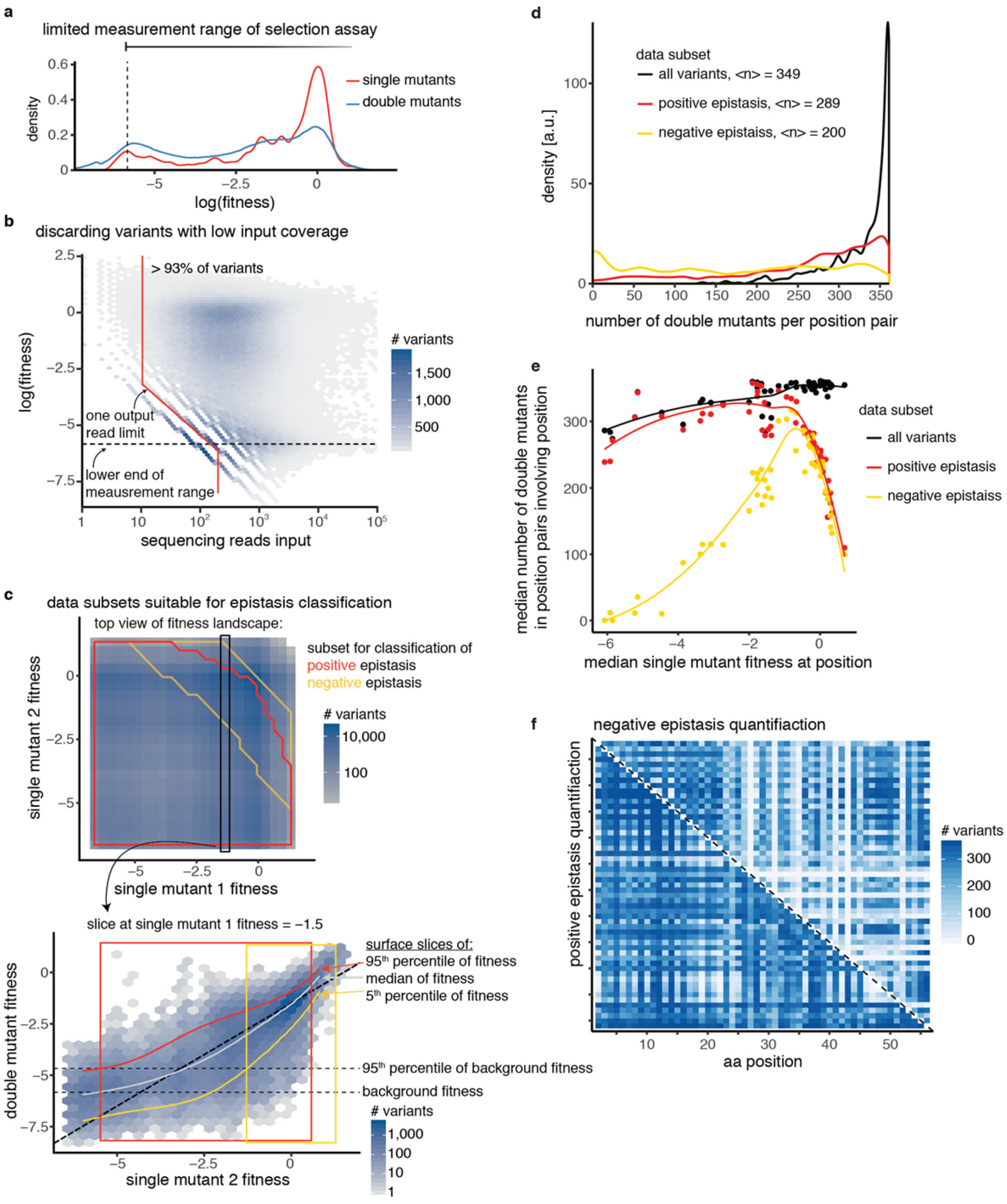
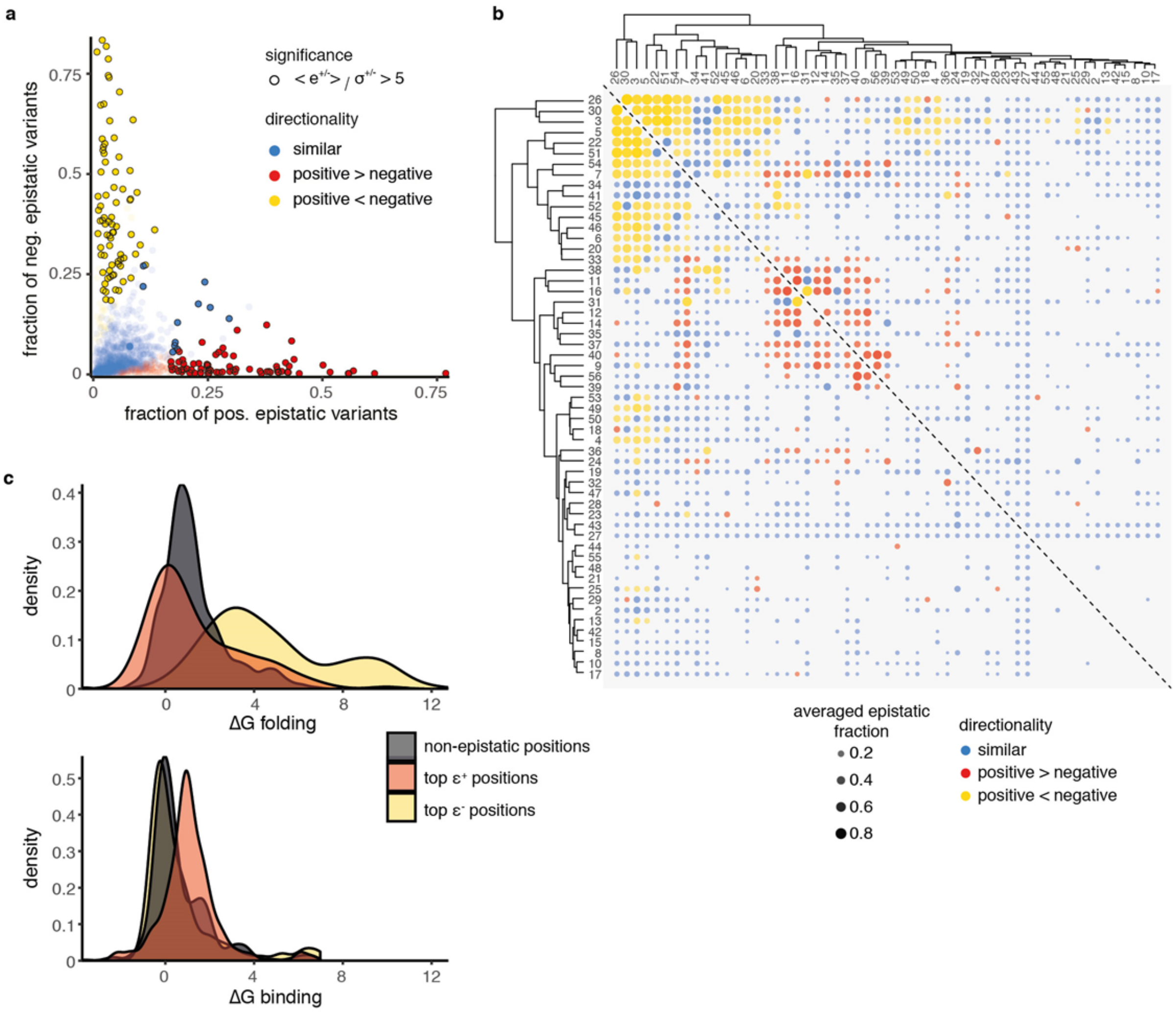


