

Predicting and interpreting large scale mutagenesis data using analyses of protein stability and conservation

Magnus H. Høie¹, Matteo Cagiada¹, Anders Haagen Beck Frederiksen¹, Amelie Stein^{1*}, Kresten Lindorff-Larsen^{1*}

*For correspondence:
amelie.stein@bio.ku.dk (AS);
lindorff@bio.ku.dk (KLL)

¹Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark

Abstract Understanding and predicting the functional consequences of single amino acid is central in many areas of protein science. Here we collected and analysed experimental measurements of effects of >150,000 variants in 29 proteins. We used biophysical calculations to predict changes in stability for each variant, and assessed them in light of sequence conservation. We find that the sequence analyses give more accurate prediction of variant effects than predictions of stability, and that about half of the variants that show loss of function do so due to stability effects. We construct a machine learning model to predict variant effects from protein structure and sequence alignments, and show how the two sources of information are able to support one another. Together our results show how one can leverage large-scale experimental assessments of variant effects to gain deeper and general insights into the mechanisms that cause loss of function.

Introduction

The ability to predict and understand the effects of amino acid changes to protein structure, stability and function plays central roles in a number of areas of protein science. For example, improving protein function or stability is key in many biotechnological applications, and the ability to understand and predict loss of function is of central importance in disease biology. Large-scale human genome sequencing efforts are revealing millions of missense variants that change the amino acid sequences of proteins, but we do not yet know the functional consequences for most of these variants.

Studies of the effects of single amino-acid changes also present opportunities to test our understanding of the protein structure-function relationship, and the interplay with the cellular environment. Generally speaking, while a large fraction of single amino-acid substitutions in a given protein are relatively well tolerated, there is a subset that has significant detrimental consequences (*Schaafsma and Vihinen, 2017; Gray et al., 2017*). Pinpointing which variants are in the detrimental group, and the biochemical and biophysical mechanisms underlying loss of fitness, is important for example for assessing pathogenicity of so-called variants of *uncertain significance* (*Richards et al., 2015*) and understanding the mechanistic origins of disease.

Recent technological advances have enabled high-throughput assays that can quantify changes in activity, stability or other protein properties of interest for thousands of variants in a single ex-

40 periment. The assays also provide large sets of data with systematic fitness profiles of variants,
41 often providing both mechanistic insight and systematic assessment of computational models for
42 predicting variant effects. Briefly, such Multiplexed Assays of Variant Effects (MAVEs, also called
43 Deep Mutational Scanning experiments) have three components: (i) generation of a DNA library
44 encoding a comprehensive set of protein variants to be tested, (ii) an assay that selects or screens
45 for function or other properties of interest, and (iii) sequencing before and after selection to deter-
46 mine how each variant fares in the assay (*Fowler and Fields, 2014*). These experiments are enabled
47 by advances in oligonucleotide synthesis and lowered cost of next-generation sequencing, and for
48 example make it possible to create and probe saturation mutagenesis libraries (i.e. exhaustive
49 generation of all 19 possible substitutions across the whole protein or region of interest). While
50 library generation and sequencing is often transferable across systems, development of assays,
51 however, usually needs to be individually tailored for the protein of interest and the function to be
52 tested. Typically, the function of the protein is coupled to a phenotype that is amenable to high-
53 throughput evaluation, such as growth or production of a fluorescent marker (*Fowler and Fields,*
54 **2014**). This enables quantification of the level of function ranging from wild-type-like, to moder-
55 ately affected or total loss of function. The need for customised assays, however, also means that
56 functional scores (sometimes called fitness scores) extracted from different MAVEs are often not
57 directly comparable to one another without further normalization.

58 The data generated by MAVEs have been shown to help predict the status of pathogenic and
59 benign variants, while also serving as useful benchmarks for computational variant classification
60 methods (*Livesey and Marsh, 2020; Frazer et al., 2020*). More generally, the data can also provide
61 detailed insight into general aspects of protein structure and function (*Gray et al., 2017; Ahler et al.,*
62 **2019b; Dunham and Beltrao, 2020; Chiasson et al., 2020; Starr et al., 2020; Cagiada et al., 2021;**
63 *Amorosi et al., 2021*). The extensive coverage of variants measured in the same assay provides a
64 rich source of data that may be used for protein design, structure prediction and identification of
65 crucial regions related to function and stability (*Stein et al., 2019*). For example, recent analyses of
66 more than 30 data sets generated by MAVE point to individual amino acids performing different
67 functions depending on their chemical environment (*Dunham and Beltrao, 2020*).

68 A given amino acid change may affect multiple properties or functions of a protein, and by
69 combining different assays it may be possible to disentangle which substitutions affect which of
70 those properties and functions. As most proteins need to be folded to function, assays probing
71 protein stability and cellular abundance have received special attention. Thus, a specific type of
72 MAVE termed VAMP-seq has been developed to probe cellular protein abundance (*Matreyek et al.,*
73 **2018**), and been shown to correlate with measurements of protein stability (*Matreyek et al., 2018;*
74 *Suiter et al., 2020*). While the detailed relationship between protein stability and abundance is
75 complicated and not fully understood (*Hingorani and Giersch, 2014; Stein et al., 2019*), we and
76 others have shown that unstable proteins are often targeted for proteasomal degradation leading
77 to lowered cellular abundance (*Nielsen et al., 2017; Chen et al., 2017; Nielsen et al., 2017; Scheller*
78 *et al., 2019; Abildgaard et al., 2019; Nielsen et al., 2020*). Thus, by combining the results from
79 a VAMP-seq experiment with a MAVE probing protein activity it is possible to distinguish between
80 variants that cause loss of function due to lowered abundance from those that change the intrinsic
81 activity of a protein (*Jepsen et al., 2020; Chiasson et al., 2020; Cagiada et al., 2021; Amorosi et al.,*
82 **2021**). These experiments and analyses suggest that a relatively large fraction of variants that cause
83 of loss-of-function are due to loss of stability and resulting degradation in the cell. Thus, it is not
84 surprising that a large number of disease-causing variants are proteasomal targets and are found
85 at low cellular levels (*Meacham et al., 2001; Yaguchi et al., 2004; Olzmann et al., 2004; Ron and*
86 *Horowitz, 2005; Yang et al., 2011, 2013; Arlow et al., 2013; Nielsen et al., 2017; Chen et al., 2017;*
87 *Nielsen et al., 2017; Scheller et al., 2019; Abildgaard et al., 2019; Nielsen et al., 2020*).

88 Although it is possible to perform multiplexed assays on multiple genes in a single experiment
89 (*Després et al., 2020; Jun et al., 2020; Hanna et al., 2021; Cuella-Martin et al., 2021*), we are still
90 far from able to probe all possible variants in all proteins by experiments. Thus, computational

91 methods are important to predict and understand variant effects, and in some cases they may
92 be even be more accurate than MAVEs for this purpose (*Jepsen et al., 2020; Frazer et al., 2020*).
93 Computational variant prediction methods make it possible to estimate effects of variants that
94 have never been seen and examine proteins that have not been studied in experiments. This
95 aspect is important for applications in clinical genetics where many variants are extremely rare
96 and may arise *de novo*. Of more practical importance, they can also bring variant effects onto a
97 common scale comparable across the proteome.

98 Most variant effect predictors are based on features extracted from evolutionary conservation
99 of homologous proteins, biophysical calculations based on structure, and general knowledge of
100 amino-acid properties (*Yue et al., 2005; Kumar et al., 2009; Adzhubei et al., 2010; Casadio et al.,
101 De Baets et al., 2012; Kircher et al., 2014; Choi and Chan, 2015; Ioannidis et al., 2016; An-*
cien et al., 2018; Wagih et al., 2018; Gerasimavicius et al., 2020; Livesey and Marsh, 2020). As
103 some of these methods have been trained and tested on classification of clinical variants, it has
104 been argued that comparison against data from MAVEs provides a useful and unbiased alterna-
105 tive to benchmark such methods (*Livesey and Marsh, 2020*). In such tests, it has been shown
106 that various sequence-based approaches, including deep-learning methods, can achieve very high
107 accuracy (*Riesselman et al., 2018; Livesey and Marsh, 2020*). Indeed, we and others have success-
108 fully applied sequence analysis and biophysical stability calculations for identification and analysis
109 of pathogenic variants, although these methods were not trained on clinical variants (*Pey et al.,
110 Yin et al., 2017; Nielsen et al., 2017; Gray et al., 2018; Scheller et al., 2019; Cline et al., 2019;
111 Abildgaard et al., 2019; Jepsen et al., 2020; Frazer et al., 2020*).

112 Experiments and computational analyses such as those discussed above are now beginning to
113 provide a consistent picture of the effects of variants on protein stability and function. Variants
114 that cause substantial loss of stability are generally found at low protein levels in the cell, and
115 thus often lead to a loss-of-function phenotype and disease. Hence, when a variant is predicted to
116 be highly destabilizing it is likely to be non-functional. The reverse, however, does not hold true.
117 Variants that do not perturb protein stability may still cause loss of function via other mechanisms,
118 such as perturbing active site residues in an enzyme or key interaction sites for binding. Such
119 effects can often be captured by evolutionary sequence analyses which are potentially sensitive to
120 all conserved molecular mechanisms that lead to loss of function.

121 Here, we aim to provide further insight into the relationship between variant effects on protein
122 stability and function, and how computational predictions of changes in thermodynamic stability
123 and analysis of sequence conservation may be used to predict variant effects. We use the Rosetta
124 software to predict changes in thermodynamic stability ($\Delta\Delta G$) (*Park et al., 2016; Leman et al., 2020;
125 Frenz et al., 2020*) using as input the structure of each protein. Similarly, we use GEMME (Global
126 Epistatic Model for predicting Mutational Effects) (*Laine et al., 2019*) to analyse multiple sequence
127 alignments and calculate a score, which we term $\Delta\Delta E$, and which captures conservation of amino
128 acids through reconstructing phylogenetic trees. To assess their predictive power on functional
129 variant consequences, we have collected 39 data sets previously generated by MAVEs on 29 pro-
130 teins, and analyse these using the $\Delta\Delta G$ and $\Delta\Delta E$ calculations. As evolution disfavours unstable
131 proteins (*Bloom et al., 2006; Echave and Wilke, 2017*), there is some correlation between the two
132 approaches, however, there are also differences, e.g. for active sites, where variants are typically
133 only identified as detrimental by evolutionary sequence analysis (*Cheng et al., 2005; Echave, 2019;
134 Jepsen et al., 2020; Cagiada et al., 2021*). We train a machine learning model that uses $\Delta\Delta G$ and
135 $\Delta\Delta E$ as input to predict variant effects as probed by MAVE experiments. We show that both compo-
136 nents of the model contribute to prediction accuracy, and use the final model to provide a global
137 view of the relationship between stability and function. In this way, our analysis of MAVE experi-
138 ments using $\Delta\Delta G$ and $\Delta\Delta E$ calculations help pinpoint which variants lose function due to loss of
139 stability. Together, our results show how loss of stability is an important contributor to loss of
140 function, and point to future improvements for predictions of variant effects.

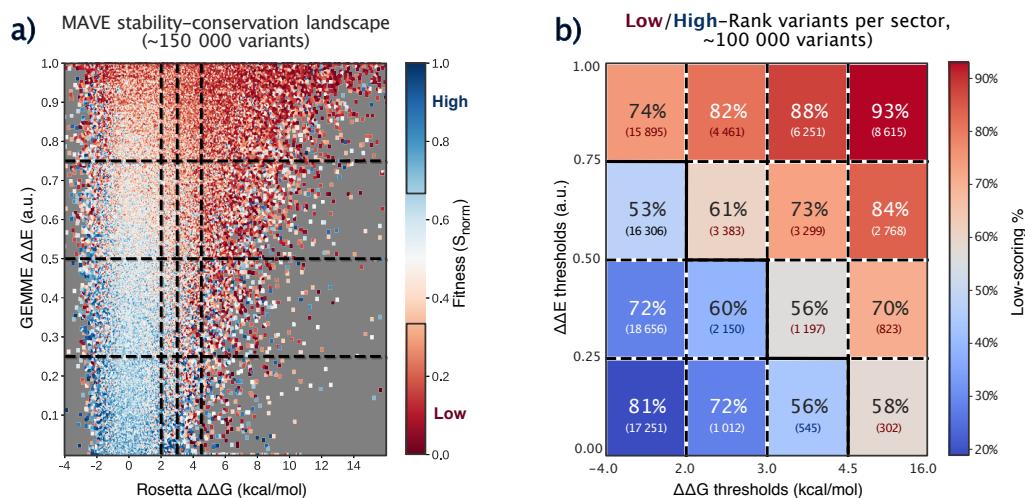


Figure 1. Stability and conservation score trends across variants. a) Analysis of the variant fitness landscape across over 150,000 variants in 39 MAVE experiments, by GEMME $\Delta\Delta E$ and Rosetta $\Delta\Delta G$ score per variant, colored by normalized MAVE fitness score. b) Percentage of high fitness (top 33 % of individual MAVE experimental scores, blue) or low fitness (bottom 33 %, red) variants per sector in the fitness landscape, and number of variants per sector. The middle tertile (33rd-66th percentile) of variants are excluded here.

