Current Biology 24, 2643–2651, November 17, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.09.072

Article

A Comprehensive Biophysical Description of Pairwise Epistasis throughout an Entire Protein Domain

C. Anders Olson,[1,*] Nicholas C. Wu,[1] and Ren Sun[1,*]
[1]Department of Molecular and Medical Pharmacology, University of California, Los Angeles, Los Angeles, CA 90095, USA

Summary

Background: Nonadditivity in fitness effects from two or more mutations, termed epistasis, can result in compensation of deleterious mutations or negation of beneficial mutations. Recent evidence shows the importance of epistasis in individual evolutionary pathways. However, an unresolved question in molecular evolution is how often and how significantly fitness effects change in alternative genetic backgrounds.
Results: To answer this question, we quantified the effects of all single mutations and double mutations between all positions in the IgG-binding domain of protein G (GB1). By observing the first two steps of all possible evolutionary pathways using this fitness profile, we were able to characterize the extent and magnitude of pairwise epistasis throughout an entire protein molecule. Furthermore, we developed a novel approach to quantitatively determine the effects of single mutations on structural stability ($\Delta\Delta G_U$). This enabled determination of the importance of stability effects in functional epistasis.
Conclusions: Our results illustrate common biophysical mechanisms for occurrences of positive and negative epistasis. Our results show pervasive positive epistasis within a conformationally dynamic network of residues. The stability analysis shows that significant negative epistasis, which is more common than positive epistasis, mostly occurs between combinations of destabilizing mutations. Furthermore, we show that although significant positive epistasis is rare, many deleterious mutations are beneficial in at least one alternative mutational background. The distribution of conditionally beneficial mutations throughout the domain demonstrates that the functional portion of sequence space can be significantly expanded by epistasis.

Introduction

Epistasis, within and between genes, is thought to play an essential role in the ability for protein sequences to evolve through neutral drift or adaptation [1, 2]. While contingencies in fitness limit pathways of divergence, permissive mutations reveal "cryptically beneficial" substitutions [3] that increase the number of acceptable mutations [4]. Epistasis can be explained in physical terms by investigating the biochemical effects of mutations singly and in combination [5]. Examples include evolution of a switch in glucocorticoid receptor-ligand specificity [6], increased hemoglobin affinity to O_2 in high-altitude deer mice [7], and antibiotic resistance in a β-lactamase variant [8], which all rely on nonadditive combinations of mutations.

*Correspondence: aolson@mednet.ucla.edu (C.A.O.), rsun@mednet.ucla.edu (R.S.)

The importance of epistasis is evident for organisms such as influenza that accumulate mutations at a high rate and adapt rapidly in response to immunological and drug pressure [9, 10]. Gong et al. demonstrated how an evolutionary pathway in influenza nucleoprotein required permissive stabilizing mutations prior to gaining certain adaptive substitutions that alone disrupted protein structure [10]. Indeed, most mutations destabilize protein structures [11, 12], and directed evolution experiments show that a large fraction of mutations are deleterious for function [13]. It was recently shown that 63 of 168 mutations chosen from a homologous protein with the same function were deleterious when substituted alone and thus that epistatic interactions are necessary to preserve function [14].

Although these examples show that epistasis is essential in individual evolutionary pathways, they do not address whether combinations of mutational fitness effects are typically epistatic. How likely is it that a mutation has the same fitness in two different genotypes? Historically, protein engineering experiments have shown that the effects of mutations on protein function are typically energetically additive [15–18]. Furthermore, next-generation sequencing technology has enabled the analysis of very large numbers of mutational pairs in experimental evolution which also show that fitness effects are usually additive [19, 20]. Here, we sought to determine whether this observation of general pairwise additivity conflicts with the apparent pervasiveness of epistasis in light of mutational sensitivity [14]. By analyzing the first two steps of all possible evolutionary pathways, we can determine the frequency of pairwise energetic nonadditivity.

Such a comprehensive analysis is necessary in order to determine how often deleterious mutations can be compensated by at least one additional mutation and, likewise, how often neutral or beneficial mutations can be negated by an additional mutation. To do this, we characterized a comprehensive fitness map of single and double mutants within protein G domain B1 (GB1) that was highly correlated to binding affinity (K_A) to immunoglobulin G fragment crystallizable (IgGFC). GB1 is well characterized structurally and is a classical model protein for folding and stability studies [21–24]. Although small, GB1 is a stable, compact, and highly soluble protein with no disulfide bonds. The structure includes an α helix packed against a four-stranded β sheet that are connected by four short loops. This extensive structural and mutagenic characterization of GB1 provided a substantial reference for validating our fitness map.

Furthermore, we were able to use the fitness map to accurately predict the effect of all nonlethal single mutations on structural stability ($\Delta\Delta G_U$). This was accomplished by identifying destabilized mutational backgrounds in which the binding data reflects a change in fraction folded upon addition of secondary mutations. Thus, our fitness map enabled us to identify common biophysical mechanisms of both negative and positive epistasis. For example, we show that exhaustion of the intrinsic stability reservoir, or threshold robustness [25–27], largely accounts for examples of significant negative epistasis. Stabilizing substitutions, which are rare, produce positive epistasis, although with a smaller magnitude compared to combinations of destabilizing mutations. We also

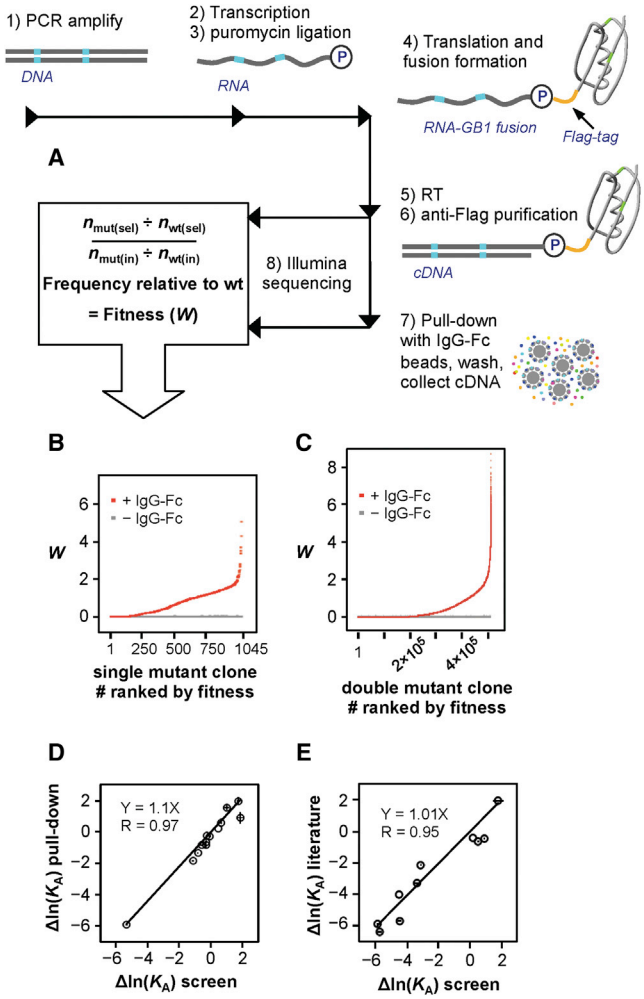


Figure 1. mRNA Display Fitness Profile Scheme

(A) A DNA library encoding all single and double mutants in GB1 was created (see also Figure S1). mRNA display was used to profile the relative binding efficiency of all variants. After a single generation of affinity enrichment, the relative fitness (*W*) for each variant was determined from the change in sequence frequency as identified by Illumina sequencing.