Supplementary Figure 1

GB1 deep mutational scanning data processing

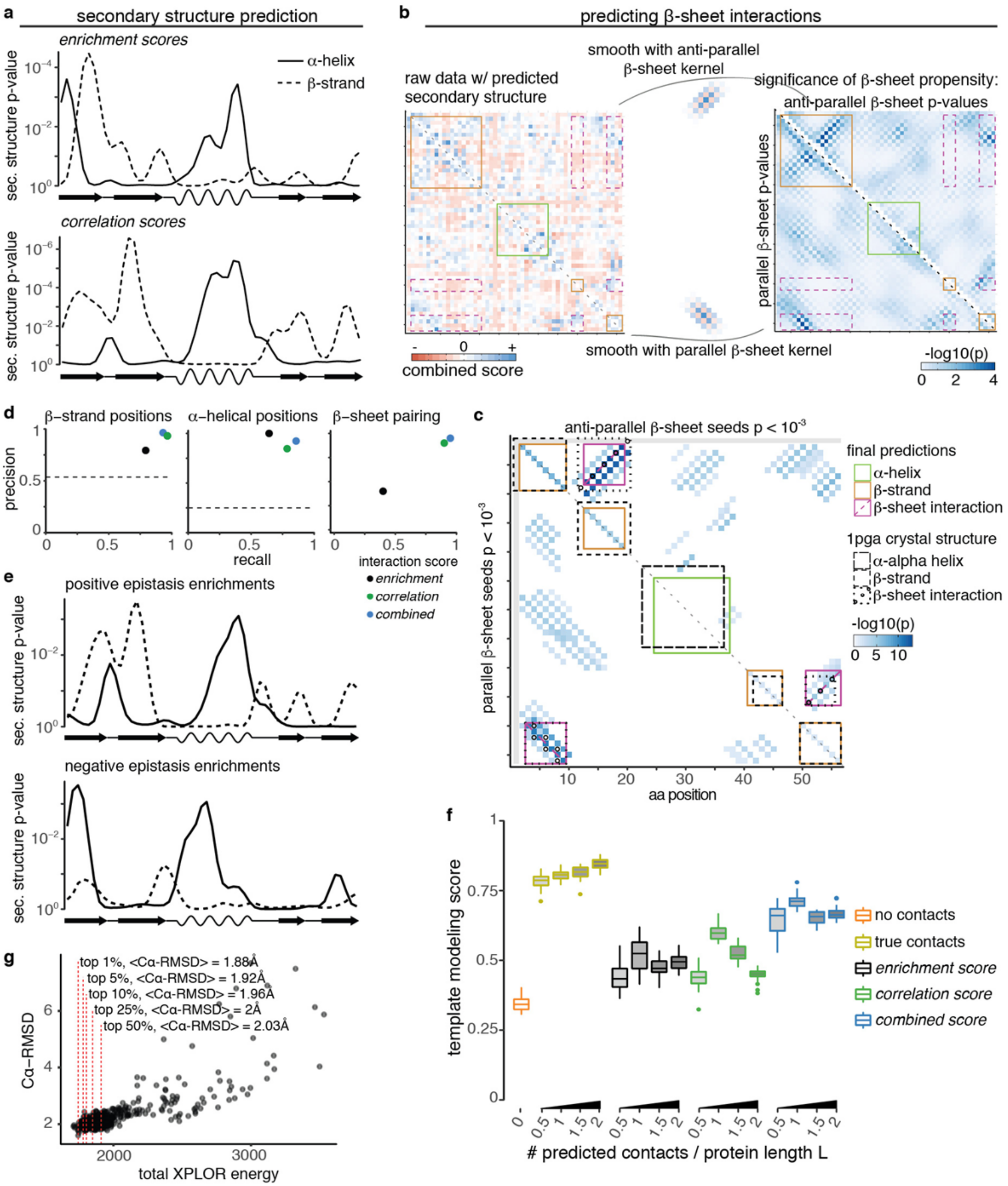
a, Distribution of fitness values for single and double mutant variants. Lower peak in distributions indicates lower limit of fitness assay measurement range (see Methods). **b,** Two-dimensional variant density showing dependency of fitness values on sequencing read counts in input library. For variants with very low coverage in the input library low fitness values cannot be accurately estimated. Red line shows sequence read cutoff used for variant inclusion (93% of variants included for downstream analysis). Horizontal dashed line indicates lower limit of fitness assay measurement range. **c**, Reliable quantification of positive and negative epistasis is limited to subsets of the data. Upper plot shows two-dimensional double mutant variant density in single mutant fitness space. Red outline shows single mutant fitness space that enables positive epistasis quantification. Yellow outline shows single mutant fitness space that enables negative epistasis quantification. Lower plot shows an example slice through fitness landscape at single mutant fitness = -1.5. Red, grey and yellow curves show slices through quantile fitness surfaces of 95th percentile, median and 5th percentile, respectively (see Figure 1b). Diagonal dashed line shows $double\ mutant\ fitness\ =\ single\ mutant\ 2\ fitness - 1.5$ (expected fitness = observed fitness). Horizontal dashed lines give lower limit of fitness assay measurement range and the 95th percentile of fitness values of variants dominated by background fitness effects. Red and yellow boxes indicate the range that includes 99% of variants within the slice that are suitable for positive or negative epistasis quantification, respectively. **d**, Distribution of number of double mutant variants across all position pair. Legend gives median number of double mutants per position pair for different data subsets. **e**, Relationship between median single mutant fitness at a position and the median number of double mutants observed in position pairs the position is involved in. Curves are loess smoothed. Across all variants, positions with stronger fitness effects show lower coverage of double mutants. Restrictions for quantification of positive epistasis additionally reduce coverage for positions with mostly neutral or positive effects. Finally, restrictions for quantification of negative epistasis strongly reduce coverage for positions with strong fitness effects, due to the lower measurement limit of the fitness assay. **f**, Number of double mutants for which positive (lower left triangle) or negative (upper right triangle) epistasis can be quantified per position pair plotted on the interaction matrix.