141 Results

142 Analysing Variant Effects from MAVEs by Calculations of Stability and Conservation

143 We first aimed to quantify how well analyses of protein stability and sequence conservation are
 144 able to capture experimental measurements of variant effects. We thus collected 39 data sets
 145 generated by MAVEs on 29 proteins from the literature (Tables S1 and S2). As we aimed to use
 146 the data in a globally-trained machine learning model we used rank normalisation to bring the
 147 original variant scores reported by the individual studies onto a common scale (s_{exp}), where $s_{\text{exp}} \sim 1$
 148 corresponds to wild-type-like activity in the experiment and $s_{\text{exp}} \sim 0$ corresponds to variants with
 149 low activity in the assay used in the MAVE.

150 For each of the 29 proteins we used Rosetta (Park et al., 2016; Leman et al., 2020; Frenz et al.,
 151 2020) to predict changes in thermodynamic stability ($\Delta\Delta G$) for each of the 19 possible variants at
 152 each of the positions resolved in the experimental structures. Here, $\Delta\Delta G = 0$ corresponds to the
 153 same stability as the wild-type protein and variants with $\Delta\Delta G > 0$ correspond to those that are
 154 predicted to destabilise the protein. We also built a multiple sequence alignment for each of the
 155 proteins and analysed these using GEMME (Laine et al., 2019). Following rank-normalisation for
 156 each protein, the resulting GEMME $\Delta\Delta E$ scores quantify how likely each of the 19 substitutions are
 157 in terms of what has been observed in the evolutionary record, with $\Delta\Delta E \approx 0$ corresponding to
 158 very conservative substitutions and $\Delta\Delta E \approx 1$ corresponding to variants that are extremely rare or
 159 absent from the alignment and hence predicted to be disruptive. In total, we thus collected triplets
 160 of s_{exp} , $\Delta\Delta G$ and $\Delta\Delta E$ for 154,808 single amino-acid variants covering 10,012 positions in the 29
 161 proteins (Fig. 1a).

162 We have previously shown that changes in protein stability ($\Delta\Delta G$) and multiple-sequence-alignment-
 163 based conservation scores ($\Delta\Delta E$) correlate with changes in cellular function or stability in selected
 164 proteins (Nielsen et al., 2017; Scheller et al., 2019; Abildgaard et al., 2019; Cagliada et al., 2021).
 165 This also holds for the 154,808 variants in the 29 proteins studied here (Fig. 1a), so that variants with
 166 large $\Delta\Delta E$ or $\Delta\Delta G$ scores tend to show low fitness (low s_{exp}). In particular, variants for which both
 167 $\Delta\Delta G$ and $\Delta\Delta E$ values are high almost always show loss of function, while those with low scores
 168 for both usually show wild-type-like fitness. As expected and discussed above, variants that do
 169 not substantially perturb stability (low $\Delta\Delta G$) can both have low and high values of s_{exp} , because

170 substitutions may affect function through other mechanisms than loss of stability and abundance.
171 From the stability-conservation landscape, however, it appears that such effects are captured by
172 the $\Delta\Delta E$ scores so that, even for variants with low $\Delta\Delta G$, conservative substitutions (low $\Delta\Delta E$) tend
173 to be associated with high fitness, and high $\Delta\Delta E$ with low fitness.

174 To quantify the power of $\Delta\Delta E$ and $\Delta\Delta G$ for classifying variants, we extracted the top and bottom
175 third of variant scores (s_{exp}) as the subsets that are most clearly associated with wild-type-like and
176 loss-of-fitness phenotypes, respectively. Next, we divided our MAVE stability-conservation land-
177 scape into a total of 16 sectors. The normalised $\Delta\Delta E$ scores have evenly distributed thresholds of
178 0.25, 0.50 and 0.75, while the $\Delta\Delta G$ thresholds are at 2.0, 3.0 and 4.5 kcal/mol (Fig. 1). Inspection of
179 the sectors confirms that extreme values of $\Delta\Delta E$ and $\Delta\Delta G$ can classify variants well into the high or
180 low fitness categories, while moderate values have lower classification power (Fig. 1b). For exam-
181 ple when $\Delta\Delta E < 0.25$ and $\Delta\Delta G < 2.0$ kcal/mol, 81% of the variants are in the high fitness category,
182 and at the opposite end, when $\Delta\Delta E > 0.75$ and $\Delta\Delta G > 4.5$ kcal/mol, 93% of the variants are in the
183 low fitness category.

184 More generally, the two-dimensional fitness landscapes illustrate the partial interdependency
185 of $\Delta\Delta E$ and $\Delta\Delta G$. Most notably, it is clear that evolution selects strongly against destabilising vari-
186 ants so that there are almost no cases with high values of $\Delta\Delta G$ and low values of $\Delta\Delta E$ (only 4%
187 of the variants have $\Delta\Delta E < 0.25$ and $\Delta\Delta G > 3.0$ kcal/mol). Also, as discussed above, low values of
188 $\Delta\Delta G$ may be associated both with high and low values of s_{exp} , whereas variants with larger values of
189 $\Delta\Delta G$ tend to have low s_{exp} . Thus, focusing on the stable variants ($\Delta\Delta G < 2.0$ kcal/mol) and only the
190 top/bottom third of the fitness scores, we find that 38% of these variants have $\Delta\Delta E > 0.5$ (Fig. 1a)
191 and are thus more likely to be non-functional. Focusing on the variants that the GEMME analysis
192 suggest are incompatible with what has been observed through evolution ($\Delta\Delta E > 0.5$), and thus
193 are more likely to be non-functional, we find that 47% are predicted to be unstable ($\Delta\Delta G > 2.0$
194 kcal/mol; Fig. 1b). Thus, in line with our previous analysis (*Cagiada et al., 2021*), these results
195 suggest that approximately half of the non-conservative substitutions are selected against due to
196 stability effects.

197 Training and Benchmarking Random Forest Models

198 Having established that there is an overall relationship between calculated values of $\Delta\Delta G$ and $\Delta\Delta E$
199 and experimental variant effects (s_{exp}) and that both independently contribute valuable information
200 (Fig. 1) we decided to train a machine learning model to predict variant effects from $\Delta\Delta G$ and $\Delta\Delta E$.
201 Before doing so, we analysed how well the data generated by the individual MAVES correlate with
202 the calculated values of $\Delta\Delta G$ and $\Delta\Delta E$ (Fig. 2a).

203 We find a broad range of Spearman correlation coefficients (r_s) between the MAVE scores (s_{exp})
204 and our calculated $\Delta\Delta G$ and $\Delta\Delta E$ scores, in line with previous observations of considerable vari-
205 ation in correlation with variant effect predictors (*Livesey and Marsh, 2020*). Correlations range
206 from low ($r_s \sim 0.1$ on BRCA1 and Calmodulin-1 (CALM1) for both $\Delta\Delta E$ and $\Delta\Delta G$) to relatively high
207 ($r_s \sim 0.6 - 0.8$) when predicting multiple independent MAVES on β -lactamase (bla) (Fig. 2a). Over-
208 all and as expected, experimental variant effects that correlate well with $\Delta\Delta G$ also correlate well
209 with $\Delta\Delta E$. In line with the results shown above (Fig. 1) and previous analyses (*Nielsen et al., 2017*;
210 *Abildgaard et al., 2019*; *Jepsen et al., 2020*; *Livesey and Marsh, 2020*; *Gerasimavicius et al., 2020*;
211 *Cagiada et al., 2021*), we find that variant effects tend to be more strongly correlated with $\Delta\Delta E$
212 than $\Delta\Delta G$. Two notable outliers to this observation are the abundance-based MAVES (VAMP-seq)
213 for PTEN (*Matreyek et al., 2018*) and NUDT15 (*Suiter et al., 2020*), which are more strongly corre-
214 lated with our analysis of protein stability than conservation (labelled as PTEN (a) and NUDT15 (a)
215 in Fig. 2a) (*Cagiada et al., 2021*). Interestingly, the VAMP-seq data for the protein TPMT (*Matreyek*
216 *et al., 2018*) is slightly more strongly correlated with $\Delta\Delta E$ than $\Delta\Delta G$. PTEN and NUDT15 variants
217 have also been assayed for their respective biochemical functions (*Mighell et al., 2018*; *Suiter et al.,*
218 *2020*), and the resulting s_{exp} scores correlate better with $\Delta\Delta E$ and less well with $\Delta\Delta G$ than the cor-
219 responding protein abundance scores (the MAVES of respective biochemical function are labelled

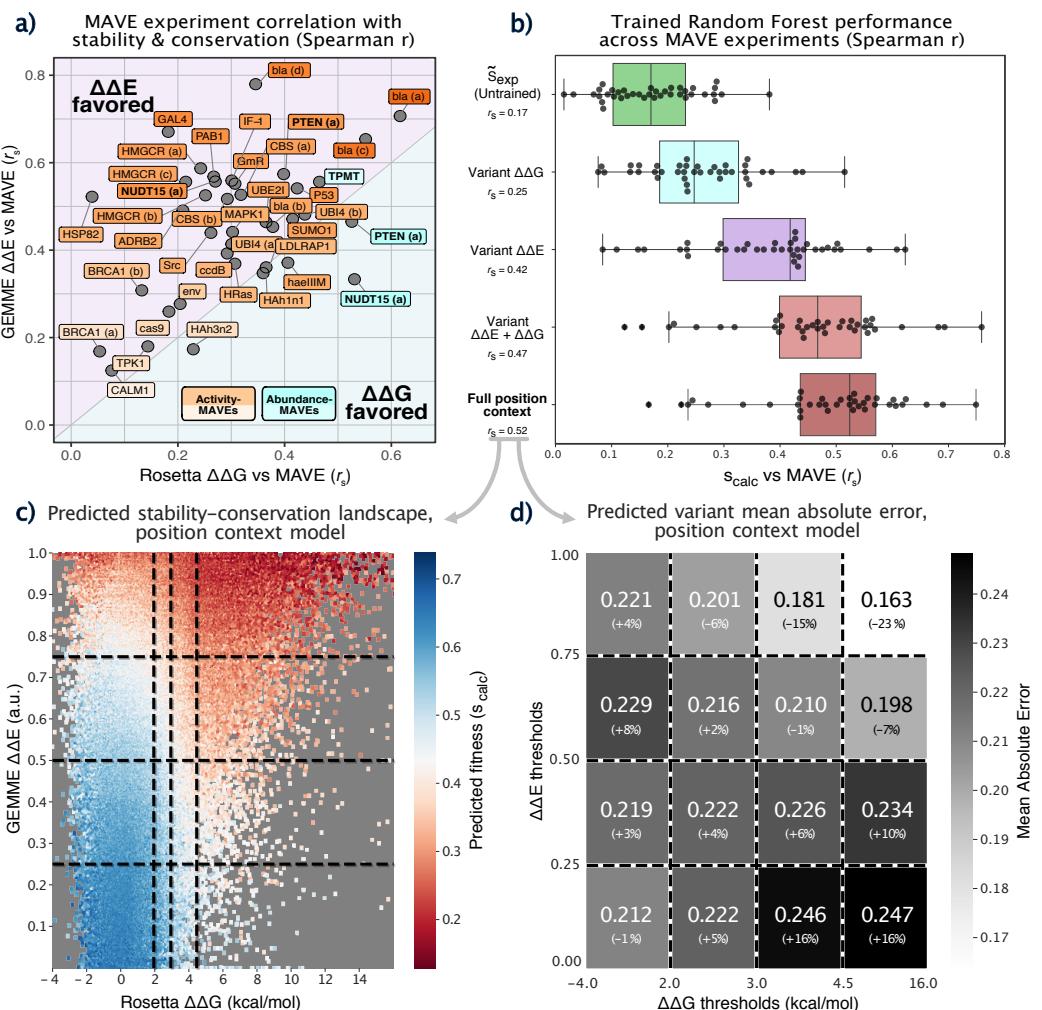


Figure 2. Correlations and predictions of variant effects. a) Spearman correlation coefficients between GEMME $\Delta\Delta E$ or Rosetta $\Delta\Delta G$, and the experimental variants effects (s_{exp}) for each of the 39 MAVEs. b) Correlation between predicted and experimental variant effects from a baseline (\tilde{s}_{exp}) and four different machine learning models. Apart from the first row, each of the rows correspond to one of four different random forest models that used different features as input (indicated by the labels and described in detail in the main text and Methods). For each class of model, each point corresponds to one of the 39 sets of data, and the correlation coefficients were calculated in a leave-one-protein-out cross-validation across 29 proteins and 39 MAVE data sets. c) Predicted fitness landscape in leave-one-protein-out cross-validation of the full position context random forest model. d) Mean absolute error (MAE) and relative difference from the mean MAE from the position context random forest model.