(B and C) The binding fitness (red dots) of each single mutant (B) and (C) all high-confidence double mutants (C; see Figure S1E). The gray dots represent the contribution to *W* from background binding to beads, which was determined from a control lacking IgG-FC.

(D) Thirteen clones were constructed and expressed in vitro with a ^{35}S-methionine label for comparison to fitness determined by the screen. Binding efficiency (see also Figure S1D) was used to estimate relative affinity (see the Experimental Procedures). Error bars represent SD from the triplicated screen (x axis) and from the pull-down when performed in triplicate (y axis).

(E) Correlation of $\Delta\ln(K_A)$ from the screen to $\Delta\ln(K_A)$ for ten variants reported in the literature (see Table S3).

See also Figure S1 and Table S1.

describe long-range positive epistasis that is pervasive within a highly conformationally dynamic network of residues. Our results confirm that epistasis is rare and also that many mutations are detrimental to function. However, this comprehensive fitness profile shows that many deleterious mutations are compensable by at least one of the numerous possible secondary mutations. Together, these results provide an empirical, biophysical description of epistasis and resolve how rare nonadditivity can contribute to the extensive divergence of protein sequences as observed in nature.

Results

High-Resolution GB1 Double-Mutant Affinity Profile

We developed a cassette saturation-mutagenic approach for assembling a library that includes all single and double mutations within the 56 residue GB1 domain (excluding Met1) (Figure 1; Figure S1 and Table S1 available online). Two technical hurdles were overcome in thus study: the ability to error correct and the ability to build a library that is focused on one or two amino acid mutations throughout the entire 55 codon random region. To enable sequencing error correction, we included internal barcodes in each cassette (Figure S1A). Linking of saturation-mutagenized cassettes was accomplished in a sequence-independent manner by using a type IIS restriction endonuclease (BciVI) (Figure S1B). After digestion, a single, degenerate M (A/C) overhang on 3′ fragments enabled specific ligation to a G, T, or K (G/T) overhang on 5′ fragments.

The use of in vitro display technologies to analyze the effects of individual mutations on binding function is well established [28, 29]. Next-generation sequencing has greatly expanded the ability to analyze mutational fitness effects quantitatively [30]. In this study, relative binding affinity of all single and nearly all double amino acid mutants to IgG-FC was characterized using mRNA display [31]. mRNA display is an in vitro genetic system in which peptides are covalently linked to the encoding mRNAs (Figure 1A) typically used to evolve novel molecular recognition tools [31, 32]. Here, we used deep sequencing combined with mRNA display to monitor the evolution of GB1 mutants in real time after one generation of affinity enrichment (Figure 1A).

By measuring the frequency of each variant before and after enrichment (Table S2), we determined relative binding efficiency, or fitness (Figures 1B, 1C, and S1D; see the Experimental Procedures). While fitness is traditionally a population-genetics term, protein fitness can be defined [30, 33, 34], and here relative fraction bound is analogous to a classical definition of relative fitness (*W*), which is the number of progeny relative to the wild-type (WT) per generation. The conditions of this screen, in which the concentration of IgG-FC is below the K_D of WT GB1, provided a large dynamic range in observed fitness effects, from 100-fold below to 8-fold higher than WT fitness (Figures 1B and 1C). Thus, our evolution experiment investigates affinity-based adaptation for improved or new function. We caution that this extremely simplified, noncompetitive evolutionary experiment has many differences in comparison to natural evolution and the relationship between affinity and in vivo fitness will not be directly correlated for many proteins, especially considering that many proteins are multifunctional. However, there are examples in natural evolution such as viral host switching that show a relationship between the affinity of host-adapted receptor binding domain variants and viral infectivity in cell culture [35].

Using a Poisson-based 90% confidence interval, we determined that the fitness effects of all 1,045 single mutants were determined with high confidence and 509,693 double mutants (95.1% of all) were characterized with high confidence (Figure S1E). Importantly, the high-confidence data set includes abundant double mutants throughout all 1,485 possible positional pairs (Figure S1E). The single generation of affinity enrichment was performed in triplicate, and Figure S1F shows that the single-mutant fitness profiles are highly correlated ($R > 0.996$ for all three comparisons). Thus, the binding, PCR, Illumina adaptor ligation, and sequencing steps are highly reproducible. Furthermore, we included a no-IgG control to

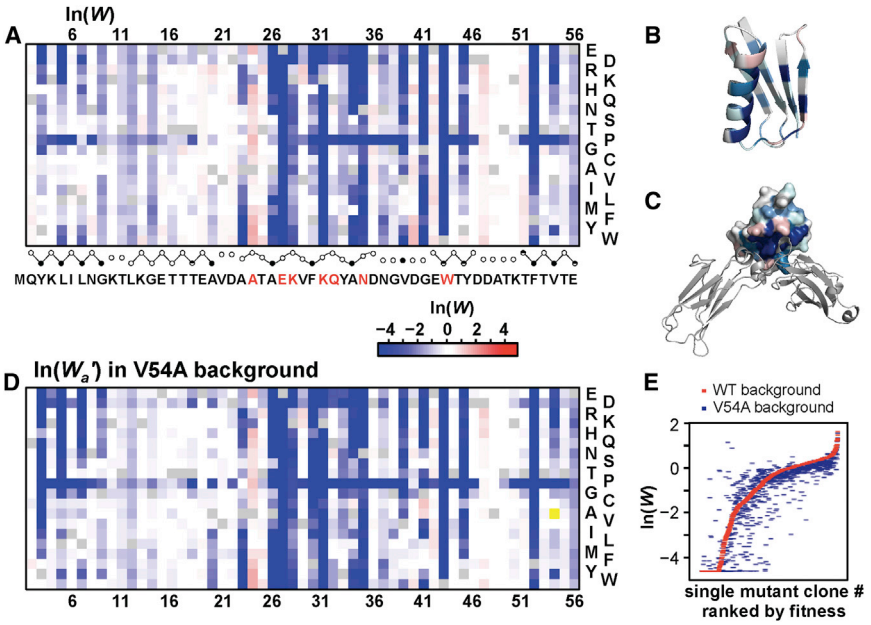


Figure 2. Affinity Profile Validation and Fitness Maps

(A) A heatmap depicting fitness of all single mutants. Residues previously determined to interact with IgG-FC [21] are highlighted in red. The fraction side-chain solvent accessibilities (closed circles, <0.1; partial circles, >0.1; open circles, >0.4) are depicted below. Circles are connected by straight or curved lines to delineate β strands and the α helix, respectively.

(B and C) Average ln(*W*) plotted on GB1 (PDB ID: 1PGA; B) [36] and the complex between protein G domain C2 (space filled) and IgG-FC (cartoon) (PDB ID: 1FCC; C) [21].

(D) A heatmap depicting fitness of all single mutants in the background of V54A.

(E) Comparison of the fitness profile to fitness effects in the background of V54A.

See also Figure S2.

show that background binding does not affect fitness calculations for any variant, including mutants known to be unfolded (Figures 1B and 1C).

We also show that *W* can be used to approximate relative affinity ($K_{A\text{-mut}}$ / $K_{A\text{-wt}}$) similar to the "shotgun scanning" approach [29] (see the Experimental Procedures). This was used to facilitate validation and enable the comparison of energetic effects to fitness effects. We show that $\Delta\ln(K_A)$ values predicted by this screen are highly correlated to those of 13 single or double mutants reconstructed and analyzed for validation by an in vitro pull-down assay (Figure 1D). Furthermore, the $\Delta\ln(K_A)$ predicted by the screen is highly correlated to that of an addition ten variants independently reported in the literature (Figure 1E and Table S3).

