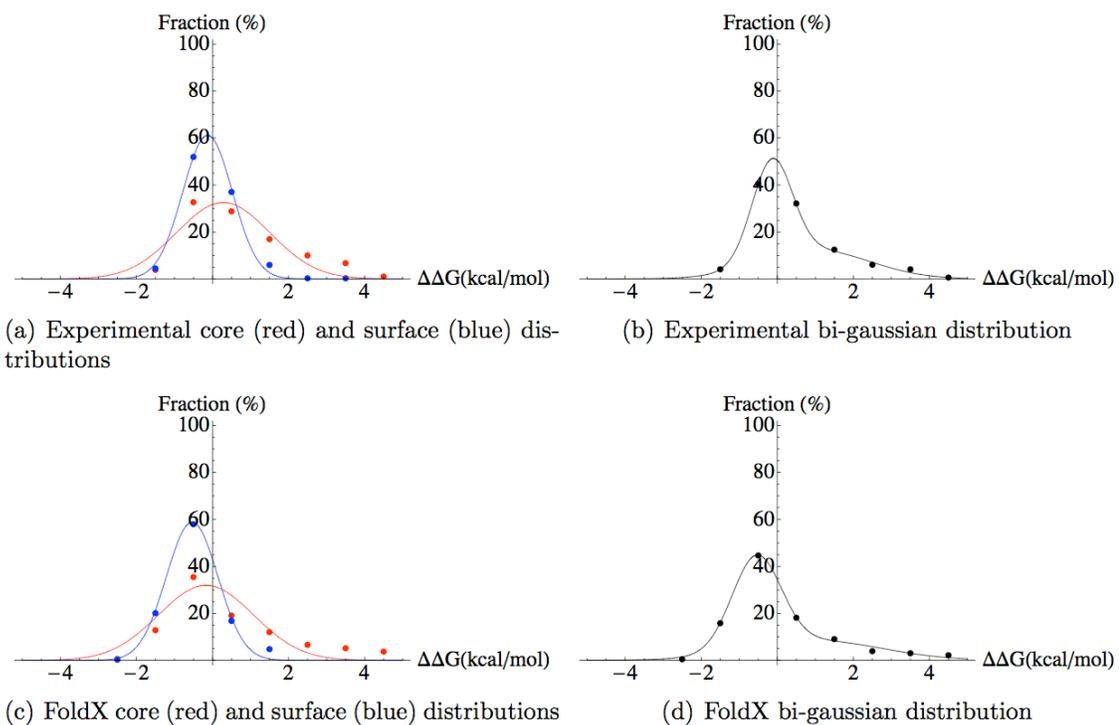
Figure 4-2: Gaussian fits of the Gβ1 mutational distribution. The single mutant dataset was calculated using FoldX3, and values corresponding to unfolded data were removed from both datasets. Equation fits are described in the methods. Positive $\Delta\Delta G$ values indicate destabilizing variants and vice versa.
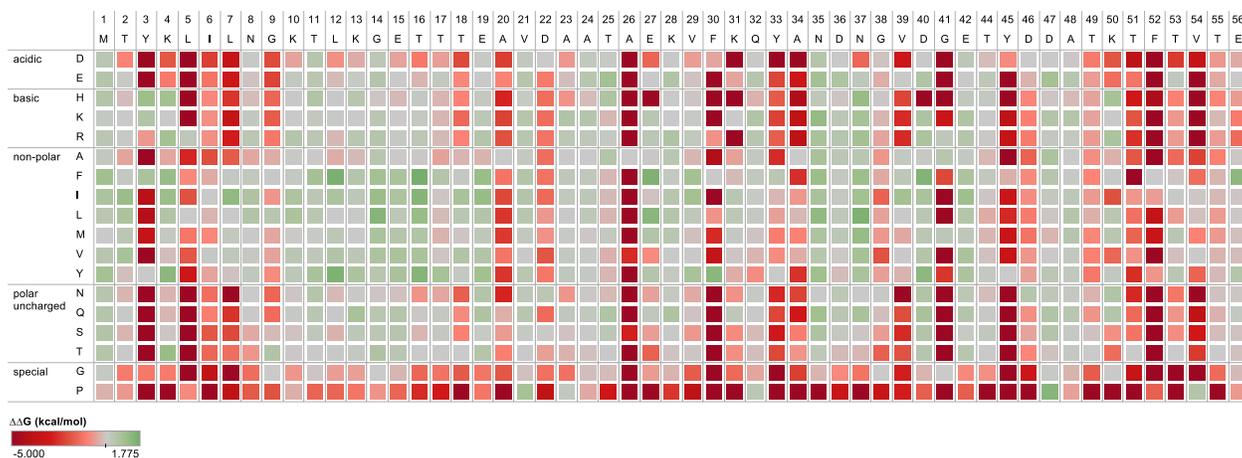
Figure 4-3: Single mutant stability landscape for the Gβ1 domain. Each mutant is colored by its ΔΔG value, where red is destabilizing and green is stabilizing. Self-identity