HHS Public Access

Author manuscript

Proteins. Author manuscript; available in PMC 2015 December 02.

Published in final edited form as:

Proteins. 2008 February 15; 70(3): 915–924. doi:10.1002/prot.21620.

New Insight into Long-Range Nonadditivity within Protein Double-Mutant Cycles

Andrei Y. Istomin[1], **M. Michael Gromiha**[2], **Oleg K. Vorov**[1], **Donald J. Jacobs**[1,*], and **Dennis R. Livesay**[3,*]

[1]Department of Physics and Optical Science, University of North Carolina at Charlotte, Charlotte, NC, USA

[2]Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan

[3]Department of Computer Science and Bioinformatics Research Center, University of North Carolina, Charlotte, NC 28223, USA

Abstract

Additivity principles in chemistry, biochemistry, and biophysics have been used extensively for decades. Nevertheless, it is well known that additivity frequently breaks down in complex biomacromolecules. Nonadditivity within protein double mutant free energy cycles of spatially close residue pairs is a generally well-understood phenomenon, whereas a robust description of nonadditivity extending over large distances remains to be developed. Here, we test the hypothesis that the mutational effects tend to be nonadditive if two structurally well-separated mutated residues belong to the same rigid cluster within the wild type protein, and additive if they are located within different clusters. We find the hypothesis to be statistically significant with *P*-values that range from 10^{-5} to 10^{-6}. To the best of our knowledge, this result represents the first demonstration of a statistically significant preponderance for nonadditivity over long distances. These findings provide new insight into the origins of long-range nonadditivity in double mutant cycles, which complements the conventional wisdom that nonadditivity arises in double mutations involving contacting residues. Consequently, these results should have far-reaching implications for a proper understanding of protein stability, structure/function analyses, and protein design.

Keywords

Free energy nonadditivity; Double mutant cycle; Protein structure; Cooperativity; Network rigidity

Introduction

Nonadditivity found in double mutant free energy cycles and associated changes in the functional properties of proteins has attracted much interest during the past three decades

*Correspondence to: Donald J. Jacobs, Department of Physics and Optical Science, University of North Carolina, 9201 University City Blvd, Charlotte, NC 28223, USA, djacobs1@uncc.edu and Dennis R. Livesay, Department of Computer Science and Bioinformatics Research Center, University of North Carolina, 9201 University City Blvd, Charlotte, NC 28223, USA, drlivesa@uncc.edu.

due to a broad range of implications derived from its studies [see e.g., (1–5)]. Understanding the origins of nonadditivity is crucial for elucidating fundamental mechanisms of protein function such as cooperativity and allostery (6,7). The ability to predict the occurrence of nonadditivity is important for biotechnology applications such as protein stabilization (8, 9) and protein design (10, 11). It is well known that changes in functional properties within double mutant cycles are usually nonadditive when the mutated residues are in direct contact with each other [for review, see (12)]. In this case, the origins of nonadditivity are rather obvious and lie in the properties of the interaction between contacting residues, which in turn are determined by amino acid type and their relative distance and orientation. Formation of hydrogen bonds between contacting residues, when possible, changes both enthalpic and entropic contributions to the unfolding free energy. When two interacting residues are simultaneously mutated to a non-interacting pair (say, by alanine-scanning), the difference away from the sum of the constituent single mutations quantifies the extent of their interaction.

Experimental investigations of double mutant cycles not only provide a direct method to probe the effective thermodynamic coupling between sites based on observed nonadditivity, they are commonly used to infer functional relationships. Moreover, because of biological importance, it is natural to expect that these functional relationships be recapitulated within the evolutionary history of a particular protein family. Coevolutionary metrics, which attempt to identify positions within a multiple sequence alignment that are covarying, have been used to predict thermodynamic coupling (13, 14). While such metrics do a reasonably good job of predicting nonadditivity within structurally proximal sites, Fodor and Aldrich (15) have convincingly shown that coevolutionary predictions of long-range nonadditivity lack statistical significance. Consequently, poor predictability for site-pairs that are structurally well separated generally persists in spite of the high-profile example to the contrary found in the PDZ binding domain by Lockless and Ranganathan (14). Furthermore, Fodor and Aldrich go on to argue that thermodynamic coupling is insensitive to protein sequence, based on their observation that nonadditivity is not limited to the subset of highly covarying residue positions, which goes against the premise of evolutionarily conserved energetic pathways championed by Ranganathan et al. (14, 16–18). Rather, they conclude that the success of coevolutionary metrics to predict nonadditivity is fortuitously well correlated with reliable predictions of structurally proximal pairs. We now turn our attention to spatially long-range nonadditivity issues.

Double mutations separated over large distances were observed to be thermodynamically coupled when the protein undergoes a substantial conformational change (3, 19). Unfortunately, this empirical observation is difficult to translate into a priori criteria to make predictions. To the best of our knowledge, no computational method has been shown to make statistically significant predictions of nonadditivity in long-range double-mutant free energy cycles, which is unfortunate because nonadditivity between structurally remote sites is certainly not an uncommon phenomenon. Moreover, thermodynamic coupling (nonadditivity property) between two distant mutants highlights the potential inadequacy of additivity principles invoked within the vast majority of theoretical models for protein folding and protein-substrate interactions. The importance of understanding the applicability of additivity principles in computational biology is comparable to, or even higher than, the

importance of understanding the accuracy of mean field approximations in quantum physics and chemistry (20). The most sophisticated computational approaches to protein dynamics, all-atom molecular dynamics simulations, rely on local additivity by employing two-body atom-atom interaction potentials. Nevertheless, despite the local additivity assumption, global nonadditivity and other nonlocal effects can be predicted (21). Furthermore, in virtually all statistical mechanical models that employ free energy decomposition schemes, additivity assumptions are not only applied to the potential energy (enthalpy) contributions, but also to entropy contributions. The only exception we are aware of that goes beyond the usual additivity assumption is the Distance Constraint Model (DCM) that augments concepts from constraint theory to a free energy decomposition scheme (22). The DCM is an all-atom statistical mechanical model that explicitly accounts for nonadditivity in conformational entropy; it has been able to successfully quantify protein stability and flexibility relationships (23–26), as well as unfolding pathways and nonadditivity effects in free energy upon structural reconstitution of thioredoxin (25). The degree of nonadditivity upon reconstitution was found to linearly correlate with the degree of mechanical rigidity within the protein (25).

