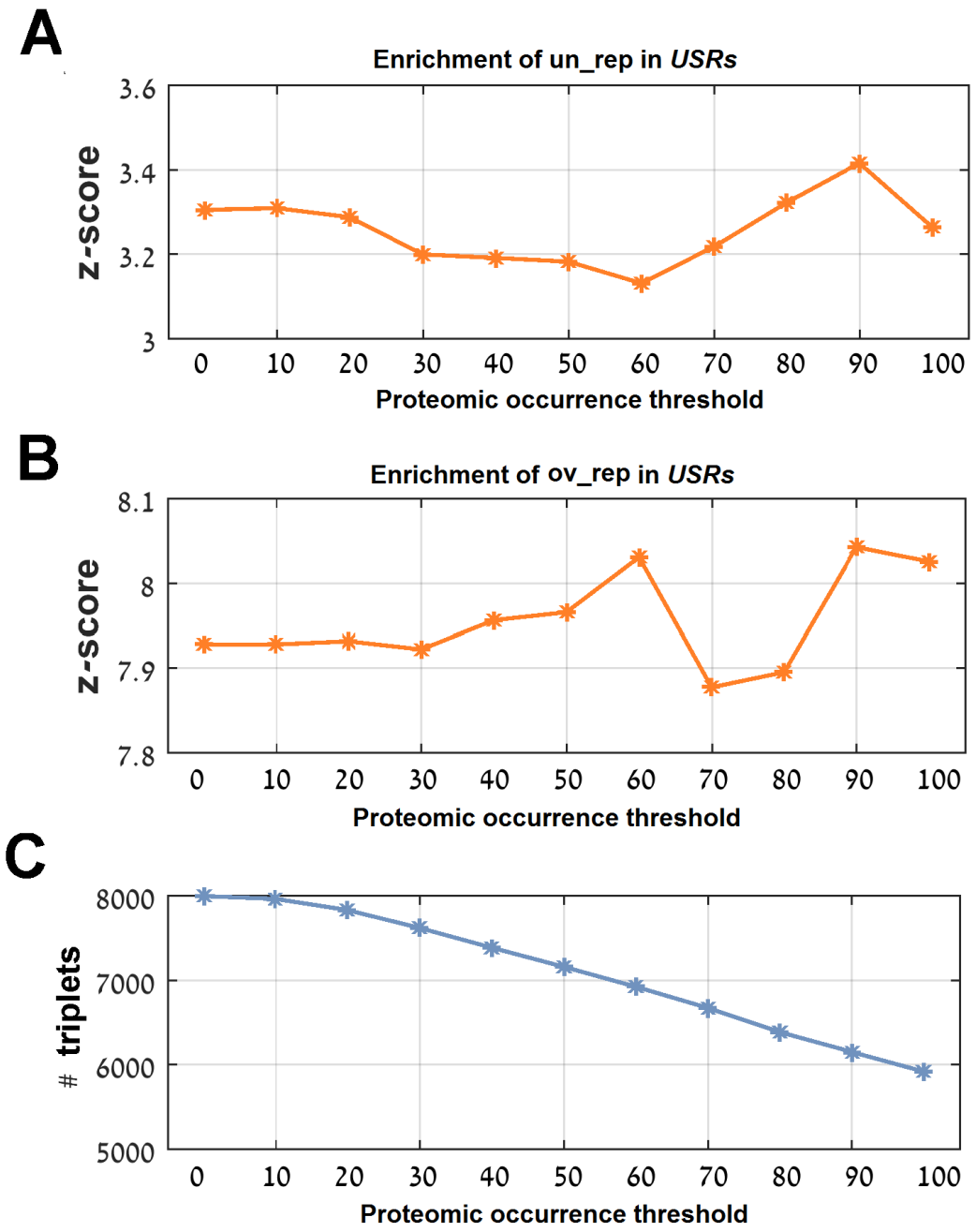

Robustness to different proteomic occurrence thresholds. The proteomic occurrence threshold is the minimum number of occurrences in the proteome below which peptide sequences are filtered. Thresholds vary between 0 and 100. **(A)** Enrichment underrepresented tripeptide sequences in the *USRs.* **(B)** Enrichment over-represented tripeptide sequences in the *USRs.* **(C)** The decrease in number of analyzed triplets over different thresholds. The 20 amino acids define a maximum number of 8,000 analyzed peptide triplets (all presented when no threshold is applied). Dipeptides sequences are not included here as they are not affected by

these thresholds (meaning that all 400 dipeptide sequences are observed for a threshold of 100).