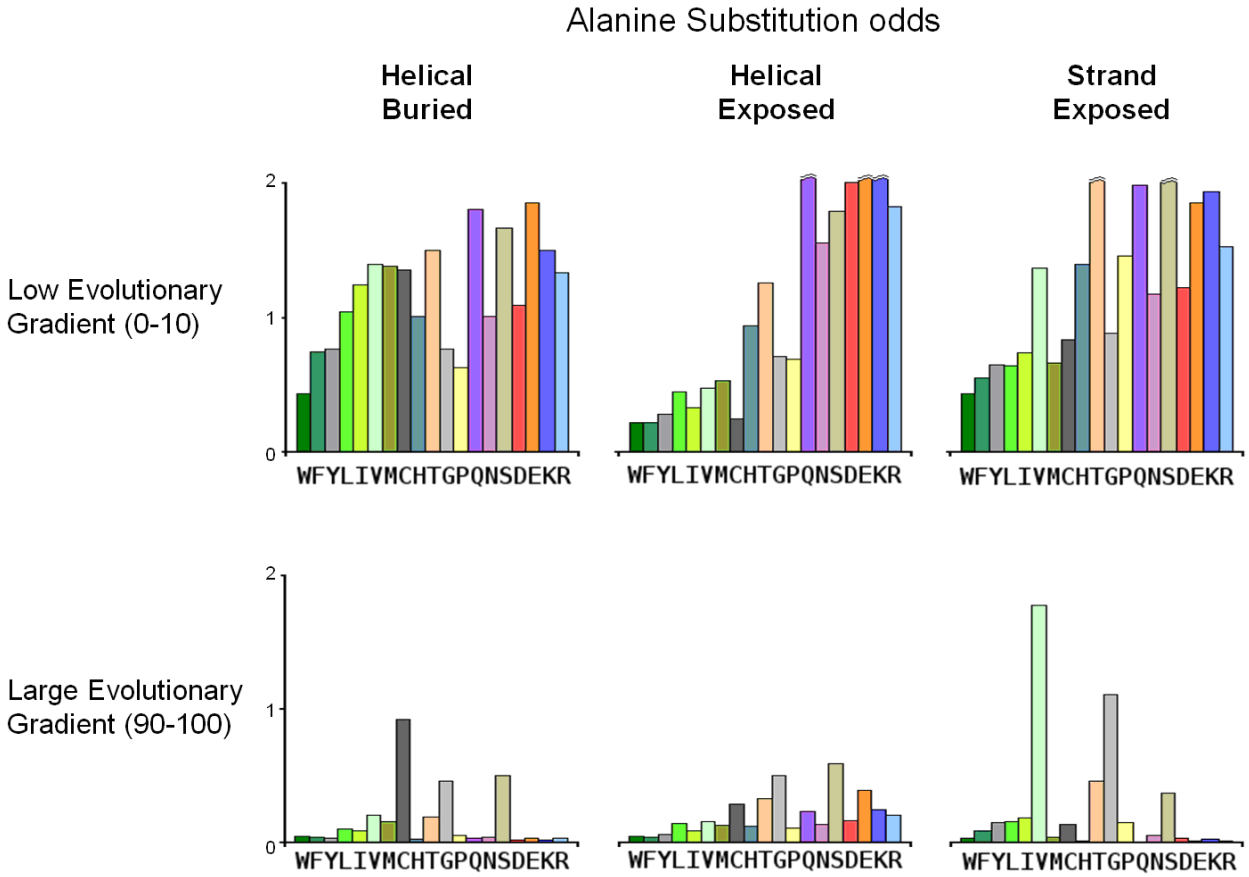
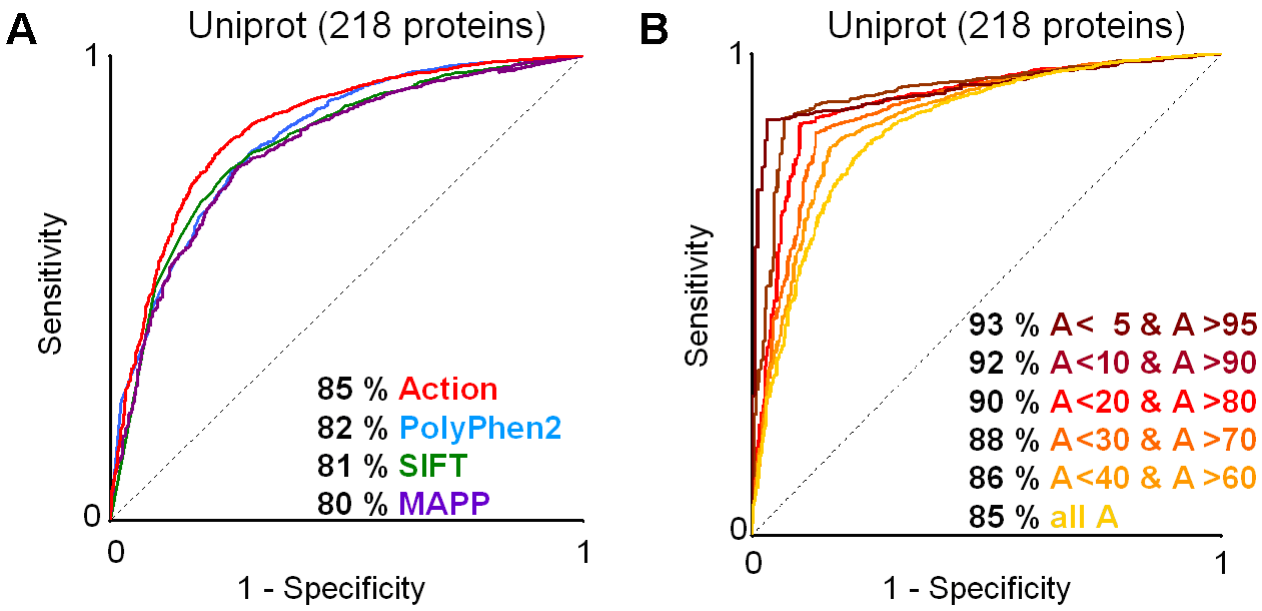
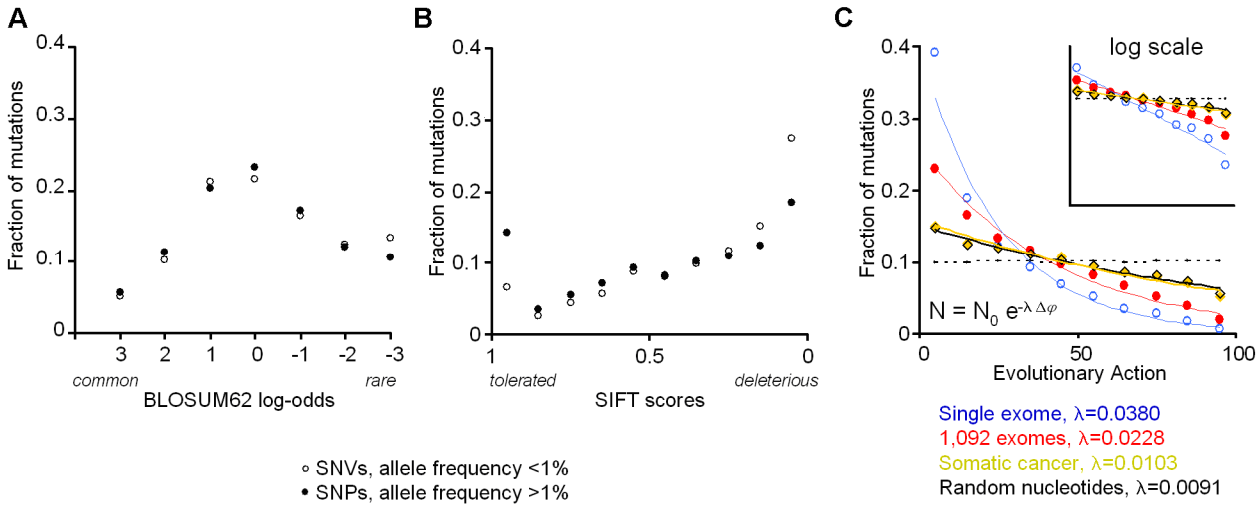
SUPPLEMENTAL MATERIAL