as PTEN (b) and NUDT15 (b) in Fig. 2a) (*Cagiada et al., 2021*). While the reasons for the low correlation between $\Delta\Delta E$ and $\Delta\Delta G$ with several of the experiments remain unclear, possible explanations include inaccuracies in stability calculations, poor sequence alignments, or experimental assays that probe properties not directly related to stability or an evolutionary-conserved function of the protein.

Next we constructed a simple baseline model that captures effects of substituting each of the 20 protein-coding amino acids for another by averaging over the normalised scores from all variants in our data set (\tilde{s}_{exp} ; Fig. S1). Similar to previous observations (*Gray et al., 2017*) this substitution matrix captures well-known biochemical patterns. In particular we observe that substitutions of hydrophobic residues—often found in the protein core—with charged or polar residues on average lead to a large loss of fitness, while changes from a polar to a hydrophobic residue on average does not cause a substantial loss of fitness. Nevertheless, although \tilde{s}_{exp} captures such chemical intuition it lacks information about structural and sequence context, and thus is overall a poor predictor of experimental variant effects (mean $r_s = 0.17$; Fig. 2b, green).

With the experimental and computational variant data we proceeded to train a set of machine learning models to predict s_{exp} from our set of available $\Delta\Delta G$, $\Delta\Delta E$ and \tilde{s}_{exp} scores. We chose to use random forest models because of their robustness to outliers and noise, minimal need to adjust model hyperparameters, as well as the possibility to easily extract information about the extent to which the different features are used in the model decision process (*Breiman, 2001; Bernard et al., 2009*). We term the values predicted by the model s_{calc} .

We used a leave-one-protein-out procedure for training, selecting one protein for validation, and training on the normalised MAVE, $\Delta\Delta E$ and/or $\Delta\Delta G$ data from all other proteins in our overall set. Thus, when more than one set of experiments had been performed on a single protein, we excluded the extra data sets during training. We assess the model by calculating the correlation between s_{calc} predicted from the resulting random forest with the s_{exp} values from all MAVEs on the protein that was left out in training, looping over all proteins one at a time.

First, we trained three random forest models using as inputs only the variant $\Delta\Delta G$, only the variant $\Delta\Delta E$, or both. In line with the variation in the correlation to the input data, we observe a range of correlation coefficients from these models, with the combined $\Delta\Delta G$ and $\Delta\Delta E$ model correlating with the normalised MAVE scores with a median $r_s = 0.47$ (Fig. 2b, orange). The $\Delta\Delta G$ -only model ($r_s = 0.25$, Fig. 2b, cyan) only performs slightly better than \tilde{s}_{exp} (green), and the $\Delta\Delta E$ -only model ($r_s = 0.42$, Fig. 2b, purple) is again closer to capturing the experimental outcomes. We note here that several of the correlation coefficients observed in these analyses (Fig. 2b) are smaller than those obtained when correlating directly with the s_{exp} values (Fig. 2a). This is because we in the random forest models aim to capture the relationship between e.g. $\Delta\Delta G$ and s_{exp} using a single model and scale, whereas the correlations in Fig. 2a effectively correspond to 39 distinct ‘models’.

As variant effects depend both on the specific substitutions but also the context within the protein, we trained a more complex ‘position-context’ model that also takes into account the scores for other substitutions at the given protein position. Specifically, in addition to the $\Delta\Delta E$ and $\Delta\Delta G$ value for the variant to be predicted, we input the entire set of 20 $\Delta\Delta G$ and 20 $\Delta\Delta E$ values at a position, representing the stability and conservation scores for all single amino-acid variants at the given position, as well as the mean score for these. We further add three features from the baseline model, $\tilde{s}_{WT \rightarrow Mut}$ (the average score when mutating from the wild-type to the specific variant (mutant) residue), $\tilde{s}_{WT \rightarrow Any}$ (the average score for substitutions from the specific wild-type amino acid to any of the other 19), and $\tilde{s}_{Any \rightarrow Mut}$ (the average score when mutating each of the 19 other amino acids to the specific variant amino-acid). The resulting position-context random forest model yields an improved performance (median $r_s = 0.52$, Fig. 2b, brown). The model also successfully recapitulates the trends in the stability-conservation landscape (Fig. 2c and Fig. S2). Looking more closely at how the model performs in different sectors of this landscape, we find that predictions generally have a similar mean absolute error (MAE) in most sectors (Fig. 2d and Fig. S3) with MAE ≈ 0.22 , with somewhat more accurate predictions for highly destabilizing ($\Delta\Delta G >$

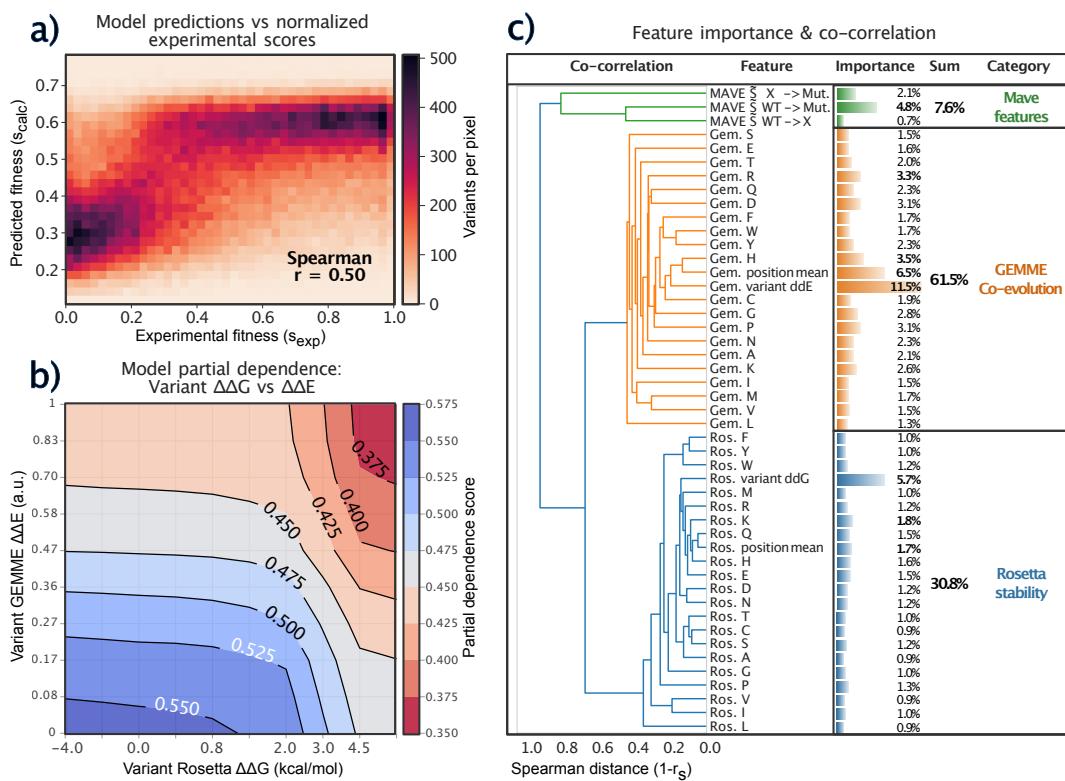


Figure 3. Interpretation of the position-context random forest model a) Density plot of predicted variant scores vs. normalized experimental fitness scores in leave-one-protein-out cross-validation. b) Partial dependence plot for the position-context model trained on all variants. Most of the variation in fitness scores is explained along the GEMME $\Delta\Delta E$ axis, with Rosetta $\Delta\Delta G$ only impacting variant scores at $\Delta\Delta G \geq 2.0$. Colours indicate the predicted partial score for the given simulated variant $\Delta\Delta G$ and $\Delta\Delta E$ score sampled across the entire data set, but does not take into account effects from the remaining features. Note the non-linearity of the scale of the axes. c) Position context model feature importance and dendrogram.

3.0) and non-conservative ($\Delta\Delta E > 0.50$) substitutions (MAE 0.16 – 0.21).

Role of Stability and Conservation in the Predictions

With the final model in hand and having shown that it recapitulates the experimentally observed stability-conservation landscape relatively well (Fig. 1a, Fig. 2c and Fig. S2), we proceed to analyse the properties of the model. Looking at the predictions of the >150,000 variant effects from the leave-one-protein-out model, we find that the position-context model predicts fitness outcomes with a Spearman correlation coefficient of 0.50 (Fig. 3a). The most dense region has a sigmoidal shape, with predictions of the experimentally-derived s_{exp} range 0.5-1.0 seeming largely indistinguishable by the model. The non-linear relationship is presumably an effect of the use of the mean square error between s_{exp} and s_{calc} as target when training the random forest model.

To gain further insight into the relative importance of the stability and conservation parameters for predicting the variant scores, we calculated the partial dependencies in the position-context random forest model. These partial dependencies quantify how the model's predictions of s_{calc} depend on the chosen features, here the $\Delta\Delta G$ and $\Delta\Delta E$ for the specific variant, marginalizing over the remaining features in the model (Molnar, 2019). For variants that at most cause a modest change in stability ($\Delta\Delta G < 2.0$ kcal/mol), the predictions are most strongly influenced by $\Delta\Delta E$. This observation is in line with the notion that for stable variants, $\Delta\Delta E$ is a relatively good predictor of s_{exp} (Fig. 1a). In contrast, for more destabilizing variants ($\Delta\Delta G > 3.0$ kcal/mol) the partial dependence vary with both $\Delta\Delta G$ and $\Delta\Delta E$, in line with the finding that both of these are useful quantities to help predict variant effects in this part of the stability-conservation landscape (Figs. 1 and 2).

291 We then proceeded to examine the impact of all features in the random forest model. We stress
292 that this so-called ‘feature importance’ only quantifies how much each feature is used overall in
293 the prediction of s_{calc} , but does not directly describe how and when these features are used (*Strobl*
294 *et al., 2007*). We calculated the correlation between each pair of features and used these to cluster
295 and build a dendrogram of the features (Fig. 3c and Fig. S4). The features fall into three overall
296 categories corresponding to $\Delta\Delta G$, $\Delta\Delta E$ and \tilde{s}_{exp} .

297 We here remind the reader that the model uses stability and conservation effects of all possible
298 substitutions at a specific position even when predicting the effect of a specific variant. Thus, for
299 example when predicting the effect of a specific isoleucine to alanine substitution, the model will
300 use as features the effect of changing that isoleucine residue to all other 19 protein-coding amino
301 acids. As expected, however, the feature-importance calculations show that the model has its
302 largest contribution from the specific substitution, with other substitutions playing a smaller role
303 (Fig. 3c). In addition to using information about the specific substitution, the average $\Delta\Delta E$ score
304 also has a large feature importance, which we take to mean that the model uses this to determine
305 whether a specific position is overall restrictive in the types of substitutions that have been seen
306 during evolution. In addition to these effects, substitutions to positively charged amino acids (e.g.
307 $\Delta\Delta E$ to arginine and histidine and $\Delta\Delta G$ to lysine) have greater than average contributions, and
308 we speculate that these values provide additional structural context on whether e.g. a position
309 is buried or not (as substitutions to large positively charged residues are generally disfavoured at
310 buried positions). We note that effects of substitutions to histidine and asparagine have previously
311 been shown to correlate most strongly with other substitutions at the same position (*Gray et al.,*
312 *2017*).