In this paper we test the hypothesis that nonadditivity in double mutant free energy cycles occurs pre-dominantly when both mutation sites are located within the same rigid region of the wild type protein structure. A hypothesis that is very similar to the one analyzed here has been first proposed by Reichmann et al. (27) to quantify the nonadditivity of changes in binding energies between protein-protein interfaces. In Ref. (27), this hypothesis was confirmed experimentally on the example of the interaction between TEM1-β-lactamase and β-lactamase inhibitor protein.

We model a three-dimensional protein structure as a set of atoms connected by chemical bonds which in turn are modeled by rigid sticks of fixed length and mathematically can be represented as constraints on the corresponding interatomic distances. We identify all sets of atoms that form rigid clusters by employing FIRST (Floppy Inclusion and Rigid Substructure Topography) software (28). The FIRST package is based on a “pebble game” algorithm (29). In less than a second of the computational time, it identifies all flexible and rigid regions in a protein structure and provides the essential information on local rigidity properties, such as whether a given atom belongs to a particular rigid cluster, over-constrained region, or region participating in a correlated motion. FIRST methodology has been successful in tackling a variety of problems, e.g., in providing a unified view of phase transitions in proteins and network glasses (30), in complementing insight from protein MD simulations (31), and is a basis for the thermodynamic DCM (22).

To quantitatively test our hypothesis, we first define a change in the free energy of unfolding upon amutation at site i,

$$\Delta\Delta G^{(i)} \equiv \Delta G^{(i)} - \Delta G^{(\mathrm{wild})}, \quad (1)$$

where $\Delta G^{(\mathrm{wild})}=G_{\mathrm{f}}^{(\mathrm{wild})}-G_{\mathrm{u}}^{(\mathrm{wild})}$ and $\Delta G^{(i)}=G_{\mathrm{f}}^{(i)}-G_{\mathrm{u}}^{(i)}$. The subscripts "f" and "u" stand for folded and unfolded states respectively, and the superscript "*i*" indicates a protein with a mutation at *i*-th residue. We also define a relative nonadditivity parameter,

$$\delta^{(ij)} \equiv \left| \frac{\Delta\Delta G^{(ij)} - (\Delta\Delta G^{(i)} + \Delta\Delta G^{(j)})}{\Delta\Delta G^{(ij)}} \right| \times 100\%. \quad (2)$$

and a relative nonadditivity threshold parameter, $\delta_{\max}^{(ij)}$.

A meaningful classification of double mutations into nonadditive and additive ones is a nontrivial and non-unique procedure because it can be performed using a variety of criteria depending on the objective. In this work, we are interested in a statistical analysis of *qualitative* manifestations of the nonadditivity phenomenon. A change in the free energy of unfolding upon a mutation is a global, cooperative property of a protein very much dependent upon its conformational state and its environment. Therefore, we are inclined to think that partitioning of mutants into nonadditive and additive for our purpose should be performed using a criterion that is based on relative (i.e., normalized) changes in the free energy, rather than only on their absolute values which are often referred to as "thermodynamic couplings". In our calculations, we shall adopt the following criterion:

$$\mathrm{Double\,Mutation}^{(ij)} \equiv \begin{cases} \mathrm{additive}, \delta^{(ij)} < \delta_{\max}^{(ij)} \\ \mathrm{nonadditive}, \delta^{(ij)} \geq \delta_{\max}^{(ij)} \end{cases}, \quad (3)$$

where we have chosen $\delta_{\max}^{(ij)}=20\%$. Under this criterion, given a free energy change of, e.g., 5.0 kcal/mol in a double mutant, a difference between free energy changes of 0.9 kcal/mol will be treated as additive, while that of 1.1 kcal/mol will be treated as nonadditive. This criterion is usually reasonable, although in cases when the free energy changes are so small that they are comparable to experimental error bars (whose common values are of the order of 0.3 kcal/mol), it may potentially lead to misassigning such mutations to nonadditive; this issue will be addressed below. We hypothesize that mutational effects tend to be nonadditive when the two mutation sites *i* and *j* are located within the same rigid cluster and additive when they are not. As follows from our results, there is indeed a statistically significant bias toward such a preferential occurrence of the nonadditivity.

Results and Discussion

Nonadditivity versus spatial proximity of mutation sites

First, we have investigated the correlation between nonadditivity and the spatial distance between mutation sites. Among 232 double mutant cycles, we have compared the preponderance of nonadditivity to occur within 112 "contact" double mutants (defined as d_{ij} ≤6 Å) to 120 "long-range" mutants (defined as d_{ij} > 6 Å). The distance between two residues is defined as the shortest distance between any heavy atom belonging to the first residue and any heavy atom belonging to the second residue. The threshold value of 6 Å was chosen upon the following considerations. In general, to ensure that two residues in a wild-

type structure are not interacting directly via van der Waals-like potentials, a threshold value of 4 Å would suffice. However, in the double mutant protein, the two new residues might be in contact, depending on the amino-acid size differences and relative orientation. The value of 6 Å was thus chosen to ensure that the two residues are in contact neither in the wild-type nor in the double mutant structure. In Fig. 1 we show a scatter plot for changes in free energies of unfolding, $\Delta\Delta G^{(ij)}$, for double mutants, and sums of free energy changes for two single mutants, $\Delta\Delta G^{(i)} + \Delta\Delta G^{(j)}$. The data are partitioned into two groups corresponding to contact (red circles) and long-range (green squares) mutants; the best linear fits to the data in these two groups are shown, with the correlation coefficient of $R = 0.84$ for contact mutants and $R = 0.97$ for long-range mutants, respectively. (The Protein Data Bank accession numbers for wild-type structures, mutated residues, numerical values for $\Delta\Delta G$, value of $\delta^{(ij)}$ in Eq. (2), and the spatial distance between mutates residues are given in Table I of the Supplementary Data.) It is clear that the slope, a, of the best linear fit to the data for contact mutants ($a = 0.54$) is less than that for long-range mutants ($a = 0.88$). This indicates a prevailing tendency that when two mutation sites are in contact in a double mutant structure, mutation at one site substantially reduces the effect of mutation at the second site on, as compared to the case if mutations only at one of the sites were made. [In (19) such a trend was termed “sub-additivity”, in contrast to the less common case of “super-additivity” when the presence of one mutation increases the effect of the second mutation.] For long-range mutants, the slope of the linear fit is still less than unity, thus indicating some “sub-additivity”. These findings are in general agreement with the conventional understanding of nonadditivity (3).

