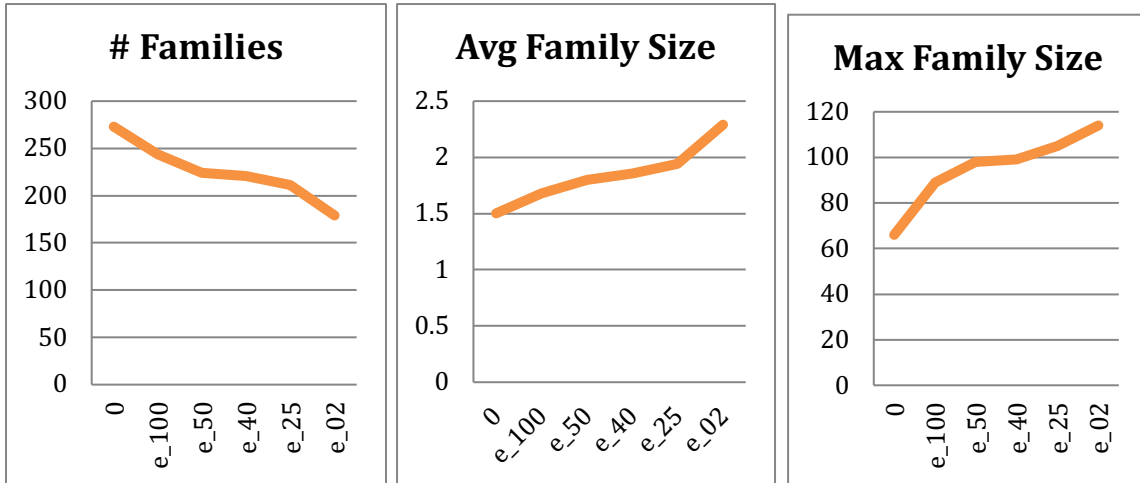
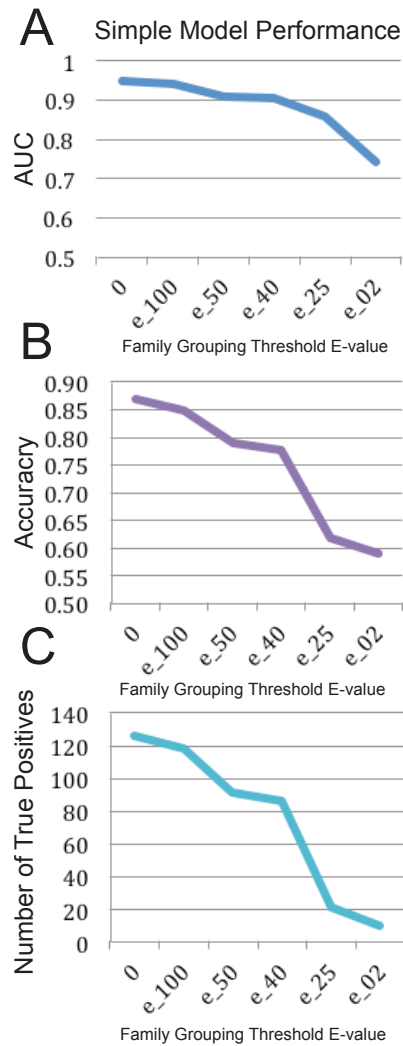


Supplemental Figures

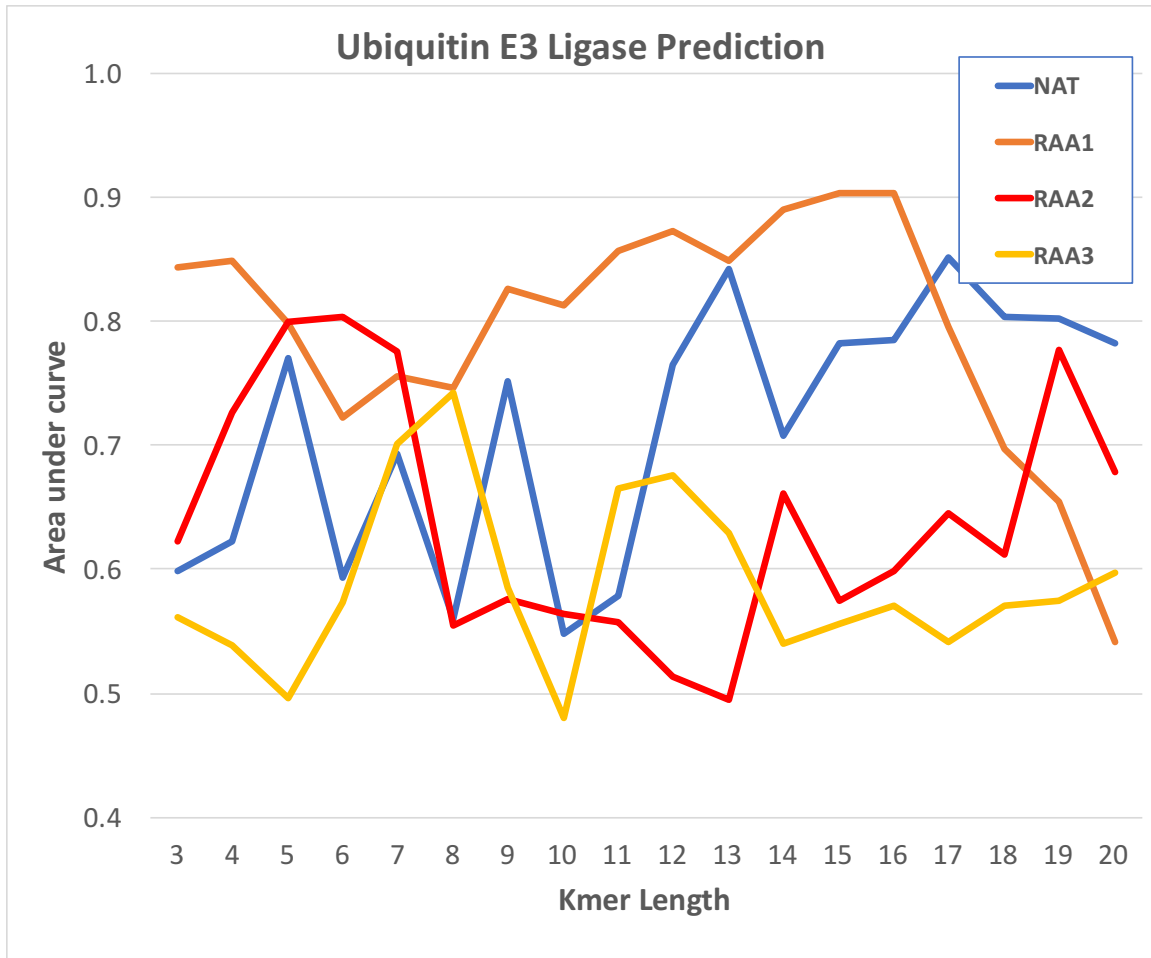


Supplemental Figure 1. Study of the effect of e-value threshold on protein families.



Supplemental Figure 2 Study of impact of family definition on initial model

performance (A) AUC performance of the initial “simple” model over families defined using various e-value thresholds. (B) Accuracy of the simple model. (C) True positives predicted by the model.



Supplemental Figure 3. Performance of reduced amino acid (RAA) models on Ub ligase family-wise classification. Models were generated for various lengths of peptide (X axis) using different RAAs (see Table 1) as described in the text. Training and testing were performed on independent sets 100 times and the average scores for each example used to calculate ROC AUCs. The plot shows that a simple grouping of amino acids by general hydrophobicity provides the best performance.