mutations, e.g., M01M, are assigned a zero value and colored gray. Mutant stability data that could not be determined are given the arbitrary value of -5.
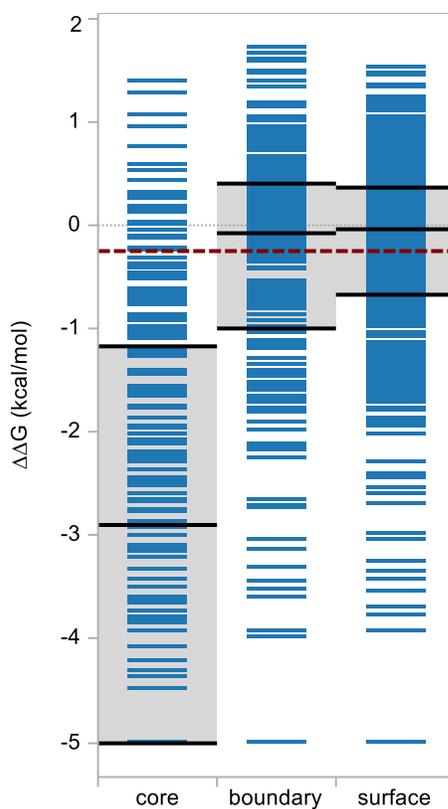
Figure 4-4: Single mutant stability distributions by RESCLASS. The ΔΔG stability distribution for each RESCLASS category is separated into quartiles. The median values for the core, boundary, and surface distributions are -2.89, -0.06, and -0.02, respectively. The red dashed line is the median value for the entire distribution, -0.25. Mutant stability data that could not be determined are given the arbitrary value of -5.