As a point of reference to compare the results presented below, we begin by testing the statistical significance of the traditional viewpoint. Using the nonadditivity threshold $\delta^{(ij)}_{\max}=20\%$ and invoking the naïve hypothesis that nonadditivity occurs within contacting mutants and additivity for long-range mutants, our 232 double mutant dataset contains 112 contact mutant pairs out of which 79 are nonadditive and 33 are additive, and 120 long-range mutant pairs out of which 70 are additive and 50 are nonadditive. If the spatial distance had no influence on the likelihood of nonadditivity, one would expect that out of 112 contact mutants there should be approximately 56 additive and 56 nonadditive ones, and out of 120 long-range mutants there should be 60 additive and 60 nonadditive ones. The probability that the observed distributions of 79+33 and 70+50 have occurred by pure chance (i.e., the *P*-value of the null hypothesis) can be calculated using a binomial distribution function because because the number of trials is finite (equal to 232) and the outcome is binary (agrees/disagrees with the hypothesis). This probability is equal to 3.4×10^{-7} and its low value demonstrates that the conventional wisdom vis-à-vis mutant additivity/nonadditivity is statistically significant.

Nonadditivity within shared rigid clusters

Next, we test our hypothesis of the preferential occurrence of nonadditivity when the two mutated residues are located within the same rigid cluster and additivity when they are located in distinct clusters. Among 232 double mutant cycles, nonadditivity (assuming 20% threshold) was found in 129. For each of the 232 double mutants, we have calculated the decomposition of the corresponding wild-type protein structure into rigid clusters and

examined whether the two mutated residues are located within the same cluster. These all-atom calculations were performed using the FIRST software that is freely available at Flexweb (28). The decomposition of protein structure into rigid clusters is determined by the constraint topology that is defined by the covalent bonding, hydrophobic contacts, salt bridges and the hydrogen bond network in a protein. Salt bridges and hydrogen bonds (both referred to as H-bonds hereafter) are defined by criteria on distances and angles between the donor, acceptor, and the corresponding hydrogen atom. Some of the H-bonds satisfying these criteria are ranked strong or weak based on their binding energy. If all H-bonds satisfying local geometric criteria are included in the network, the protein structure generally appears rigid throughout. Conversely, if none of these cross-linking H-bonds are included, the protein is predicted to be without "structure" in the normal sense of the word. We have thus performed such calculations for a range of values of the H-bond energy cut-off, E_c, from −0.05 to −5.00 kcal/mol. Because the rigid cluster decompositions are calculated at an all-atom level, the answer to the question whether or not a pair of residues (within the wild type protein) belong to the same cluster is not unique. Here we have adopted the following *criterion*: If each of the two residues shares *n* or more atoms (including heavy and hydrogen atoms located in the backbone or sidechain) with the same rigid cluster, they are considered to belong to the same cluster; otherwise, they are considered to belong to different clusters.

In Fig. 2(A), we present fractions of double mutant cycles supporting our hypothesis, i.e., the fraction of nonadditive mutants with mutation sites located within the same cluster, *and* the fraction of additive mutants with mutation sites belonging to different clusters. Due to the purely mechanistic nature of the current approach, there is no possibility of defining a "correct" value of the H-bond energy cut-off, E_c, uniquely. Therefore, our results are shown as functions of E_c for six different values of the minimum number of shared atoms, $n = 1, 2, .., 6$. It appears that when there are shared atoms, most frequently those are N, H_N, C_α, H_α, C, and O atoms along the backbone. One sees that the *right side* of the graph (E_c close to −0.05 kcal/mol) corresponds to *almost completely rigid* structures where almost any two residues belong to the same large cluster. On the contrary, the *left side* (E_c close to −5.00 kcal/mol) corresponds to *very flexible* structures where almost any two residues belong to different small clusters. The gray-colored region in Fig. 2(A) represents the range of E_c within which the fractions of both additive and nonadditive double mutants are greater than 50% for all values of *n*. Outside of this range, our hypothesis is not satisfied (although it is trivially satisfied for *either* additive *or* nonadditive mutants separately). In Fig. 2(B), we show the *P*-values against the null hypothesis within the range of E_c highlighted in Fig. 2(A). The results for the values of E_c at which the *two fractions are equal* are summarized in Table I along with the corresponding *P*-values. In all cases considered, these *P*-values have the order of magnitude 10^{-4} to 10^{-5}, thus indicating that the observed support of our hypothesis has statistical significance. Note that the common threshold of statistical significance is 5.0×10^{-2}. The results of our analyses for two other values of the nonadditivity threshold parameter, $\delta_{\max}^{(ij)}$, equal to 15% and 25%, are similar to those above (cf. Figs. 1 and 2 in the Supplementary Data).

In order to further elucidate the correlation between nonadditivity and rigidity, we have performed similar analyses separately for the long-range ($d_{ij} > 6$ Å) and contact ($d_{ij} \leq 6$ Å)

mutation site pairs. In Figs. 3(A) and 4(A) we respectively show the calculated fractions of long-range and contact double mutants that agree with the nonadditivity-rigidity correlation hypothesis. One sees that the for long-range mutants [Fig. 3(A)] there is a clear bias of approximately 10% or more toward agreement with our hypothesis, and the *P*-values against the null hypothesis are as low as 10^{-5} to 10^{-6} [cf. Fig. 3(B) and Table II]. Note that the best of these values are only about an order of magnitude from the significance of the traditional distance-only viewpoint. Because there are fewer observations, the *P*-values in Fig. 3 should be larger than those in Fig. 2 if one assumes that our hypothesis applies equally to long-range and contact mutants. The fact that they are instead smaller indicates that the hypothesis mostly applies to the long-range mutants. Conversely, no such statistically significant bias is found for contacting mutant sites (Fig. 4). It appears that for contact mutants the two residues belong to the same rigid cluster in the physically meaningful range of E_c regardless of whether they are additive or not. Similar results were found using the threshold distance of 4 Å to discriminate the long-range and contact mutants. To further check the validity of our results, we have performed an independent calculation using a slightly refined criterion for classifying double mutations as nonadditive or additive. Double mutations were classified as nonadditive only if $\delta^{(ij)} \geq \delta^{(ij)}_{\max}$, $|\Delta\Delta G^{(ij)}| > 0.3$ kcal/mol, and $|\Delta\Delta G^{(i)} + \Delta\Delta G^{(j)}| > 0.3$ kcal/mol; they were classified as additive otherwise. This allowed us to avoid the possibility of mistakenly assigning nonadditivity in cases when experimental values for free energy changes were too small in absolute magnitude to be reliable. We have found that for -1.50 kcal/mol $\leq E_c \leq -1.15$ kcal/mol our hypothesis is satisfied for all $n = 1-6$, and the corresponding *P*-values are in the range $3.3 \times 10^{-4} \leq P \leq 4.7 \times 10^{-2}$ (cf. Fig. 3 in Supplementary Data).