Supplementary Figure 2

Positive and negative epistasis enrichments across position pairs

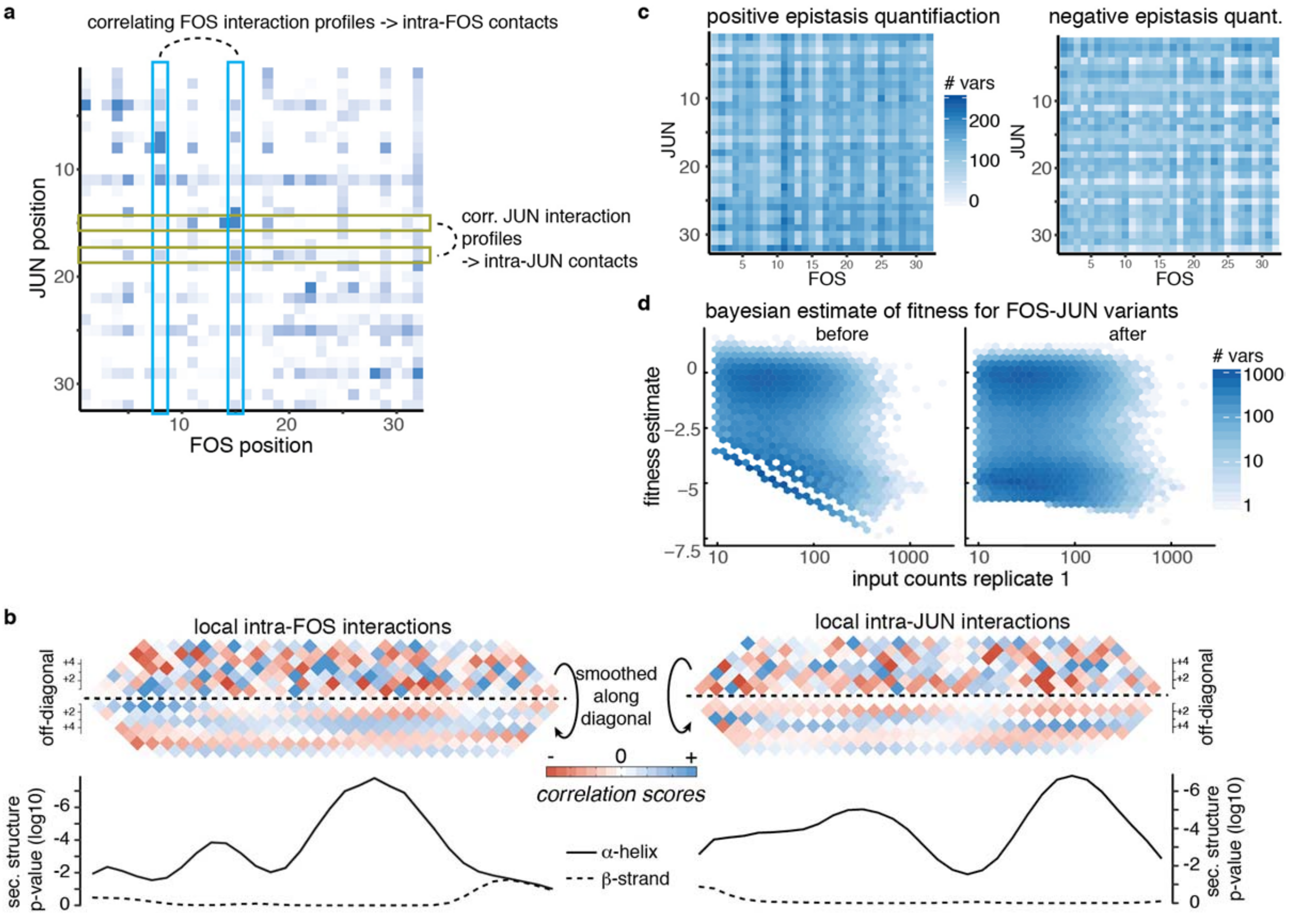
a, Position pair-wise fractions of positive and negative epistatic variants. Black circles mark position pairs with highly significant fractions (either positive or negative, or both); red dots: positive epistatic fraction significantly larger than negative epistatic fraction; yellow dots: negative epistatic fraction significantly larger than positive epistatic fraction; blue dots: no significant differences. **b**, Hierarchical cluster analysis of epistatic fraction patterns. Positions are clustered according to the Euclidean distance of their mean epistatic fractions (weighted average of positive and negative epistatic fractions, weights are inverse variances of fractions in resampling runs) to all other positions. Note that directionality of interactions (positive or negative fractions more significant) was not used for clustering but only marked post-analysis. Clustering shows two highly interconnected clusters of positions that interact mostly positively or negatively within each cluster but hardly any strong interactions are observed between the two clusters (with exception of positions 7 and 54). **c**, Cross-referencing positions involved in epistatic interactions with inferred free energy values of GB1 stability and Immunoglobulin G binding (Otwinowski, J., *Mol Biol Evol*. **10**, 2345-2354, 2018) shows that positions involved in the top negative epistatic interactions are strongly enriched for contributions to stability (upper panel), while positions involved in the top positive epistatic interactions are enriched for contributions to binding (lower panel), as previously noticed (Olson, C.A. *et al*., Current Biology, **24**, 2643-2651, 2014). “Top e+ positions”: Positions being involved in at least 3 of the top 28 positive epistatic interactions. “Top e- positions”: Positions being involved in at least 3 of the top negative epistatic interactions. Non-epistatic positions: all other positions in the protein. This suggests that epistasis arises mostly from functional or energetic couplings between positions. Moreover, the periodicity of local negative epistatic interactions provide evidence for a shift towards an alternative three-helical conformation that has been previously reported for this sequence family (Supplementary Figure 3e; Alexander, P. et al., *PNAS.* **106**, 21149-21154, 2009). This modular organization of epistasis is thus reminiscent of the concept of energetically-coupled protein sectors identified from patterns of sequence co-evolution (Lockless, S.W. & Ranganathan, R., *Science*. **286**, 295-299, 1999).



