Searching for a mechanistic description of pairwise epistasis in protein systems

Jonathan Barnes[1], Craig Miller[1], and F. Marty Ytreberg[1]

[1]University of Idaho

June 7, 2021

Abstract

When two or more amino acid mutations occur in protein systems, they can interact in a non-additive fashion termed epistasis. One way to quantify epistasis between mutation pairs in protein systems is by using free energy differences: = ??$G_{1,2}$ - (??G_1 + ??G_2) where ??G refers to the change in the Gibbs free energy, subscripts 1 and 2 refer to single mutations in arbitrary order and 1,2 refers to the double mutant. In this study, we explore possible biophysical mechanisms that drive pairwise epistasis in both protein-protein binding affinity and protein folding stability. Using the largest available datasets containing experimental protein structures and free energy data, we derived statistical models for both binding and folding epistasis ( ) with similar explanatory power (R^2) of 0.299 and 0.258, respectively. These models contain terms and interactions that are consistent with intuition. For example, increasing the Cartesian separation between mutation sites leads to a decrease in observed epistasis for both folding and binding. Our results provide insight into factors that contribute to pairwise epistasis in protein systems and their importance in explaining epistasis. However, the low explanatory power indicates that more study is needed to fully understand this phenomenon.

Title:	Searching for a mechanistic description of pairwise epistasis in protein systems.
First Author:	Jonathan E. Barnes University of Idaho Department of Physics jonathan@barnes.science
Co-author	Craig R. Miller University of Idaho Department of Biological Sciences crmiller@uidaho.edu
Co-author	F. Marty Ytreberg University of Idaho Department of Physics ytreberg@uidaho.edu
Data Availability Statement	All data is available via a public repository: https://github.com/YtrebergPatelLab/EpistasisStats

Title:	Searching for a mechanistic description of pairwise epistasis in protein systems.
Funding Statement	The research was supported by the Center for Modeling Complex Interactions sponsored by the National Institute of General Medical Sciences (https://www.nigms.nih.gov) under award number P20 GM104420 and the National Science Foundation (https://www.nsf.gov) EPSCoR Track-II under award number OIA1736253. Computer resources were provided in part by the Institute for Bioinformatics and Evolutionary Studies Computational Resources Core sponsored by the National Institutes of Health (NIH P30 GM103324). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Conflict of Interest Disclosure	There are no conflicts of interest

Searching for a mechanistic description of pairwise epistasis in protein systems

Jonathan E. Barnes[1,3*], Craig R. Miller[2,3], F. Marty Ytreberg[1,3,4]

1 Department of Physics, University of Idaho, Moscow, ID, United States of America, 2 Department of Biological Sciences, University of Idaho, Moscow, ID, United States of America 3 Institute for Modeling Collaboration and Innovation, University of Idaho, Moscow, ID, United States of America, 4 Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID, United States of America *jonathan@barnes.science

Acknowledgements

The research was supported by the Center for Modeling Complex Interactions sponsored by the National Institute of General Medical Sciences (https://www.nigms.nih.gov) under award number P20 GM104420 and the National Science Foundation (https://www.nsf.gov) EPSCoR Track-II under award number OIA1736253. Computer resources were provided in part by the Institute for Bioinformatics and Evolutionary Studies Computational Resources Core sponsored by the National Institutes of Health (NIH P30 GM103324). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

When two or more amino acid mutations occur in protein systems, they can interact in a non-additive fashion termed epistasis. One way to quantify epistasis between mutation pairs in protein systems is by using free energy differences: = ??$G_{1,2}$ - (??G_1 + ??G_2) where ??G refers to the change in the Gibbs free energy, subscripts 1 and 2 refer to single mutations in arbitrary order and 1,2 refers to the double mutant. In this study, we explore possible biophysical mechanisms that drive pairwise epistasis in both protein-protein binding affinity and protein folding stability. Using the largest available datasets containing experimental protein structures and free energy data, we derived statistical models for both binding and folding epistasis ( ) with similar explanatory power (R^2) of 0.299 and 0.258, respectively. These models contain terms and interactions that are consistent with intuition. For example, increasing the Cartesian separation between mutation sites leads to a decrease in observed epistasis for both folding and binding. Our results provide insight into factors that contribute to pairwise epistasis in protein systems and their importance in explaining epistasis. However, the low explanatory power indicates that more study is needed to fully understand this phenomenon.

Keywords: epistasis, binding affinity, folding stability, non-additivity, statistical modeling

1. Introduction

Multiple amino acid mutations can interact in biological systems, leading to nonadditive effects termed epistasis. While a general understanding of the concept of epistasis has existed for many years, the prevalence of epistasis, or its importance in biological systems, is still a matter of debate[1–5]. Some believe it is a major force in evolution, either by constraining the available pathways for systems to evolve, by counteracting mutations that reduce fitness through compensatory effects, or by contributing to a more rugged fitness landscape[6–18]. Others have explored the epistatic effect between sets of beneficial mutations, finding that epistasis is pervasive and a key aspect of adaption, but leading to diminishing returns or negative epistasis[10,19,20,20–22]. Other studies using RNA viruses have shown that epistasis is prevalent and likely a mechanism for their evolution[23–28]. Epistasis has also been shown to be a likely contributing factor to drug and antibody resistance of influenza A, HIV-1 and other pathogens[12,24,29,30], and for general disease susceptibility in humans[31]. Finally, the complexity that epistasis provides in understanding mutation effects must be accounted for in protein engineering and design[32–35].

For pairs of simultaneous mutations in proteins (we will refer to these as "double mutations"), epistasis can be expressed in terms of free energy differences:

$= ??G_{1,2}$ - $(??G_1 + ??G_2)$ (EQ 1)

Where $??G_{1,2}$ corresponds to the change in the folding or binding free energy due to the double mutation, and $??G_1 + ??G_2$ refers to the sum of the constituent single mutation free energy changes. This nonadditivity can be caused by direct interactions between mutational sites, or by indirect effects such as conformational perturbations. Epistasis is positive when the double mutant is more stabilizing than the sum of the constituent singles (< 0) and negative when the double mutant is more destabilizing than the sum of the constituent singles (> 0).

Despite its importance to understanding biological systems, a comprehensive mechanistic picture of the drivers of epistasis in proteins is not known. An early attempt to explain epistasis mechanisms is a study by Wells[36]; they concluded that features like separation distance, electrostatic interactions, and conformational perturbations were likely contributors. However, this conclusion was based on a small data set containing a total of 12 folding and binding systems, with less than 75 total multiple mutations. More recent studies have examined specific protein systems like TEM-1 β-lactamase[37,38]and the IgG-binding domain of protein G[39], finding pervasive negative epistasis. Long-range epistasis has also received attention Gromiha et al. proposed that distant residues that are part of a specific local group (they defined this as a rigid cluster) could lead to epistasis[40]. Other researchers have used tools like molecular dynamics to analyze if networks of interactions can mediate long-range epistasis[41]. Classification systems have also been developed. Jemimah et al. used structural features to build a model to classify whether mutational pairs would be additive (i.e., not epistatic)[42]. These previous studies provide a basis for understanding possible contributors to epistasis and some even offer predictive capability, however they do not provide a complete understanding of epistasis mechanisms and their interactions.

In this study, we determine biophysical drivers of pairwise epistasis in protein systems and rank their contribution to the observed epistasis, (EQ 1). We used protein structural data, protein-protein binding affinities, and protein folding stabilities from the largest, most diverse datasets currently available. We explored possible relationships between the observed epistasis and features that are intrinsic to both the proteins and the mutated residues. A statistical model selection procedure was performed to determine the features that are most important to explaining the observed epistasis. The models determined for binding and folding have similar and modest predictive power. Both models contain similar features that include separation distance and charge interactions. Our work serves as a stepping stone to further our understanding of the biophysical drivers of epistasis, and to build future models with more complex features and interactions.

