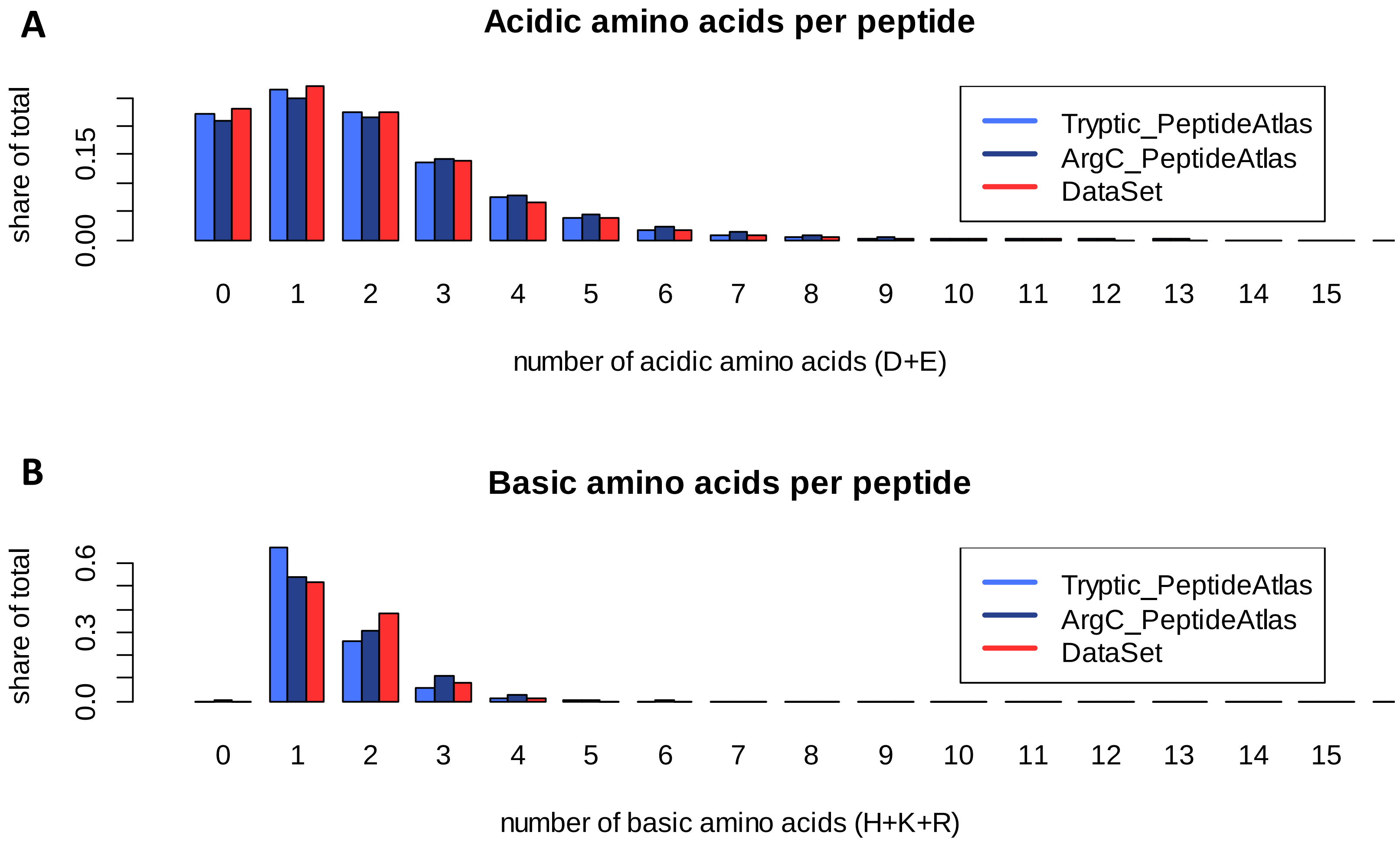

Supplemental Figure 6: For all proteins covered by the PeptideAtlas (PA) repository (October 2017) the most N-terminal peptides were extracted if they had semi-trypsin or Arg-C specificity (C-terminal specific, N-terminal flexible), corresponding to 11,282 PA trypsin and 7,868 PA Arg-C peptides.

(A) Number of acidic amino acids (D+E) per peptide. Whereas there is virtually no difference between the trypsin PeptideAtlas data and our dataset, the Arg-C PA data show a slight trend towards higher numbers of acidic amino acids. (B) Number of basic amino acids (H+K+R) per peptide. The distribution is well in line with the theoretical net charge states and the length distribution in Supplemental Figures 3 and 5 and again may represent iTRAQ-related issues with highly charged peptides, discussed in Supplemental Figure 3.