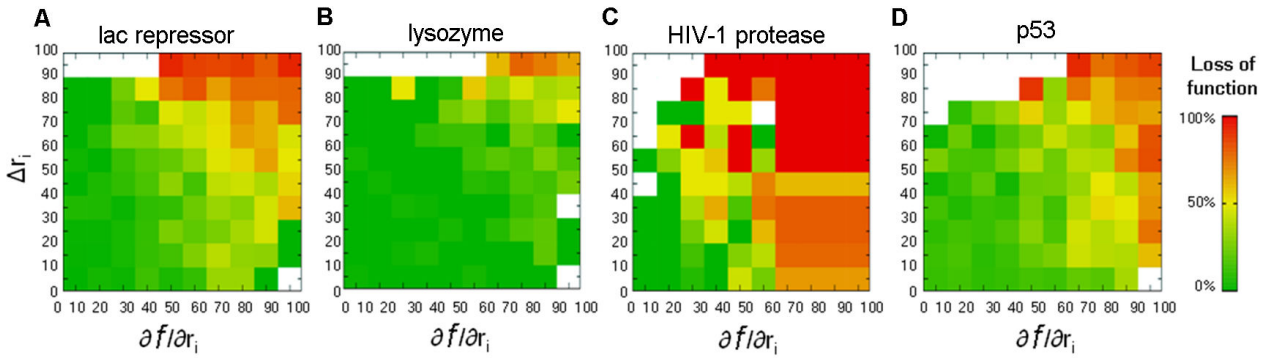
Supplemental Figure S1. The evolutionary gradient dependent substitution odds of alanine also depend on solvent accessibility and secondary structure. The upper row shows the odds of alanines with the smallest evolutionary gradient being substituted to any of the other 19 amino acids when the substituted position is buried and in a helix (left), exposed and in a helix (center), or exposed in a strand (right). The lower row shows how these same odds change if the substituted position has a large gradient. The amino acid types are labeled with single-letter code.