Figure 2A depicts ln(*W*) as a heatmap for all 19 single mutants at each position. The average ln(*W*) values per position are displayed on GB1 structures alone or in complex with IgG-FC (Figures 2B and 2C). As expected, core residues are sensitive to substitution [37], indicating severe structural destabilization or that small changes in structure that accommodate core volume changes might adversely affect binding affinity [38, 39]. Surface residues that are sensitive to mutation correlate with alanine scanning mutagenesis [40] and clarify relative importance for ligand recognition [21]. However, beneficial and detrimental surface mutations are found throughout the domain, thus highlighting the importance of such comprehensive screens for characterizing the sequence determinants of functionality [28, 30, 41–44]. For example, alanine scanning could not uncover the importance of position Thr25, where acidic substitutions are highly deleterious and basic substitutions are highly beneficial while Ala is neutral [40] (Figures 1E and 2A).

The double-mutant fitness landscape is depicted as a heatmap showing all high-confidence double mutants (up to 361) for all 1,485 positional pairs (Figure S2A). The comprehensive nature of this screen enables an alternative approach to interpret this data by showing the fitness of all substitutions (*a*) in alternative mutational backgrounds (*b*). For example, we show the fitness effects of all single mutants (W_a') in the background of V54A (Figures 2D and 2E). V54A alone is functionally neutral; however, certain positions become more sensitive to mutations, whereas others change from deleterious to beneficial, notably at position Gly 41 (Figure 2D; vide infra). Furthermore, we show how the fitness of all mutations will change in the background of highly adaptive mutations, such as A24E, which is observed in nature (Figures S2B–S2D). In this background, the functional test is less stringent (i.e., the K_D of A24E is closer to [IgG]), and thus the distribution of fitness effects (DFE) [45] shifts significantly (Figures S2E–S2G), as it would in a less stringent test of fitness (Figure S2H).

Frequency and Proximity of Epistatic Interactions

Figures 2D and 2E show that certain mutations display a change in fitness in combination with V54A and are thus epistatic. Various models can be used to determine whether combinations of mutations display epistasis (ε) [19, 46, 47]. The difference between the fitness in Figure 2D and the fitness in Figure 2A produces one measure of epistasis [$\varepsilon = \ln(W_a') - \ln(W_a)$], which is identical to the relative epistasis model described by Kahn et al. [$\varepsilon = \ln(W_{ab}) - \ln(W_a) - \ln(W_b)$] [46]. We show that the relative model is suitable for the highly adaptive landscape of this experiment (Figure S3A). Here, epistasis (ε) refers to the relative model unless stated otherwise.

We displayed ε for all observed double mutants (Figure 3A) and the average ε for all substitutions at each pairwise positional combination (Figure 3B) as a heatmap. In addition to the 509,693 high-confidence variants, 7,585 variants were unambiguous in sign or significance in ε, resulting in characterization of 96.5% of all pairs. The 90% confidence interval (see above and the Experimental Procedures) was used to minimize epistasis resulting from very low-fitness double-mutants, which could display very large fold change in observed compared to expected fitness due to statistical noise. All 1,485 pairwise positional combinations are represented (Figure S1E), thereby providing a comprehensive description of epistasis throughout the entire protein molecule.

Generally, mutational pairs interact additively or nearly additively, and thus strongly epistatic pairs are rare (Figures 3C and S3B–S3G). This observation is in agreement with two recent large-scale analyses of epistasis [19, 20]. It is worth noting that although only a fraction of all double mutants display

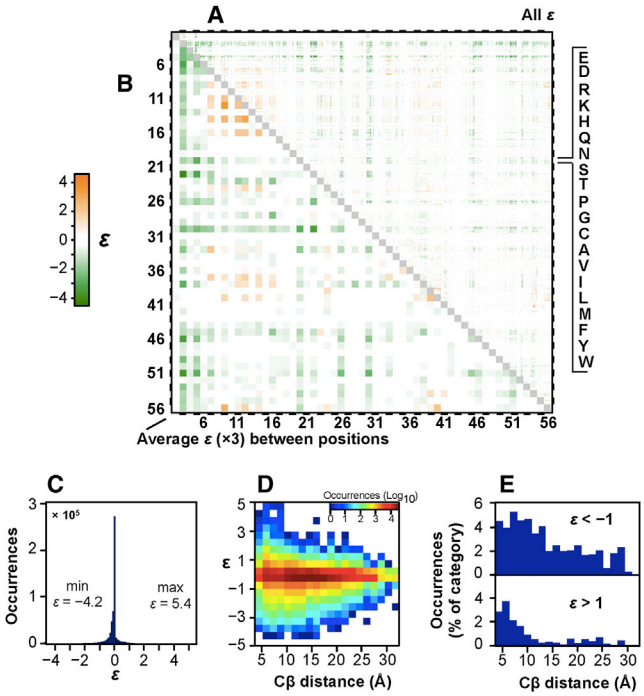


Figure 3. Pairwise Epistasis Map throughout GB1

(A) A heatmap depicting epistasis for 517,278 double mutants (96.5% of all possible). The amino acid order is listed top to bottom and is the same left to right. Each of the 1,485 pairs display 19 × 19 sequence variants.

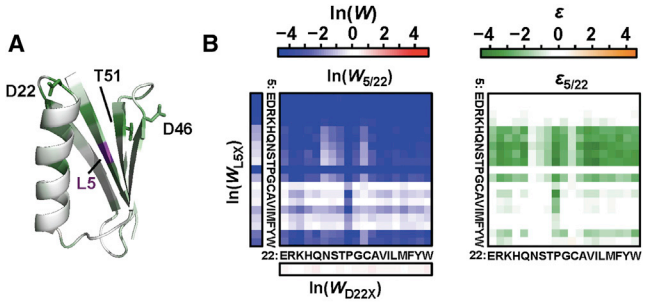
(B) Average ε for all sequence combinations at each positional combination multiplied by a factor of 3 in order to match the range of the color bar.

(C) Histogram showing extent of epistasis (increments of 0.1).

(D) 2D histogram relating ε to Cβ distances in 1PGA [36].

(E) $|\varepsilon| > 1$ as a percentage of total occurrences in each Cβ distance bin.

See also Figure S3.



C

Mutant (*a*/*b*)	ln(W_a)	ln(W_b)	W_{ab}	ln(W_{ab}) (CI low)	ε (CI low)
L5T/A26G	−0.54±0.13	0.14±0.006	0.008±0.005	<−4.61	<−4.21
Y3H/L5T	0.14±0.08	−0.54±0.13	0.006±0.0009	<−4.61	<−4.21
L5T/I6G	−0.54±0.13	0.13±0.11	0.004±0.0027	<−4.61	<−4.20
L5T/D22I	−0.54±0.13	0.36±0.04	0.011±0.0040	<−4.36	<−4.18
L7P/T51D	−0.65±0.03	0.21±0.07	0.006±0.009	<−4.61	<−4.16

Figure 4. Positions Display Negative Epistasis in General Independent of Amino Acid Combination

(A) Average ε between position 5 and all other positions in GB1 (1PGA) [36]. Distant surface residues that demonstrate negative epistasis are labeled.