Supplementary Figure 3

Secondary and tertiary structure prediction for GB1 domain

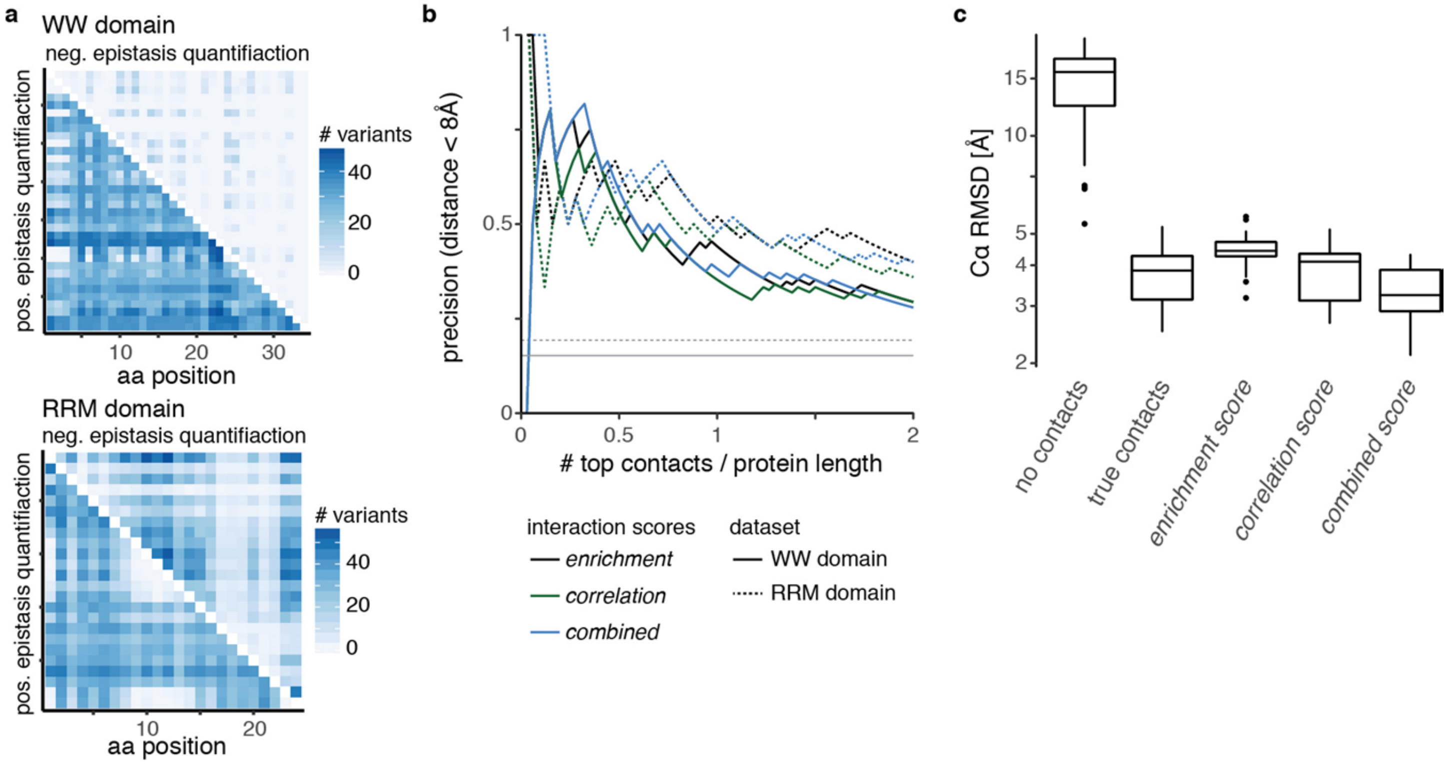
a, Secondary structure propensity derived from kernel smoothing for *enrichment* (upper*)* and *correlation scores* (lower) (one-sided permutation test). **b**, Detecting beta sheet pairing with two-dimensional kernel smoothing. Left plot shows raw combined score interaction matrix, with secondary structure element predictions (see Figure 3a,b) marked as squares along the diagonal (orange – beta strand, green – alpha helix). Off-diagonal purple rectangles show potential regions of beta sheet pairing. Right plot: calculation of beta sheet pairing propensity with beta sheet kernels (one-sided permutation test). Upper right triangle shows anti-parallel beta sheet propensity. Lower left shows parallel beta sheet propensity. **c**, Matrix of aggregated propensities of beta sheet pairing stretches (upper right – anti-parallel, lower left – parallel, $p < 10^{-3}$, one-sided permutation test, Fisher's method for aggregation of p-values) and the predictions for beta sheet pairing and secondary structure elements derived from them. In brief, predictions are performed by picking the highest propensity stretch that is consistent with predicted beta strands, if necessary modifying beta strand predictions (e.g. introducing an initially not predicted split between beta strands 1 and 2), then disregarding all stretches that conflict with the picked top-stretch. This procedure is repeated until no more beta sheet stretches with aggregated propensity $P < 10^{-3}$ are left. Finally, beta strand predictions are updated such that only positions involved in a beta sheet interaction are retained. Reference elements from crystal structure are shown as comparison (lower triangle – parallel beta sheets, upper triangle – anti-parallel beta sheet, diagonal –secondary structure elements). **d**, Precision and recall for beta strand, alpha helix and beta sheet predictions from *enrichment*, *correlation* and *combined scores* (in comparison to crystal structure). Dashed lines for beta strand and alpha helical positions give random expectation. Random expectation for beta sheet pairing precision is below 1%. **e**, Secondary structure propensities derived from local positive or negative epistatic enrichments (one-sided permutation test). The upper panel shows secondary structure propensity derived from positive epistatic interactions, which are in line with secondary structure elements in the GB1 crystal structure (PDB entry 1pga). The lower panel shows secondary structure propensity derived from negative epistatic interactions, which are devoid of beta strand signals and instead show a three-helical pattern, which is reminiscent of the three-helical structure of the protein G A domain that binds albumin (Alexander, P. et al., *PNAS.* **106**, 21149-21154, 2009). **f**, Template modeling score of top 5% structural models (n = 25) compared to crystal structure 1pga and the dependency on number of predicted contacts used. “No contacts” – only restraints for secondary structure predicted by PSIPRED. “True contacts’’ – restraints derived from 0.5-2*L contacts (linear sequence separation greater than 5 aa, random subset), secondary structure elements and beta sheet interactions from crystal structure. All other: restraints derived from top 0.5-2*L contacts, secondary structure element and beta sheet interaction predictions from the three interaction scores, as indicated by color. Boxplots: Box covers 1st to 3rd quartile of the data, with middle bar indicating median. Whiskers extend at maximum to 1.5 times the inter quartile range away from the box. **g**, Accuracy estimates of structural simulations are robust to top model cut-off. Shown are accuracies ($C\alpha$ root-mean-square deviation) of all 500 structural models generated from combined score-derived restraints compared to GB1 reference structure as function of internal XPLOR energies, used to rank models. Average $C\alpha - RMSD$ values of different inclusion cut-offs for top models are shown.