2. Methods

2.1: Curating experimental data

Experimental binding affinity data was obtained from SKEMPI v2.0[43]and folding stability data from ProTherm 4[44]. Since the focus of our study is pairwise epistasis, we extracted a subset of the data consisting of all instances where there was data for a double mutant and the corresponding constituent singles. For both folding and binding data, values were converted to kcal/mol. A temperature of 298 K was used if not specified in the dataset. Averages were calculated for mutations that included multiple free energy values. The attributes in the resulting curated folding and binding datasets used in our study include the PDB ID, protein complex name, the mutation(s), and either binding or folding free energy values. The total number of data points for double mutants with constituent single mutants were 572 from 58 protein-protein complexes for binding, and 204 from 30 protein systems for folding. Epistasis was calculated for each double mutation data point using EQ 1, that is, by taking the difference between the free energy change due to the double mutation and the sum of the free energy changes due to the constituent single mutations. Protein structures used for analysis were acquired from the RCSB Protein Data Bank (PDB)[45].

2.2: Extracting Features as Possible Drivers of Epistasis

For electrostatics, and other categorical features described below, the explicit wildtype-mutant pairs are henceforth denoted separated by a semicolon for simplicity: wt_1wt_2;mut_1mut_2.

Amino Acid Properties

To investigate the effect of electrostatics on epistasis, we classified amino acids as positively charged (+), negatively charged (-), or neutral (0). To incorporate every wildtype-mutant pair state would be infeasible due to overparameterization, as it would result in 3^4=81 possible categories (++;–, ++;-+, ++;+-, +-;–, . . .). To avoid overparameterization, we explored various abstractions of this data, incorporating this into our model selection process (detailed below). The resulting charge contribution was given by a simplified charge-interaction scheme with pairs belonging to one of three categories: attractive (+- or -+, denoted "A"), repulsive (– or ++, denoted "R"), and neutral (all other cases, denoted "0"). The reverse of each wildtype-mutant states were classified as the same (e.g. 0;A = A;0), resulting in four categories: 0A, 0R, AR, and 00 to capture all possible electrostatic interactions. Note that the AR case was not present in either dataset.

To include the change in size for the constituent amino acids we used the van der Waals volume in . To capture the net effect due to the change in size for both sites we used the metric (referred to as size_{net}.

size_{net} = | size_{m1} - size_{wt1}| + | size_{m2} - size_{wt2}| (EQ 2)

where wt and m correspond to the wildtype and mutant amino acids, respectively, and 1 and 2 denote the amino acid sites in an arbitrary order. Under this scheme, if one or both sites undergo a large/small change in volume occupancy the corresponding metric will be large/small respectively, even if they are in opposing directions.

To include the effect of hydrophobicity, each residue is classified as either "H" for hydrophobic or "P" for polar. Using all possible 16 categories would be possible, but risk overfitting. We instead found the following abstraction: a boolean value ("0" or "1") that denotes whether the net hydrophobicity of the pair changed upon mutation. For example, HP;PH would give 0 since the net hydrophobicity remained the same. By contrast, PP;HP or PP;HH would both give 1, since the net hydrophobic state changed upon mutation.

Structural Properties

Separation distance was defined as the Cartesian separation between the alpha carbons for each mutational site. This Euclidean distance $r$was calculated using the x, y, z coordinates for the mutation sites via the standard formula:

$$r = \sqrt{(x_{wt1} - x_{wt2})^2 + (y_{wt1} - y_{wt2})^2 + (z_{wt1} - z_{wt2})^2}$$

(EQ 3)

Secondary structure information was included by considering whether a given mutational site was located in an alpha helix ("H"), beta sheet ("S"), or loop ("L"). Secondary structure content was determined using a PyMol script[46]. As with other categorical features overparameterization may be a concern, though in this case the explicit consideration only has nine possible cases. We tested the possible abstractions, ranging from explicit consideration of the structures at each site (e.g., HL,LL,LS,...) to the simplest case of a boolean value denoting whether both sites belong to the same type of structure ("0") or different structures ("1").

We also considered the effect of solvent accessible surface area (SASA): a metric describing whether a residue is exposed or buried. To calculate the SASA, we first prepared the PDB files using *pdbfixer* from the OpenMM software suite[47], to add missing residues, replace non-standard residues with their standard equivalents, and add missing hydrogens. The repaired structures were then processed with FoldX[48]to generate mutations using the *BuildModel* command. DSSP v 3.0.0[49]was then used to calculate the absolute SASA ($SASA_{abs}$) for each residue of interest. Both absolute and relative SASA were considered, relative SASA ($SASA_{rel}$) was calculated using the empirical max accessible surface area (ASA_{max}) generated by Tien et al[50]via the formula:

$SASA_{rel} = SASA_{abs}$ / ASA_{max}. (EQ 4)

Since SASA changes affect both wildtype and mutant residues, we used a modified version of EQ 2 replacing $size_{net}$ with SASA.

We also included classification information. For binding, we included the type of protein-protein complex broken into five categories, based on the information provided in the SKEMPI v2.0 database: antibody-antigen (AB/AG), T cell receptor-peptide bound major histocompatibility complex (TCR/pMHC), Cytokine-Cytokine receptor (Cyto/Cyto), GTPase-other, and non-specific protein-protein interaction (Pr/PI) which functioned as the reference category for the statistical models. We also included a boolean value indicating whether or not the mutational sites occur on the same ("0") or different ("1") protein chains, as sites which occur on the same chain may have a different effect on binding than if they occur on opposing chains. For folding, we included the system size given by the total number of residues acquired from the PDB.

2.3: Statistical Analysis

To analyze the relationship between epistatic effect and separation distance, we conducted a likelihood ratio test that compares a null model (where separation distance is unrelated to epistasis) against an alternative model (where epistasis decays with increasing separation). More precisely, we defined the null model to be that epistasis values are sampled from a normal distribution that is independent of the separation between residues. For the alternative model, epistasis values are sampled from a normal (same mean as the null case) with a standard deviation that decays exponentially as a function of separation according to a $e^{? \ r}$ where r is the separation between residue site alpha-carbons (EQ 3) and a and ? are the curve's parameters estimated by maximum likelihood for the dataset. This maximum likelihood was determined by a grid-search method, considering all possible a and ?, taking the resulting model with the largest likelihood. The likelihood ratio is given by the ratio of the log of the two likelihoods of the data under the two models:

$$\Lambda(r) = \frac{\mathcal{L}(\theta_0|r)}{\mathcal{L}(\theta_1|r)} \rightarrow log(\Lambda(r)) = log(\mathcal{L}(\theta_0|r)) - log(\mathcal{L}(\theta_1|r))$$

(EQ 5)

where $\mathcal{L}$, θ_1, θ_2 refers to the likelihood, log is the natural logarithm, and correspond to the null and alternative models respectively. Small values of Λ indicate that the alternative model has more explanatory power than the null. We first calculated the likelihood ratio for the experimental data, Λ_{exp}. In order to determine statistical significance of Λ_{exp} we then obtained the distribution of Λ under the null through parametric

simulation. Specifically, we simulated datasets using the mean and standard deviation of the experimental epistasis data. We then repeated the fitting exercise used on the real dataset for the simulated dataset, using the same separation data, and calculated Λ. This process was repeated 1000 times to obtain the distribution of Λ under the null: Λ_{sim}. The p-value for the test was then calculated as the proportion of Λ_{sim} less than or equal to Λ_{exp}.