(B) Binding fitness and epistasis for all 361 combinations of substitutions for Leu5 and Asp22. Leu5 is a critical core residue that is sensitive to mutation, whereas Asp22 is part of a helix-stabilizing N-capping motif near the binding surface, where substitutions are generally well tolerated for binding function.

(C) The most significant values of negative epistasis are listed. Each pair includes mutations that are expected to destabilize the structure but alone do not unfold the protein or significantly disrupt affinity. A Poisson-based 90% confidence interval was used to generate an upper boundary for binding fitness, thereby enabling a conservative estimate of epistasis.

See also Figure S4.

$|\varepsilon| > 1$ (~4%), there are nonetheless thousands of such epistatic pairs (Figures S3B and S3C). We also show that epistasis is similarly rare when calculated using another common epistasis model, the product model (Figures S3E and S3F) [19]. Importantly, due to the low frequency or small magnitude of epistatic effects, the observed double-mutant DFE is nearly identical to the expected distribution (Figures S3H–S3L). Although lethal double mutants are slightly more frequent than would be predicted based on a model without epistasis (Figure S3H), this demonstrates the predictability of the distribution of multiple mutant fitness effects in this adaptive landscape. We also show that observed relative epistasis in this experiment closely matches a model of energetic nonadditivity (scaled by −1 / RT; Figures S3M and S3N) [18]. The differences resulting from the nonlinear nature of the relationship between fraction bound and affinity are numerous but relatively small (Figure S3M). This would not be observed in a test of mutational robustness for well-adapted proteins, as is depicted by the DFE in Figure S2H.

As expected, strongly epistatic pairs tend to be close in space, although very large negative epistasis ($\varepsilon < -3$) can be long range (Figure 3D). However, most neighboring residues do not display either form of epistasis (Figure 3E). Even considering interactions within 6 Å (Cβ), only 8.1% display $|\varepsilon| > 1$. Thus, for many mutations, binding fitness is independent of the background in which one appears. For positions that are energetically coupled, double mutants might be predicted to display either negative or positive epistasis depending on the physicochemical nature of the two amino acid substitutions. To highlight an example, the maps showing ln(*W*) and ε for all 361 amino acid combinations at positions 32 and 36 are enlarged (Figures S3O and S3P). However, an interesting observation from Figures 3A and 3B is that some positional combinations, including long-range combinations, display either negative or positive epistasis in general.

General Negative Epistasis throughout GB1

We wanted to determine what mechanism could explain general patterns of epistasis independent of specific amino acid identities. For example, core mutations, such as those at position 5 (Leu), display general negative epistasis with other positions throughout the domain (Figures 3B and 4A). In addition to general negative epistasis between substitutions at position 5 and other core positions, long-distance negative epistasis occurs between position 5 and surface positions within the stable β3-β4 loop [22, 48], as well as substitutions for Asp22, a helical capping residue (Figure 4A). Figure 4B highlights ln(*W*) and ε for all 361 amino acid combinations within positions 5 and 22. The threshold robustness model [25–27] (Figures S4A–S4C) may explain the pervasive negative epistasis exhibited between these and similar residues. Most proteins are marginally stable [11, 12] yet withstand destabilizing mutations that do not significantly decrease the fraction of folded protein. However, when two such destabilizing substitutions combine, the stability "reservoir" can be exhausted thus resulting in a decrease in the fraction of

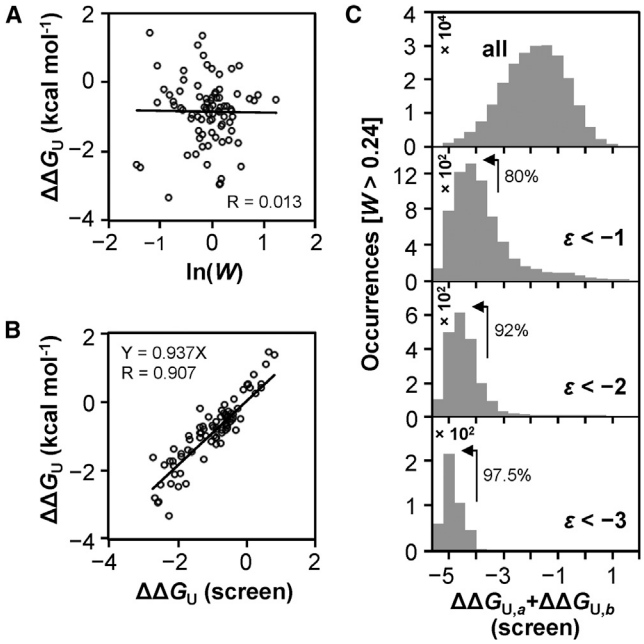


Figure 5. Relationship between Structural Stability Effects and Epistasis

(A) Comparison of ln(W) to free energy of unfolding relative to the WT ($\Delta\Delta G_U$) reported in the literature.

(B) The predicted thermodynamic stability of 82 single mutants compared to $\Delta\Delta G_U$ values reported in the literature (see Table S4). $\Delta\Delta G_U$ predicted by the screen are median values identified by estimating the change in fraction of unfolded protein in five destabilized mutant backgrounds. This analysis was limited to variants displaying $W > 0.24$ (709 of 1,045), as lower fitness values did not produce sufficient dynamic range to measure decreased structural stability.

(C) Histograms of $\Delta\Delta G_{U,a} + \Delta\Delta G_{U,b}$ showing how the distribution of predicted stability changes as the magnitude of negative epistasis increases. The percentage of each epistasis category resulting from combinations of significantly destabilized mutations (left of the arrow) is listed.

See also Figure S5 and Table S4.

native protein and a concomitant loss in function (Figures S4A–S4C). Thus, additive stability effects produce nonadditive functional effects. This model is consistent with the observation that large values of negative epistasis can be long range (Figures 3D and 3E). The threshold robustness model is also consistent with the observation that the combination of buried polar residues at position 5 and substitutions at 22 that abolish a helical capping motif display some of the largest values of negative epistasis observed (Figures 4C and S4D).

Structural Stability and Functional Epistasis

We further examined to what extent structural effects could account for examples of either negative or positive epistasis. To do this, we developed a method to estimate change in free energy of unfolding ($\Delta\Delta G_U$) for single mutants from the binding data. We found that ln(W) is uncorrelated to $\Delta\Delta G_U$ reported in the literature as expected for destabilizing mutations that remain folded at the screen temperature (Figures 5A and S4A and Table S4). However, for partially unfolded mutants, addition of a second mutation will increase the fraction unfolded (f_U) if destabilizing and, conversely, will increase the native fraction (f_N) if stabilizing. As noted above, the threshold robustness effect can be explained as additive stability effects that produce nonadditive functional effects (Figure S4). We hypothesized that certain mutants might be identified that satisfy the condition $W_a = f_{N,a}$, and if these backgrounds are generally noninteracting other than through stability effects, we can estimate $f_{N,ab} = W_{ab} / W_b$. The predicted $f_{N,ab}$ can then be used to estimate structural stability of single mutants (b) by $\Delta\Delta G_U = -RT \times \ln(f_{U,ab} / f_{N,ab}) + RT \times \ln(f_{U,a} / f_{N,a})$.