Supplementary Figure 4

FOS-JUN interaction dataset properties

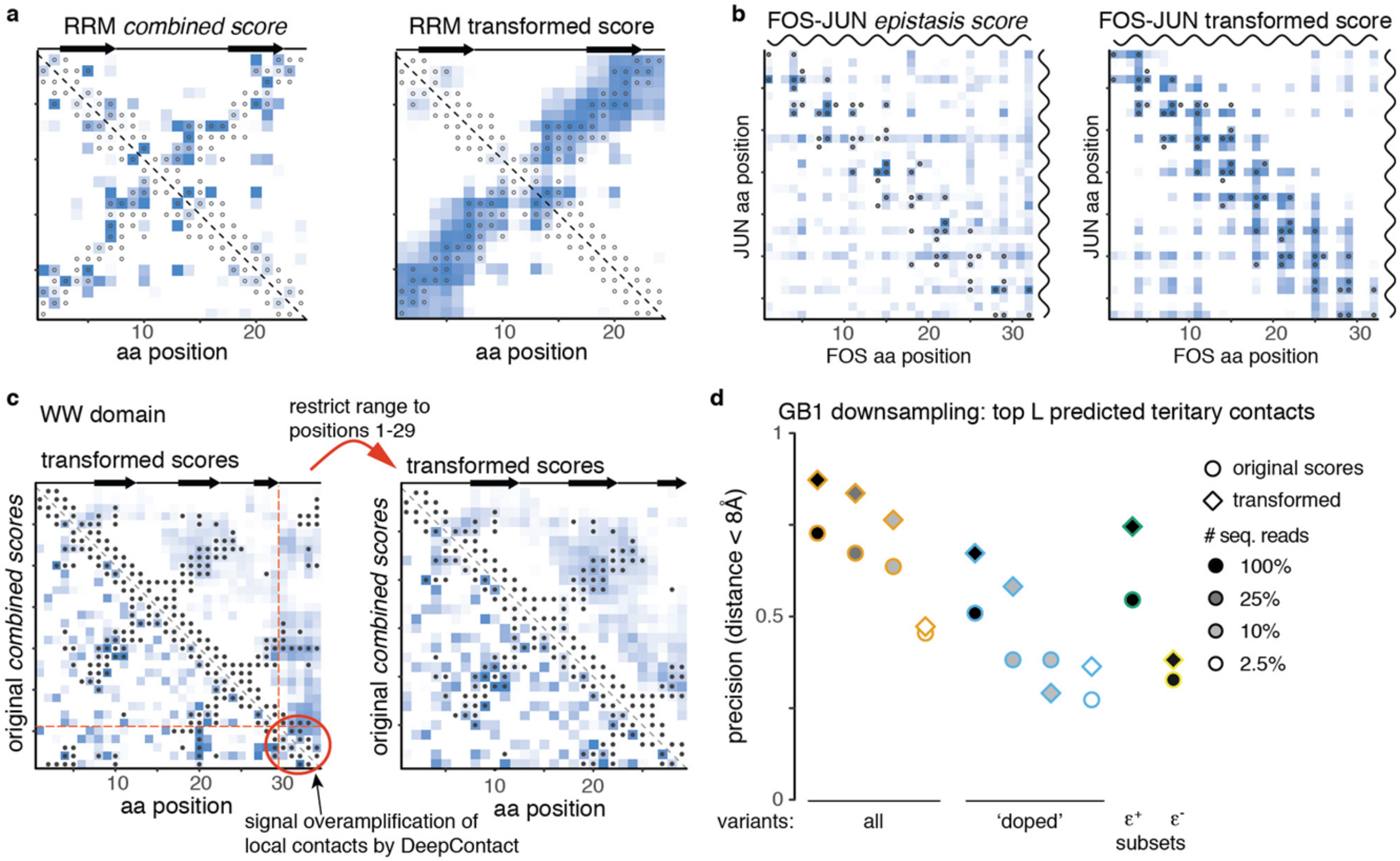
a, Learning about intra-molecular contacts in FOS or JUN from epistatic pattern correlations. Column-wise correlation of epistatic patterns of the trans interaction score map serve to calculate intra-FOS *correlation scores* and thus reveal relationships between positions in FOS. Likewise, row-wise correlation of epistatic patterns reveal relationships between positions in JUN. **b**, Local interactions in intra-molecular *correlation scores* reveal secondary structures of protein interaction partners. Upper panels: Data above diagonal shows *correlation score* data close to the diagonal, i.e. local interactions. Data below the diagonal is smoothed with a Gaussian kernel to reveal interaction periodicity. Lower panels: Secondary structure propensities derived from kernel smoothing (one-sided permutation test, see also Figure 3a-c). **c**, Number of double mutants for which positive (left) or negative (right) epistasis can be quantified per position pair plotted on the trans-interaction matrix of the FOS-JUN interaction. **d**, Bayesian estimation of fitness values in FOS-JUN interaction data. Mutants with low input sequencing coverage display limited measurement range and many dropouts (~15% of variants without reads in output). Left panel shows original fitness distribution as function of input coverage in selection replicate 1, right panel shows Bayesian estimates of fitness as function of input coverage in selection replicate 1.