The experimental dataset we use (cf. Table I in Supplementary Data) contains redundancy in some sense, as it includes entries on double mutant cycles where several identical mutants were studied under different experimental conditions. In order to test that it is not the bias of our dataset that results in the preferential occurrence of long-range nonadditivity described above, we have performed a set of independent calculations using a reduced dataset, in which 66 records have been eliminated and 166 unique, randomly chosen mutants were retained (cf. Table II in Supplementary Data). The results for fractions of nonadditive and additive long-range mutants supporting our hypothesis calculated using the reduced dataset (cf. Fig. 4 in Supplementary Data) are very similar to those presented above, although the corresponding *P*-values for -1.55 kcal/mol $\leq E_c \leq -1.30$ kcal/mol have increased to $5.0 \times 10^{-4} \leq P \leq 1.0 \times 10^{-2}$ due to fewer datapoints. The *P*-value of the naïve hypothesis of nonadditivity-distance correlation on the reduced dataset has similarly increased (from 3.4×10^{-7} to 3.2×10^{-5}) for the same reason.

These results are encouraging as they demonstrate that structure rigidity plays a statistically significant role in determining when long-range nonadditivity occurs, which has almost comparable statistical significance to the usual contact-pair discriminating criterion. Moreover, we did not find any statistical significance in understanding long-range nonadditivity for double mutants using residue burial or interaction density properties. These findings raise an interesting question: Can nonadditivity in structurally well separated double mutant free energy cycles be predicted from a single static three-dimensional protein

structure? Unfortunately, the rigidity criterion from FIRST used in this work is not sufficiently complete to provide reliable predictions. While our observation regarding distinctions between additive and nonadditive double mutations is statistically significant, at present it has very limited predictive power. In particular, the bias toward the preferential occurrence of nonadditivity versus additivity has a small magnitude ranging from 7% to 11%. Therefore, predictions using FIRST based on a single rigid cluster decomposition will not be reliable. A few factors possibly decreasing the reliability are: (*i*) the same cut-off H-bond energy, E_c, was used for all proteins in the dataset, without optimizing E_c for each protein structure; (*ii*) the variability in how the constraint topology is defined through hydrophobic, H-bond, and torsion constraints; (*iii*) only wild type protein rigidity characteristics are considered without concern for the structural perturbation the mutation will induce; and (*iv*) no account of electrostatic influences. In future work, we will attempt to improve the predictive power by employing a thermodynamic Gibbs ensemble of rigid cluster decompositions using the DCM, which provides thermodynamically averaged statistical information on residue to residue rigidity correlations that are conveyed in molecular cooperativity plots (25, 26).

Finally, because our hypothesis relates nonadditivity to the location of mutation sites within wild-type structures but not to types of amino acid substitutions, we have performed an additional independent set of calculations on a further reduced “consensus” dataset derived from the one given in Table I in the Supplementary Data. In this consensus dataset, each entry represents a double mutant with generic (i.e., arbitrary) substitutions at sites *i* and *j* in a given PDB structure. Each entry was assigned a consensus value “additive” or “nonadditive” based on the prevalence of additive or nonadditive mutants (using a 50% majority rule). When all six atoms are considered (i.e., the condition on two residues belonging to the same rigid cluster is the most stringent) and -1.55 kcal/mol $\leq E_c \leq -1.30$ kcal/mol, the fractions of mutants supporting our hypothesis are 8–12%, although the corresponding *P*-values have increased to $2.1\times10^{-2} \leq P \leq 3.8\times10^{-2}$ due to even fewer datapoints (24 nonadditive and 31 additive long-range mutants, data not shown).

The initial distance-only hypothesis about the occurrence of nonadditivity has better statistical significance than the nonadditivity-rigidity correlation hypothesis introduced here. This is not surprising as the former has been known for many years and its physical basis is straightforward. As such, we suggest it represents ”the low hanging fruit.” On the other hand, only our new hypothesis provides a rationale for the existence of long-range nonadditivity which complements the conventional distance-only viewpoint.

Molecular examples

In Fig. 5 we present four examples of wild-type protein structures (T4 lysozyme, cytochrome c, serine proteinase inhibitor CI-2, and acidic fibroblast growth factor). Each is colored according to their rigid cluster decompositions calculated for $E_c = -1.35$ kcal/mol, meaning regions of different color represent distinct rigid clusters (red coloring indicates mutation sites). T4 lysozyme (A) is primarily composed on a single rigid cluster with few disjoint flexible regions near its surface. The free energy change within a particular T4 lysozyme double mutant (V111I, F153L; both sites are located within the same rigid cluster)

is more than twice the sum of the constituent single mutations ($\delta^{(ij)} = 55\%$, see Table I in Supplementary Data). This mutant represents an example of "super-additivity" (19), or more precisely "super-nonadditivity," meaning that the observed stability change is greater than the sum of the single mutations. This result is an interesting exception to the more common "sub-additive" pairs (19) that tend to reduce the effect of a second mutation when the first is already present (note: this is discussed above as the explanation for why the slopes within Fig. 1 are less than unity). This result suggests that supper-nonadditivity occurs in this instance because both mutation sites are buried inside the hydrophobic core and both single and double mutations destabilize the structure (i.e., $\Delta\Delta G < 0$), resulting in a cooperative disruption of the large rigid cluster.

A similar large nonadditivity ($\delta^{(ij)} = 120\%$) is observed in the (Q37R, E43Y) cytochrome c double mutant [Fig. 5(B)]. Here, one site is located in a loop region, whereas the other is found in an α-helix belonging to the same rigid cluster. In this case, however, the sum of the constituent single mutation effects is larger than the effect of the double mutant, following the trend of "sub-additive" (19) mutations in Fig. 1. This is possibly due to the fact that both sites are solvent-exposed and both single and double mutations have stabilizing effects (i.e., $\Delta\Delta G > 0$). While the parallel between solvent accessibility and super-vs. sub-nonadditivity in these two examples is inline with the intuitive expectation, no statistically robust correlations between the two are observed across the entire dataset (data not shown).

Juxtaposed to the nonadditivity scenarios above, the effects of spatially well-separated mutations (P25A, A35G) in serine proteinase inhibitor CI-2 [Fig. 5(C)] and (L44F, F108Y) in fibroblast growth factor [Fig. 5(D)] located within different rigid clusters are essentially additive, $\delta^{(ij)} = 0.38\%$ and $\delta^{(ij)} = 10.3\%$, respectively (see Table I in Supplementary Data). In both cases, the sum of the effects of single mutations is only slightly greater than the effect of a double mutation, indicating that thermodynamic coupling across distinct rigid substructures is limited. While beyond the scope of the work presented here, this result suggests that mutation(s) bridging distinct rigid substructures into one may be able to confer nonadditivity to other pairs therein, whereas the pair would be additive in the absence of the bridging mutation(s).

