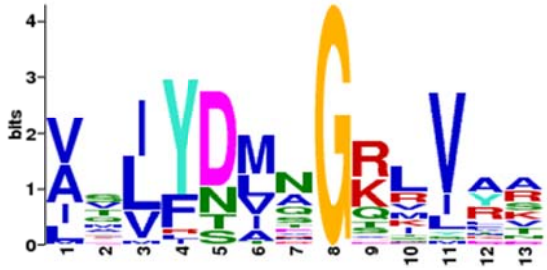
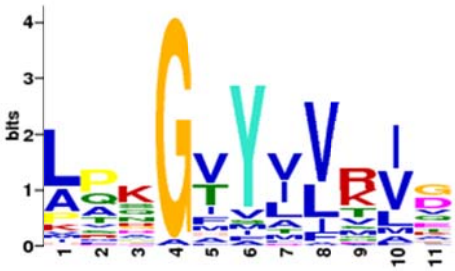
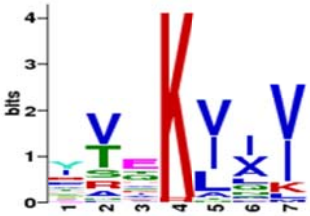
A.



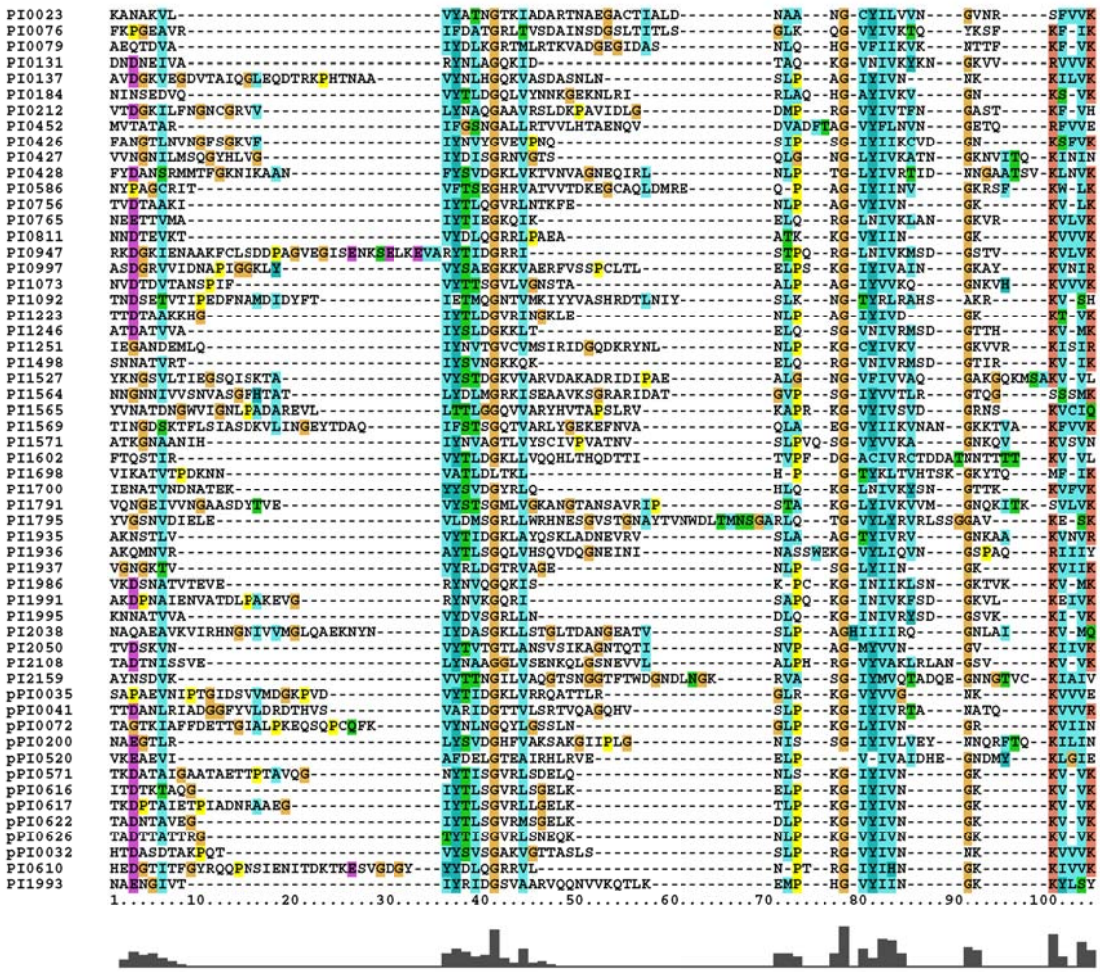
B.