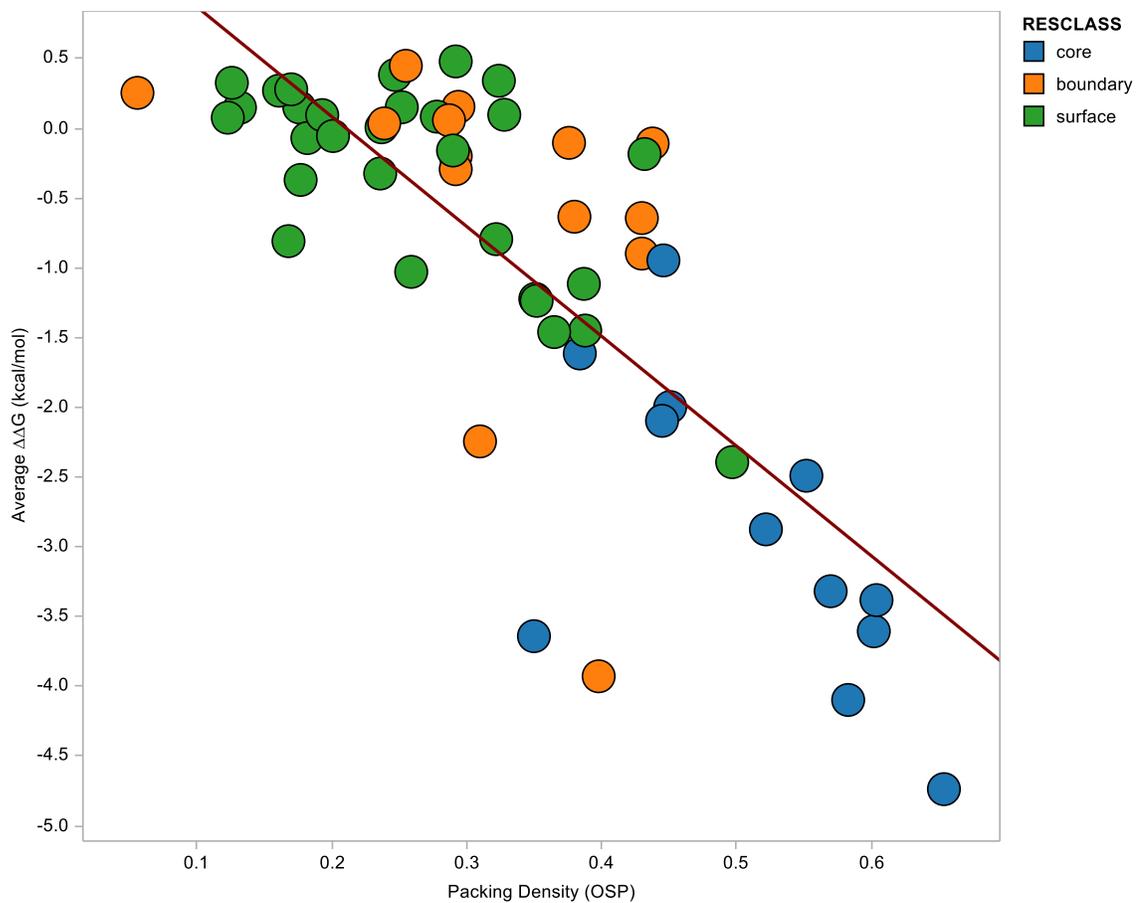
Figure 4-5: Packing density is linearly correlated with ΔΔG averaged by position. Each data point represents a position in the Gβ1 domain, and is colored by RESCLASS. The equation for the red trend line is $y = -7.9x + 1.6$, with an r^2 of 0.62.
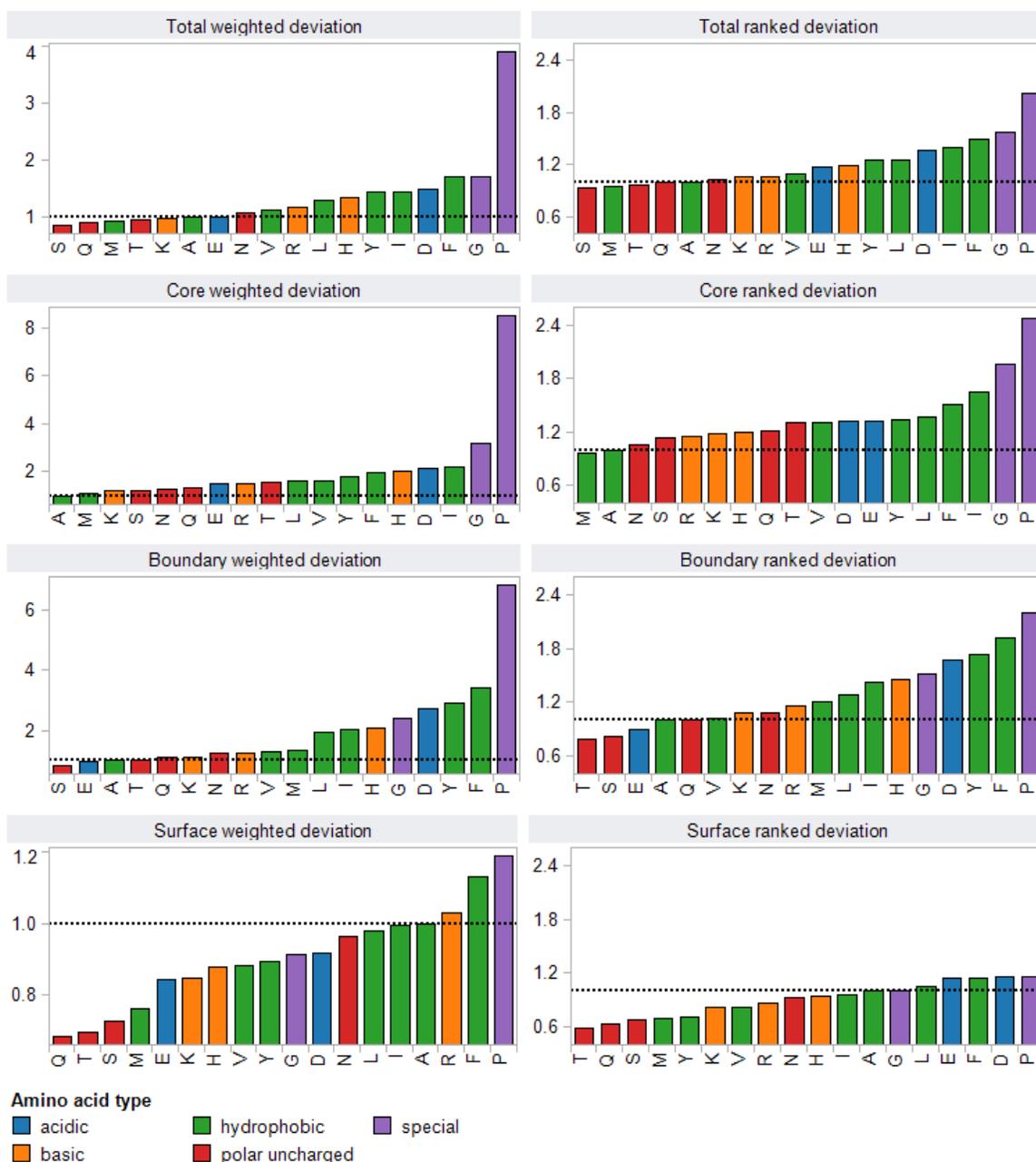
Figure 4-6: Amino acid scanning mutagenesis. The amino acid that best matches the stability at every position in the Gβ1 domain was determined using both ranking