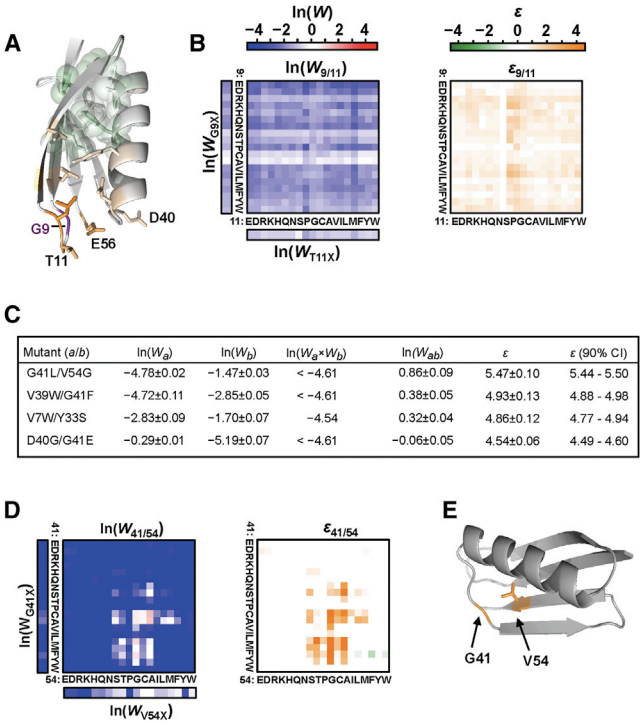
The large number of GB1 variants characterized in the literature provided a substantial reference to identify stability effects from the binding data. An automated analysis was performed that identified multiple background mutations (a) that generated $\Delta\Delta G_U$ values that correlate well with the values found in the literature. These backgrounds therefore satisfy the conditions stated above. This method is limited however to mutants (b) with sufficient fitness to produce a dynamic range in observed fitness (ab; those with $W_b < 0.24$ correlated poorly). An average of the values generated from five reference backgrounds (Y3A, Y3C, L5N, L5S, and F30N) was highly correlated to 82 $\Delta\Delta G_U$ values published in the literature with a slope of 0.94 and a correlation coefficient (R) of 0.907. We note that this correlation is very good considering variability in experimental $\Delta\Delta G_U$ calculations illustrated by Potapov et al. [49]. They show that the correlation (R) between 406 pairs of $\Delta\Delta G_U$ values reported for identical protein variants is 0.86.

In order to estimate the importance of the threshold robustness effect in shaping the GB1 double-mutant fitness landscape, we estimated the structural stability of double mutants by summing $\Delta\Delta G_U$ and determined the number of occurrences of significant negative epistasis for different predicted double mutant stabilities (Figure 5C). For some combinations, $\Delta\Delta G_U$ will not be additive (Figures S5A–S5C and Table S5), which can mitigate the threshold robustness effect. However, as negative epistasis becomes more significant, double mutants predicted to be very unstable account for most occurrences (Figure 5C). For example, 97.5% of all double mutants displaying $\varepsilon < -3$ are predicted to be at least 4 kcal mol^{-1} less stable than WT GB1. Thus, our study empirically demonstrates the extent of the threshold robustness model in functional epistasis.

It is also expected that positive epistasis will arise from stability effects. This will occur if a stabilizing mutation increases the fraction of native protein in the background of a highly destabilized mutant that is partially unfolded at room temperature. Stabilizing mutations have been shown to set the stage for evolution by permitting adaptive mutations that are destabilizing alone [10, 50, 51]. However, it is known that stabilizing mutations are lower in frequency and in magnitude compared to destabilizing mutations [12], and this is corroborated by our screen (Figure S5D). Thus, additive effects from the smaller number and magnitudes of stabilizing mutations overall contribute less to epistasis in comparison with additive effects from combinations of two destabilizing substitutions (Figures S5E–S5J).

General Positive Epistasis within a Dynamic Region

In addition to general negative epistasis, it is also apparent that combinations of mutations within a smaller group of positions display positive epistasis on average (Figure 3B). One position is A24, which shows that positive epistasis is correlated with low-fitness positions and negative epistasis is correlated with high-fitness positions (Figure S6A; see also Figures S2C and S2D). Other positions that display positive epistasis in general include residues within the β1-β2 loop (7, 9, 11), β strand 2 (12, 14, 16), C-terminal end of the α helix through the α-β3 loop (33, 37, 38, 40), and C-terminal β4 residues



Mutant (*a*/*b*)	ln(W_a)	ln(W_b)	ln(W_a×W_b)	ln(W_{ab})	ε	ε (90% CI)
G41L/V54G	−4.78±0.02	−1.47±0.03	< −4.61	0.86±0.09	5.47±0.10	5.44 - 5.50
V39W/G41F	−4.72±0.11	−2.85±0.05	< −4.61	0.38±0.05	4.93±0.13	4.88 - 4.98
V7W/Y33S	−2.83±0.09	−1.70±0.07	−4.54	0.32±0.04	4.86±0.12	4.77 - 4.94
D40G/G41E	−0.29±0.01	−5.19±0.07	< −4.61	−0.06±0.05	4.54±0.06	4.49 - 4.60

Figure 6. Positions that Display General Positive Epistasis Independent of Amino Acid Combination

(A) Average ε for all pairwise combinations with position 9. Glu56, which couples the two dynamic loops through H bonds, is highlighted.

(B) Epistasis and binding fitness for all 361 combinations of substitutions for G9 and T11, which are located within the highly dynamic β1-β2 loop.

(C) The 20 most significant examples of positive epistasis include double mutants from four pairwise positional combinations. The double mutants displaying the largest value of positive epistasis per positional pair are listed. These combinations include neighboring residues within or at the edge of the conformationally dynamic region that overall demonstrates pervasive positive epistasis (see Figures 3A and 3B). For calculating epistasis, we limited expected fitness by $\ln(W_a \times W_b) \geq \ln(0.01)$ to minimize spurious epistasis values for lethal or nearly lethal double mutants resulting from nonmeaningful predicted fitness values below the background ($\sim W = 0.01$).

(D) Fitness and epistasis for all double mutants including positions 41 and 54.

(E) Exchanging volume from core residue Val54 to Gly41 demonstrates the most extreme value of positive epistasis.

See also Figure S6.

(54, 56) (Figures 3B and S6B–S6D). These residues participate in a network of residues that undergo correlated conformational dynamics [52–54]. Remarkably, the pattern of general positive epistasis seen in Figure 3B is very similar to that of correlated NH bond vector motions modeled by Lange et al. [54].

Combinations of mutations within the 12 positions listed in the dynamic region account for 49% of epistasis values >1, while accounting for only 4.4% of all pairs. Figure 6A shows the structure of GB1 depicting the average epistasis between substitutions for Gly9 and mutations at all other positions. This region directly contacts IgG-FC through a main-chain H bond between the Val39 carbonyl oxygen and Asn434 on IgG [21]. This loop is coupled to the β1-β2 loop through H bonds between the C-terminal Glu56 carboxylate with the Asp40 and Lys10 amides [36] (Figure 6A). The dynamic region extends through β strand 2, which is located on the opposite side of GB1 relative to the IgG-FC binding surface. Note that several mutations within this dynamic region also display slightly negative epistasis on average with substitutions in the protein core (Figures 3B and 6A). This is consistent with the threshold robustness effect as such substitutions are predicted to decrease the stability of the structure.

Many of the residues in the coupled, dynamic region of GB1 are generally sensitive to substitution (Figure 2C). For example, all 361 possible combinations of substitutions for G9 and T11 are highlighted (Figure 6B). The data show that when one mutation reduces fitness, an additional mutation in this region imparts a diminished negative effect. We constructed G9A, T11A, and the double mutant G9A/T11A to validate this epistatic effect (Figure S3A). This validation also confirms that subtle changes in amino acid identity in this region can have a significant effect on binding fitness from a distance (Figures 2A and 2C).