313 Adding up the individual contributions of the features from the three classes (\tilde{s}_{exp} , $\Delta\Delta E$ and $\Delta\Delta G$)
314 we find that the $\Delta\Delta E$ features have roughly twice the ‘importance’ as $\Delta\Delta G$ features, with the total
315 importance being substantially greater than from \tilde{s}_{exp} (Fig. 3c). Looking at the values for the specific
316 variant predicted, the values are 11.8%, 5.7% and 4.8% for $\Delta\Delta E$, $\Delta\Delta G$ and \tilde{s}_{exp} , respectively, though
317 we note caveats in interpreting these numbers very precisely (*Strobl et al., 2007*). Thus, in line with
318 analysing simpler models using only a subset of the features (Fig. 2b) this analysis shows that all
319 three classes of features contribute to the performance of the final model, with $\Delta\Delta E$ carrying the
320 greatest weight.

321 A Global Map of Variant Effects

322 The results above further support previous observations that analyses of sequence conservation
323 are overall more informative than stability calculations to predict variant affects as probed by most
324 MAVE experiments (*Jepsen et al., 2020; Livesey and Marsh, 2020; Gerasimavicius et al., 2020; Cagiada et al., 2021*). Nevertheless, they also show that stability calculations can help improve overall
326 prediction accuracy and provide mechanistic insight into the relative roles of stability and conser-
327 vation in explaining variant effects. Specifically, as argued previously (*Stein et al., 2019; Jepsen*
328 *et al., 2020; Cagiada et al., 2021*), we can use $\Delta\Delta E$ calculations as a proxy for capturing a broad
329 range of effects on biological function, and $\Delta\Delta G$ as the subset that involves specifically stability
330 and abundance.

331 We thus used the data that we collected to create global maps of variant effects, and analysed
332 these in terms of fitness, stability and conservation. Inspired by previous work (*Dunham and Bel-*
333 *trao, 2020*), we visualized the 6500 amino acid positions in our data set with more than 15 values
334 of s_{exp} and corresponding values of $\Delta\Delta G$ and $\Delta\Delta E$ (Fig. 4). Specifically, we used UMAP dimension-
335 reduction (*McInnes et al., 2018*) to represent all positions in a two-dimensional map where
336 positions with similar profiles of s_{exp} , $\Delta\Delta G$ and $\Delta\Delta E$ are located close to one another (see Meth-
337 ods). Finally, we colour-coded this map using the predicted and calculated position-averaged fit-
338 ness scores, the position-averaged values of $\Delta\Delta G$ and $\Delta\Delta E$, values for \tilde{s}_{exp} and the relative solvent
339 accessibility of the position (RSA) (Fig. 4).

340 As expected from the fact that the model was trained to predict the experimental data, the

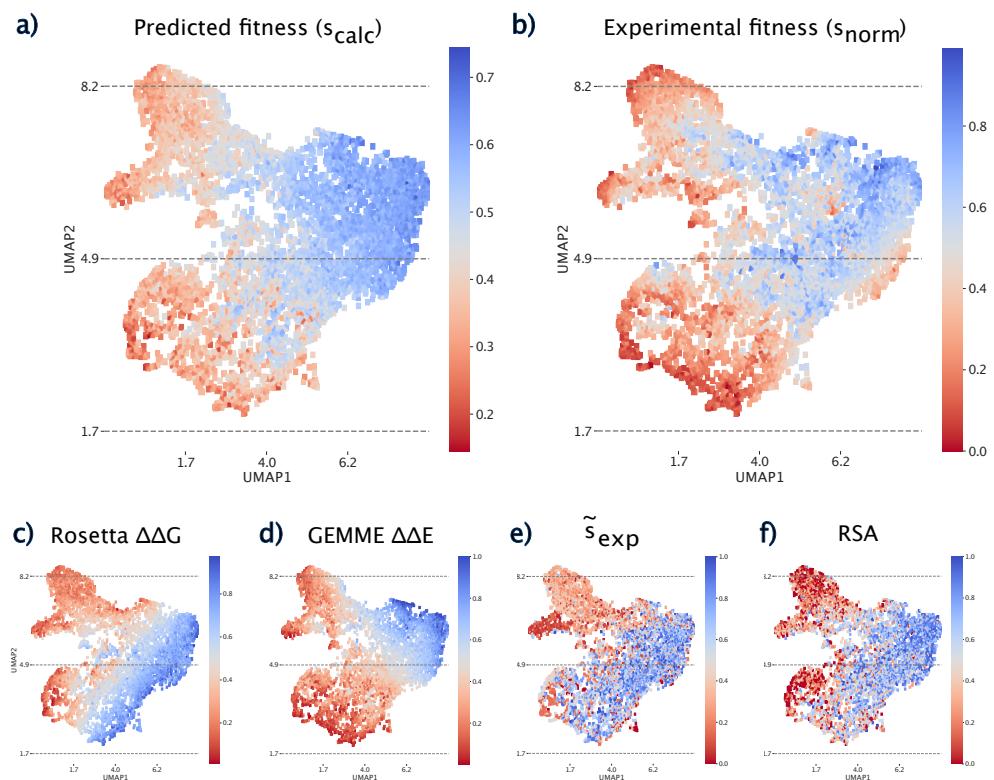


Figure 4. UMAP projection of over 6500 positions with data for at least 15 variants in the MAVEs. The maps are coloured by different properties and are locally-averaged using a convolutional kernel (see Methods). All scores were normalized between 0 and 1 to ease comparisons. a) Position-averaged predicted variant scores from the position-context random forest model. b) Position-averaged experimental fitness scores. c) Position-averaged $\Delta\Delta G$ values. d) Position-averaged $\Delta\Delta E$ values. e) \tilde{s}_{exp} . f) Relative solvent accessibility from protein structure.

341 map coloured by s_{calc} (Fig. 4a) and s_{exp} (Fig. 4b) closely resemble one another, though with a few
342 differences highlighting the imperfection of the model. In particular, the maps reveal two regions
343 (top left and bottom left of the maps) that are enriched in low-fitness variants. The map of stability
344 effects (Fig. 4c) show a gradual change moving from the top left to the bottom right, and illustrates
345 that in particular the top left part of the map corresponds to positions where variants on average
346 are destabilizing. In contrast, the map coloured by evolutionary scores (Fig. 4d) reveal a gradient
347 from the bottom left to the top right, and identifies many of the same regions as the experimental
348 map in terms of low scores. Comparing the maps coloured by the $\Delta\Delta G$ and $\Delta\Delta E$ scores reveal
349 two regions (top left, and part of the bottom left) with amino acid positions where it appears that
350 variants cause loss of function due to loss of stability. Comparison with the map indicating solvent
351 accessibility (Fig. 4f) shows clearly that many of these positions are buried inside the protein struc-
352 ture. The same comparisons also show that a group of positions near the bottom are enriched in
353 positions where variants lose function in ways that are not due to stability, but that can nonethe-
354 less be discovered using analyses of the evolutionary record. Finally, looking at the map coloured
355 by the average \tilde{s}_{exp} scores (Fig. 4e) shows similarities to both the stability map (Fig. 4c) and solvent
356 accessibility map (Fig. 4f), in line with the fact that these are buried positions and that \tilde{s}_{exp} captures
357 basic physico-chemical aspects of amino acid chemistry and protein structure (Fig. S1).

358 Discussion

359 We have analysed the relationship between experimental measurements of variant effects on pro-
360 tein function and computational analysis of protein stability and conservation. We collected over
361 150,000 measurements from multiplexed assays of variant effect in 29 proteins, and compared
362 them with predictions of changes in protein stability ($\Delta\Delta G$) and evolutionary conservation ($\Delta\Delta E$).
363 Our goal was two-fold. First, we aimed to examine how well these computed scores could pre-
364 dict variant effects, and second we wanted to shed further light on how often changes in protein
365 stability may perturb protein function.

366 In general, and in line with previous observations, we find that our analysis of conservation of
367 sequences through evolution ($\Delta\Delta E$) is more strongly correlated with the experimental measure-
368 ments than predictions of changes in thermodynamic stability ($\Delta\Delta G$) (Fig. 2a). We note here that
369 although each of the 39 experiments provides a systematic and comprehensive analysis of variant
370 effects, they were conducted using very different assays for selection/screening and in different
371 organisms. Together with uncertainties in the experimental and predicted values, this likely ex-
372 plains the substantial variation that we observe in agreement between experimental scores and
373 computed values. Thus, we defer in general from analysing individual proteins and variants and
374 focus on the global agreement. As also previously observed (Cagiada et al., 2021), the VAMP-seq
375 experiments that specifically probe protein abundance tend to be more strongly correlated with
376 stability effects than measurements probing protein activity. Finally, we note that while discrep-
377 ancies between experimental measurements of variant effects and computational predictions can
378 point to shortcomings in the prediction models, these may also reveal aspects of the protein's
379 function that the assay was not sensitive to.

380 Our global stability-conservation landscape (Fig. 1a) reveals the interdependency of stability
381 and conservation. As evolution tends to disfavour unstable proteins, we find that variants with high
382 $\Delta\Delta G$ tend to have large values of $\Delta\Delta E$ and generally low fitness (low s_{exp}). The reverse relationship,
383 however, is not true because being stable is a necessary, but not sufficient, criterion for being
384 functional. Examining the stable variants (low $\Delta\Delta G$) we find that our analysis of the evolutionary
385 record ($\Delta\Delta E$) can be used to predict with reasonable accuracy whether a variant retains function
386 or not (Fig. 1b). In our recent analysis of data probing activity and abundance in NUDT15 and PTEN,
387 we showed that about half of the variants that lose function, do so because they become unstable
388 and are found at low abundance. In line with this, we find that roughly half of the variants that have
389 high $\Delta\Delta E$ scores also have high values of $\Delta\Delta G$, and hypothesize that the remaining variants (low

390 $\Delta\Delta G$, high $\Delta\Delta E$) lose function because of substitutions in functionally important residues (*Cagiada et al., 2021*).

392 Based on the partial correlation between both $\Delta\Delta E$ and $\Delta\Delta G$ with s_{exp} we built a prediction
393 model with these parameters as input. The final position-context model reaches an accuracy that
394 surpasses that of models that solely use $\Delta\Delta E$ or $\Delta\Delta G$ (Fig. 2b). This suggests that although $\Delta\Delta E$ cal-
395 culations to some extent capture stability effects, they can still be improved by explicitly including
396 calculated values of $\Delta\Delta G$. Indeed, the model is most accurate for variants that both $\Delta\Delta E$ and $\Delta\Delta G$
397 suggest would be non-functional (Fig. 2d).

398 The random forest model recapitulates key aspects of the global stability-conservation land-
399 scape (Fig. S2) including the interdependency of $\Delta\Delta G$ and $\Delta\Delta E$. A more detailed analysis of the
400 random forest model shows that $\Delta\Delta E$ indeed is the more informative quantity when $\Delta\Delta G$ is small,
401 whereas both $\Delta\Delta G$ and $\Delta\Delta E$ contribute to accuracy when $\Delta\Delta G$ is large (Fig. 3b). This is also re-
402 flected in the feature importance analysis which shows that overall the various $\Delta\Delta E$ terms con-
403 tribute roughly twice the importance of the $\Delta\Delta G$ terms among the decision trees (Fig. 3c). Similar
404 effects are also seen in our global maps of variant effects (Fig. 4), which reveal an almost orthog-
405 onal gradient of scores for $\Delta\Delta G$ and $\Delta\Delta E$, and indicate how the combination of these two maps
406 contribute to the final predicted scores. We also show that including contextual information about
407 effects of other substitutions at the same position further increases correlation with experimen-
408 tal scores. We speculate that this context captures additional information about the structural,
409 biochemical and functional requirements at this position.