and actual kcal/mol weighted deviations (see Methods). The results are normalized to alanine incorporation (black dotted line) to compare against typical functional scanning methodology. The charts describe the overall data and RESCLASS categories and are sorted from the best to worst match. Amino acids are colored by physiochemical type.
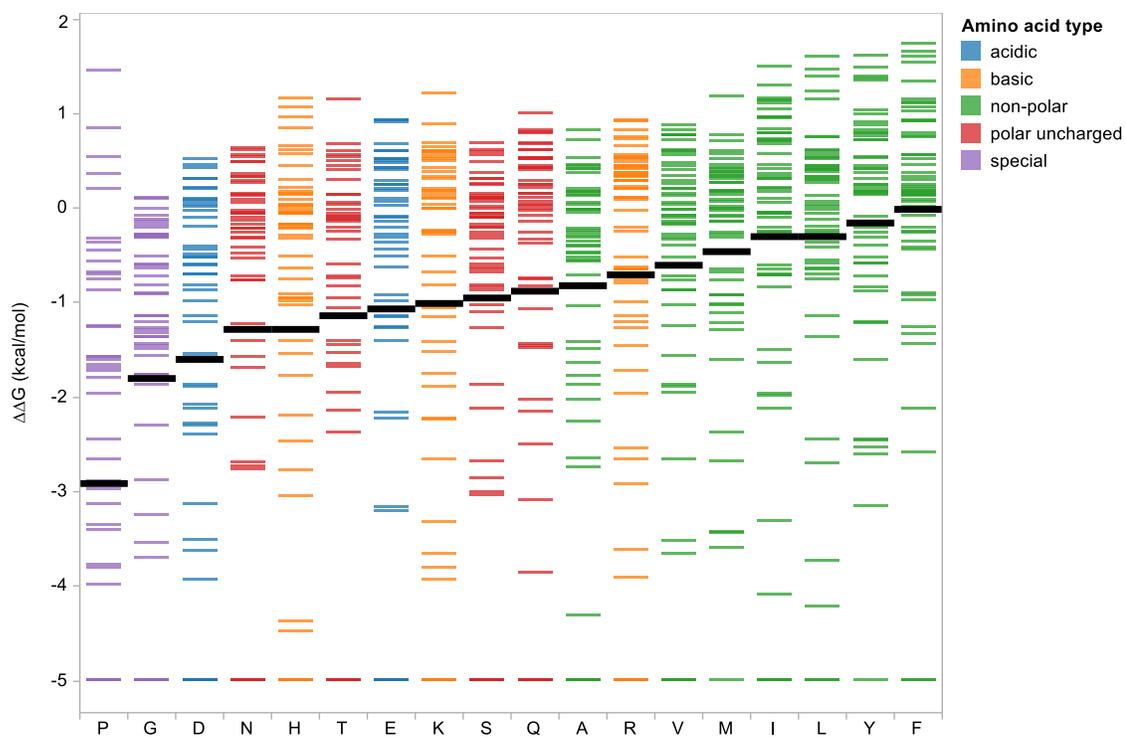
Figure 4-7: Stability distribution of Gβ1 by mutant amino acid. Incorporated amino acids are sorted by the average ΔΔG stability effect of that mutation (black reference line) over the Gβ1 domain. Amino acids are colored by physiochemical type.