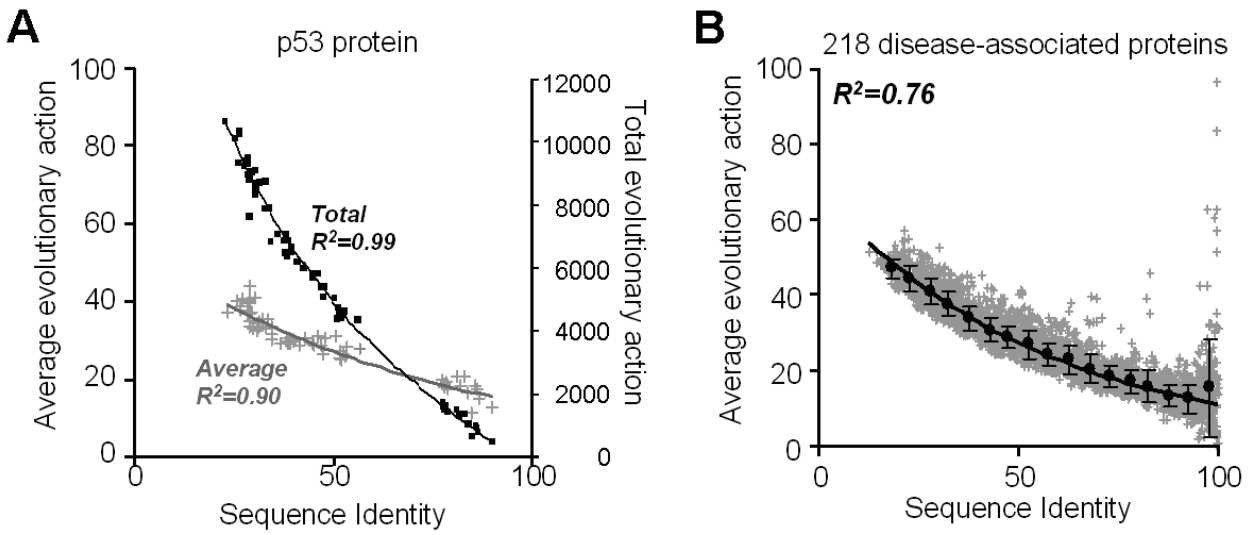
Supplemental Figure S2. The Area Under the receiver operating characteristic Curves (AUC) of the relative sensitivity and specificity to separate mutations associated with disease or not. The dataset consists of 9,347 mutations found on 218 genes, selected so that each gene harbors both disease-causing and benign mutations (see methods). (*A*) The AUC of the action was compared to these of Polyphen2, SIFT, and MAPP methods. (*B*) The AUC of the action (A) was recalculated for fractions of the data of gradually increasing prediction confidence.



Supplemental Figure S3. Distributions of predicted impacts for human protein coding mutations. (*A*) The fractions of SNVs (open symbols) and SNPs (closed symbols) from the 1000 Genomes Project for different Blosum62 log-odd values and (*B*) SIFT score predictions. The Blosum62 and the SIFT scores are reversed in x-axis, in order to indicate more disruptive mutations to the right. (*C*) The fraction of mutations over ten bins of action for: all polymorphisms of the 1000 Genomes Project (in red); polymorphisms of a single exome (in blue); somatic cancer mutations obtained from the TCGA portal (in yellow); and non-synonymous mutations obtained by the translation of random nucleotide changes following the standard genetic code (black line). The colored lines correspond to exponential fits of the data with the same color and they have R^2 values of 0.96 to 0.97. Random amino acid switches have a flat distribution (dotted black line).



Supplemental Figure S4. The fraction of loss of function mutations as a function of Δr_i and $\partial f/\partial r_i$. The color maps present the fraction of loss of function mutations in a scale from green (0%) to red (100%) in (*A*) 4,041 lac repressor mutations in *E. coli,* (*B*) 2,015 lysozyme mutations in bacteriophage T4 (*C*) 336 HIV-1 protease mutations and (*D*) 2,314 p53 mutations. White blocks indicate the lack of data for these bins.



Supplemental Figure S5. The action differences between homologous protein sequences. The average action of amino acid differences per residue (left y-axis) and the total action (panel a, right y-axis) between (*A*) human p53, (*B*) 218 disease-associated human proteins, paired to each one of their homologs in other species, plotted against the sequence identity. The average action data were fit by exponential lines with R^2 values of 0.9 and 0.76, respectively, and the total action data by a logarithmic line with R^2 value of 0.99.