Linear statistical models were used to determine the biophysical features that are best able to explain the observed epistasis. The absolute value of the epistasis, , was used as a response variable for our model building. The choice to use the absolute value was necessary to ensure a monotonic relationship between the features and the response variable, as assumed when using linear models. . One could imagine analyzing positive and negative epistasis separately; however, this was not possible due to small sample sizes. All features described above were considered in a standard model selection procedure, including all pairwise interactions terms. For any features where we considered more than one level of abstraction, only one level was included in any given model. To evaluate model performance, the corrected Akaike information criterion (AICc) was used. The corrected criterion was chosen over the standard AIC due to the potential for overfitting models that contain a large number of terms given a small amount of data[51]. Models were generated and tested using R software[52]by considering all permutations of abstracted and non-abstracted features. Model selection was performed using a modified form of stepAIC from the MASS[53]package to perform forward and backward selection based on AICc (further verified by the AICc function of AICcmodavg[54]and compared to standard AIC). Forward selection explores model space by starting with a term-less model and systematically adding terms to find the model with the best value for a given criterion. Conversely, backward selection starts with the complete full-term model and removes terms to find the best model. This model selection process was performed twice with randomized input terms to avoid potential ordering bias (terms treated differently based on their position in the initial list) and the lowest AICc values were compared for consistency. Once we verified that there was no ordering bias, the model with the lowest AICc for both binding and folding was used for further analysis.

To rank the importance of features present in the final statistical models for their effect on epistasis we compared R^2values with and without each feature and it's interactions. Features with larger explanatory power of the observed epistasis will have a larger change in R^2 when removed.

2.4 - Quantification of Experimental Error and Model Validation

In order to develop a model for epistasis, it is important to quantify how much of the observed epistasis could be attributed to error, or noise, in the experimental data. Quantification of overall error is based on the error in three values (??$G_{1,2}$, ??G_1, ??G_2), each of which were determined using a broad range of techniques and conditions from diverse studies (e.g., 60+ for binding). A survey of six studies that contained some of the largest observed epistasis for binding showed the experimental standard error for ??G to be in the range 0.05 - 0.3 kcal/mol[55–57]. However, some studies do explicitly include the error for epistasis (frequently termed the coupling energy). For example, in the case of barnase-barstar, Schreiber et al., reports errors in from 0.2 - 0.39 kcal/mol across 33 mutation pairs[58]and Goldman et al. reports an error of 0.3 kcal/mol across 13 pairs for an Idiotype-AntiIdiotype Protein-Protein complex[59]. There are outliers, such as the study from Pielak, et al. with six mutational pairs in the Iso-1-cytochrome C Peroxidase complex[60]found to have an error range of 0.4 - 1.0 kcal/mol with an average error of 0.75 kcal/mol for six samples; an unusually large error. In summary, the reported error for our curated binding and folding datasets are in the range of 0.2 - 1.0 kcal/mol, with mean around 0.4 kcal/mol. For the remainder of this study, we will use a slightly more conservative estimated error of 0.5 kcal/mol to quantify the amount of observed epistasis.

Since our binding and folding data comes from many different protein systems collected by a diversity of methodologies and laboratories, there is an inherent imbalance in the quantity and quality of data for each system. To test the robustness of our model to this bias, we applied a modified "leave-one-out" procedure. We randomly removed 10% of the protein systems and their data, creating a subset from the remaining 90% of systems. The model selection procedure was performed on this subset to generate a new model. This process of removing 10% of the systems and running model selection was repeated 100 times. The resulting

100 subset models were analyzed and compared to determine which terms appeared, their frequency of appearance, and average performance or ranking when present in a model.

3. Results

To build a statistical model for epistasis in proteins we used data for binding curated from SKEMPI v2.0 (572 mutation pairs), and for folding curated from ProTherm 4 (204 mutation pairs). We first considered the extent to which epistasis was present in our data set. To determine this, we defined an epistasis cutoff; values where | | is larger than the cutoff are considered epistatic, and other values are not. Ideally, the cutoff would be chosen based on the experimental error or uncertainty, however, given that our data come from a broad spectrum of methods and sources, this is not possible to determine for the dataset as a whole (see supplemental Figure S1 for the dataset divisions with various cutoffs).

Figure 1 shows the free energy change of the double mutant as a function of the sum of individual free energies for both binding and folding datasets with a cutoff of 0.5 kcal/mol. Both figures 1 and 2 show that epistasis is present in binding and folding. In both datasets there is a marked trend for large sums of constituent single mutations (sum in EQ 1) to correspond to a double mutant with free energy falling below the 1:1 line (i.e., more stabilizing that predicted by additivity). The opposite is true for constituent mutations with smaller sums.

After ascertaining the extent to which epistasis is present in our data, we investigated how well the separation between mutation sites could explain the epistatic effect. Figure 2 shows the relationship between separation distance and the observed epistasis for binding (top) and folding (bottom). Both show the general expected trend of less epistasis as separation increases. Both also show a larger number of data points for distances with the largest values, or spread in (around 6-10).

Figure 3 shows our analysis to determine whether the apparent decrease in epistasis with increasing separation distance (Figure 2) is due to an actual relationship or a consequence of the larger number of data points at small distances. Figure 3A shows null model ($\sigma(\ )$ is not a function of r) and alternative model ($\sigma(\ )$ exponentially decreases as a function of r) for the likelihood ratio analysis. Figure 3B shows the simulated distribution of the likelihood ratio, Λ,from the analysis with 1000 samples. The experimentally observed likelihood ratio is well outside the distribution of null ratios given by the label "EXP" and has a value of -5.50 compared with the tail of the simulation distribution minimum of -4.59. In simple terms, this results in a p-value of $p < 1/1000$ ($p<0.001$) in strong support of the alternative model.

Table 1 shows a summary of the binding (1A) and folding (1B) statistical models for epistasis in protein systems. Both models have similar predictive power in the range of 25-30%. The final selected binding model contains all features that we considered except for hydrophobicity (seven features, 28 terms including interactions) and depends on $SASA_{abs}$ and secondary structure in addition to binding specific features like the complex type. The folding model is simpler (five features, 12 terms with interactions), and depends on hydrophobicity and $SASA_{rel}$. Features are listed in order according to their relative contribution to the explanatory power of the full model. That is, the highest-ranked feature is the one whose removal leads to the greatest reduction in R^2. For the binding epistasis model, the largest contributor was the complex type, with a change in R^2 of 0.128 upon removal followed by charge with a change in R^2 of 0.078 upon its removal. The remaining terms each contribute ˜5% or less to the predictive power of the binding model. For the folding epistasis model, the largest contributor was hydrophobicity with a change in R^2 of 0.151 upon removal, followed by both size and charge with similar contributions (change in R^2 of 0.0765 and 0.0695 with their removal respectively). The remaining terms each contribute ˜4.5% or less to the predictive power of the folding model.

Table 2 shows the results of 100 trials of our "leave-10%-out" robustness test where 10% of the available systems were randomly removed. These results show that both of our full models are highly robust – with the binding model being slightly more robust than the folding model. All terms present in the full models are present in the leave-10%-out analysis, most occurring in all trials. Additionally, the mean ranks of most terms are identical to the full-data binding model with more variance in the folding model. Graphical

representation of our robustness tests shown in supplemental Figure S2.