Conclusions

We report here a statistically significant association between nonadditivity within double mutant free energy cycles and structural correlations in the location of the mutation sites. By calculating an all-atom decomposition of each protein structure into rigid clusters based on constraint topologies, we have found that there is a statistically significant bias toward nonadditivity when the two mutation sites are located within the same rigid cluster, and occurrence of additivity when they are located within different clusters. The *P*-values supporting our nonadditivity-rigidity correlation hypothesis vary from 10^{-4} to 10^{-5}, indicating strong statistical significance. Most interestingly, these observations become even more significant when considering long-range mutants by themselves (*P*-values: 10^{-5} to 10^{-6}). Conversely, no preferential occurrence is observed when contact mutations are considered alone. To the best of our knowledge, this result represents the first report of a statistically significant prevalence of nonadditivity within long-range double mutant free

energy cycles. The observed preferential occurrence of long-range nonadditivity within the same rigid cluster should have far-reaching implications as it allows for a new interpretation of nonadditivity whose likelihood can be assessed by a simple and computationally efficient approach.

It should be pointed out that we expect the nonadditivity trends found in this work may also be captured by other mechanical descriptions of protein structure, e.g., by normal mode analysis within elastic network models (ENMs). In normal mode analysis within a residue-level Gaussian or anisotropic ENM (which requires knowledge of only the wild-type protein structure), additivity would be expected when the two mutations are made in regions whose collective low-frequency vibrations are described by two independent sets of displacement eigenvectors; nonadditivity would be more likely when both mutations are in a region whose collective low-frequency vibrations are described by a single set of almost collinear eigenvectors. The basis for this expectation is the fact that distance-constraint-based descriptions can be viewed as a limiting case of ENMs with covalent bonds being modeled by springs of infinite stiffness. ENMs thus effectively take into account the rigidity topology within a protein structure, which was found in this work to be an important determinant of the long-range nonadditivity.

Materials and methods

Double mutant cycle free energy of unfolding changes, $\Delta\Delta G$, were drawn from the ProTherm database (32). The following experimental conditions were imposed to be identical for single and double mutants: temperature, pH, measurement method (DSC, CD, etc.), and the method of protein unfolding (thermal, urea, etc.). In total, 232 matching combinations of double mutant cycles met the uniformity criteria above, in 33 different proteins with known wild-type structures. Utilizing a 20% nonadditivity threshold, nonadditivity was found in 129 cycles, whereas additivity was found in 103 cycles. Out of the 232 double mutants, 120 were classified as “long-range”, meaning the distance between mutation sites was greater than 6 Å.

Decompositions of the 33 protein structures (taken from PDB) into rigid clusters were performed using the FIRST software (28), which is an implementation of a network rigidity analysis algorithm for investigating protein rigidity and flexibility (29). Within each protein structure, decompositions were calculated for 100 values of the hydrogen bond cut-off energy, ranging from −0.05 kcal/mol to −5.00 kcal/mol thus producing 3300 rigid cluster decompositions.

Prior to rigidity calculations, the missing hydrogen atoms were inserted into each of the 33 protein structures using the MOE software (Chemical Computing Group) and structures were then optimized to reduce the potential energy using MOE under the Amber99 all-atom force field (33). The *P*-values for the null hypotheses were calculated as products of the corresponding *P*-values for nonadditive and additive mutants,

$$P=P_{non}P_{add}, \quad (4)$$

where

$$P_{non} = \sum_{m=N_{non}}^{N_{non}^{(tot)}} p_{N_{non}^{(tot)}}(m), \quad (5)$$

$$P_{add} = \sum_{m=N_{add}}^{N_{add}^{(tot)}} p_{N_{add}^{(tot)}}(m), \quad (6)$$

In the equations above, N_{non} and N_{add} are the numbers of nonadditive and additive mutants supporting the hypothesis, $N_{non}^{(tot)}$ and $N_{add}^{(tot)}$ are the total numbers of nonadditive and additive mutants, and $p_n(m)$ is the binomial distribution,

$$p_n(m) = \frac{n!}{2^n m!(n-m)!}. \quad (7)$$

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant sponsor: This work is supported by NIH grant R01 GM073082-01A1 to DJJ and DRL.

Anthony Fodor is acknowledged for insightful discussions related to this work. Key to FIRST is the use of a graph-rigidity algorithm that is claimed in U.S. Patent 6,014,449, which has been assigned to the Board of Trustees Michigan State University. Used with permission. We thank the Center for Biomedical Engineering Systems for partial support of this project.

References

1. Jencks WP. On the attribution and additivity of binding energies. Proc Natl Acad Sci USA. 1981; 78:4046–4050. [PubMed: 16593049]
2. Ackers GK, Smith FR. Effects of site-specific amino acid modification on protein interactions and biological function. Annu Rev Biochem. 1985; 54:597–629. [PubMed: 3896127]
3. Wells JA. Additivity of mutational effects in proteins. Biochemistry. 1990; 29:8509–8517. [PubMed: 2271534]
4. Green SM, Shortle D. Patterns of nonadditivity between pairs of stability mutations in staphylococcal nuclease. Biochemistry. 1993; 32:10131–10139. [PubMed: 8399139]
5. Gregoret LM, Sauer RT. Additivity of mutant effects assessed by binomial mutagenesis. Proc Natl Acad Sci USA. 1993; 90:4246–4250. [PubMed: 8483940]
6. Huang Z, Wagner CR, Benkovic SJ. Nonadditivity of mutational effects at the folate binding site of Escherichia coli dihydrofolate reductase. Biochemistry. 1994; 33:11576–11585. [PubMed: 7918371]
7. Mildvan AS. Inverse thinking about double mutants of enzymes. Biochemistry. 2004; 43:14517–14520. [PubMed: 15544321]
8. Skinner MM, Terwilliger TC. Potential use of additivity of mutational effects in simplifying protein engineering. Proc Natl Acad Sci USA. 1996; 93:10753–10757. [PubMed: 8855252]