Supplementary Figure 5

RRM and WW domain dataset properties

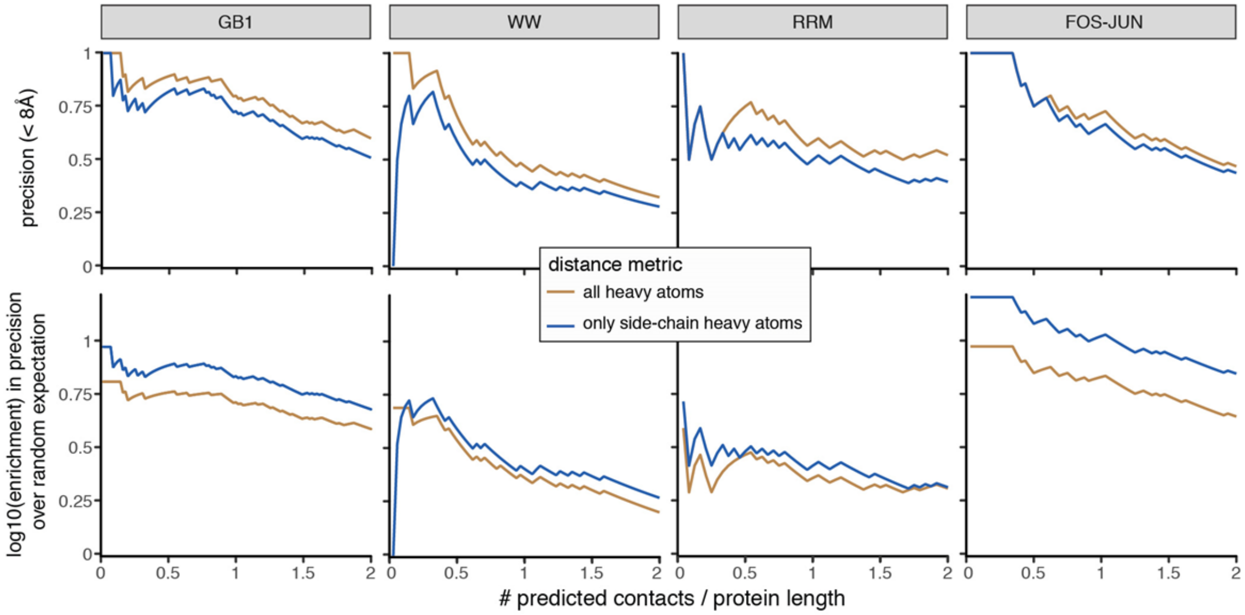
a, Number of double mutants for which positive (lower left triangle) or negative (upper right triangle) epistasis can be quantified per position pair plotted on the interaction matrix for WW (left) and RRM (right) domains. **b**, Precision of interaction scores to predict direct contacts (distance < 8Å in reference structure) as a function of top scoring position pairs for WW and RRM domain interaction scores. Grey horizontal lines give random expectation. Only position pairs with linear sequence separation greater than 5 amino acids are considered. **c**, Accuracy ($C\alpha$ root-mean-square deviation) of top 5% structural models (n = 25) of the WW domain (core positions 6-29) generated from deep mutational scanning data derived restraints compared to reference structure (PDB entry 1k9q). Structural models were generated in XPLOR-NIH by simulated annealing with restraints derived from top 17 tertiary contacts and secondary structure elements predicted by PSIPRED. No beta sheet pairing information was used. Note that despite similar precision of predicted contacts, *enrichment* and *combined score*-derived WW domain structural models are more accurate than *enrichment score*-derived models. Boxplots: Box covers 1st to 3rd quartile of the data, with middle bar indicating median. Whiskers extend at maximum to 1.5 times the inter quartile range away from the box.