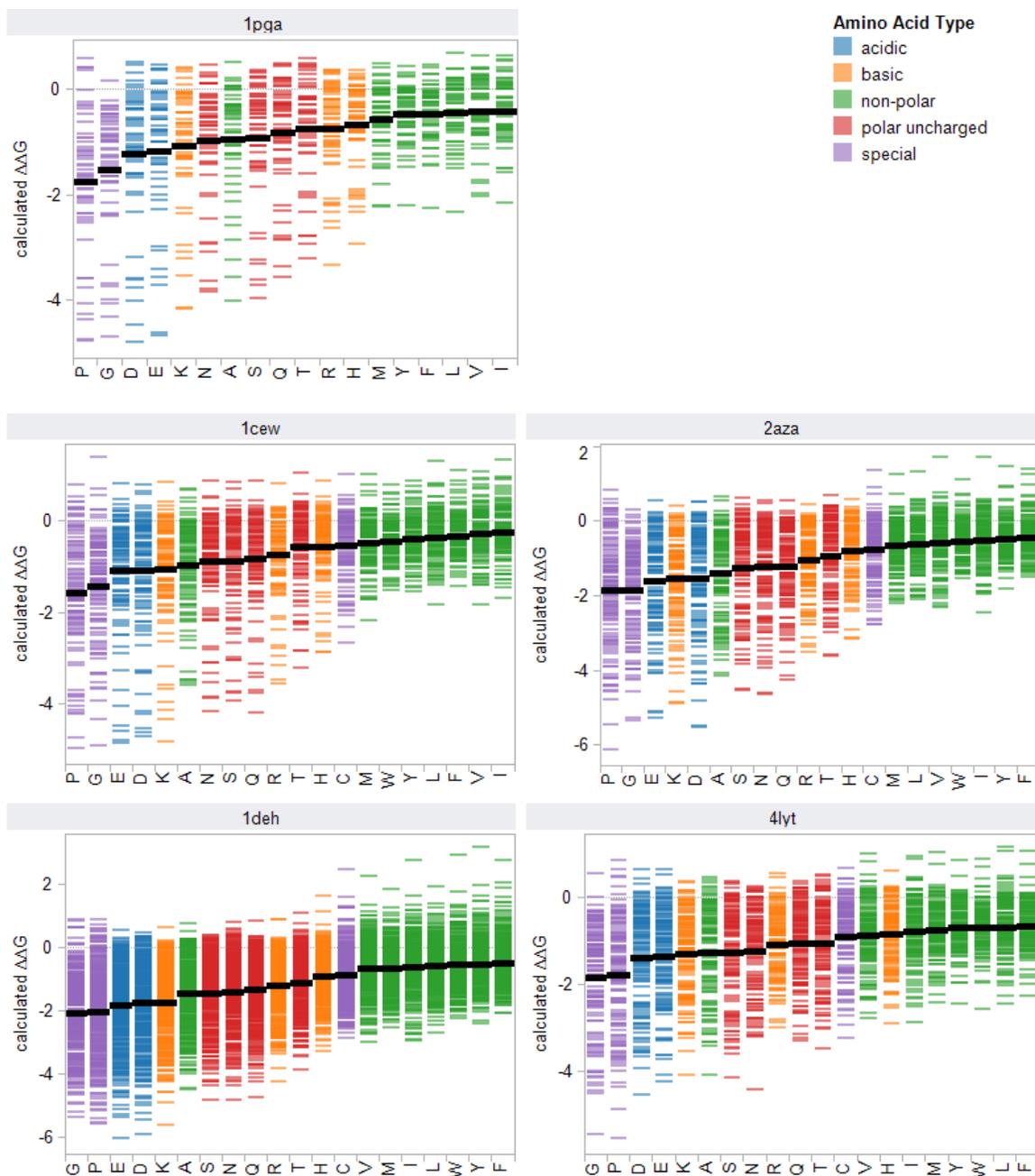
Figure 4-8: Calculated stability distributions by mutant amino acid. Single mutant stability data was calculated with Popmusic2 for the Gβ1 domain (1pga), along with four other proteins: cystatin (1cew), azurin (2aza), alcohol dehydrogenase (1deh), and lysozyme (4lyt). Incorporated amino acids are sorted by the average ΔΔG stability effect of that mutation (black reference line) over each domain. Amino acids are colored by physiochemical type.