Figure 4 further illustrates the results of our statistical model for epistasis in binding. For charge, the subcategory for interactions involving an attractive pairing (0A) contains the most strongly epistatic mutations. While mutations in this subcategory cover a broad range of values, many tend towards positive epistasis; the largest value belongs to this subcategory. Neutral or constant charge states (00) show a near normal distribution centered on zero with some low levels of epistasis. Changes involving a repulsive interaction (0R) contain the least number of data and have a narrow distribution, with fewer large values for in either direction. For the complex type category, the antibody-antigen subcategory shows the most epistasis, including the most positive. TCR/pMHC also contains a large amount of positive epistasis. Cytokine-cytokine is the only subcategory with a negative mean suggesting that mutations in this subcategory tend to have negative epistasis. Generic protein-protein complexes show similar behavior to the neutral charge category; centered on zero, broad spread, but low numbers of epistatic data points. Other categorical features are shown in supplemental Figure S3.

4. Discussion

Before building linear models we first determined the extent to which our datasets contain meaningful epistasis. That is, considering there is uncertainty in the data, where should we draw the line between epistatic and non-epistatic values of ; We estimated (see Methods) that the error for both datasets fall between 0.2 - 1.0 kcal/mol with an average around 0.5 - 0.6 kcal/mol. From this, we estimated a cutoff of 0.5 kcal/mol, i.e., $| | > 0.5$ kcal/mol are considered epistatic. There are further limitations of our dataset; the data is not from randomized studies. Instead, the experiments were generally conducted in a targeted fashion with a priori knowledge of function. This may explain why we find more positive epistasis (more stabilizing that additivity predicts) than negative epistasis (more destabilizing that additivity predicts) as shown in Figure 1. Alternatively, it is possible that more positive epistasis is present in the data because negative epistasis could lead to protein misfolding or non-binding events in the experiments. The former reasoning is an artifact of how the data was generated, and the latter is related to biophysical features of the proteins; both carry different implications for the dataset and warrant future work.

Separation distance is the most intuitive feature expected to contribute to epistasis, because residues that are near each other are more likely to interact than those far apart. Simple comparisons show a decreasing spread of epistasis with increasing distance (Figure 2). The folding data show this most strongly with a sharp peak around the shortest separation distances of approximately 6 , dropping to near zero at larger distances. The binding data show a possible peak around 10 , however, the trend is not as clear. Additionally, with binding there is a paucity of data from 25 to 40 with only one data point around 40 . Our tests using likelihood ratio methods (Figure 3) confirm that separation does play a role in epistasis for both binding and folding. Our alternative model (width of possible values depends on separation) was a better explanation than the null model (no relationship between separation and), with a p-value of $p < 0.001$ in the case of binding, and $p < 0.002$ in the case of folding. The importance of separation is also illustrated in our models (Table 1) where both folding and binding models have negative coefficients for separation distance. In the folding model, a 10 increase in separation between residues results in a decrease in epistasis of 0.416. In the binding model, the effect of separation alone is an order of magnitude less than the folding model and has less significance in the model (p=0.8074). Instead, the effect of separation in the binding model is most strongly characterized by the interaction with charge. With charge alone, changes involving attractive pairings show an increase in epistatic effect whereas changes involving a repulsive pairing show a decrease. The interaction between charge and separation contributes an opposing effect: as separation between residues increases, changes involving attractive and repulsive pairings cause a decrease and increase in epistatic effect respectively. Intuitively, as separation between charged residues, regardless of categorization, increases the net effect of charge on epistasis tends towards zero ($\Delta_{\text{charge}} + \Delta_{\text{charge:separation}} \tilde{} 0$).

In addition to separation distance, amino acid size is present in both models. Size is another feature one might intuitively expect to contribute to epistasis: large absolute changes in size imply that voids are created when residues change from larger to smaller, or that smaller to larger residues create steric clashes. In both

models the coefficient is positive (increases in occur with change in size) with more of an effect in the case of folding (on the order of 10^{-2} vs 10^{-3} in the case of binding). Size interaction terms differ between binding and folding. In the case of binding, when there are changes in size that occur on different protein chains, there is a reduction in epistasis. Otherwise, for all complex types, changes in size lead to an increase in epistasis, most strongly with Antibody-antigen complexes. For folding, size_{net}interacts with hydrophobicity and SASA_{rel} leading to decreases in epistasis. This will be discussed further with the features specific to the folding model.

In the case of both binding and folding, there are unique features that contribute significantly to the observed epistasis. In the case of binding, these elements only apply to binding interactions such as the type of complex (defined by function) and whether both mutations occur on the same side of the binding interaction. Complex type is the most significant contributor to the observed epistasis ([?]R^2= 0.17) with most complexes showing less epistatic effect compared to the reference category of generic protein-protein complexes. There is an exception with Cytokine-cytokine complexes that shows a small increase in epistasis with a coefficient of +0.4677. The interaction side is a smaller contributor compared to complex type ([?]R^2 = 0.0465), with a slight increase in epistatic effect when mutations occur on opposite sides of the binding interaction. This is consistent with intuition; if both mutations are near the binding interface and on opposite sides, they are more likely to directly interact, or propagate effects at the interface. Additional features that contribute to epistasis in binding are secondary structure and SASA_{abs}. Secondary structure has a minor contribution, with a slight increase in epistatic effect when residues belong to different secondary structure types. This is counterbalanced by an interaction with separation distance, where residues that occur in different secondary structures, and are also far apart, lead to a decrease in epistatic effect. This could be due to direct interactions between sites; if they are close together but belong to different secondary structures, they can change these structures either directly or indirectly. This is less likely to happen if they are further apart. SASA_{abs} is the penultimate feature in the model ranking with a very small coefficient (-0.006). This implies that changes in the total exposed surface area due to the two mutations lead to small reductions in the epistatic effect.

Unique to the folding model, hydrophobicity is present, and is the strongest contributor to epistasis with a [?]R^2 of 0.1506. Changes in the net hydrophobicity lead to an increase in the observed epistasis. This is consistent with other studies that have shown that hydrophobicity contributes to predicting folding stabilities with double mutations[61]. Most of the other terms present in the folding model interact with hydrophobicity leading to a stronger effect on epistasis, and a reduction when paired with changes in size, and changes in charge involving attractive interactions.

Since our statistical models for both binding and folding explain approximately 25-30% of the observed epistasis, an important question is: what explains the other 70-75%? We believe the answer lies in dynamical properties that are beyond the scope of what we investigated here. Protein complexes are not static objects, thus static features like those considered in this study are only likely to capture some of the true physical effect they can have on these systems. While a tool like molecular dynamics could potentially help address this question, given the number of mutations and systems considered here, the computational cost would be unreasonably large and will be left as a topic for future study.

Given the size of our datasets, and the imbalanced nature of the data in terms of protein systems, we performed a "leave-10%-out" validation procedure to test the robustness of our models and determine whether there are system-specific effects (see Table 2). We found that our binding model was very robust; all terms appearing in the full model were also present in the validation trials effectively 100% of the time (the least significant term, secondary structure, was missing from three trials). The mean rank was also consistent between the validation trials and the full model ranking for the three most significant terms, the 4th and 5th are switched but close enough to be within a margin of error, the 6th and 7th were also consistently ranked. The folding model was slightly less robust. The effect of hydrophobicity was very robust being ranked first in the full model and appearing in 99 of the 100 validation trials with a mean rank of one. The remaining folding model terms appear between 96% to 100% of the time, however their mean rankings are generally inconsistent with their full model rank, indicating that while they are important to explaining epistasis we

cannot be as certain of their relative contribution.