9. Amin N, Liu AD, Ramer S, Aehle W, Meijer D, Metin M, Wong S, Gualfetti P, Schellenberger V. Construction of stabilized proteins by combinatorial consensus mutagenesis. Prot Engineering Des Sel. 2004; 17:787–793.
10. Sandberg WS, Terwilliger TC. Engineering multiple properties of a protein by combinatorial mutagenesis. Proc Natl Acad Sci USA. 1993; 90:8367–8371. [PubMed: 8378307]
11. Matsuura T, Yomo T, Trakulmalemsai S, Ohashi Y, Yamamoto K, Urabe I. Nonadditivity of mutational effects on the properties of catalase I and its application to efficient directed evolution. Prot Engineering. 1998; 11:789–795.
12. Mildvan AS, Weber DJ, Kuliopulos A. Quantitative interpretations of double mutations of enzymes. Arch Biochem Biophys. 1992; 294:327–340. [PubMed: 1567189]
13. Horovitz A, Bochkareva ES, Yifrach O, Girshovich AS. Prediction of an Inter-residue Interaction in the Chaperonin GroEL from Multiple Sequence Alignment is Confirmed by Double-mutant Cycle Analysis. J Mol Biol. 1994; 238:133–138. [PubMed: 7908986]
14. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science. 1999; 286:295–299. [PubMed: 10514373]
15. Fodor A, Aldrich R. On evolutionary conservation of thermodynamic coupling in proteins. J Biol Chem. 2004; 279:19046–19050. [PubMed: 15023994]
16. Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol. 2003; 10:59–69. [PubMed: 12483203]
17. Hatley ME, Lockless SW, Gibson SK, Gilman AG, Ranganathan R. Allosteric determinants in guanine nucleotide-binding proteins. Proc Natl Acad Sci USA. 2003; 100:14445–14450. [PubMed: 14623969]
18. Shulman A, Larson C, Mangelsdorf D, Ranganathan R. Structural determinants of allosteric ligand activation in RXR heterodimers. Cell. 2004; 116:417–429. [PubMed: 15016376]
19. LiCata VJ, Ackers GK. Long-range, small magnitude nonadditivity of mutational effects in proteins. Biochemistry. 1995; 34:3133–3139. [PubMed: 7880807]
20. Dill KA. Additivity principles in biochemistry. J Biol Chem. 1997; 272:701–704. [PubMed: 8995351]
21. Rod TH, Radkiewicz JL, Brooks CL III. Correlated motion and the effect of distal mutations in dihydrofolate reductase. Proc Natl Acad Sci USA. 2003; 100:6980–6985. [PubMed: 12756296]
22. Jacobs DJ, Dallakyan S, Wood GG, Heckathorne A. Network rigidity at finite temperature: Relationships between thermodynamic stability, the nonadditivity of entropy, and cooperativity in molecular systems. Phys Rev E. 2003; 68:061109, 1–21.
23. Livesay DR, Dallakyan S, Wood GG, Jacobs DJ. A flexible approach for understanding protein stability. FEBS Lett. 2004; 576:468–476. [PubMed: 15498582]
24. Jacobs DJ, Dallakyan S. Elucidating protein thermodynamics from the three-dimensional structure of the native state using network rigidity. Biophys J. 2005; 88:903–915. [PubMed: 15542549]
25. Jacobs DJ, Livesay DR, Hules J, Tasayco M-L. Elucidating quantitative stability/flexibility relationships within thioredoxin and its fragments using a distance constraint model. J Mol Biol. 2006; 358:882–904. [PubMed: 16542678]
26. Livesay DR, Jacobs DJ. Conserved quantified stability/flexibility relationships (QSFR) in an orthologous RNase H pair. Proteins. 2006; 62:130–143. [PubMed: 16287093]
27. Reichmann D, Rahat O, Albeck S, Meged R, Dym O, Schreiber G. The modular architecture of protein-protein binding interfaces. Proc Natl Acad Sci USA. 2005; 102:57–62. [PubMed: 15618400]
28. Thorpe, MF., et al. Flexweb: Analysis of Flexibility in Biomolecules and Networks. http://flexweb.asu.edu
29. Jacobs DJ, Rader AJ, Thorpe MF, Kuhn LA. Protein flexibility predictions using graph theory. Proteins. 2001; 44:150–165. [PubMed: 11391777]
30. Rader AJ, Hespenheide BM, Kuhn LA, Thorpe MF. Protein unfolding: Rigidity lost. Proc Natl Acad Sci USA. 2002; 99:3540–3545. [PubMed: 11891336]

31. Gohlke H, Kuhn LA, Case DA. Change in Protein Flexibility Upon Complex Formation: Analysis of Ras-Raf Using Molecular Dynamics and a Molecular Framework Approach. Proteins. 2004; 56:322–337. [PubMed: 15211515]

32. Kumar MD, Bava KA, Gromiha MM, Prabakaran P, Kitajima K, Uedaira H, Sarai A. ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. Nucleic Acids Res. 2006; 34:D204–206. Database issue. [PubMed: 16381846]

33. Ponder JW, Case DA. Force fields for protein simulations. Adv Prot Chem. 2003; 66:27–85.

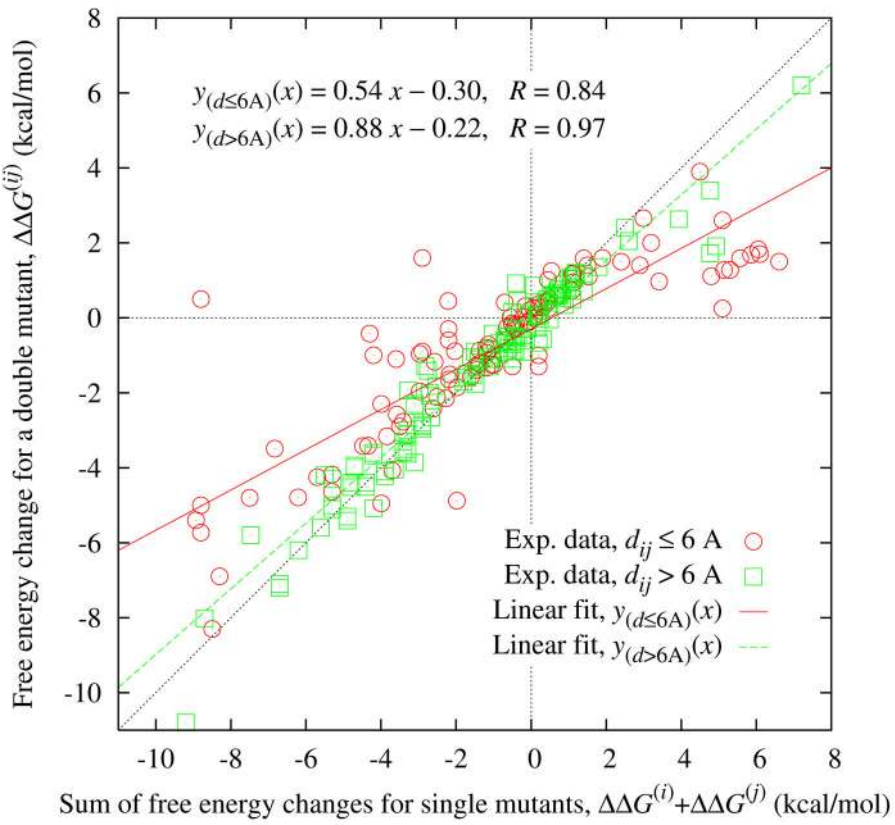

Fig. 1.

