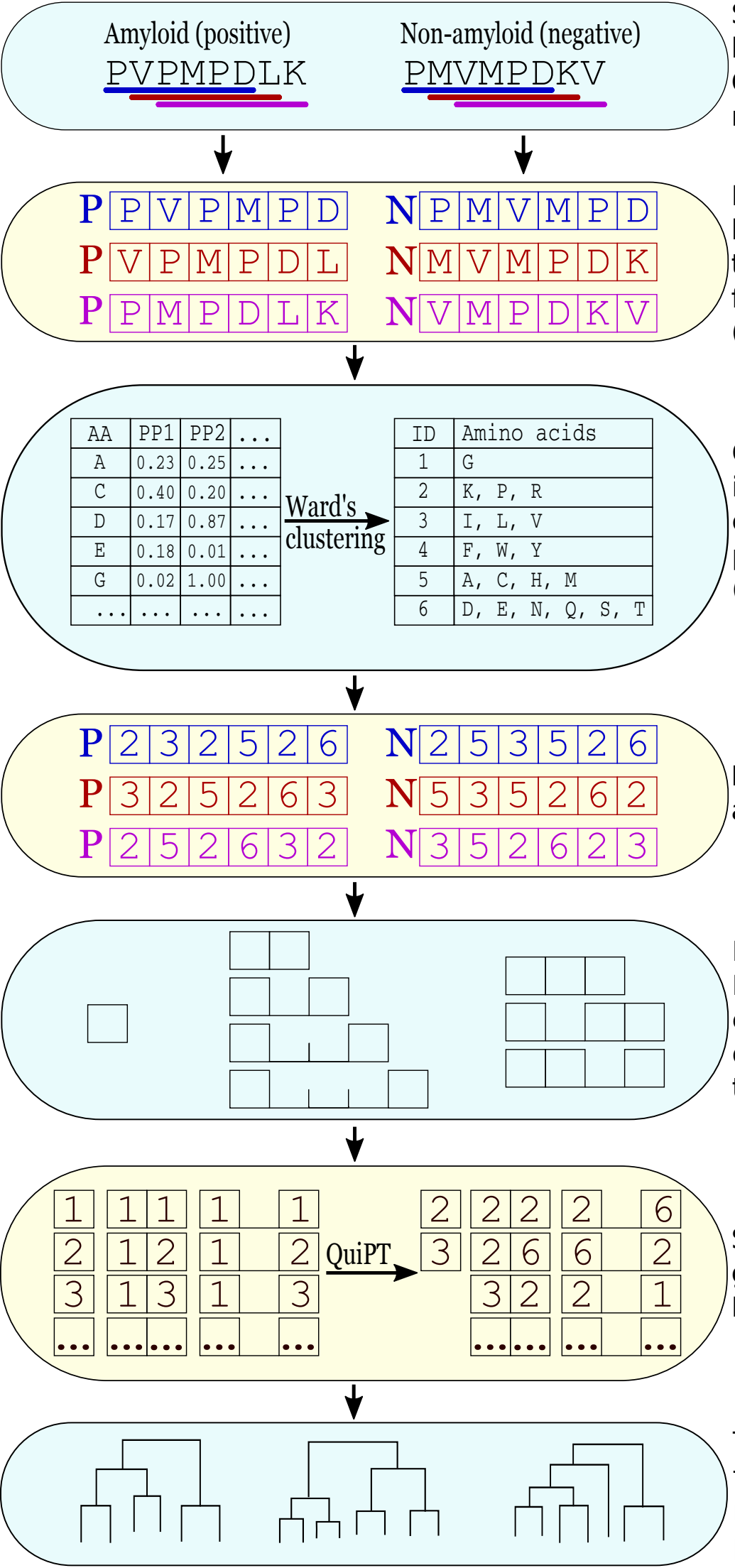

Source data: peptides with known amyloidicity status. Overlapping hexamers are marked by horizontal lines.

Extraction of overlapping hexamers with ascribed the amyloid status taken from their source peptide (P-positive, N - negative).

Clusterization of amino acids into an encoding using a combination of various physicochemical properties (PP).

Reduction of the amino acid alphabet in hexamers.

Extraction of n-grams. From each hexamer, we extracted continuous and discontinous n-grams with the length n = 1, 2 or 3.

Selection of informative n-grams with Quick Permutation Test (QuiPT).

Training of a random forest classifier using the n-grams selected in the previous step.