Supplementary Figure 6

Deep learning improves contact prediction from deep mutagenesis data

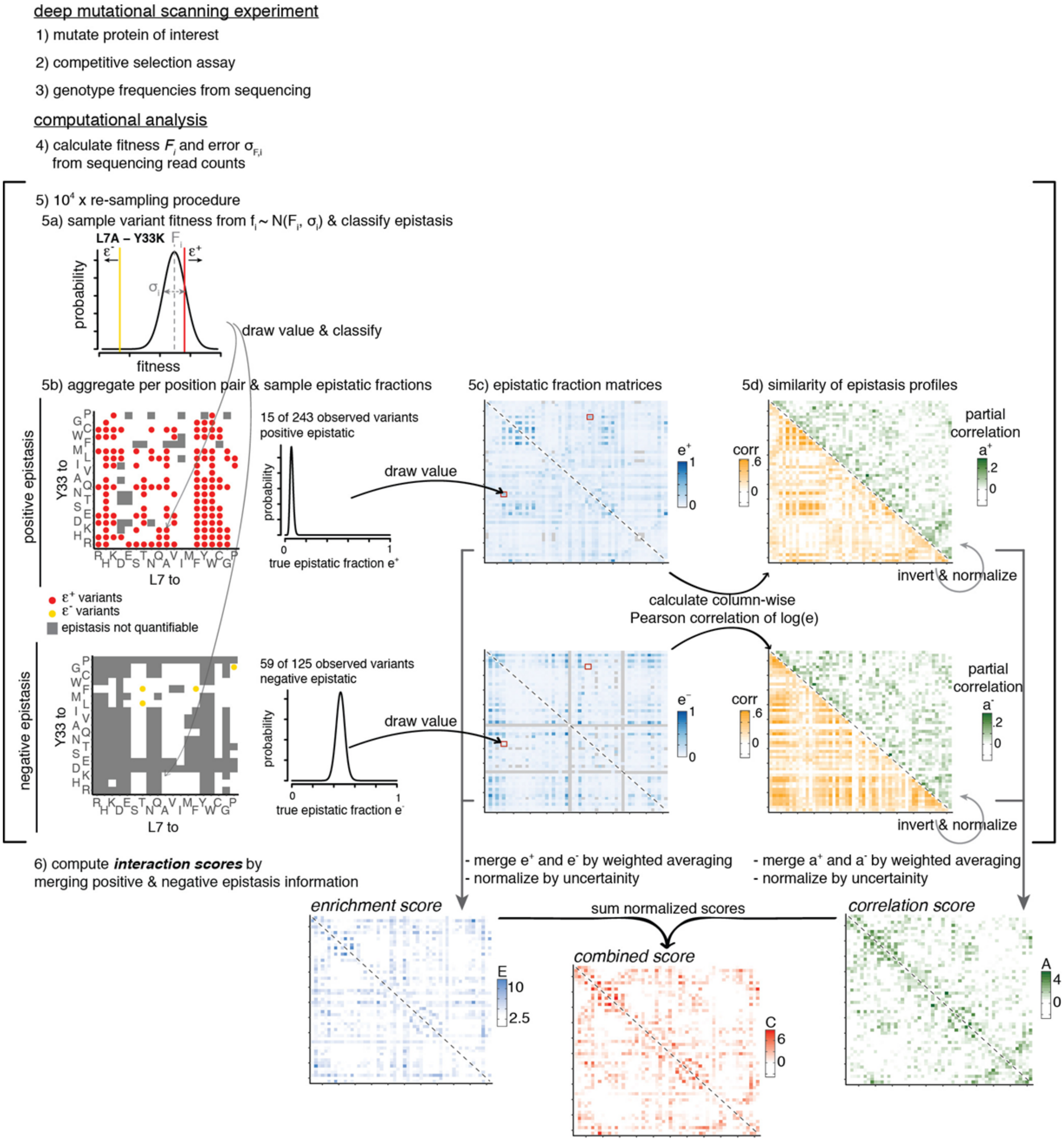
a, RRM domain *combined score* interaction map before (left) and after (right) *DeepContact* transformation. **b**, FOS-JUN trans-interaction *enrichment score* interaction map before (left) and after (right) *DeepContact* transformation. **c**, Left: Full WW domain *combined score* interaction map before (lower left) and after (upper right) DeepContact transformation. DeepContact amplifies a signal from local contacts in the C-terminal region of the domain, thus concentrating the strongest transformed signal in this region. Removing positions 30-34 removes this artefact (right plot). a-c: Grey dots show contacts (distance < 8Å) in reference structure. **d**, Change in precision of top L *combined score* contacts for down-sampled GB1 domain datasets after DeepContact transformation.