Scatter plot of changes in the free energy of unfolding, $\Delta\Delta G$, in 232 double mutant cycles. Experimental data corresponding to 112 double mutants with d_{ij} ≤6 Å are indicated by (red) circles, while those corresponding to **120** cycles with d_{ij} > 6 Å are indicated by (green) squares. The linear fits to those data are shown by full (red) and dashed (green) lines, respectively. Dotted lines are shown to guide the eye. In regression equations, *x* and *y* stand for abscissa and ordinate axis quantities, respectively; *R* is the correlation coefficient.

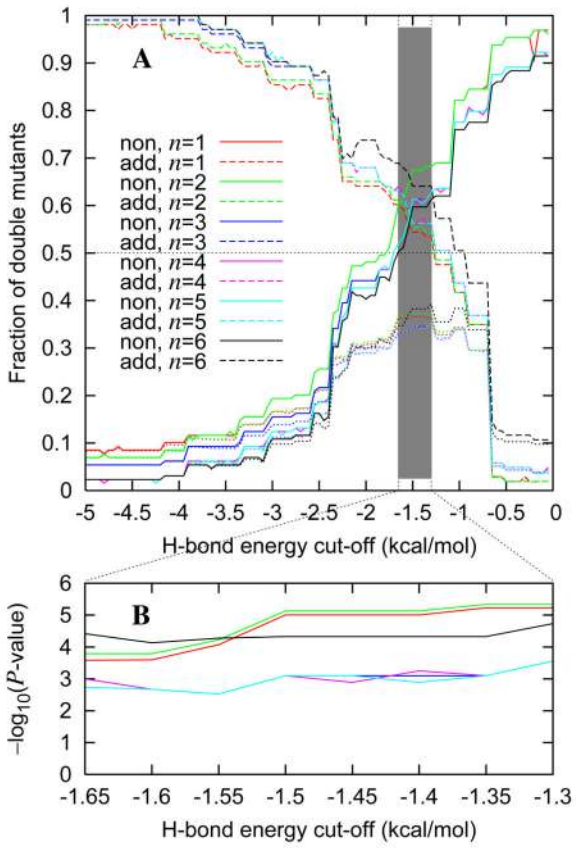

Fig. 2.

(A) Fractions of double mutants satisfying the nonadditivity-rigidity correlation hypothesis. Solid lines represent the fraction of nonadditive mutants with mutation sites located within the same rigid cluster, whereas dashed lines represent the fraction of additive mutants with mutation sites located in different clusters. Dotted lines represent the product of the two fractions. Two residues are considered to be within the same cluster if each of them shares *n* or more atoms with the cluster; otherwise, they are considered to be in different clusters; predictions for different *n* are differentiated by color. The gray band indicates the range of values of the hydrogen bond energy cut-off, E_c, in which the fractions of both nonadditive *and* additive mutants are greater than 0.5 and thus support our hypothesis for all *n*. Outside of this range, the hypothesis is not satisfied. (B) Negatives of decimal logarithms of the corresponding P-values for given *n*, in the range of E_c colored gray in (A). The line coloring in (B) is the same as in (A).

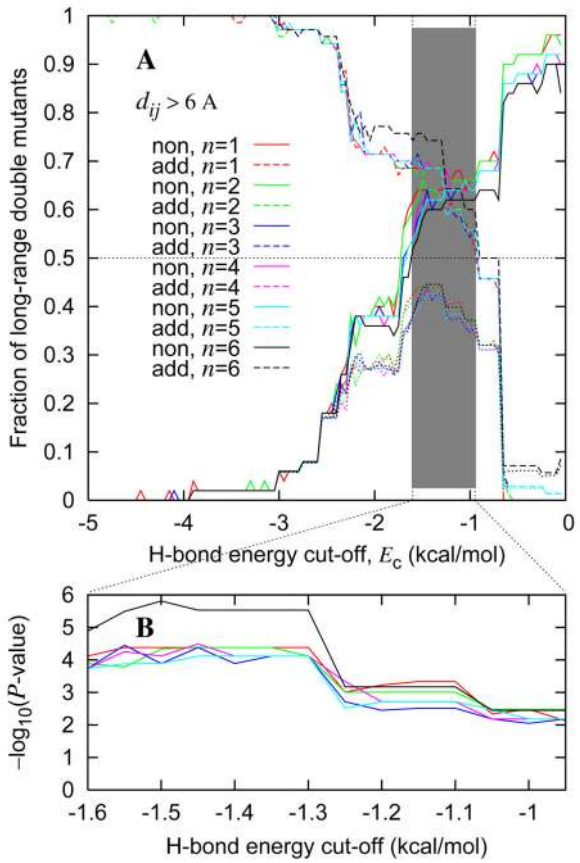

Fig. 3.

Same as Fig. 2 but for long-range mutants only (d_{ij} > 6 Å). Note that the lowest ° P-values in (B) correspond to maxima of the dotted curves in (A) in the range −1.6 kcal/mol ≤E_c ≤ −1.3 kcal/mol, which is very similar to the results in Fig. 2(B).