410 There are several possibilities for future improvements of the prediction methods used here.
411 First, a number of methods have recently been developed to analyse sequence information (*Ries-
412 selman et al., 2018; Alley et al., 2019; Frazer et al., 2020; Hsu et al., 2021*) and could be tested
413 instead of the GEMME method that we used. Second, while the Rosetta method that we used to
414 predict $\Delta\Delta G$ is among the most accurate methods for stability prediction, it is computationally ex-
415 pensive and requires access to protein structures. Thus, future work is needed on assessing the
416 utility of predicted structures (both from template-free and template-based methods), as well as
417 testing and developing more computationally efficient methods for predicting stability effects.

418 Another area for further development is to analyse and improve predictions in areas of inter-
419 mediate values of $\Delta\Delta G$ and $\Delta\Delta E$, where errors in the predicted values would have a greater impact.
420 While our results overall conform to the expected relationship between $\Delta\Delta E$ and $\Delta\Delta G$, more subtle
421 effects can modulate this relationship. For example, it has been observed that active site residues
422 may actually be destabilizing (*Shoichet et al., 1995*), and more detailed stability calculations may
423 thus aid in detecting such effects.

424 The types of analyses presented here may also be used to improve or interpret experiments.
425 For example, when constructing an experimental assay it might be useful to compare experimental
426 variant effects to calculations of $\Delta\Delta E$ to examine whether the assay is sensitive to evolutionarily-
427 conserved functions. Also, it has recently been demonstrated that biophysical ambiguities prevent
428 accurate predictions of how substitutions combine to affect phenotype (*Li and Lehner, 2020*), and
429 we suggest that biophysical models and predictions of variant effects may help alleviate some of
430 these ambiguities. Finally, a random forest model combining sequence and structural features has
431 been used to predict pathogenicity of human missense variants (*Ponsoni et al., 2020*).

432 In summary our large scale analysis contributes to the mounting evidence for the important role
433 for loss of stability in loss of function. In addition to the improvements observed when combining
434 conservation analysis with structure-based stability predictions, our analyses also help pinpoint
435 those variants that lose function due to stability. Such information helps provide mechanistic in-
436 sight into specific variants and proteins, but may also be a starting point for developing therapies.
437 In particular, variants that lose function due to loss of abundance, but whose intrinsic function is
438 not otherwise perturbed, would be ideal targets for approaches that aim to restabilize or other-
439 wise restore protein levels (*Balch et al., 2008; Kampmeyer et al., 2017; Stein et al., 2019; Henning
440 et al., 2021*).

441 **Methods**

442 $\Delta\Delta G$ calculations using Rosetta

443 We searched the Protein Data Bank (*Rose et al., 2010*) using the BLASTp webserver (*McEntyre J.*

444 *2002*) with default settings using as query the sequences for which the MAVE was performed. If

445 only a part of a protein is covered by mutagenesis data, we only searched for structures of that

446 part. We chose structures by balancing highest available coverage and resolution, selecting struc-

447 tures solved using X-ray crystallography when available. All calculations were carried out using the

448 Rosetta version with GitHub SHA 28f338acfb3bfd87048b38a04772486975dc83fa from July 2, 2020.

449 We first relaxed the structures using the *relax* application and the following flags:

```
450   -fa_max_dis 9
451   -relax:constrain_relax_to_start_coords
452   -ignore_unrecognized_res
453   -missing_density_to_jump
454   -nstruct 1
455   -relax:coord_constrain_sidechains
456   -relax:cartesian
457   -beta
458   -score:weights beta_nov16_cart
459   -ex1
460   -ex2
461   -relax:min_type lbfgs_armijo_nonmonotone
462   -flip_HNQ
463   -no_optH false
```

464 Subsequently, we carried out saturation mutagenesis to calculate $\Delta\Delta G$ for each single amino
465 acid substitution using the Cartesian $\Delta\Delta G$ protocol and the *beta_nov16_cart* energy function with
466 three iterations as previously described (*Park et al., 2016; Frenz et al., 2020*). Flags for the $\Delta\Delta G$
467 calculations were:

```
468   -fa_max_dis 9.0
469   -ddg::dump_pdbs false
470   -ddg:iterations 3
471   -score:weights beta_nov16_cart
472   -missing_density_to_jump
473   -ddg:mut_only
474   -ddg:bbnbrs 1
475   -beta_cart
476   -ex1
477   -ex2
478   -ddg::legacy true
479   -optimize_proline true
```

480 Scores from the three iterations were averaged. Values of $\Delta\Delta G$ in Rosetta Energy Units were
481 divided by 2.9 to bring them onto a scale corresponding to kcal/mol (Frank DiMaio, University of
482 Washington; personal correspondence; *Jepsen et al. (2020)*).

483 Evolutionary conservation analysis using GEMME

484 We calculated evolutionary conservation scores ($\Delta\Delta E$) using GEMME, a global epistatic model for

485 predicting mutational effects (*Laine et al., 2019*), based on a multiple sequence alignment of ho-

486 mologs of each protein of interest. We used the sequence used in the MAVE experiment as input to

487 HHblits (version 2.0.15 and settings -e 1e-10 -i 1 -p 40 -b 1 -B 20000) to search UniRef30_2020_03_hhsu-

488 (*Mirdita et al., 2017; UniProt Consortium, 2019; Steinegger et al., 2019*). We de-gapped the align-
489 ment with respect to the MAVE sequence, removed sequences with 50% gaps and used the output
490 alignment as input to GEMME, with default settings. Finally, we rank-normalized the output $\Delta\Delta E$
491 scores and scaled them to a [0,1] scale.

492 Collecting and normalizing data from MAVEs

493 We downloaded 39 data sets that had been generated by MAVEs from publicly available reposi-
494 tories including MAVEdb (*Esposito et al., 2019*) and a compilation by *Livesey and Marsh (2020)* (see
495 Tables S1 and S2). We used the variant fitness scores as presented in the original publication includ-
496 ing possible normalization. Three data sets (P53, MAPK1 and Src) showed a reverse relationship
497 between variant scores and $\Delta\Delta G$ and $\Delta\Delta E$ and were therefore reversed to match our convention
498 of high scores for wild-type-like activity and low scores for low activity. We then rank-normalised
499 and scaled these scores to a [0,1] range. We term these normalized scores s_{exp} .

500 We next merged each data set from the MAVEs with the corresponding $\Delta\Delta E$ and $\Delta\Delta G$ scores
501 aligning the sequences based on using Biopython, `pairwise2.align.globalds(target_seq.upper(),`
502 `seq.upper(), MatrixInfo.blosum62, -3, -1)`.

503 Baseline substitution model

504 For each MAVE data set, we calculated a 20×20 amino-acid matrix containing the average of each
505 the 400 possible Wild-type \rightarrow Mutant variant scores (e.g. the average s_{exp} for all valine to alanine
506 substitutions). We then calculated a global substitution score matrix (\tilde{s}_{exp}) by averaging these ma-
507 trices.

508 Features and Random Forest models

509 For each variant, we extracted in total 47 computational features. Depending on the model, each
510 s_{exp} was matched with its variant $\Delta\Delta G$ and $\Delta\Delta E$ score, as well as the 20 $\Delta\Delta G$ and 20 $\Delta\Delta E$ values cor-
511 responding to all available scores at the respective position, plus the mean $\Delta\Delta G$ and $\Delta\Delta E$ scores
512 for the position. We note that this results in some redundancy in the use of the data, but simplifies
513 the data structure in the model. We also included three global features from the baseline sub-
514 stitution model: (i) The mean MAVE score $\tilde{s}_{WT \rightarrow Mut}$, (ii) $\tilde{s}_{WT \rightarrow Any}$ was calculated as the mean for any
515 substitution from the given wild-type amino acid, and (iii) $\tilde{s}_{Any \rightarrow Mut}$ as the mean for mutating to the
516 specified amino-acid from any wild-type. Thus, these three features correspond to (i) an entry in
517 \tilde{s}_{exp} as well as the (ii) row-average, and (iii) column-average of \tilde{s}_{exp} (Fig. S1).

518 We trained the Random Forest models using the RandomForestRegressor in Scikit-Learn (*Pe-*
519 *dregosa et al., 2011*), with a mean-squared-error loss function, 150 trees and a minimum of 15
520 samples per leaf.

521 We first trained two models, using the variant $\Delta\Delta G$ or $\Delta\Delta E$ score only (1 feature). Next we
522 trained a combined model using both features. Training of the position-context model was per-
523 formed with all 47 available features. We iteratively trained the model and evaluated validation
524 set performance in a leave-one-protein-out cross-validation, with the removal of all validation data
525 sets for the selected protein from the training data set for each training round.

526 Model analysis

527 To extract feature importance from a globally-trained model, we trained a new random forest
528 model using all 47 features and all available data sets. The feature co-correlation dendrogram was
529 constructed by first calculating the Spearman correlation between each pair of features, which in
530 turn was converted into a distance as $1 - r_s$, and used as input to build the dendrogram using SciPy
531 (*Virtanen et al., 2020*). We measured partial dependence using PDPbox (*Jiangchun, 2018*).

532 Relative solvent accessibility

533 We used DSSP (*Kabsch and Sander, 1983*) to calculate the relative solvent accessibility (RSA) using
534 the structures that we also used for the Rosetta calculations. The RSA was used in analysis but not
535 in training.

536 UMAP projection of MAVE positions

537 To map variant scores to positions, we removed positions with fewer than 15 experimental s_{exp}
538 scores from the data set, and calculated position means for s_{exp} , $\Delta\Delta E$, $\Delta\Delta G$, $\tilde{s}_{\text{WT} \rightarrow \text{Mut}}$ and RSA. Any
539 missing s_{exp} scores were then imputed using `sklearn.impute.SimpleImputer` from the amino-acid
540 mean across all extracted positions, and all scores were then normalized to a 0–1 range. For com-
541 putational efficiency we reduced the data set to 20 features using PCA before projecting the data
542 into two dimensions (UMAP1 and UMAP2), using `UMAP-learn` with default settings (*McInnes et al.,*
543 **2018**).

544 Data availability

545 Scripts and data to repeat our analyses are available via [https://github.com/KULL-Centre/papers/
546 tree/master/2021/ML-variants-Hoie-et-al](https://github.com/KULL-Centre/papers/tree/master/2021/ML-variants-Hoie-et-al)

547 **Acknowledgement**

548 Our research is supported by the PRISM (Protein Interactions and Stability in Medicine and Ge-
549 nomics) centre funded by the Novo Nordisk Foundation (NNF18OC0033950, to A.S. and K.L.-L.)
550 and the Lundbeck Foundation (R272-2017-4528, to A.S.)