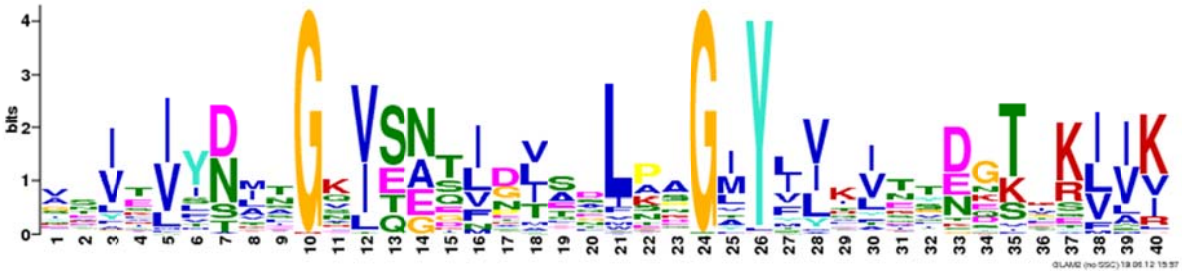
C.



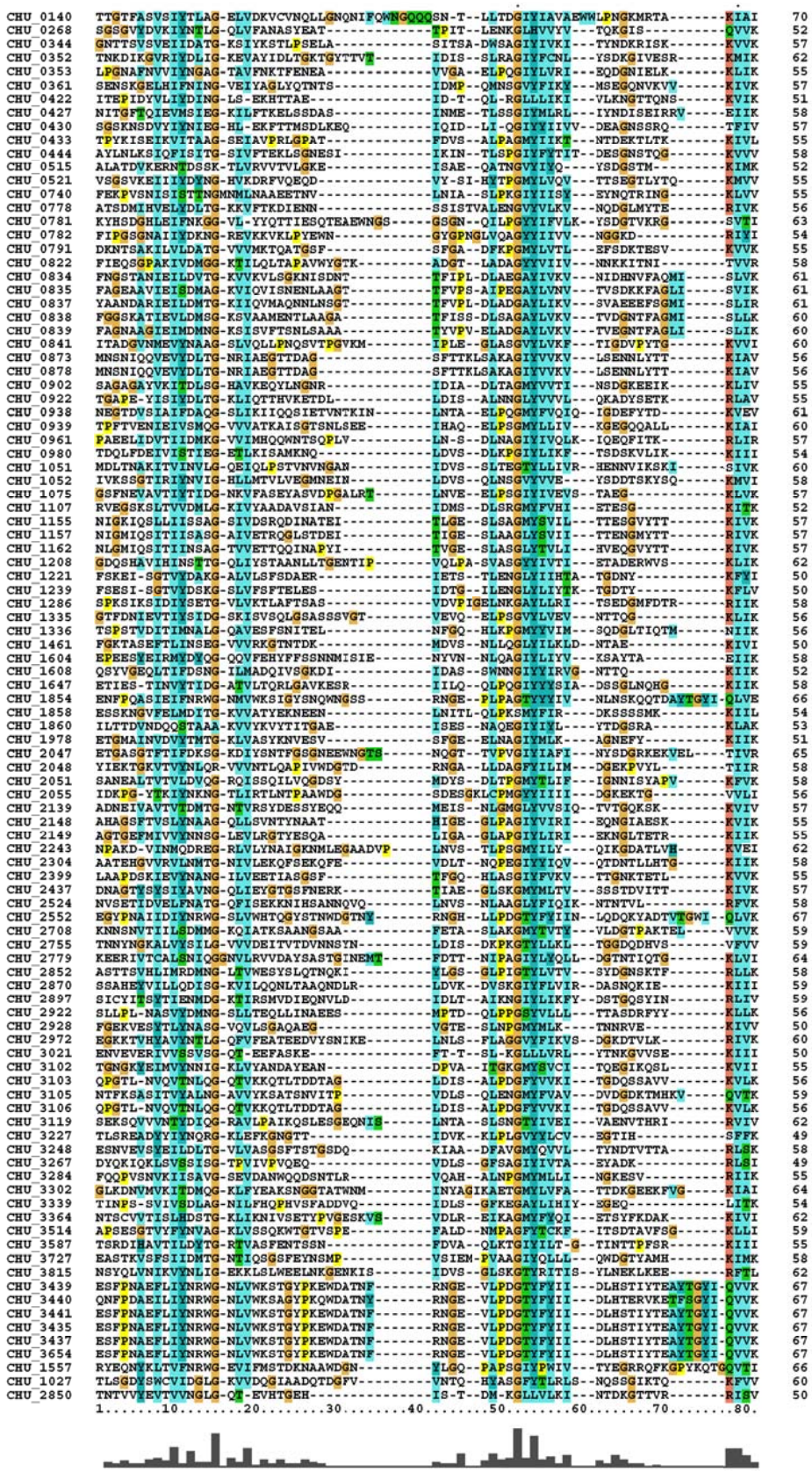
Supplementary Figure 1. Protein logo representations of the individual motifs present in the last 80 amino acids of the C-terminal region of 30 experimentally identified CTD proteins from *Porphyromonas gingivalis* and *Tannerella forsythia* produced by MEME (http://meme.sdsc.edu). (A) Motif B (B) Motif D (C) Motif E.



Supplementary Figure 2. Multiple sequence alignment of the C-terminal region of 54 predicted *P. intermedia* CTD proteins, PI0610 and PI1993. Alignment was extracted from GLAM2 analysis of the 56 *P. intermedia* protein sequences. Parameters used for the GLAM2 analysis were –b (maximum number of aligned columns) = 100, -z (minimum number of sequences in the alignment) = 56 and default for the remaining parameters. *P. intermedia* protein sequences were from the Oralgen Sequence Databases (www.oralgen.lanl.gov).



Supplementary Figure 3. The *C. hutchinsonii* CTD motif derived from 95 HMM predicted CTD proteins in *C. hutchinsonii* (using HMM from *P. gingivalis* and *T. forsythia*). The motif was discovered using GLAM2 with parameters set to: –b (maximum number of aligned columns) = 100, -z (minimum number of sequences in the alignment) = 95 and default for the remaining parameters. *C. hutchinsonii* protein sequences were sourced from NCBI Genbank (www.ncbi.nlm.nih.gov/genbank).



Supplementary Figure 4. Multiple sequence alignment of the C-terminal region of 102 *C. hutchinsonii* proteins. The alignment was constructed from 95 predicted CTD proteins and nine *C. hutchinsonii* large proteins (two of which were in the 95 predicted CTD proteins). Alignment was extracted from GLAM2 analysis of the 102 *C. hutchinsonii* protein sequences. Parameters used for the GLAM2 analysis were –b (maximum number of aligned columns) = 100, -z (minimum number of sequences in the alignment) = 102 and default for the remaining parameters. *C. hutchinsonii* protein sequences were sourced from NCBI Genbank (www.ncbi.nlm.nih.gov/genbank).