Variant	Number of Residues	% Helix[a]	% Strand[a]	Average OSP[b]
1pga	56	25	43	0.332
1cew	108	20	48	0.314
2aza	129	16	36	0.368
4lyt	129	41	11	0.376
1deh	374	28	25	0.392

[a] Secondary structure was determined through DSSP.
[b] Residue packing density (OSP) was averaged over each protein.

Table 4-2: Bioinformatics statistics for selected proteins

OSP percentile rank	Average ΔΔG[a] by quartiles	Average ΔΔG[a] by halves
1–25	0.382	0.449
26–50	0.516	
51–75	0.097	-0.094
76–100	-0.285	

[a] Calculated over Y/F/I/L amino acids only

Table 4-3: Comparing the average ΔΔG of hydrophobic mutations by OSP

	Popmusic2	FoldX3	Rosetta (no bb min)[a]	Rosetta (cst bb min)[b]	Rosetta (full bb min)[c]
All	0.56/825	0.35/825	0.26/819	0.45/825	0.44/825
All w/o clashes	-/-	0.52/742	0.35/747	0.62/810	0.61/824
+VolΔ	0.46/489	0.42/417	0.37/428	0.56/476	0.52/488
-VolΔ	0.6/340	0.53/329	0.35/323	0.62/338	0.64/340
Core	0.28/127	0.31/73	0.3/75	0.26/115	0.22/127
Boundary	0.55/221	0.65/217	0.63/213	0.72/219	0.71/221
Surface	0.54/477	0.41/452	0.35/459	0.57/476	0.57/476
NP➔NP	0.43/125	0.57/91	0.52/85	0.60/118	0.62/124
NP➔P	0.49/163	0.62/125	0.34/139	0.64/156	0.53/163
P➔NP	0.66/249	0.47/240	0.42/236	0.69/248	0.68/249
P➔P	0.58/288	0.41/286	0.58/287	0.52/288	0.53/288

All entries are tuples of correlation coefficient (r) and number of data points (n); NP: nonpolar; P: polar.

[a] No backbone minimization after repacking

[b] Constrained backbone minimization after repacking

[c] Unconstrained backbone minimization after repacking

Table 4-4: Algorithm performance by linear correlation

	Popmusic2	FoldX3	Rosetta (no bb min)	Rosetta (cst bb min)	Rosetta (full bb min)
All	0.62/935	0.59/935	0.61/918	0.61/935	0.60/935
+Vol∆	0.52/527	0.52/527	0.52/510	0.53/527	0.51/527
-Vol∆	0.75/412	0.69/412	0.72/412	0.71/412	0.70/412
Core	0.78/204	0.84/204	0.86/187	0.83/204	0.82/204
Boundary	0.57/238	0.57/238	0.53/238	0.53/238	0.52/238
Surface	0.58/493	0.50/493	0.55/493	0.53/493	0.54/493
NP➔NP	0.61/147	0.68/147	0.61/140	0.65/147	0.64/147
NP➔P	0.66/210	0.72/210	0.68/200	0.74/210	0.67/210
P➔NP	0.61/272	0.57/272	0.64/272	0.61/272	0.60/272
P➔P	0.62/306	0.49/306	0.53/306	0.51/306	0.52/306

All entries are tuples of fraction correct and number of data points (n); NP: nonpolar; P: polar.