551 **References**

- 552 **Abildgaard AB**, Stein A, Nielsen SV, Schultz-Knudsen K, Papaleo E, Shrikhande A, Hoffmann ER, Bernstein I,
553 Gerdes AM, Takahashi M, Ishioka C, Lindorff-Larsen K, Hartmann-Petersen R. Computational and cellular
554 studies reveal structural destabilization and degradation of MLH1 variants in Lynch syndrome. *Elife*. 2019
555 Nov; 8.
- 556 **Adkar B**, Tripathi A, Sahoo A, Bajaj K, Goswami D, Chakrabarti P, Swarnkar M, Gokhale R, Varadarajan R. Protein Model Discrimination Using Mutational Sensitivity Derived from Deep Sequencing. *Structure*. 2012;
557 20(2):371–381. doi: [10.1016/j.str.2011.11.021](https://doi.org/10.1016/j.str.2011.11.021).
- 559 **Adzhubei IA**, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method
560 and server for predicting damaging missense mutations. *Nature methods*. 2010; 7(4):248–249.
- 561 **Ahler E**, Register AC, Chakraborty S, Fang L, Dieter EM, Sitko KA, Vidadala RS, Trevillian BM, Golkowski
562 M, Gelman H, et al. A Combined Approach Reveals a Regulatory Mechanism Coupling Src's Kinase Ac-
563 tivity, Localization, and Phosphotransferase-Independent Functions. *Molecular Cell*. 2019; 74(2). doi:
564 [10.1016/j.molcel.2019.02.003](https://doi.org/10.1016/j.molcel.2019.02.003).
- 565 **Ahler E**, Register AC, Chakraborty S, Fang L, Dieter EM, Sitko KA, Vidadala RSR, Trevillian BM, Golkowski M,
566 Gelman H, Stephany JJ, Rubin AF, Merritt EA, Fowler DM, Maly DJ. A Combined Approach Reveals a Regula-
567 tory Mechanism Coupling Src's Kinase Activity, Localization, and Phosphotransferase-Independent Functions.
568 *Mol Cell*. 2019 Apr; 74(2):393–408.e20.
- 569 **Alley EC**, Khimulya G, Biswas S, Alquraishi M, Church GM. Unified rational protein engineering with sequence-
570 based deep representation learning. *Nature Methods*. 2019; 16(12):1315–1322. doi: [10.1038/s41592-019-0598-1](https://doi.org/10.1038/s41592-019-0598-1).
- 572 **Amorosi CJ**, Chiasson MA, McDonald MG, Wong LH, Sitko KA, Boyle G, Kowalski JP, Rettie AE, Fowler DM, Dun-
573 ham MJ. Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. *bioRxiv*. 2021;
574 .
- 575 **Ancien F**, Pucci F, Godfroid M, Rooman M. Prediction and interpretation of deleterious coding variants in terms
576 of protein structural stability. *Scientific reports*. 2018; 8(1):1–11.
- 577 **Arlow T**, Scott K, Wagenseller A, Gammie A. Proteasome inhibition rescues clinically significant unstable
578 variants of the mismatch repair protein Msh2. *Proceedings of the National Academy of Sciences*. 2013;
579 110(1):246–251.
- 580 **Balch WE**, Morimoto RI, Dillin A, Kelly JW. Adapting proteostasis for disease intervention. *science*. 2008;
581 319(5865):916–919.

- 582 **Bandaru P**, Shah NH, Bhattacharyya M, Barton JP, Kondo Y, Cofsky JC, Gee CL, Chakraborty AK, Kortemme T,
583 Ranganathan R, et al. Deconstruction of the Ras switching cycle through saturation mutagenesis. *eLife*. 2017
584 Jul; <https://elifesciences.org/articles/27810>.
- 585 **Bernard S**, Heutte L, Adam S. Influence of hyperparameters on random forest accuracy. In: *International
586 workshop on multiple classifier systems* Springer; 2009. p. 171–180.
- 587 **Bloom JD**, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. *Proceedings of the
588 National Academy of Sciences*. 2006; 103(15):5869–5874. doi: [10.1073/pnas.0510098103](https://doi.org/10.1073/pnas.0510098103).
- 589 **Breiman L**. Random forests. *Machine learning*. 2001; 45(1):5–32.
- 590 **Brenan L**, Andreev A, Cohen O, Pantel S, Kamburov A, Cacchiarelli D, Persky N, Zhu C, Bagul M, Goetz E, et al.
591 Phenotypic Characterization of a Comprehensive Set of MAPK1 /ERK2 Missense Mutants. *Cell Reports*. 2016;
592 17(4):1171–1183. doi: [10.1016/j.celrep.2016.09.061](https://doi.org/10.1016/j.celrep.2016.09.061).
- 593 **Cagiada M**, Johansson KE, Valanciute A, Nielsen SV, Hartmann-Petersen R, Yang JJ, Fowler DM, Stein A, Lindorff-
594 Larsen K. Understanding the origins of loss of protein function by analyzing the effects of thousands of
595 variants on activity and abundance. *Mol Biol Evol*. 2021 Mar; .
- 596 **Casadio R**, Vassura M, Tiwari S, Fariselli P, Luigi Martelli P. Correlating disease-related mutations to their effect
597 on protein stability: A large-scale analysis of the human proteome. *Human Mutation*. 2011; 32(10):1161–
598 1170.
- 599 **Chen L**, Brewer MD, Guo L, Wang R, Jiang P, Yang X. Enhanced degradation of misfolded proteins promotes
600 tumorigenesis. *Cell reports*. 2017; 18(13):3143–3154.
- 601 **Cheng G**, Qian B, Samudrala R, Baker D. Improvement in protein functional site prediction by distinguishing
602 structural and functional constraints on protein family evolution using computational design. *Nucleic Acids
603 Research*. 2005; 33(18):5861–5867.
- 604 **Chiasson MA**, Rollins NJ, Stephany JJ, Sitko KA, Matreyek KA, Verby M, Sun S, Roth FP, DeSloover D, Marks DS,
605 et al. Multiplexed measurement of variant abundance and activity reveals VKOR topology, active site and
606 human variant impact. *Elife*. 2020; 9:e58026.
- 607 **Choi Y**, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and
608 indels. *Bioinformatics*. 2015; 31(16):2745–2747.
- 609 **Cline MS**, Babbi G, Bonache S, Cao Y, Casadio R, Cruz X, Díez O, Gutiérrez-Enríquez S, Katsonis P, Lai C, et al.
610 Assessment of blind predictions of the clinical significance of BRCA1 and BRCA2 variants. *Human Mutation*.
611 2019; 40(9):1546–1556. doi: [10.1002/humu.23861](https://doi.org/10.1002/humu.23861).
- 612 **Cuella-Martin R**, Hayward SB, Fan X, Chen X, Huang JW, Taglialatela A, Leuzzi G, Zhao J, Rabadan R, Lu C,
613 et al. Functional interrogation of DNA damage response variants with base editing screens. *Cell*. 2021;
614 184(4):1081–1097.
- 615 **Dandage R**, Pandey R, Jayaraj G, Rai M, Berger D, Chakraborty K. Differential strengths of molecular de-
616 terminants guide environment specific mutational fates. *PLOS Genetics*. 2018; 14(5). doi: [10.1371/journal.pgen.1007419](https://doi.org/10.1371/journal.pgen.1007419).
- 618 **De Baets G**, Van Durme J, Reumers J, Maurer-Stroh S, Vanhee P, Dopazo J, Schymkowitz J, Rousseau F. SNPeffec
619 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic acids research*.
620 2012; 40(D1):D935–D939.
- 621 **Deng Z**, Huang W, Bakkalbasi E, Brown NG, Adamski CJ, Rice K, Muzny D, Gibbs RA, Palzkill T. Deep Sequencing
622 of Systematic Combinatorial Libraries Reveals B-Lactamase Sequence Constraints at High Resolution. *Journal
623 of Molecular Biology*. 2012 Sep; [https://www.sciencedirect.com/science/article/pii/S0022283612007711?
624 via=ihub](https://www.sciencedirect.com/science/article/pii/S0022283612007711?via=ihub).
- 625 **Després PC**, Dubé AK, Seki M, Yachie N, Landry CR. Perturbing proteomes at single residue resolution using
626 base editing. *Nature communications*. 2020; 11(1):1–13.
- 627 **Doud MB**, Bloom JD. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemag-
628 glutinin. *MDPI*. 2016 Jun; <https://www.mdpi.com/1999-4915/8/6/155>.
- 629 **Dunham A**, Beltrao P. Exploring amino acid functions in a deep mutational landscape. *BioRxiv*. 2020; .

- 630 **Echave J.** Beyond stability constraints: a biophysical model of enzyme evolution with selection on stability and
631 activity. *Molecular biology and evolution*. 2019; 36(3):613–620.
- 632 **Echave J**, Wilke CO. Biophysical Models of Protein Evolution: Understanding the Patterns of Evolutionary
633 Sequence Divergence. *Annual Review of Biophysics*. 2017; 46(1):85–103. doi: 10.1146/annurev-biophys-
634 070816-033819.
- 635 **Esposito D**, Weile J, Shendure J, Starita LM, Papenfuss AT, Roth FP, Fowler DM, Rubin AF. MaveDB: an open-
636 source platform to distribute and interpret data from multiplexed assays of variant effect. *Genome biology*.
637 2019; 20(1):1–11.
- 638 **Firnberg E**, Labonte JW, Gray JJ, Ostermeier M. Comprehensive, High-Resolution Map of a Genes Fitness Land-
639 scape. OUP Academic. 2014 Feb; <https://doi.org/10.1093/molbev/msu081>.
- 640 **Fowler DM**, Fields S. Deep mutational scanning: a new style of protein science. *Nat Methods*. 2014 Aug;
641 11(8):801–807.
- 642 **Frazer J**, Notin P, Dias M, Gomez A, Brock K, Gal Y, Marks D. Large-scale clinical interpretation of genetic variants
643 using evolutionary data and deep learning. *bioRxiv*. 2020; .
- 644 **Frenz B**, Lewis SM, King I, DiMaio F, Park H, Song Y. Prediction of Protein Mutational Free Energy: Benchmark
645 and Sampling Improvements Increase Classification Accuracy. *Front Bioeng Biotechnol*. 2020 Oct; 8:558247.
- 646 **Gerasimavicius L**, Liu X, Marsh JA. Identification of pathogenic missense mutations using protein stability
647 predictors. *Scientific Reports*. 2020 Sep; 10(1):15387. <https://doi.org/10.1038/s41598-020-72404-w>, doi:
648 10.1038/s41598-020-72404-w.
- 649 **Giacomelli AO**, Yang X, Lintner RE, McFarland JM, Duby M, Kim J, Howard TP, Takeda DY, Ly SH, Kim E, et al.
650 Mutational processes shape the landscape of TP53 mutations in human cancer. *Nature News*. 2018 Sep;
651 <https://www.nature.com/articles/s41588-018-0204-y>.
- 652 **Gray VE**, Hause RJ, Fowler DM. Analysis of Large-Scale Mutagenesis Data To Assess the Impact of Single Amino
653 Acid Substitutions. *Genetics*. 2017; 207(1):53–61. doi: 10.1534/genetics.117.300064.
- 654 **Gray VE**, Hause RJ, Luebeck J, Shendure J, Fowler DM. Quantitative missense variant effect prediction using
655 large-scale mutagenesis data. *Cell systems*. 2018; 6(1):116–124.
- 656 **Haddox HK**, Dingens AS, Bloom JD. Experimental Estimation of the Effects of All Amino-Acid Mutations to
657 HIV's Envelope Protein on Viral Replication in Cell Culture. *PLOS Pathogens*. 2016; 12(12). doi: 10.1371/jour-
658 nal.ppat.1006114.
- 659 **Hanna RE**, Hegde M, Fagre CR, DeWeirdt PC, Sangree AK, Szegletes Z, Griffith A, Feeley MN, Sanson KR, Baidi
660 Y, et al. Massively parallel assessment of human variants with base editor screens. *Cell*. 2021; 184(4):1064–
661 1080.
- 662 **Henning NJ**, Boike L, Spradlin JN, Ward CC, Belcher B, Brittain SM, Hesse M, Dovala D, McGregor LM, McKenna
663 JM, et al. Deubiquitinase-Targeting Chimeras for Targeted Protein Stabilization. *bioRxiv*. 2021; .
- 664 **Hingorani KS**, Giersch LM. Comparing protein folding in vitro and in vivo: foldability meets the fitness chal-
665 lenge. *Current opinion in structural biology*. 2014; 24:81–90.
- 666 **Hsu C**, Nisonoff H, Fannjiang C, Listgarten J. Combining evolutionary and assay-labelled data for protein fitness
667 prediction. *bioRxiv*. 2021; .
- 668 **Ioannidis NM**, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D,
669 et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *The American
670 Journal of Human Genetics*. 2016; 99(4):877–885.
- 671 **Jacquier H**, Birgy A, Nagard HL, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit E, Poulain J, Barnaud G, et al.
672 Capturing the mutational landscape of the beta-lactamase TEM-1. *PNAS*. 2013 Jul; [https://www.pnas.org/
673 content/early/2013/07/17/1215206110](https://www.pnas.org/content/early/2013/07/17/1215206110).
- 674 **Jepsen MM**, Fowler DM, Hartmann-Petersen R, Stein A, Lindorff-Larsen K. Classifying disease-associated vari-
675 ants using measures of protein activity and stability. In: *Protein Homeostasis Diseases* Elsevier; 2020.p. 91–107.
- 676 **Jiang RJ**. Exhaustive Mapping of Missense Variation in Coronary Heart Disease-related Genes. *TSpace*. 2019
677 Nov; <http://hdl.handle.net/1807/98076>.