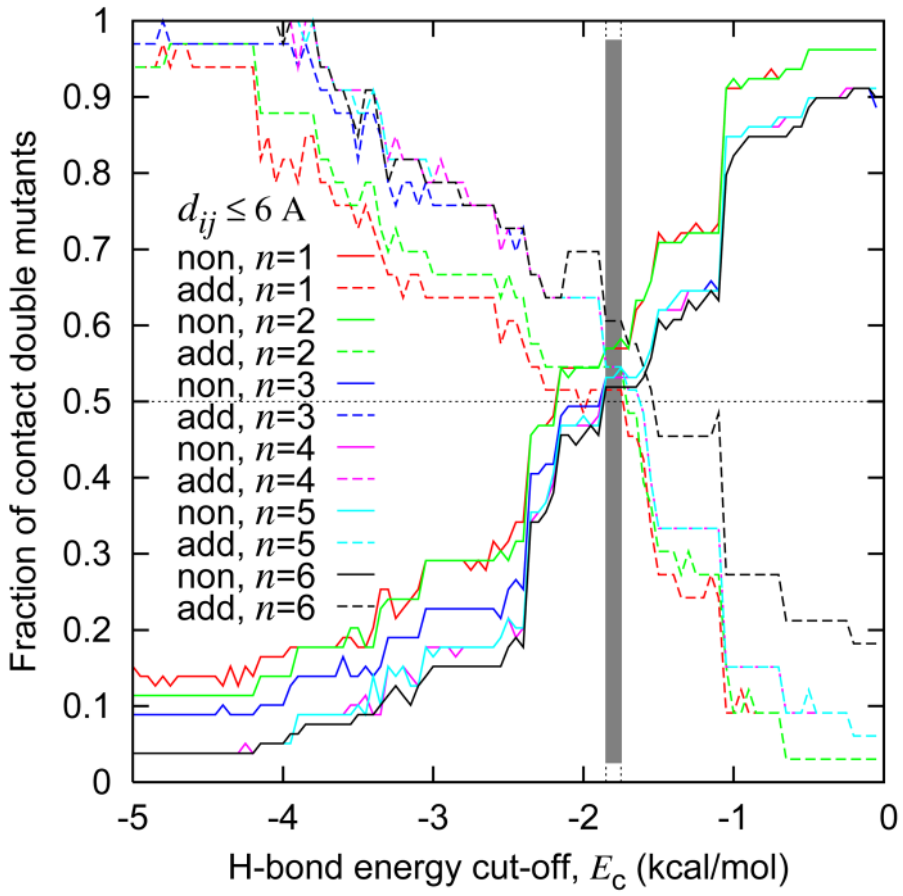

Fig. 4.

Same as Fig. 2(A) but for contact mutants only (d_{ij} ≤6 Å). The region of E_c in which our hypothesis is supported by both nonadditive and additive fractions for all n is very narrow and the statistical significance of this support is low; thus, the corresponding P-values are not shown.

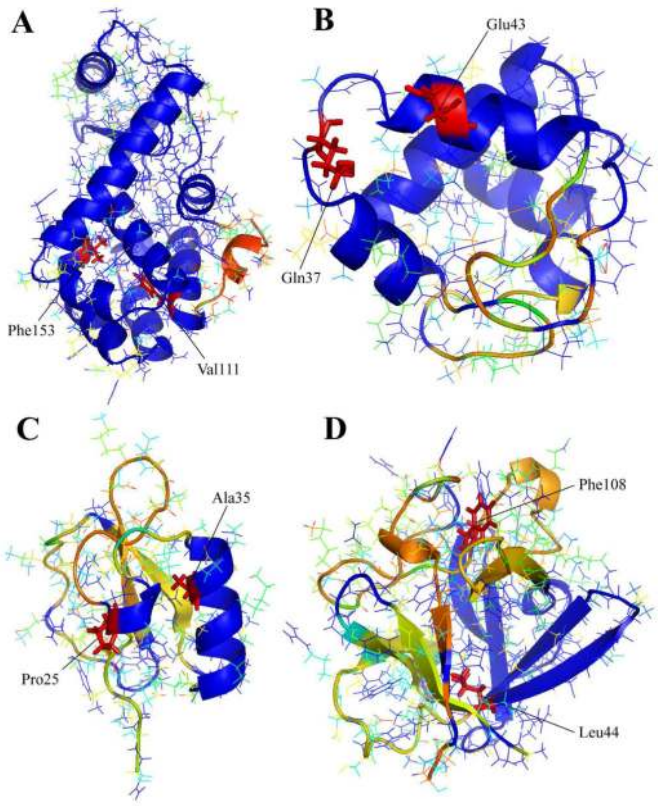

Fig. 5.

Examples of protein structures colored according to their rigid cluster decompositions. Different colors correspond to distinct rigid clusters; mutation sites are indicated by red coloring. (A) T4 Lysozyme (2LZM); (B) Cytochrome c C551 (451C); (C) Serine proteinase inhibitor CI-2 (2CI2); (D) Acidic fibroblast growth factor (2AFG). In (A) and (B), the spatially well-separated mutation sites are located in the same rigid clusters and the mutations resulted in nonadditivity of free energy changes, while in (C) and (D) the mutation sites are located within different clusters and mutation effects were additive.

Table I

Examples of data shown in Fig. 2 for values of E_c at which the fractions of nonadditive and additive double mutants supporting our hypothesis are equal or have closest values. Here N_{non} and N_{add} are the numbers of nonadditive and additive mutants supporting the hypothesis. The total numbers of these mutants are $N_{non}^{(tot)}=129$ and $N_{add}^{(tot)}=103$, respectively.

n	E_c (kcal/mol)	N_{non}	N_{add}	P_{non}	P_{add}	$P = P_{non}P_{add}$
1	−1.65	78	62	1.09×10^{-2}	2.41×10^{-2}	2.62×10^{-4}
2	−1.60	79	62	6.69×10^{-3}	2.41×10^{-2}	1.61×10^{-4}
3	−1.50	79	58	6.69×10^{-3}	1.18×10^{-1}	7.92×10^{-4}
4	−1.50	79	58	6.69×10^{-3}	1.18×10^{-1}	7.92×10^{-4}
5	−1.50	79	58	6.69×10^{-3}	1.18×10^{-1}	7.92×10^{-4}
6	−1.30	79	66	6.69×10^{-3}	2.77×10^{-3}	1.85×10^{-5}

Table II

Summary of the data shown in Fig. 3 (long-range mutants, d_{ij} > 6 Å). The total numbers of long-range mutants are $N_{non}^{(tot)}=50$ and $N_{add}^{(tot)}=70$, respectively.

n	E_c(kcal/mol)	N_{non}	N_{add}	P_{non}	P_{add}	$P = P_{non}P_{add}$
1	[−1.55, −1.3]	32	48	3.25×10^{-2}	1.27×10^{-3}	4.13×10^{-5}
2	[−1.45, −1.35]	32	48	3.25×10^{-2}	1.27×10^{-3}	4.13×10^{-5}
3	[−1.35, −1.3]	31	48	5.94×10^{-2}	1.27×10^{-3}	7.57×10^{-5}
4	[−1.4, −1.3]	31	48	5.94×10^{-2}	1.27×10^{-3}	7.57×10^{-5}
5	[−1.45, −1.3]	31	48	5.94×10^{-2}	1.27×10^{-3}	7.57×10^{-5}
6	[−1.45, −1.3]	30	50	1.01×10^{-1}	2.92×10^{-5}	2.96×10^{-6}