A limitation in the current study, that is also a limitation for all similar studies, is the lack of comprehensive, diverse, and unbiased datasets. Given the challenges associated with measuring binding or folding free energies for a large number of mutants, these datasets are built with narrow focus and small sample sizes. Such databases tend to be biased toward systems of particular interest. Additionally, they will not contain mutations that result in a nonviable protein or system. This does not make the data any less relevant since in nature proteins must be viable, and thus we should expect similar results (e.g., the preponderance of positive epistasis observed in this study). If we want to understand the nature of epistasis at the level of protein stability, we need to study it across more protein systems in a more systematic fashion. To build a truly predictive model of epistasis, dynamic properties would need to be considered and a larger, more representative sample of data would need to be accessible.

5. Conclusion & Future Work

In this study we investigated possible mechanisms and determined statistical models for pairwise epistasis in proteins based on the largest, most diverse, experimental data available. Mechanistic features were investigated that are intrinsic to the mutating amino acids (e.g., charge, hydrophobicity) or to the proteins (e.g. secondary structure, distance between mutational sites). Using a model selection procedure we ranked these features by their power in explaining the observed epistasis. The resulting models for both binding and folding had similar explanatory power of 25-30% and were composed of similar high-ranked features. The features included in both models were charge, separation distance, and residue size. The largest contributing features were complex type for binding, and hydrophobicity for folding. Our results shed some light on the mechanisms for pairwise epistasis in proteins, and highlights the need for larger datasets. Our study also suggests that development of a truly predictive model for epistasis will likely require difficult to ascertain features such as conformational changes, bond formation, and other propagated mutational effects.

Data availability:

All data and scripts used for the analysis in this manuscript are available at the Ytreberg-Patel lab Github repository:*https://github.com/YtrebergPatelLab/EpistasisStats*

References

1. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet*. 2002;11(20):2463-2468. doi:10.1093/hmg/11.20.2463

2. Whitlock MC, Phillips PC. MULTIPLE FITNESS PEAKS AND EPISTASIS. :31.

3. Moore JH. A global view of epistasis. *Nat Genet*. 2005;37(1):13-14. doi:10.1038/ng0105-13

4. Mackay TF, Moore JH. Why epistasis is important for tackling complex human disease genetics. *Genome Med*. 2014;6(6):125. doi:10.1186/gm561

5. Sanjuan R, Elena SF. Epistasis correlates to genomic complexity. *Proc Natl Acad Sci*. 2006;103(39):14402-14405. doi:10.1073/pnas.0604543103

6. Natarajan C, Inoguchi N, Weber RE, Fago A, Moriyama H, Storz JF. Epistasis Among Adaptive Mutations in Deer Mouse Hemoglobin. *Science*. 2013;340(6138):1324-1327. doi:10.1126/science.1236862

7. Salverda MLM, Dellus E, Gorter FA, et al. Initial Mutations Direct Alternative Pathways of Protein Evolution. Zhang J, ed. *PLoS Genet*. 2011;7(3):e1001321. doi:10.1371/journal.pgen.1001321

8. Draghi JA, Plotkin JB. SELECTION BIASES THE PREVALENCE AND TYPE OF EPISTASIS ALONG ADAPTIVE TRAJECTORIES: SELECTION BIASES EPISTASIS ALONG ADAPTIVE TRAJECTORIES.*Evolution*. 2013;67(11):3120-3131. doi:10.1111/evo.12192

9. Kouyos RD, Silander OK, Bonhoeffer S. Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol Evol*. 2007;22(6):308-315. doi:10.1016/j.tree.2007.02.014

10. Rokyta DR, Joyce P, Caudle SB, Miller C, Beisel CJ, Wichman HA. Epistasis between Beneficial Mutations and the Phenotype-to-Fitness Map for a ssDNA Virus. Malik HS, ed. *PLoS Genet*. 2011;7(6):e1002075. doi:10.1371/journal.pgen.1002075

11. Gong LI, Suchard MA, Bloom JD. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife*. 2013;2. doi:10.7554/eLife.00631

12. Bloom JD, Gong LI, Baltimore D. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. *Science*. 2010;328(5983):1272-1275. doi:10.1126/science.1187816

13. Lozovsky ER, Chookajorn T, Brown KM, et al. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc Natl Acad Sci*. 2009;106(29):12025-12030. doi:10.1073/pnas.0905922106

14. Bridgham JT, Ortlund EA, Thornton JW. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*. 2009;461(7263):515-519. doi:10.1038/nature08249

15. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis. *Science*. 2007;317(5844):1544-1548. doi:10.1126/science.1142819

16. Weinreich DM. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*. 2006;312(5770):111-114. doi:10.1126/science.1123539

17. Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA. Epistasis as the primary factor in molecular evolution. *Nature*. 2012;490(7421):535-538. doi:10.1038/nature11510

18. Kvitek DJ, Sherlock G. Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape. Zhang J, ed. *PLoS Genet*. 2011;7(4):e1002056. doi:10.1371/journal.pgen.1002056

19. Chou H-H, Chiu H-C, Delaney NF, Segre D, Marx CJ. Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. *Science*. 2011;332(6034):1190-1192. doi:10.1126/science.1203799

20. Wei X, Zhang J. Patterns and Mechanisms of Diminishing Returns from Beneficial Mutations. Agashe D, ed. *Mol Biol Evol*. 2019;36(5):1008-1021. doi:10.1093/molbev/msz035

21. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*. 2014;344(6191):1519-1522. doi:10.1126/science.1250939

22. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. *Science*. 2011;332(6034):1193-1196. doi:10.1126/science.1203801

23. Shapiro B, Rambaut A, Pybus OG, Holmes EC. A Phylogenetic Method for Detecting Positive Epistasis in Gene Sequences and Its Application to RNA Virus Evolution.*Mol Biol Evol*. 2006;23(9):1724-1730. doi:10.1093/molbev/msl037

24. Sanjuán R, Cuevas JM, Moya A, Elena SF. Epistasis and the Adaptability of an RNA Virus.*Genetics*. 2005;170(3):1001-1008. doi:10.1534/genetics.105.040741

25. Burch CL, Chao L. Epistasis and Its Relationship to Canalization in the RNA Virus φ6.*Genetics*. 2004;167(2):559-567. doi:10.1534/genetics.103.021196

26. Michalakis Y. EVOLUTION: Epistasis in RNA Viruses. *Science*. 2004;306(5701):1492-1493. doi:10.1126/science.1106677

27. da Silva J, Coetzer M, Nedellec R, Pastore C, Mosier DE. Fitness Epistasis and Constraints on Adaptation in a Human Immunodeficiency Virus Type 1 Protein Region. *Genetics*. 2010;185(1):293-303. doi:10.1534/genetics.109.112458

28. Sanjuan R, Moya A, Elena SF. The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc Natl Acad Sci*. 2004;101(43):15376-15379. doi:10.1073/pnas.0404125101

29. Bonhoeffer S. Evidence for Positive Epistasis in HIV-1. *Science.* 2004;306(5701):1547-1550. doi:10.1126/science.1101786

30. Trindade S, Sousa A, Xavier KB, Dionisio F, Ferreira MG, Gordo I. Positive Epistasis Drives the Acquisition of Multidrug Resistance. Zhang J, ed. *PLoS Genet.* 2009;5(7):e1000578. doi:10.1371/journal.pgen.1000578

31. Moore JH. The Ubiquitous Nature of Epistasis in Determining Susceptibility to Common Human Diseases. *Hum Hered.* 2003;56(1-3):73-82. doi:10.1159/000073735

32. Swint-Kruse L. Using Evolution to Guide Protein Engineering: The Devil IS in the Details. *Biophys J.* 2016;111(1):10-18. doi:10.1016/j.bpj.2016.05.030

33. Melero C, Ollikainen N, Harwood I, Karpiak J, Kortemme T. Quantification of the transferability of a designed protein specificity switch reveals extensive epistasis in molecular recognition. *Proc Natl Acad Sci.* 2014;111(43):15426-15431. doi:10.1073/pnas.1410624111