- 678 Jiangchun L. Python Partial Dependence Plot Toolbox. <https://github.com/SauceCat/PDPbox>. 2018; .
- 679 Jones EM, Lubock NB, Venkatakrishnan A, Wang J, Tseng AM, Paggi JM, Latorraca NR, Cancilla D, Satyadi M, Davis
680 JE, et al. Structural and functional characterization of G protein-coupled receptors with deep mutational
681 scanning. *eLife*. 2020; 9. doi: [10.7554/elife.54895](https://doi.org/10.7554/elife.54895).
- 682 Jun S, Lim H, Chun H, Lee JH, Bang D. Single-cell analysis of a mutant library generated using CRISPR-guided
683 deaminase in human melanoma cells. *Communications biology*. 2020; 3(1):1–12.
- 684 Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and
685 geometrical features. *Biopolymers*. 1983; 22(12):2577–2637. <https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.360221211>, doi: <https://doi.org/10.1002/bip.360221211>.
- 687 Kampmeyer C, Nielsen SV, Clausen L, Stein A, Gerdes AM, Lindorff-Larsen K, Hartmann-Petersen R. Blocking
688 protein quality control to counter hereditary cancers. *Genes, Chromosomes and Cancer*. 2017; 56(12):823–
689 831.
- 690 Kelsic ED, Chung H, Cohen N, Park J, Wang HH, Kishony R. RNA Structural Determinants of Optimal Codons
691 Revealed by MAGE-Seq. *Cell Systems*. 2016; 3(6). doi: [10.1016/j.cels.2016.11.004](https://doi.org/10.1016/j.cels.2016.11.004).
- 692 Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the
693 relative pathogenicity of human genetic variants. *Nature genetics*. 2014; 46(3):310–315.
- 694 Kitzman JO, Starita LM, Lo RS, Fields S, Shendure J. Massively parallel single-amino-acid mutagenesis. *Nature
695 Methods*. 2015; 12(3):203–206. doi: [10.1038/nmeth.3223](https://doi.org/10.1038/nmeth.3223).
- 696 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function
697 using the SIFT algorithm. *Nature protocols*. 2009; 4(7):1073.
- 698 Laine E, Karami Y, Carbone A. GEMME: a simple and fast global epistatic model predicting mutational effects.
699 *Molecular biology and evolution*. 2019; 36(11):2604–2619.
- 700 Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. Deep mutational scanning of
701 hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *PNAS*. 2018 Aug; <https://www.pnas.org/content/115/35/E8276>.
- 703 Leman JK, Weitzner BD, Lewis SM, Adolf-Bryfogle J, Alam N, Alford RF, Aprahamian M, Baker D, Barlow KA, Barth
704 P, et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nature methods*.
705 2020; 17(7):665–680.
- 706 Li X, Lehner B. Biophysical ambiguities prevent accurate genetic prediction. *Nature communications*. 2020;
707 11(1):1–11.
- 708 Livesey BJ, Marsh JA. Using deep mutational scanning to benchmark variant effect predictors and identify
709 disease mutations. *Mol Syst Biol*. 2020 Jul; 16(7):e9380.
- 710 Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE, Kircher M, Khechaduri A, Dines JN, Hause
711 RJ, Bhatia S, Evans WE, Relling MV, Yang W, Shendure J, Fowler DM. Multiplex assessment of protein variant
712 abundance by massively parallel sequencing. *Nat Genet*. 2018 Jun; 50(6):874–882.
- 713 McEntyre J OJ. The NCBI Handbook, The BLAST Sequence Analysis Tool. Sunnyvale, CA: Bethesda (MD) National
714 Center for Biotechnology Information (US); 2002. <https://www.ncbi.nlm.nih.gov/books/NBK21097/>.
- 715 McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction.
716 arXiv preprint arXiv:180203426. 2018; .
- 717 Meacham GC, Patterson C, Zhang W, Younger JM, Cyr DM. The Hsc70 co-chaperone CHIP targets immature
718 CFTR for proteasomal degradation. *Nature cell biology*. 2001; 3(1):100–105.
- 719 Melamed D, Young DL, Gamble CE, Miller CR, Fields S. Deep mutational scanning of an RRM domain of the Sac-
720 charomyces cerevisiae poly(A)-binding protein. *RNA*. 2013; 19(11):1537–1551. doi: [10.1261/rna.040709.113](https://doi.org/10.1261/rna.040709.113).
- 721 Mighell TL, Evans-Dutson S, O’Roak BJ. A Saturation Mutagenesis Approach to Understanding PTEN Lipid Phos-
722 phatase Activity and Genotype-Phenotype Relationships. *Am J Hum Genet*. 2018 May; 102(5):943–955.
- 723 Mirdita M, von den Driesch L, Galiez C, Martin MJ, Söding J, Steinegger M. UniClust databases of clustered and
724 deeply annotated protein sequences and alignments. *Nucleic acids research*. 2017; 45(D1):D170–D176.

- 725 **Mishra P**, Flynn J, Starr T, Bolon DA. Systematic Mutant Analyses Elucidate General and Client-Specific Aspects
726 of Hsp90 Function. *Cell Reports*. 2016; 15(3):588–598. doi: [10.1016/j.celrep.2016.03.046](https://doi.org/10.1016/j.celrep.2016.03.046).
- 727 **Molnar C**. Interpretable Machine Learning, Ch. 5.1 - Partial Dependence Plot (PDP). ISBN: 9780244768522,
728 Creative Commons; 2019. <https://christophm.github.io/interpretable-ml-book/>.
- 729 **Nielsen SV**, Schenstrøm SM, Christensen CE, Stein A, Lindorff-Larsen K, Hartmann-Petersen R. Protein desta-
730 bilization and degradation as a mechanism for hereditary disease. In: *Protein Homeostasis Diseases* Elsevier;
731 2020.p. 111–125.
- 732 **Nielsen SV**, Stein A, Dinitzen AB, Papaleo E, Tatham MH, Poulsen EG, Kassem MM, Rasmussen LJ, Lindorff-
733 Larsen K, Hartmann-Petersen R. Predicting the impact of Lynch syndrome-causing missense mutations from
734 structural calculations. *PLoS Genet*. 2017 Apr; 13(4):e1006739.
- 735 **Olzmann JA**, Brown K, Wilkinson KD, Rees HD, Huai Q, Ke H, Levey AI, Li L, Chin LS. Familial Parkinson's disease-
736 associated L166P mutation disrupts DJ-1 protein folding and function. *Journal of Biological Chemistry*. 2004;
737 279(9):8506–8515.
- 738 **Park H**, Bradley P, Greisen Jr P, Liu Y, Mulligan VK, Kim DE, Baker D, DiMaio F. Simultaneous optimization of
739 biomolecular energy functions on features from small molecules and macromolecules. *Journal of chemical
740 theory and computation*. 2016; 12(12):6201–6212.
- 741 **Pedregosa F**, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R,
742 Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine
743 Learning in Python. *Journal of Machine Learning Research*. 2011; 12:2825–2830.
- 744 **Pey AL**, Stricher F, Serrano L, Martinez A. Predicted effects of missense mutations on native-state stability
745 account for phenotypic outcome in phenylketonuria, a paradigm of misfolding diseases. *Am J Hum Genet*.
746 2007 Nov; 81(5):1006–1024.
- 747 **Ponzoni L**, Peñaherrera DA, Oltvai ZN, Bahar I. Rhapsody: Predicting the pathogenicity of human missense
748 variants. *Bioinformatics*. 2020; 36(10):3084–3092.
- 749 **Richards S**, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. Stan-
750 dards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the
751 American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics
752 in medicine*. 2015; 17(5):405–423.
- 753 **Riesselman AJ**, Ingraham JB, Marks DS. Deep generative models of genetic variation capture the effects of
754 mutations. *Nature Methods*. 2018; 15(10):816–822. doi: [10.1038/s41592-018-0138-4](https://doi.org/10.1038/s41592-018-0138-4).
- 755 **Rockah-Shmuel L**, Toth-Petroczy A, Tawfik DS. Systematic Mapping of Protein Mutational Space by Prolonged
756 Drift Reveals the Deleterious Effects of Seemingly Neutral Mutations. *PLOS Computational Biology*. 2015;
757 11(8). doi: [10.1371/journal.pcbi.1004421](https://doi.org/10.1371/journal.pcbi.1004421).
- 758 **Ron I**, Horowitz M. ER retention and degradation as the molecular basis underlying Gaucher disease hetero-
759 geneity. *Human molecular genetics*. 2005; 14(16):2387–2398.
- 760 **Rose PW**, Beran B, Bi C, Bluhm WF, Dimitropoulos D, Goodsell DS, Prlić A, Quesada M, Quinn GB, Westbrook
761 JD, et al. The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic acids research*. 2010;
762 39(suppl_1):D392–D401.
- 763 **Schaafsma GC**, Vihinen M. Large differences in proportions of harmful and benign amino acid substitutions
764 between proteins and diseases. *Human mutation*. 2017; 38(7):839–848.
- 765 **Scheller R**, Stein A, Nielsen SV, Marin FI, Gerdes AM, Di Marco M, Papaleo E, Lindorff-Larsen K, Hartmann-
766 Petersen R. Toward mechanistic models for genotype-phenotype correlations in phenylketonuria using pro-
767 tein stability calculations. *Hum Mutat*. 2019 Apr; 40(4):444–457.
- 768 **Shoichet BK**, Baase WA, Kuroki R, Matthews BW. A relationship between protein stability and protein function.
769 *Proceedings of the National Academy of Sciences*. 1995; 92(2):452–456. <https://www.pnas.org/content/92/2/452>, doi: [10.1073/pnas.92.2.452](https://doi.org/10.1073/pnas.92.2.452).
- 770 **Spencer JM**, Zhang X. Deep mutational scanning of *S. pyogenes* Cas9 reveals important functional domains.
772 *Nature News*. 2017 Dec; <https://www.nature.com/articles/s41598-017-17081-y>.

- 773 **Starita LM**, Young DL, Islam M, Kitzman JO, Gullingsrud J, Hause RJ, Fowler DM, Parvin JD, Shendure J, Fields
774 S, et al. Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. *Genetics*. 2015 Jun; <https://www.genetics.org/content/200/2/413>.
- 776 **Starr TN**, Greaney AJ, Hilton SK, Ellis D, Crawford KH, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC,
777 et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and
778 ACE2 binding. *Cell*. 2020; 182(5):1295–1310.
- 779 **Stein A**, Fowler DM, Hartmann-Petersen R, Lindorff-Larsen K. Biophysical and Mechanistic Models for Disease-
780 Causing Protein Variants. *Trends Biochem Sci*. 2019 Jul; 44(7):575–588.
- 781 **Steinegger M**, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. HH-suite3 for fast remote homology
782 detection and deep protein annotation. *BMC bioinformatics*. 2019; 20(1):1–15.
- 783 **Stiffler M**, Hekstra D, Ranganathan R. Evolvability as a Function of Purifying Selection in TEM-1 B-Lactamase.
784 *Cell*. 2015; 160(5):882–892. doi: [10.1016/j.cell.2015.01.035](https://doi.org/10.1016/j.cell.2015.01.035).
- 785 **Strobl C**, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations,
786 sources and a solution. *BMC bioinformatics*. 2007; 8(1):1–21.
- 787 **Suiter CC**, Moriyama T, Matreyek KA, Yang W, Scaletti ER, Nishii R, Yang W, Hoshitsuki K, Singh M, Trehan A,
788 Parish C, Smith C, Li L, Bhojwani D, Yuen LYP, Li CK, Li CH, Yang YL, Walker GJ, Goodhand JR, et al. Massively
789 parallel variant characterization identifies NUDT15 alleles associated with thiopurine toxicity. *Proc Natl Acad
790 Sci U S A*. 2020 Mar; 117(10):5394–5401.
- 791 **UniProt Consortium**. UniProt: a worldwide hub of protein knowledge. *Nucleic acids research*. 2019;
792 47(D1):D506–D515.
- 793 **Virtanen P**, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser
794 W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E,
795 Carey CJ, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*. 2020;
796 17:261–272. doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- 797 **Wagih O**, Galardini M, Busby BP, Memon D, Typas A, Beltrao P. A resource of variant effect predictions of single
798 nucleotide variants in model organisms. *Molecular systems biology*. 2018; 14(12):e8430.
- 799 **Weile J**, Sun S, Cote AG, Knapp J, Verby M, Mellor JC, Wu Y, Pons C, Wong C, Lieshout Nv, et al. A framework
800 for exhaustively mapping functional missense variants. *Molecular Systems Biology*. 2017 Dec; <https://www.embopress.org/doi/full/10.1525/msb.20177908>.
- 802 **Yaguchi H**, Ohkura N, Takahashi M, Nagamura Y, Kitabayashi I, Tsukada T. Menin missense mutants associ-
803 ated with multiple endocrine neoplasia type 1 are rapidly degraded via the ubiquitin-proteasome pathway.
804 *Molecular and cellular biology*. 2004; 24(15):6569–6580.
- 805 **Yang C**, Asthagiri AR, Iyer RR, Lu J, Xu DS, Ksendzovsky A, Brady RO, Zhuang Z, Lonser RR. Missense mutations in
806 the NF2 gene result in the quantitative loss of merlin protein and minimally affect protein intrinsic function.
807 *Proceedings of the National Academy of Sciences*. 2011; 108(12):4980–4985.
- 808 **Yang C**, Huntoon K, Ksendzovsky A, Zhuang Z, Lonser RR. Proteostasis modulators prolong missense VHL
809 protein activity and halt tumor progression. *Cell reports*. 2013; 3(1):52–59.
- 810 **Yin Y**, Kundu K, Pal LR, Moult J. Ensemble variant interpretation methods to predict enzyme activity and as-
811 sign pathogenicity in the CAGI4 NAGLU (Human N-acetyl-glucosaminidase) and UBE2I (Human SUMO-ligase)
812 challenges. *Human Mutation*. 2017; 38(9):1109–1122. doi: [10.1002/humu.23267](https://doi.org/10.1002/humu.23267).
- 813 **Yue P**, Li Z, Moult J. Loss of protein structure stability as a major causative factor in monogenic disease. *Journal
814 of molecular biology*. 2005; 353(2):459–473.