In some cases, combinations of substitutions in the dynamic region result in dramatic reversal of lethal fitness to positive fitness (Figure 6C). The most extreme example, G41L/V54G, results in the exchange of volume from the C-terminal core residue Val54 to the α-β3 loop (Figures 6D and 6E). However, how the loop conformation can change to accommodate this swap is not intuitive either by manual inspection or through computational analysis using the parameters described by Kellogg et al. [55]. Interestingly, a highly diverged homolog of protein G of identical length demonstrates the sequence variation 41L/54G (Figure S6E). Furthermore, analysis by EVfold shows that this is a highly coevolving pair [56]. In summary, this analysis has uncovered an important role for residues involved in a dynamic network in contributing to GB1 function and identified how nonadditivity in this region, in some cases extreme, affects the double mutant fitness landscape.

Impact of Epistasis on Adaptive Pathways

While context independent fitness effects generally dominate the mutational landscape of GB1 (Figures 3C–3E), epistasis may promote or limit mutational walks in sequence space (Figure S7A). While most pairs do not display large positive epistasis, there are 37,405 pairs (7.2%) that display ε > 0.15. We wanted to determine how positive epistatic effects are distributed throughout the domain. We calculated fitness for each single mutant in all alternative backgrounds [$\ln(W_a') = \ln(W_{ab}) - \ln(W_b)$] and show the range between the highest [if $\ln(W_{ab}) > -2$] and lowest values in comparison to the fitness in the WT background (Figure 7A). Many deleterious mutations display significantly improved fitness in at least one of the 1,026 possible non-WT backgrounds. In fact, of the 678 single mutations that are deleterious in the WT background, the fitness of 429 can reverse in sign [$\ln(W_a') > 0$] and are thus compensable. Even considering only beneficial double mutants [$\ln(W_{ab}) > 0$], more than one-third of the deleterious mutations (240 of 678) reverse in sign in at least one alternative mutational background and are therefore "cryptically beneficial" (Figures 7B and S7C).

Discussion

With next-generation sequencing, the number of sequence variants in highly diverse populations can be counted before and after laboratory-designed tests of fitness, thereby quantifying evolution [30, 34, 41–44, 57–60]. An important question related to such studies is how often the observed fitness effects would change in the background of other mutations.

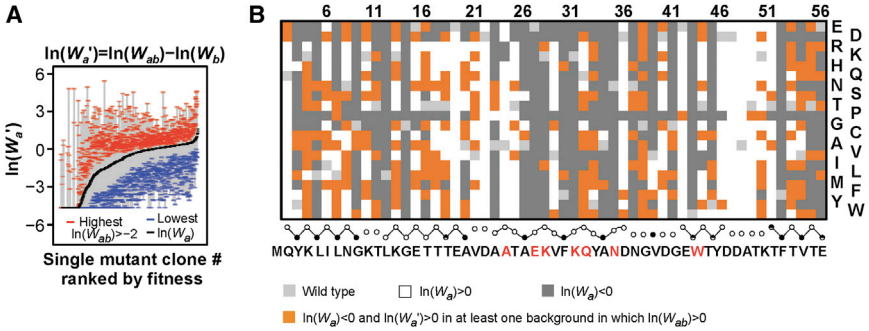


Figure 7. The Fitness Effects of Many Mutations Change Dramatically Depending on the Background in which They Appear

(A) The range is bound by a blue dash for the lowest fitness in any of the 1,026 possible alternate backgrounds and by a red dash for the highest fitness. The highest fitness values are limited to double mutants displaying $\ln(W_{ab}) > -2$.

(B) A map showing deleterious single mutants that are beneficial in at least one alternative mutational background even while limiting double-mutant fitness greater than the WT (orange).

See also Figure S7.

Fields and colleagues have demonstrated the ability to characterize thousands of single and double mutants in segments of protein domains and thus make important conclusions on the frequency and nature of epistasis in protein function [19, 20]. In this article, we observe the first two steps of all evolutionary pathways in the recognition of IgG by GB1 and therefore observe how fitness profiles change in all alternative mutational backgrounds. This comprehensive analysis determines how often deleterious effects are compensated and beneficial effects are negated for all mutations in an entire protein molecule.

The fitness profile in Figure 1B and the DFE in Figure S2E show that the stringency of this fitness challenge is analogous to evolution of new function. It is well understood that the highest affinity possible will often not be selected for function in vivo [28]. However, the beneficial mutations we identify in vitro are found in natural protein G homologs (Figure S2B), and one homolog that does not benefit from tandem duplication has seven mutations that are all adaptive in this screen (see Figure S2B). Furthermore, there are ligand pairs that demonstrate a functional demand for exceptional affinity [61], including for IgG binding proteins similar to GB1 [62]. Such an adaptive landscape as described in this experiment could possibly be analogous to natural evolution in viral receptor host switching. For example, mutations found after adaptation of SARS from civet to human show enhanced affinity to receptor in vitro, and those mutations enhanced viral infectivity in cell culture [35]. Furthermore, affinity-based adaptation can occur if ligand concentrations decrease, for example, as observed in increased affinity for O_2 in high-altitude deer mouse hemoglobin [7].

We show common biophysical mechanisms for both negative and positive epistasis, including how additive stability effects produce functional epistasis. Although the environment of the cell will modulate the concentration of functional protein compared to what is observed in vitro [63], there is a clear relationship between protein stability and fitness in cells and viruses [10, 27, 64]. The cooperative nature of protein folding creates an inherently epistatic effect from additive stability effects [25–27]. In this experiment, additive effects of destabilizing mutations account for nearly all examples of very large negative epistasis. That destabilizing mutations are both more common and larger in magnitude compared to stabilizing mutations [12] explains why there are more significant negative epistatic effects compared to positive epistatic effects in this experiment. Stabilizing mutations might display stronger epistatic effects in vivo, however, by counteracting degradation or aggregation, such as been demonstrated in β-lactamase evolution [8].

Furthermore, we demonstrated that long-range deleterious fitness effects throughout a dynamic region are not additive and therefore that mutations in this region display positive epistasis in general. These observations mirror results from extensive characterization of PDZ domains, which also display long-range mutational sensitivity in dynamic regions [43, 65, 66]. This effect can be exploited for allosteric modulation in nature or through engineering [67, 68]. The most substantial occurrences of positive epistasis were found in the region between positions in which two highly deleterious mutations combine to produce neutral or beneficial double mutants. A similar "hot spot" of epistasis predicted to produce a conformational switch that removes unfavorable interactions was also seen in an exceptionally high-throughput mutagenic study of an RRM domain [20].

The results of this paper reconcile observations about the importance of epistasis in adaptive evolution despite the rarity of it. We can see that while it should not be expected that mutations have different fitness in alternative backgrounds, most mutations can have a very different effect in at least one alternative genetic background. Cryptically beneficial mutations [$\ln(W_a') > 0$ and $\ln(W_{ab}) > 0$] are found throughout 43 positions in the 55-residue domain. Furthermore, while the WT is optimal at 17 positions, compensatory mutations reveal beneficial mutations within ten of these 17 positions even when limiting $\ln(W_{ab}) > 0$. Thus, while sign epistasis limits pathways of adaptation, it at the same time facilitates sequence change in light of mutational sensitivity.

Experimental Procedures

See the Supplemental Experimental Procedures for complete details for library construction, mRNA display and affinity enrichment, sequencing, data analysis, and validation.