34. Miton CM, Tokuriki N. How mutational epistasis impairs predictability in protein evolution and design: How Epistasis Impairs Predictability in Enzyme Evolution.*Protein Sci.* 2016;25(7):1260-1272. doi:10.1002/pro.2876

35. Reetz MT. The Importance of Additive and Non-Additive Mutational Effects in Protein Engineering. *Angew Chem Int Ed.* 2013;52(10):2658-2666. doi:10.1002/anie.201207842

36. Wells JA. Additivity of mutational effects in proteins. *Biochemistry.* 1990;29(37):8509-8517. doi:10.1021/bi00489a001

37. Dellus-Gur E, Elias M, Caselli E, et al. Negative Epistasis and Evolvability in TEM-1 β-Lactamase—The Thin Line between an Enzyme's Conformational Freedom and Disorder. *J Mol Biol.* 2015;427(14):2396-2409. doi:10.1016/j.jmb.2015.05.011

38. Gonzalez CE, Ostermeier M. Pervasive Pairwise Intragenic Epistasis among Sequential Mutations in TEM-1 β-Lactamase. *J Mol Biol.* 2019;431(10):1981-1992. doi:10.1016/j.jmb.2019.03.020

39. Olson CA, Wu NC, Sun R. A Comprehensive Biophysical Description of Pairwise Epistasis throughout an Entire Protein Domain. *Curr Biol.* 2014;24(22):2643-2651. doi:10.1016/j.cub.2014.09.072

40. Istomin AY, Gromiha MM, Vorov OK, Jacobs DJ, Livesay DR. New insight into long-range non-additivity within protein double-mutant cycles. *Proteins Struct Funct Bioinforma.* 2007;70(3):915-924. doi:10.1002/prot.21620

41. Yu H, Dalby PA. Coupled molecular dynamics mediate long- and short-range epistasis between mutations that affect stability and aggregation kinetics.*Proc Natl Acad Sci.* 2018;115(47):E11043-E11052. doi:10.1073/pnas.1810324115

42. Jemimah S, Gromiha MM. Exploring additivity effects of double mutations on the binding affinity of protein-protein complexes. *Proteins Struct Funct Bioinforma.* 2018;86(5):536-547. doi:10.1002/prot.25472

43. Justina Jankauskaite, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, Iain H Moal. "SKEMPI 2.0: An updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation". *Bioinformatics.* bty635. doi:10.1093

44. Sarai A, Uedaira H, Bava KA, Kitajima K, Gromiha MM. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.* 2004;32(suppl_1):D120-D121. doi:10.1093/nar/gkh082

45. Berman HM, Westbrook J, Feng Z, et al. The Protein Databank. *Nucliec Acids Res.* 2000;28:235-242.

46. *The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.*

47. Eastman P, Swails J, Chodera JD, et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. Gentleman R, ed. *PLOS Comput Biol*. 2017;13(7):e1005659. doi:10.1371/journal.pcbi.1005659

48. The FoldX web server: an online force field. *Nucleic Acids Res*. 2005;33(Web Server):W382-W388. doi:10.1093/nar/gki387

49. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22(12):2577-2637. doi:10.1002/bip.360221211

50. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. Maximum Allowed Solvent Accessibilites of Residues in Proteins. *PLoS ONE*. 2013;8(11). doi:10.1371/journal.pone.0080635

51. Burnham KP, Anderson DR, Burnham KP. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. Springer; 2002.

52. R Core Team.*R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2020. https://www.R-project.org/

53. Venables WN, Ripley BD. *Modern Applied Statistics with S*. Fourth. Springer; 2002. http://www.stats.ox.ac.uk/pub/MASS4

54. Mazerolle MJ.*AICcmodavg: Model Selection and Multimodel Inference Based on (Q)AIC(c)*.; 2020. https://cran.r-project.org/package=AICcmodavg

55. Jacksod SE, Fersht AR. Contribution of Residues in the Reactive Site Loop of Chymotrypsin Inhibitor 2 to Protein Stability and Activityt. :8.

56. Pons J, Rajpal A, Kirsch JF. Energetic analysis of an antigen/antibody interface: Alanine scanning mutagenesis and double mutant cycles on the hyhel-10/lysozyme interaction. *Protein Sci*. 1999;8(5):958-968. doi:10.1110/ps.8.5.958

57. Li Y, Li H, Smith-Gill SJ, Mariuzza RA. Three-Dimensional Structures of the Free and Antigen-Bound Fab from Monoclonal Antilysozyme Antibody HyHEL-63^{+} , $^{++}$. *Biochemistry*. 2000;39(21):6296-6309. doi:10.1021/bi000054l

58. Schreiber G, Fersht AR. Energetics of protein-protein interactions: Analysis ofthe Barnase-Barstar interface by single mutations and double mutant cycles.*J Mol Biol*. 1995;248(2):478-486. doi:10.1016/S0022-2836(95)80064-6

59. Goldman ER, Dall'Acqua W, Braden BC, Mariuzza RA. Analysis of Binding Interactions in an Idiotope-Antiidiotope Protein-Protein Complex by Double Mutant Cycles $^{+}$. *Biochemistry*. 1997;36(1):49-56. doi:10.1021/bi961769k

60. Pielak GJ, Wang X. Interactions between Yeast Iso-1-cytochrome *c* and Its Peroxidase^{+}. *Biochemistry*. 2001;40(2):422-428. doi:10.1021/bi002124u

61. Huang L-T, Gromiha MM. Reliable prediction of protein thermostability change upon double mutation from amino acid sequence. *Bioinformatics*. 2009;25(17):2181-2187. doi:10.1093/bioinformatics/btp370

Tables:

1A: Binding Model

Feature	Categorical Breakdown	Removal ([?]R^2)	Coefficient
Full Model	572	0.2991	
Intercept			0.6994
Complex Type		0.1275	
AB/AG	66 (8)		-0.8401

Cyto/Cyto	69 (4)		0.4677
GTPase/other	85 (7)		-0.4134
TCR/pMHC	58 (9)		-1.0090
Pr/PI	294 (30)		0.0000
Charge		0.0778	
0	450		0.0000
0A	69		1.0195
0R	53		-0.7424
Separation		0.0522	-0.0025
Interaction Side		0.0464	0.5889
1	399		
0	173		
$Size_{net}$		0.0427	0.0022
$SASA_{abs}$		0.0327	-0.0060
Secondary Structure		0.0162	0.5837
1	315		
0	257		
Binding Model Interaction Terms	Binding Model Interaction Terms	Binding Model Interaction Terms	Binding Model
Feature1:Feature2			Coefficient
Separation:Secondary Structure			-0.0331
$Size_{net}$:Interaction Side			-0.0052
Separation:Charge			
0A			-0.0882
0R			0.0454
$SASA_{abs}$:Complex Type			
AB/AG			-0.0140
Cyto/Cyto			0.0072
GTPase/other			0.0072
TCR/pMHC			0.0062
$Size_{net}$:Complex Type			
AB/AG			0.0130
Cyto/Cyto			0.0023
GTPase/other			0.0008
TCR/pMHC			0.0060
Interaction Side:Charge			
0A			0.7289
0R			0.1460
Interaction Side: Complex Type			
AB/AG			0.1376
Cyto/Cyto			-1.4297
GTPase/other			-0.3189
TCR/pMHC			0.0062

1B: Folding Model

Feature	Categorical Breakdown	Removal ([?]R^2)	Coefficient
Full Model	204	0.2578	
Intercept			-0.0607
HP		0.1506	0.0746
0	133		
1	71		