815 Supporting Information

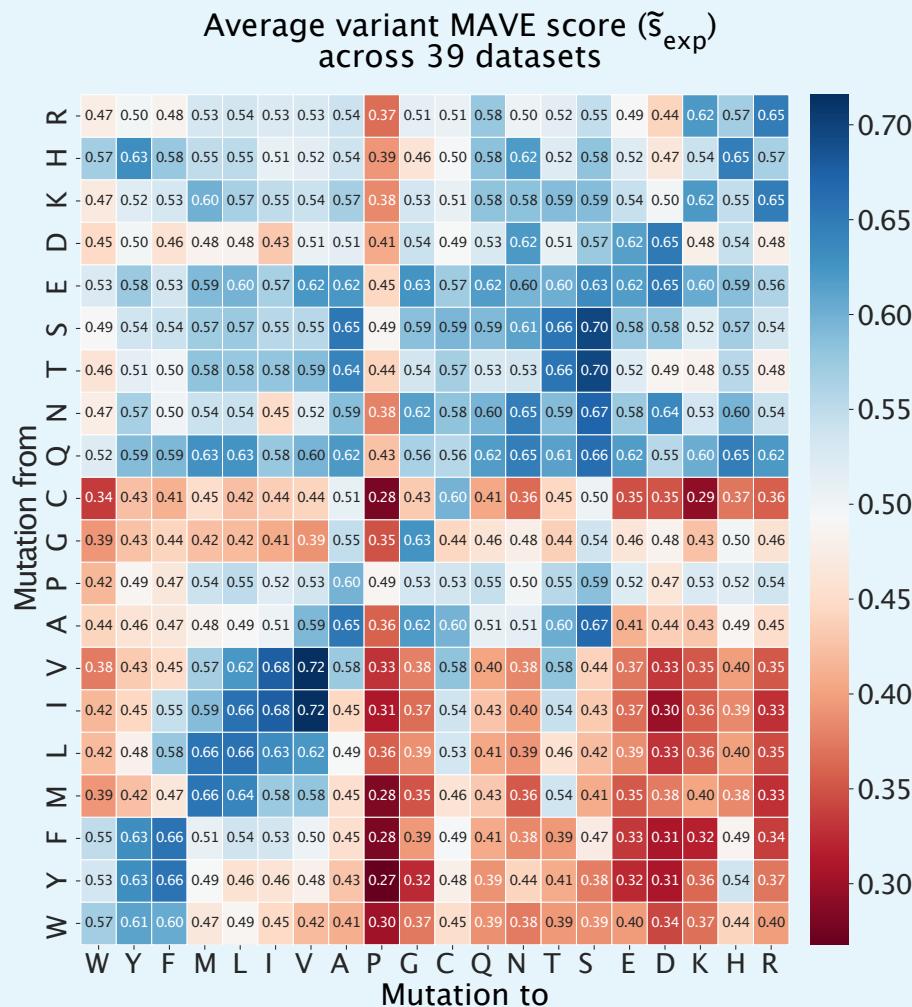
Name	Organism	Assay	Predicted	GEMME	Rosetta	MAVE WT → Mut	PDB	MAVE reference
bla (a) (P62593)	E. coli	Antibiotic resistance	0.81	0.71	0.62	0.48	1ZG4_A	Firmberg et al 2014
bla (c) (P62593)	E. coli	Antibiotic resistance	0.75	0.65	0.55	0.49	1ZG4_A	Stifler et al 2015
bla (d) (P62593)	E. coli	Antibiotic resistance	0.69	0.78	0.35	0.29	1ZG4_A	Jacquier et al 2013
GAL4 (P04386)	Yeast	Two-hybrid assay	0.66	0.67	0.18	0.36	3COQ_A	Kitzman et al 2015
PAB1 (P04147)	Yeast	Competitive growth assay	0.62	0.57	0.27	0.48	6R5K_D	Melamed et al 2013
NUDT15 (b) (Q9NV35)	Human	Drug sensitivity, FACS	0.62	0.56	0.27	0.25	5BON_A	Suiter et al 2019
TPMT (P51580)	Human	Abundance, GFP fluorescence	0.61	0.56	0.47	0.38	2H11_A	Matreyek et al 2018
PTEN (b) (P60484)	Human	Competitive growth assay	0.60	0.57	0.40	0.36	1D5R_A	Mihell et al 2018
P53 (P04637)	Human	Competitive growth assay, P53 agonists	0.60	0.54	0.42	0.28	4QO1_B	Giacomelli et al 2018
GmR (Q53396 template)	E. Coli	Antibiotic resistance	0.57	0.55	0.31	0.25	6BVC_A	Dandageet al 2018
HMGCR (a) (P04035)	Human	Functional complementation, atorvastatin	0.57	0.59	0.24	0.32	1DQ8_A	Jiang et al, 2019
infA (P69222)	E. Coli	Competitive growth assay	0.56	0.56	0.30	0.33	1AH9_A	Kelsic et al 2016
CBS (a) (P35520)	Human	Competitive growth assay	0.55	0.53	0.32	0.35	4COO_A	Sun et 2020
SUMO1 (P63165)	Human	Competitive growth assay, yeast	0.55	0.47	0.42	0.42	1WYW_B	Weile et al 2017
PTEN (a) (P60484)	Human	Abundance, GFP fluorescence	0.54	0.46	0.53	0.40	1D5R_A	Matreyek et al 2018
HMGCR (c) (P04035)	Human	Functional complementation, rosuvastatin	0.53	0.56	0.21	0.29	1DQ8_A	Jiang et al, 2019
UBE2I (P63279)	Human	Competitive growth assay, yeast	0.53	0.46	0.37	0.40	2GRN_A	Weile et al 2017
HMGCR (b) (P04035)	Human	Functional complementation, control	0.53	0.53	0.25	0.29	1DQ8_A	Jiang et al, 2019
bla (b) (P62593)	E. coli	Antibiotic resistance	0.52	0.45	0.38	0.29	1ZG4_A	Deng et al 2012
CBS (b) (P35520)	Human	Competitive growth assay	0.52	0.52	0.29	0.31	4COO_A	Sun et 2020
ADRB2 (P07550)	Human	Competitive growth assay	0.51	0.49	0.21	0.33	2R4R_A	Jones et al 2019
haelliM (P20589)	H. Aegyptius	Competitive growth assay	0.50	0.37	0.41	0.39	3UBT_A	Rockah-Shmuel et al 2015
Src (b) (P12931)	S. Cerevisiae	Kinase activity, catalytic domain	0.48	0.44	0.26	0.29	2H8H_A	Ahler et al 2019
UBI4 (a) (POCG63)	Yeast	Competitive growth assay	0.48	0.48	0.44	0.44	3OLM_D	Roscoe et al 2013
MAPK1 (P28482)	Human	Competitive growth assay, doxycycline	0.47	0.44	0.30	0.22	4QTA_A	Brenan et al 2016
HSP82 (P02829)	Yeast	Competitive growth assay	0.47	0.52	0.04	0.26	2CG9_A	Mishra et al 2016
ccdB (P62554)	E. coli	Reverse survival assay	0.45	0.39	0.29	0.39	1VUB_A	Adkare et al 2012
UBI4 (b) (POCG63)	Yeast	FACS-sorting	0.44	0.41	0.30	0.47	3OLM_D	Roscoe & Bolon, 2014
HA-HIN1 (A0A2Z5-U3Z0)	Influenza virus	Competitive replication assay	0.44	0.35	0.36	0.41	6MYA_A	Doud & Bloom, 2016
NUDT15 (a) (Q9NV35)	Human	Abundance, GFP fluorescence	0.44	0.33	0.53	0.37	5BON_A	Suiter et al 2019
LDLRAP1 (Q5SW96)	Human	Two-hybrid assay, OBFC1 interaction	0.43	0.36	0.37	0.41	3SO6_A	Jiang et al, 2019
Hras (P01112)	Human	Two-hybrid assay	0.43	0.37	0.31	0.36	5E95_A	Bandaru et al 2017

816
817
818
819
820
Supporting Table 1. Overview of experimental data generated by MAVE (part 1 shows 31/39 data sets). The Spearman correlation coefficients between experimental scores and the matching variant GEMME $\Delta\Delta E$, Rosetta $\Delta\Delta G$ and substitution averages ($\tilde{s}_{WT \rightarrow mut}$) are shown.

BRCA1 (b) (P38398)	Human	Yeast two-hybrid assay, phage	0.38	0.31	0.13	0.24	1JM7_A	Starita et al, 2015
env (P03377)	HIV virus	Competitive replication assay	0.33	0.28	0.20	0.33	6B0N_G	Haddox et al 2016
cas9 (Q99ZW2)	S. Pyrogenes	Survival assay	0.27	0.26	0.18	0.20	4CMP_A	Spencer & Zhang, 2017
HA-H3N2 (AOA097-PF60)	Influenza virus	Competitive replication assay	0.24	0.17	0.23	0.26	2YP7_A	Lee et al 2018
TPK1 (Q9H3S3)	Human	Competitive growth assay, yeast	0.24	0.18	0.14	0.28	3S4Y_A	Weile et al 2017
BRCA1 (a) (P38398)	Human	Competitive growth assay, HAP1 cells	0.22	0.17	0.05	0.20	1JM7_A	Starita et al, 2015
CALM1 (P0DP23)	Human	Competitive growth assay, yeast	0.17	0.12	0.08	0.23	4HEX_A	Weile et al 2017

821
822
823
824
825

Supporting Table 2. Overview of experimental data generated by MAVE (part 2 shows 8/39 data sets). The Spearman correlation coefficients between experimental scores and the matching variant GEMME $\Delta\Delta E$, Rosetta $\Delta\Delta G$ and substitution averages ($\tilde{s}_{WT \rightarrow mut}$) are shown.



826
827
828
829
830
831
832

Supporting Figure 1. Average scores (after normalization) across 39 data sets generated by MAVEs (see Methods). Lower scores (red) indicate lower fitness, with normalized experimental scores ranging from 0–1. Substitutions from hydrophobic and typically buried residues to polar or charged residues tend to cause substantial loss of fitness. Substitutions to proline and from cysteine are (on average) also particularly disruptive. These values can be used as a baseline to predict variant effects (Fig. 2b).