Calculation of Structural Stability Effects

In order to estimate change in fraction folded, we assumed there will be mutational backgrounds (*a*) in which $W_{ab} / W_b = f_{N,ab}$. This can occur if the reference mutations are partially folded but neutral in the native state, if the test mutant (*b*) is fully folded in the native state, and if the two mutations do not interact functionally (only through additive thermodynamic stability effects). Thus, these conditions mean, given that the observed *W* is a product of the fraction folded and fitness of the native state ($W = f_N \times W_N$), that the background mutations must satisfy $f_{N,a} = W_a$ ($W_{N,a} = 1$) and the test mutants (*b*) must satisfy $f_{N,b} = 1$. Therefore, for pairs that are energetically additive, $W_{N,ab} = W_b$. Substituting into $W_{ab} = f_{N,ab} \times W_{N,ab}$ gives us $f_{N,ab} = W_{ab} / W_b$. We automatically converted the 82 test mutants from the literature (Table S4) into $f_{N,ab}$ using all suitable backgrounds ($0 < W_a < 1$) and then into relative free energy of unfolding. At equilibrium, $k_F \times [U] = k_{Un} \times [N]$, and thus $f_{U,ab} / f_{N,ab} = k_{Un} / k_F$, which by definition equals K_{Un}, and therefore $\Delta\Delta G_U = -RT \times \ln(f_{U,ab} / f_{N,ab}) + RT \times \ln(f_{U,a} / f_{N,a})$. Numerous substitutions at positions Y3, L5, F30, and A26 produced highly correlated data. The average $\Delta\Delta G_U$ values from the top five backgrounds from positions 3, 5, and 30 (Y3A, Y3C, L5N, L5S, and F30N) produced highly correlated data with a slope close to 1.

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2014.09.072.

Author Contributions

C.A.O. designed the experiments, performed the experiments, analyzed the data, and wrote the manuscript. N.C.W. created all scripts, analyzed the data, and revised the manuscript. R.S. designed experiments, provided intellectual support, and revised the manuscript.

Acknowledgments

R.S. was supported by the NIH (R21AI110261 and R01AI085583). C.A.O. was supported by NCI Cancer Education Grant R25 CA 098010. N.C.W. was supported by a UCLA MBI Whitcome grant. We thank X. Li for technical advice. We thank T.T. Wu, L.Q. Al-Mawsawi, T.T. Takahashi, R.W. Roberts, and J. Lloyd-Smith for comments.

Received: June 5, 2014
Revised: August 6, 2014
Accepted: September 25, 2014
Published: October 16, 2014

References

1. Breen, M.S., Kemena, C., Vlasov, P.K., Notredame, C., and Kondrashov, F.A. (2012). Epistasis as the primary factor in molecular evolution. Nature *490*, 535–538.
2. Kimura, M. (1991). Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. Proc. Natl. Acad. Sci. USA *88*, 5969–5973.
3. Weinreich, D.M., Watson, R.A., and Chao, L. (2005). Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution *59*, 1165–1174.
4. Povolotskaya, I.S., and Kondrashov, F.A. (2010). Sequence space and the ongoing expansion of the protein universe. Nature *465*, 922–926.
5. Harms, M.J., and Thornton, J.W. (2013). Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat. Rev. Genet. *14*, 559–571.
6. Bridgham, J.T., Ortlund, E.A., and Thornton, J.W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature *461*, 515–519.
7. Natarajan, C., Inoguchi, N., Weber, R.E., Fago, A., Moriyama, H., and Storz, J.F. (2013). Epistasis among adaptive mutations in deer mouse hemoglobin. Science *340*, 1324–1327.
8. Weinreich, D.M., Delaney, N.F., Depristo, M.A., and Hartl, D.L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. Science *312*, 111–114.
9. Kryazhimskiy, S., Dushoff, J., Bazykin, G.A., and Plotkin, J.B. (2011). Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet. *7*, e1001301.
10. Gong, L.I., Suchard, M.A., and Bloom, J.D. (2013). Stability-mediated epistasis constrains the evolution of an influenza protein. Elife *2*, e00631.
11. Bloom, J.D., Raval, A., and Wilke, C.O. (2007). Thermodynamics of neutral protein evolution. Genetics *175*, 255–266.
12. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L., and Tawfik, D.S. (2007). The stability effects of protein mutations appear to be universally distributed. J. Mol. Biol. *369*, 1318–1332.
13. Romero, P.A., and Arnold, F.H. (2009). Exploring protein fitness landscapes by directed evolution. Nat. Rev. Mol. Cell Biol. *10*, 866–876.
14. Lunzer, M., Golding, G.B., and Dean, A.M. (2010). Pervasive cryptic epistasis in molecular evolution. PLoS Genet. *6*, e1001162.
15. Wells, J.A. (1990). Additivity of mutational effects in proteins. Biochemistry *29*, 8509–8517.
16. Sandberg, W.S., and Terwilliger, T.C. (1993). Engineering multiple properties of a protein by combinatorial mutagenesis. Proc. Natl. Acad. Sci. USA *90*, 8367–8371.
17. Gregoret, L.M., and Sauer, R.T. (1993). Additivity of mutant effects assessed by binomial mutagenesis. Proc. Natl. Acad. Sci. USA *90*, 4246–4250.
18. LiCata, V.J., and Ackers, G.K. (1995). Long-range, small magnitude nonadditivity of mutational effects in proteins. Biochemistry *34*, 3133–3139.
19. Araya, C.L., Fowler, D.M., Chen, W., Muniez, I., Kelly, J.W., and Fields, S. (2012). A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. Proc. Natl. Acad. Sci. USA *109*, 16858–16863.
20. Melamed, D., Young, D.L., Gamble, C.E., Miller, C.R., and Fields, S. (2013). Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. RNA *19*, 1537–1551.
21. Sauer-Eriksson, A.E., Kleywegt, G.J., Uhlén, M., and Jones, T.A. (1995). Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. Structure *3*, 265–278.
22. McCallister, E.L., Alm, E., and Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. Nat. Struct. Biol. *7*, 669–673.
23. Malakauskas, S.M., and Mayo, S.L. (1998). Design, structure and stability of a hyperthermophilic protein variant. Nat. Struct. Biol. *5*, 470–475.
24. Wunderlich, M., Max, K.E., Roske, Y., Mueller, U., Heinemann, U., and Schmid, F.X. (2007). Optimization of the gbeta1 domain by computational design and by in vitro evolution: structural and energetic basis of stabilization. J. Mol. Biol. *373*, 775–784.
25. Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C., and Arnold, F.H. (2005). Thermodynamic prediction of protein neutrality. Proc. Natl. Acad. Sci. USA *102*, 606–611.
26. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D.S. (2006). Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. Nature *444*, 929–932.
27. Tokuriki, N., and Tawfik, D.S. (2009). Stability effects of mutations and protein evolvability. Curr. Opin. Struct. Biol. *19*, 596–604.
28. Pál, G., Kouadio, J.L., Artis, D.R., Kossiakoff, A.A., and Sidhu, S.S. (2006). Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. J. Biol. Chem. *281*, 22378–22385.
29. Weiss, G.A., Watanabe, C.K., Zhong, A., Goddard, A., and Sidhu, S.S. (2000). Rapid mapping of protein functional epitopes by combinatorial alanine scanning. Proc. Natl. Acad. Sci. USA *97*, 8950–8954.
30. Fowler, D.M., Araya, C.L., Fleishman, S.J., Kellogg, E.H., Stephany, J.J., Baker, D., and Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. Nat. Methods *7*, 741–746.
31. Roberts, R.W., and Szostak, J.W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. USA *94*, 12297–12302.
32. Olson, C.A., Nie, J., Diep, J., Al-Shyoukh, I., Takahashi, T.T., Al-Mawsawi, L.Q., Bolin, J.M., Elwell, A.L., Swanson, S., Stewart, R., et al. (2012). Single-round, multiplexed antibody mimetic design through mRNA display. Angew. Chem. Int. Ed. Engl. *51*, 12449–12453.
33. Soskine, M., and Tawfik, D.S. (2010). Mutational effects and the evolution of new protein functions. Nat. Rev. Genet. *11*, 572–582.
34. Firnberg, E., Labonte, J.W., Gray, J.J., and Ostermeier, M. (2014). A comprehensive, high-resolution map of a gene's fitness landscape. Mol. Biol. Evol. *31*, 1581–1592.
35. Wu, K., Peng, G., Wilken, M., Geraghty, R.J., and Li, F. (2012). Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. J. Biol. Chem. *287*, 8904–8911.
36. Gallagher, T., Alexander, P., Bryan, P., and Gilliland, G.L. (1994). Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. Biochemistry *33*, 4721–4729.
37. Lesk, A.M., and Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. J. Mol. Biol. *136*, 225–270.
38. Di Nardo, A.A., Larson, S.M., and Davidson, A.R. (2003). The relationship between conservation, thermodynamic stability, and function in the SH3 domain hydrophobic core. J. Mol. Biol. *333*, 641–655.
39. Xu, J., Baase, W.A., Baldwin, E., and Matthews, B.W. (1998). The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. Protein Sci. *7*, 158–177.
40. Sloan, D.J., and Hellinga, H.W. (1999). Dissection of the protein G B1 domain binding site for human IgG Fc fragment. Protein Sci. *8*, 1643–1648.
41. Wu, N.C., Young, A.P., Al-Mawsawi, L.Q., Olson, C.A., Feng, J., Qi, H., Chen, S.H., Lu, I.H., Lin, C.Y., Chin, R.G., et al. (2014). High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. Sci. Rep. *4*, 4942.

