



Fig. 4. Cluster analysis of enamel pellicle peptide sequences identifies enamel-binding residues. A network of pairwise pellicle peptide sequence alignments was clustered, with edge weights given by alignment scores (increasing edge width and green to violet color) and threshold cutoff corresponding to that of the maximum Matthews correlation coefficient in Figure 2c. Protein names that appear multiple times indicate alternate peptides derived from the same protein. Multiple sequence alignments for each cluster are analyzed: the estimated importance of each alignment column relative to the entire set is shown as blue bars; and that of each residue to its cluster is shown in white to green; bold font marks the top 25th percentile of scores. This analysis identifies the mutual information in the pellicle peptides that encodes the function they have in common: enamel binding. All comparisons were calculated with the pellitrix amino acid substitution matrix.