Supplementary Figure 7

Genetic interactions are mediated by amino acid side-chains

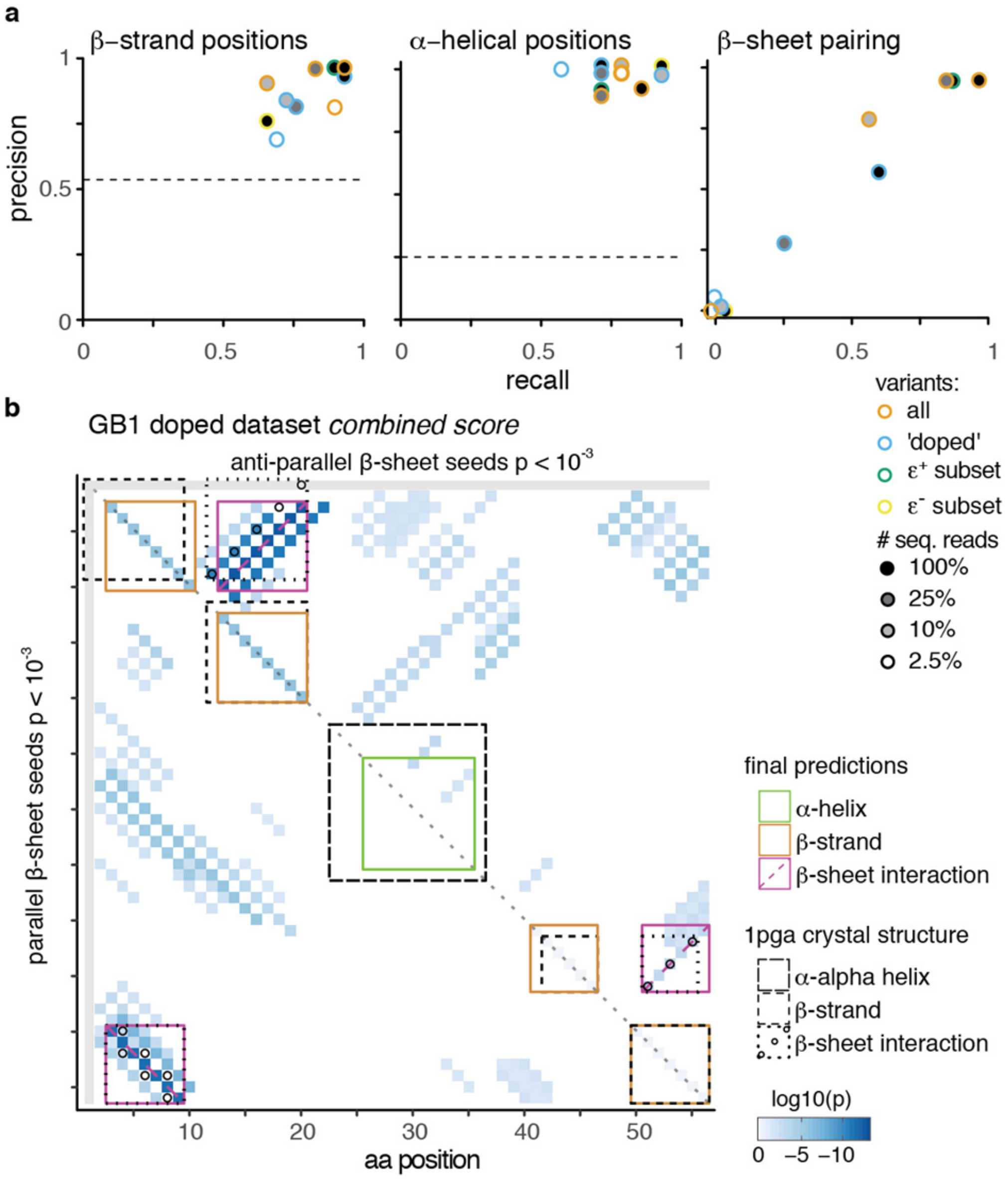
Differences in contact prediction precision and enrichment over random expectation for all heavy atom or 'only side-chain heavy atom' distance metrics (from combined scores [GB1, WW, RRM] or enrichment scores [FOS-JUN]). As expected, using all heavy atoms (including backbone heavy atoms) increases precision of predicted contacts, by about 10%. Restricting distance measurements to side-chain heavy atoms, however, increases precision over random expectation, often by more than 2-fold (note the log10-scale), indicating that side-chain distances are more informative for epistatic interactions. This, together with the fact that secondary structures can be detected based on their periodic side-chain geometries, suggests that epistatic interactions are mostly mediated by side-chain interactions.



Supplementary Figure 8

Deep mutational sequencing data to contact prediction workflow

Overview of workflow to predict interacting position pairs from deep mutational scanning datasets (see Methods and Results).



Supplementary Figure 9

Secondary structure prediction from down-sampled GB1 datasets

a, Precision and recall for beta strand, alpha helix and beta sheet predictions derived from *combined scores* of down-sampled GB1 datasets (in comparison to reference structure 1pga). Dashed lines for beta strand and alpha helical positions give random expectation. Random expectation for beta sheet pairing precision is below 1%. Note that some coinciding data points were slightly moved for better identifiability. **b**, Beta sheet pairing predictions for the 'doped' GB1 dataset with 100% sequencing read coverage. Beta sheet pairing between beta strands 1 and 2 is predicted in correct anti-parallel direction, but exact pairing of positions is off by 2; thus precision and recall of beta sheet pairing for doped GB1 dataset drops to ~60% (see panel a, right plot).