



Supp. Figure S1. Scoring the Magnitude of Effect of Amino Acid Substitutions. The expected input for an unweighted prediction is the protein sequence and substitution whereas the expected input for a weighted prediction is the SwissProt/TrEMBL protein ID and substitution. Next, protein domain annotations from the SUPERFAMILY and Pfam databases are made. In addition, if an unweighted prediction is requested, an *ab initio* HMM is built from the alignment of homologous sequences collected as part of the *JackHMMER* algorithm. The amino acid substitution is then mapped onto the corresponding HMM match states where the information gain (as measured by the Kullback-Leibler divergence from the SwissProt/TrEMBL amino acid composition) is then calculated. This is then used to deduce the most informative HMM and a prediction is made accordingly.