42. Roscoe, B.P., Thayer, K.M., Zeldovich, K.B., Fushman, D., and Bolon, D.N. (2013). Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J. Mol. Biol. *425*, 1363–1377.
43. McLaughlin, R.N., Jr., Poelwijk, F.J., Raman, A., Gosal, W.S., and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. Nature *491*, 138–142.
44. Qi, H., Olson, C.A., Wu, N.C., Ke, R., Loverdo, C., Chu, V., Truong, S., Remenyi, R., Chen, Z., Du, Y., et al. (2014). A quantitative high-resolution genetic profile rapidly identifies sequence determinants of hepatitis C viral fitness and drug sensitivity. PLoS Pathog. *10*, e1004064.
45. Eyre-Walker, A., and Keightley, P.D. (2007). The distribution of fitness effects of new mutations. Nat. Rev. Genet. *8*, 610–618.
46. Khan, A.I., Dinh, D.M., Schneider, D., Lenski, R.E., and Cooper, T.F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. Science *332*, 1193–1196.
47. Mani, R., St Onge, R.P., Hartman, J.L., 4th, Giaever, G., and Roth, F.P. (2008). Defining genetic interaction. Proc. Natl. Acad. Sci. USA *105*, 3461–3466.
48. Blanco, F.J., Rivas, G., and Serrano, L. (1994). A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. Nat. Struct. Biol. *1*, 584–590.
49. Potapov, V., Cohen, M., and Schreiber, G. (2009). Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. Protein Eng. Des. Sel. *22*, 553–560.
50. Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein stability promotes evolvability. Proc. Natl. Acad. Sci. USA *103*, 5869–5874.
51. Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D.S. (2008). How protein stability and new functions trade off. PLoS Comput. Biol. *4*, e1000002.
52. Markwick, P.R., Bouvignies, G., and Blackledge, M. (2007). Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy. J. Am. Chem. Soc. *129*, 4724–4730.
53. Clore, G.M., and Schwieters, C.D. (2004). Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. Biochemistry *43*, 10678–10691.
54. Lange, O.F., Grubmüller, H., and de Groot, B.L. (2005). Molecular dynamics simulations of protein G challenge NMR-derived correlated backbone motions. Angew. Chem. Int. Ed. Engl. *44*, 3394–3399.
55. Kellogg, E.H., Leaver-Fay, A., and Baker, D. (2011). Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins *79*, 830–838.
56. Marks, D.S., Colwell, L.J., Sheridan, R., Hopf, T.A., Pagnani, A., Zecchina, R., and Sander, C. (2011). Protein 3D structure computed from evolutionary sequence variation. PLoS ONE *6*, e28766.
57. Deng, Z., Huang, W., Bakkalbasi, E., Brown, N.G., Adamski, C.J., Rice, K., Muzny, D., Gibbs, R.A., and Palzkill, T. (2012). Deep sequencing of systematic combinatorial libraries reveals β-lactamase sequence constraints at high resolution. J. Mol. Biol. *424*, 150–167.
58. Thyagarajan, B., and Bloom, J.D. (2014). The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. Elife *3*, e03300.
59. Traxlmayr, M.W., Hasenhindl, C., Hackl, M., Stadlmayr, G., Rybka, J.D., Borth, N., Grillari, J., Rüker, F., and Obinger, C. (2012). Construction of a stability landscape of the CH3 domain of human IgG1 by combining directed evolution with high throughput sequencing. J. Mol. Biol. *423*, 397–412.
60. Hietpas, R.T., Jensen, J.D., and Bolon, D.N. (2011). Experimental illumination of a fitness landscape. Proc. Natl. Acad. Sci. USA *108*, 7896–7901.
61. Schreiber, G., Buckle, A.M., and Fersht, A.R. (1994). Stability and function: two constraints in the evolution of barstar and other proteins. Structure *2*, 945–951.
62. Nitsche-Schmitz, D.P., Johansson, H.M., Sastalla, I., Reissmann, S., Frick, I.M., and Chhatwal, G.S. (2007). Group G streptococcal IgG binding molecules FOG and protein G have different impacts on opsonization by C1q. J. Biol. Chem. *282*, 17530–17536.
63. Hingorani, K.S., and Gierasch, L.M. (2014). Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge. Curr. Opin. Struct. Biol. *24*, 81–90.
64. Mayer, S., Rüdiger, S., Ang, H.C., Joerger, A.C., and Fersht, A.R. (2007). Correlation of levels of folded recombinant p53 in escherichia coli with thermodynamic stability in vitro. J. Mol. Biol. *372*, 268–276.
65. Gianni, S., Haq, S.R., Montemiglio, L.C., Jürgens, M.C., Engström, A., Chi, C.N., Brunori, M., and Jemth, P. (2011). Sequence-specific long range networks in PSD-95/discs large/ZO-1 (PDZ) domains tune their binding selectivity. J. Biol. Chem. *286*, 27167–27175.
66. Petit, C.M., Zhang, J., Sapienza, P.J., Fuentes, E.J., and Lee, A.L. (2009). Hidden dynamic allostery in a PDZ domain. Proc. Natl. Acad. Sci. USA *106*, 18249–18254.
67. Reynolds, K.A., McLaughlin, R.N., and Ranganathan, R. (2011). Hot spots for allosteric regulation on protein surfaces. Cell *147*, 1564–1575.
68. Gunasekaran, K., Ma, B., and Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? Proteins *57*, 433–443.