[a] No backbone minimization after repacking

[b] Constrained backbone minimization after repacking

[c] Unconstrained backbone minimization after repacking

Table 4-5: Algorithm performance by fraction correct
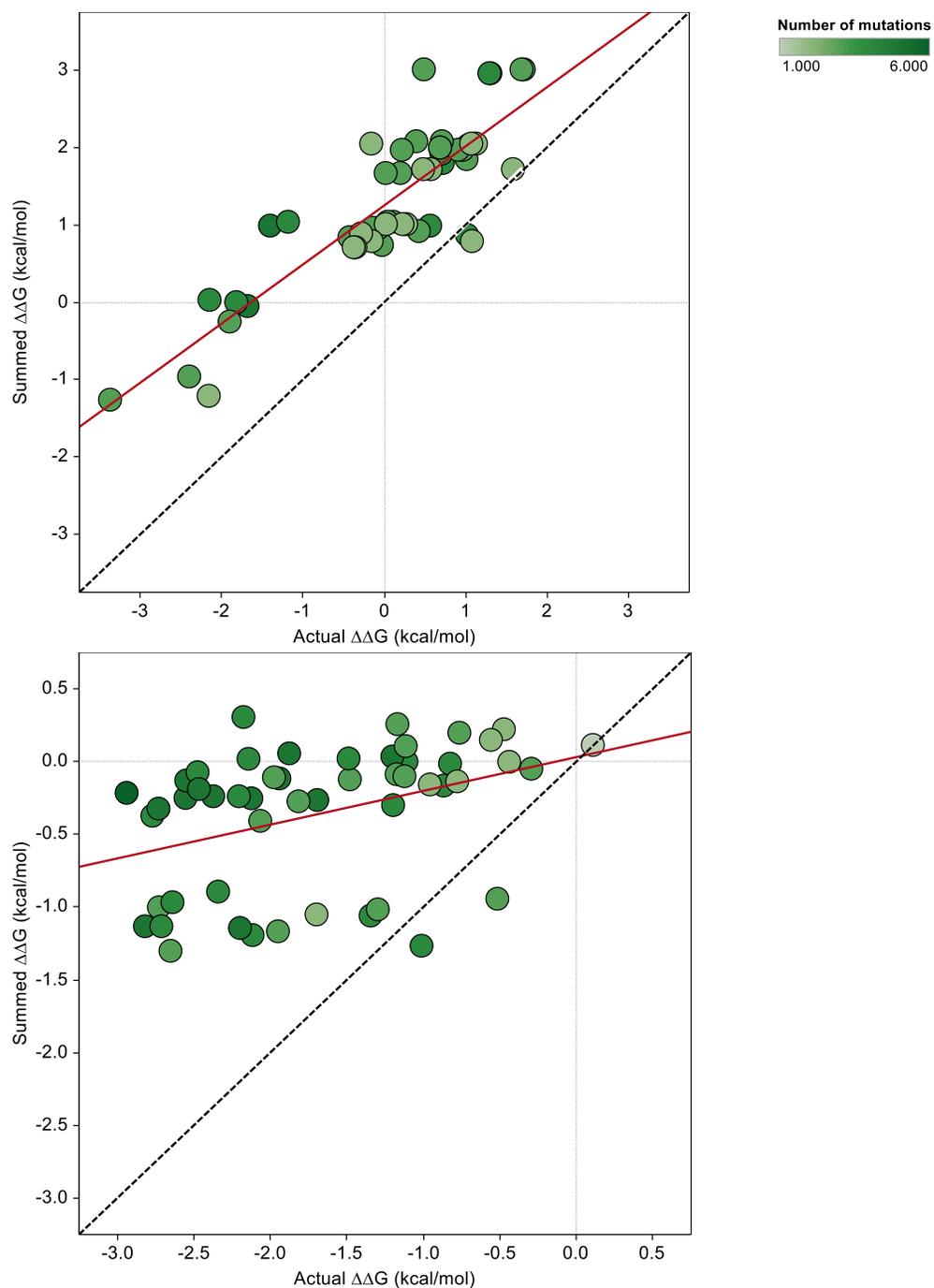
Figure 4-9: Complex additivity in core and surface mutation libraries. Actual ΔΔG is plotted against the sum of single mutation ΔΔG values for core (top) and surface (bottom) mutational libraries. Each data point is colored by the number of mutations from wild-type it carries. The black dashed line is $y = x$, and serves as the indicator for perfect additivity. The r^2 for the red trend lines are 0.74 (top) and 0.15 (bottom).

Variant	Identity	Actual $\Delta\Delta G^a$	Sum $\Delta\Delta G^b$
6-fold	K04T L12Y E27L F30Y N35K E56F	-0.10	8.07
8-fold	6-fold G14L D47P	0.57	10.98
16-fold	8-fold M01F T02I L07I T16I E19I N37L D40F E42I	unfolded	21.21

[a] Determined stability through experiment; units in kcal/mol
[b] Determined stability by summing the $\Delta\Delta G$ of the individual single mutants; units in kcal/mol

Table 4-6: Identity and stability of additive variants