$Size_{net}$		0.0765	0.0120
Charge		0.0695	
0	174		0.0000
0A	13		1.8631
0R	17		-0.7356
Separation		0.0446	-0.0416
$SASA_{rel}$		0.0383	0.3269
Folding Model Interaction Terms	Folding Model Interaction Terms	Folding Model Interaction Terms	Folding Model In
Feature1:Feature2			Coefficient
$SASA_{rel}$:$Size_{net}$			-0.0437
$SASA_{rel}$:Separation			0.1577
$Size_{net}$:HP			-0.0103
$SASA_{rel}$:HP			2.7022
HP:Charge			
0A			-2.0467
0R			0.8380

Table 1: Summaries of epistasis models for binding (A) and folding (B). The leftmost column (column one) contains features. Any categorical abstractions are listed directly below the category with right justification. Column two gives the specific number of mutation pairs for a given category, where applicable. For complex type specifically, the number of complexes of that type are indicated in parentheses. Column three is the change in R^2(ΔR^2), i.e., how much poorer the model fits the data after removing this feature. In the case of the full model, column three is the R^2. Removal of a feature also removes all subcategories and any interaction terms involving the feature. Column four lists coefficients for the feature/interaction term in the full model. The rightmost column contains p-values for the features, and features within a given category.

2A Binding Validation

Feature	Mean Rank	Average [?]R^2	Number of Models (/100)	In Full Model
Complex Type	1.04	0.134	100	Yes
Charge	2.01	0.082	100	Yes
Separation	3.66	0.0560	100	Yes
$Size_{net}$	4.45	0.047	100	Yes
Interaction side	4.66	0.046	100	Yes
SASA	5.64	0.0380	100	Yes
Secondary Structure	6.68	0.022	97	Yes
Hydrophobicity	7.469	0.018	32	No
2B Folding Validation	2B Folding Validation	2B Folding Validation	2B Folding Validation	2B Folding Validation
Feature	Mean Rank	Average [?]R^2	Number of Models (/100)	In Full Model
Hydrophobicity	1.212	0.15	99	Yes
Charge	3.083	0.082	96	Yes
Secondary Structure	3.213	0.0810	47	Yes
$Size_{net}$	3.22	0.074	100	Yes
Separation	3.98	0.058	100	Yes
SASA	4.897	0.041	97	Yes
Number Residues	5.269	0.0380	26	No

Table 2: Results from 100 trials of our "leave-10%-out" model robustness test for binding (A) and folding (B). The feature is indicated by the first column. The second column indicates the average rank across all trials the given feature appeared in, lower numbers suggest more robust features. The third column indicates

the average [?]R^2 from all trials the feature appeared in (higher numbers suggest more robustness), the fourth column indicates the total number of trials a given feature occurred in out of 100 possible, the fifth column indicates whether the feature was present in the full model, and the last column indicates the rank of the feature in the full model.

Figure legends:

Figure 1: Epistasis scatterplots for binding (A) and folding (B). Both figures use a cutoff of 0.5 kcal/mol and show data characterized as no epistasis (black), positive epistasis (blue), and negative epistasis (red).

Figure 2: Observed epistasis as a function of alpha-carbon separation between mutation sites for binding (A) and folding (B). Black indicates no-epistasis using our cutoff of 0.5 kcal/mol, and blue and red indicate positive and negative epistasis, respectively.

Figure 3: (A) Comparison between the alternative (left) and null (right) models for epistatic effect, , as a function of separation distance, r . Results of log(likelihood) ratio test for separation distance with 1000 samples for simulated data for binding affinity (B) and folding stability (C). These plots show the alternative model is a significantly better explanation of the data than the null model.

Figure 4: Comparison of binding model of epistasis for the categories of charge (A) and complex type (B). The mean value for a given subcategory is indicated by a black dot. The barplots show the histograms within the categories. In parenthesis is the number of mutation pairs belonging to each category. For the complex type, the number of complexes belonging to each category are shown in square brackets.

Figure 1:

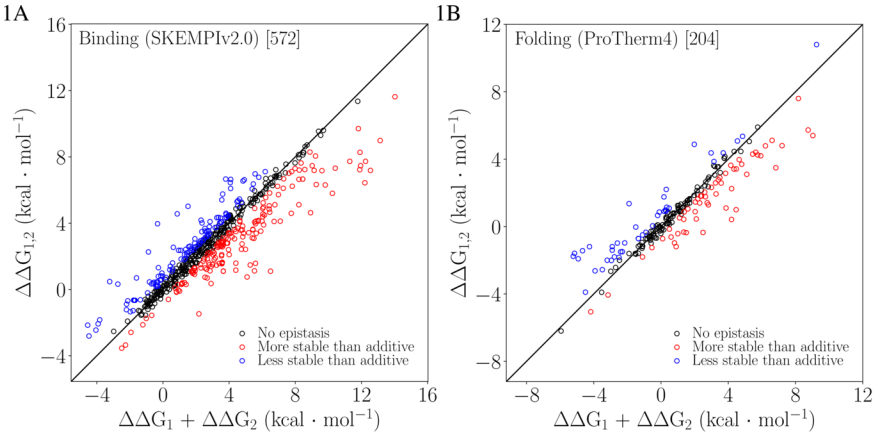


Figure 2:

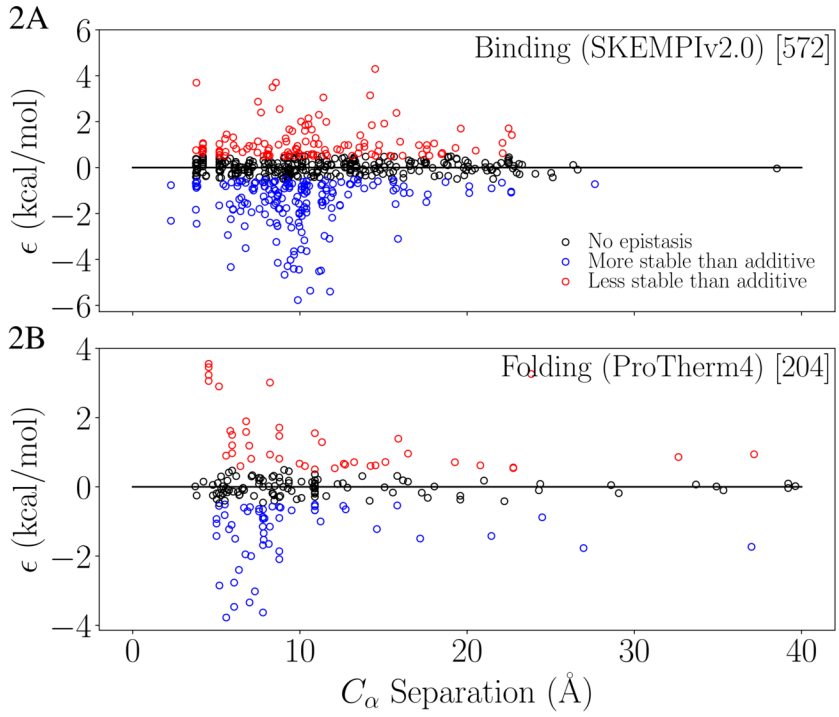


Figure 3:

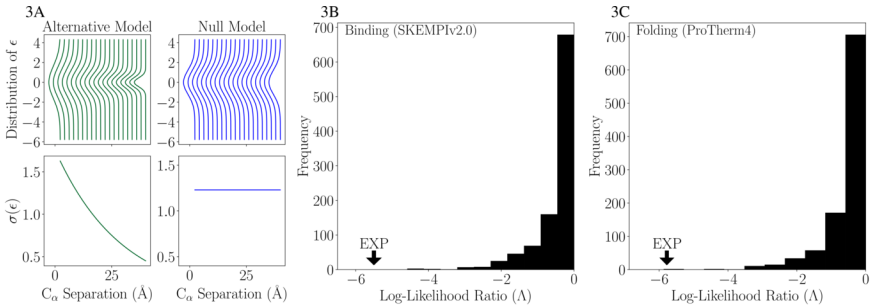
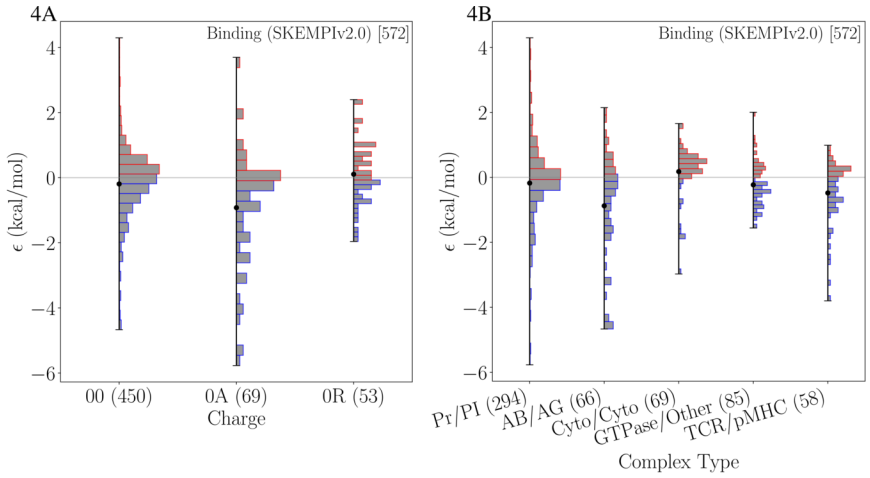


Figure 4:



Supplemental figure legends:

Figure S1: Comparison of cutoffs for epistasis for both binding affinity (left) and folding stability (right). Purple bars correspond to mutants characterized as epistatic while black indicates non-epistatic. Within the sign breakdown, blue indicates negative epistasis and red indicates positive.

Figure S2: Leave-10-percent-out test results for both binding (left) and folding (right) that demonstrate our models are robust. The top plots show the ?R^2 (y-axis) for the 100 runs with complexes removed (run indicated on x-axis). The color coding corresponds to a given feature. Abstracted features were combined to a single heading for that feature for simplicity. The bottom plots indicate the range of ?R^2 given by an error bar. The y-axis (and color coding) indicates the feature.

Figure S3: Comparison of epistasis by subcategory for categorical or boolean features. The mean value for a given subcategory is indicated by a black dot. The barplots show the histograms within the categories. In parenthesis is the number of mutation pairs belonging to each category. For the complex type, the number of complexes belonging to each category are shown in square brackets.

Figure S1:

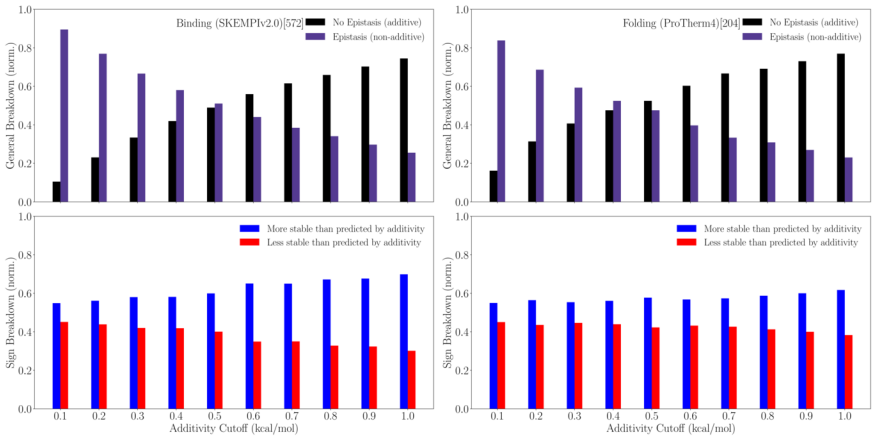


Figure S2:

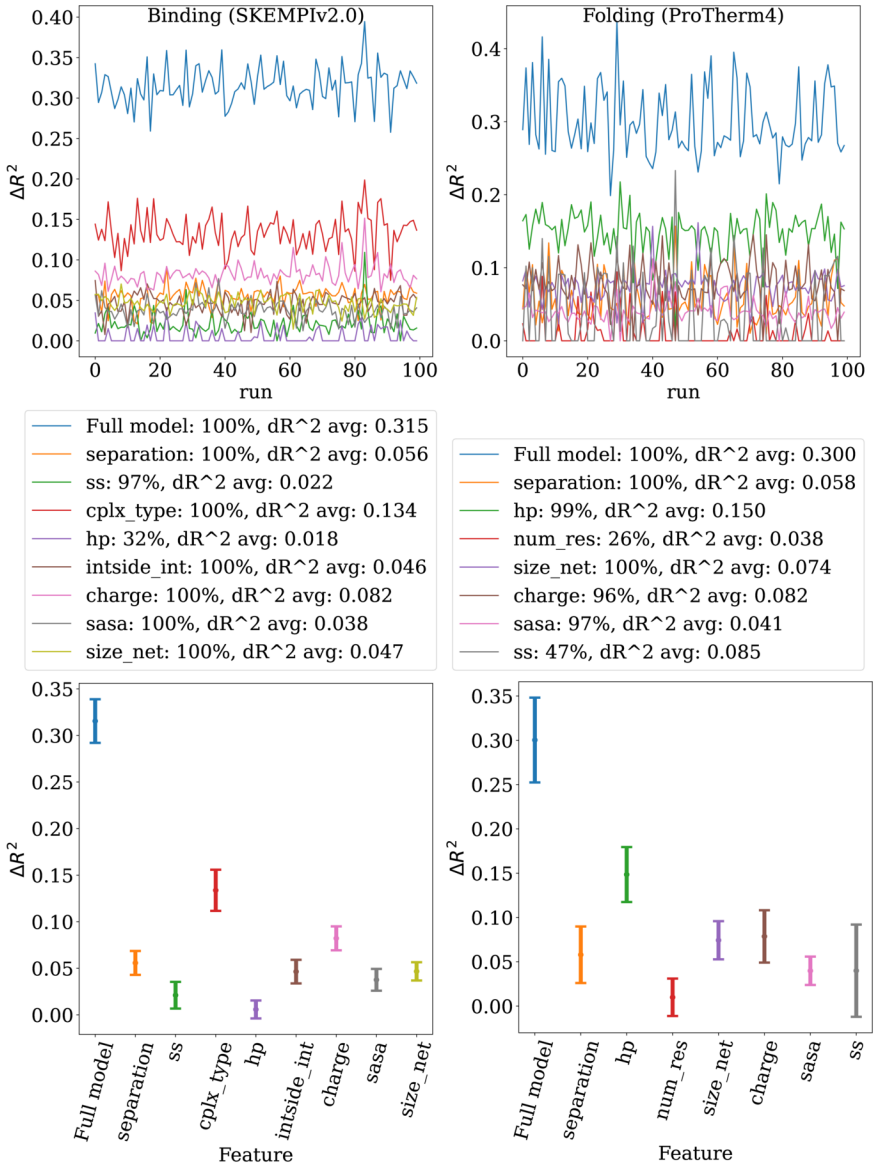


Figure S3: