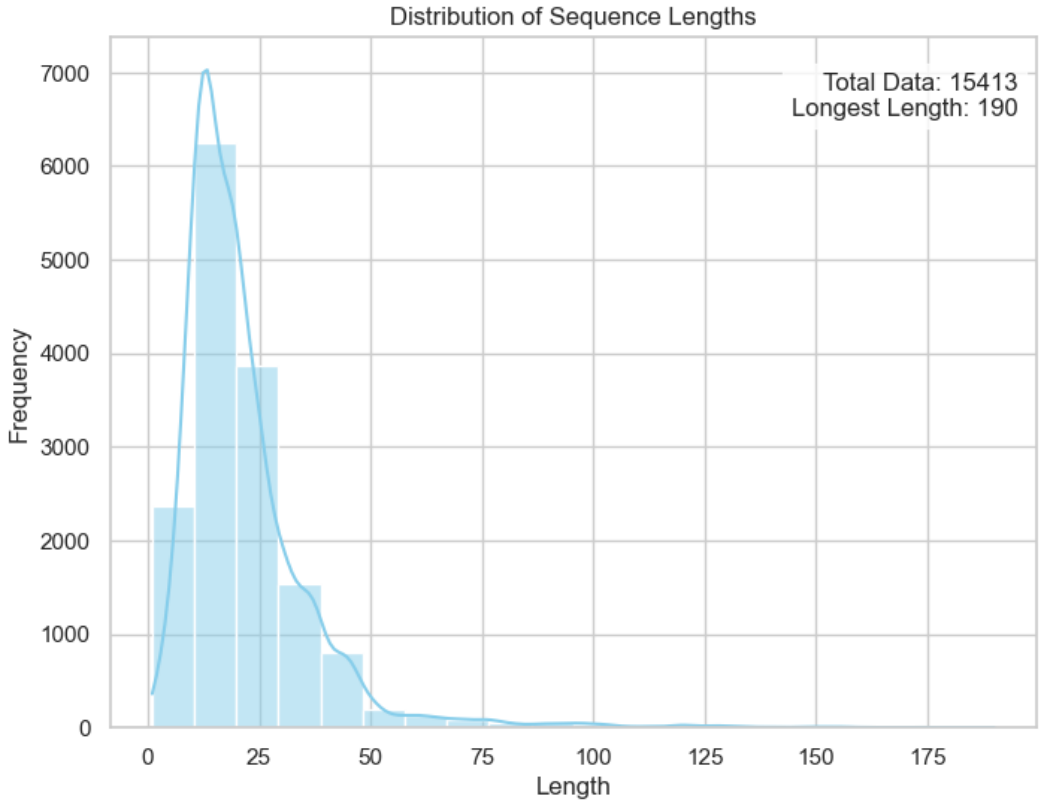
Antimicrobial Peptides Sequence Data Analysis - Results & Visualizations

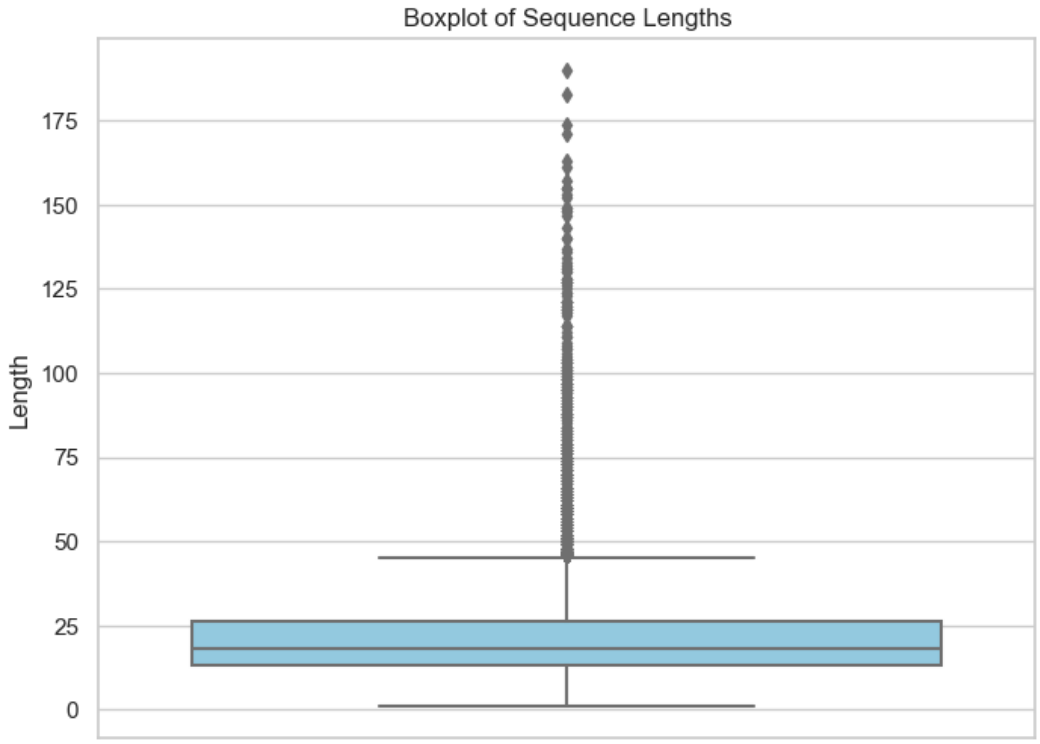
1. Sequence Length Distribution

The distribution of sequence lengths shows a right-skewed pattern with a majority of peptides below 40 amino acids. Total sequences analyzed: 15,413. The longest sequence has 190 residues.



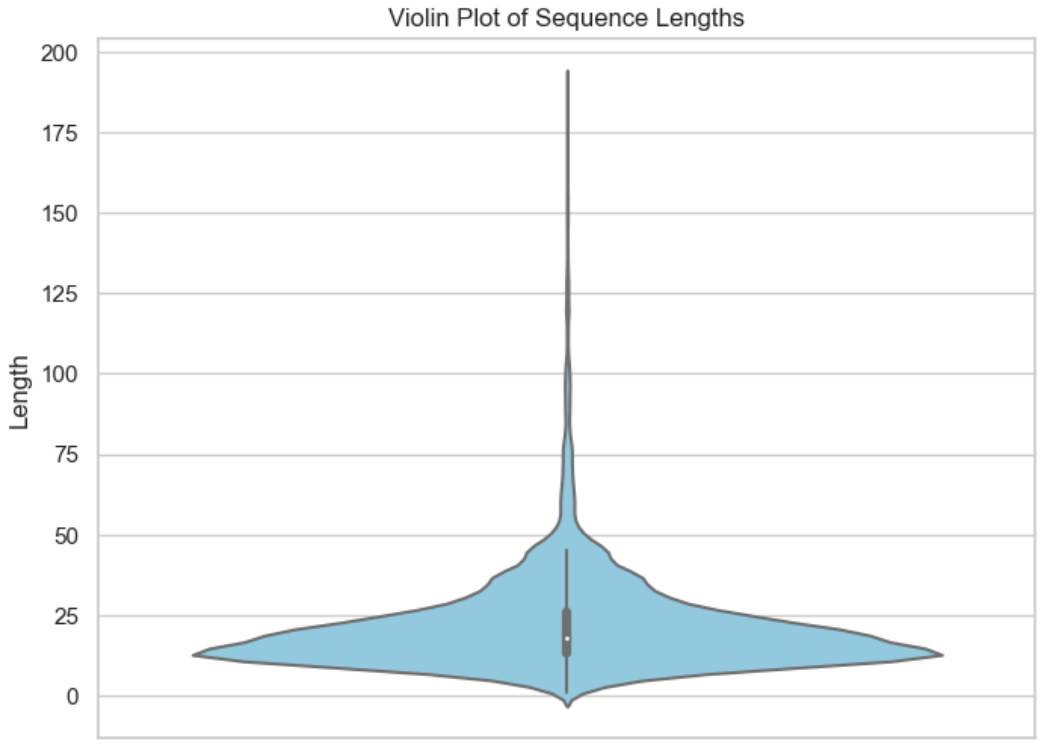
2. Boxplot of Sequence Lengths

The boxplot reveals the interquartile range of peptide lengths and highlights the presence of significant outliers beyond 100 residues.



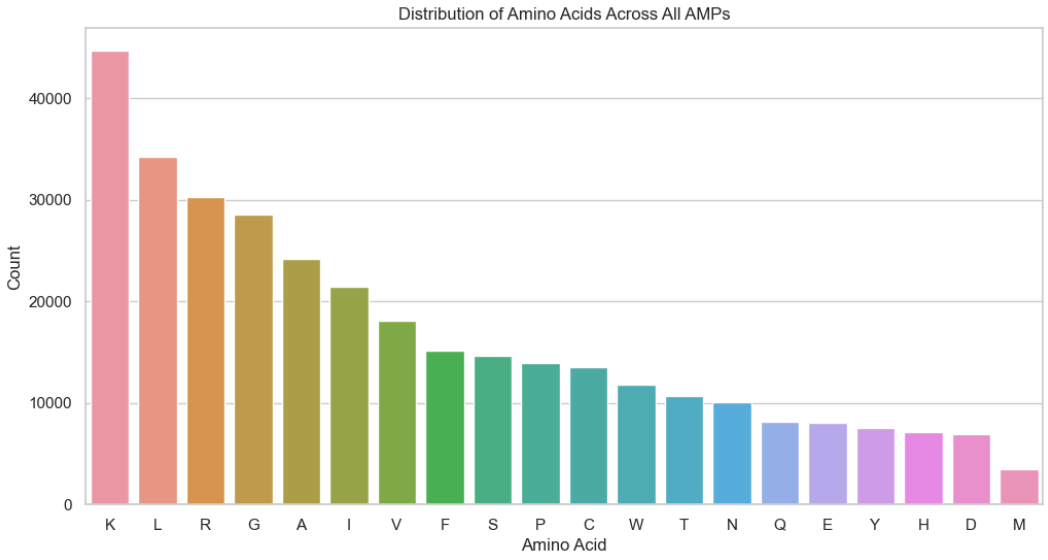
3. Violin Plot of Sequence Lengths

The violin plot shows the density distribution of peptide lengths, reinforcing the concentration around shorter sequences.



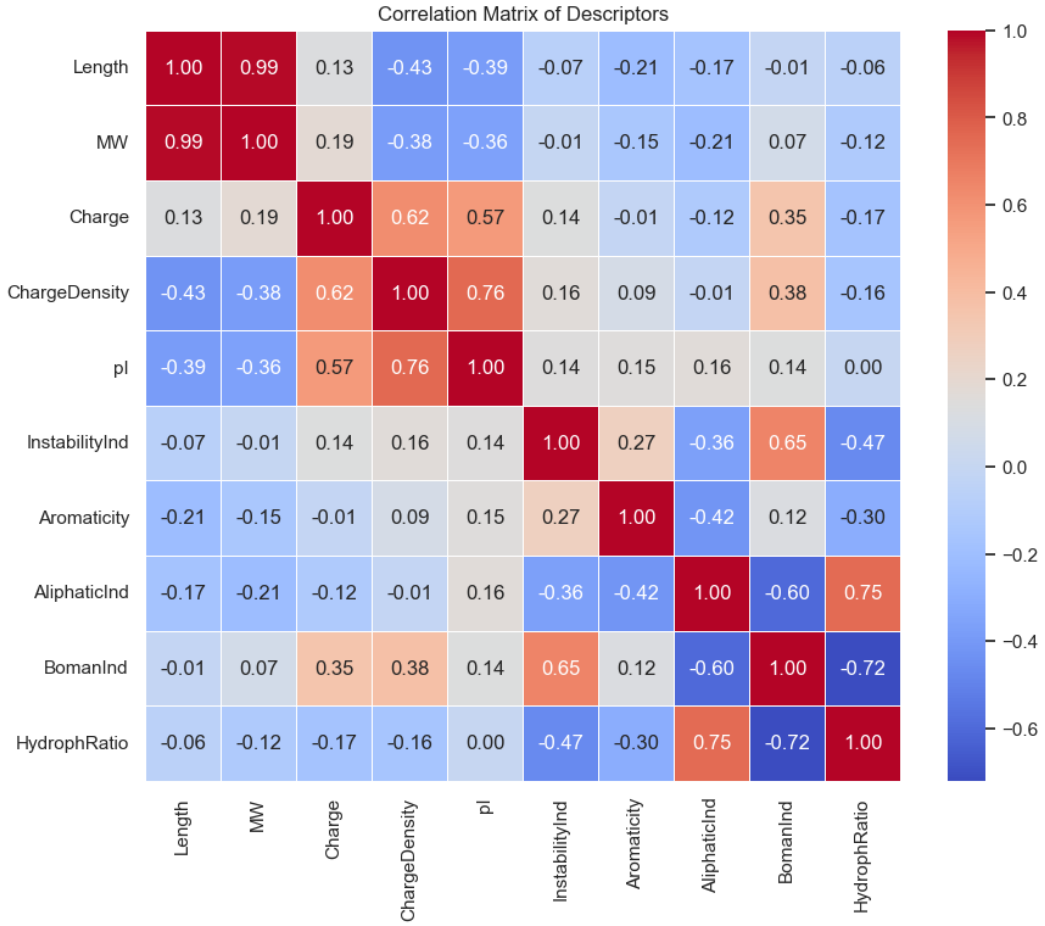
4. Amino Acid Composition

Lysine (K), Leucine (L), and Arginine (R) are the most frequent amino acids in the AMP dataset, indicating their importance in antimicrobial functionality.



5. Descriptor Correlation Matrix

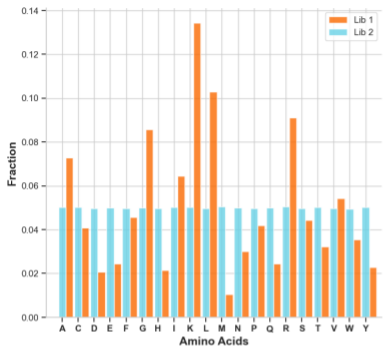
The correlation matrix reveals strong positive correlations between molecular weight (MW) and sequence length, as well as between charge-related properties.

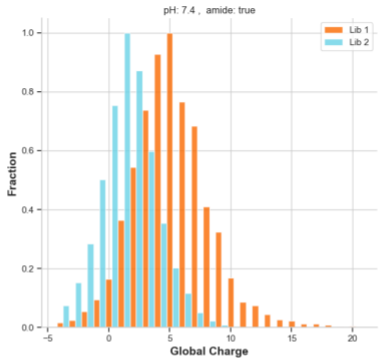


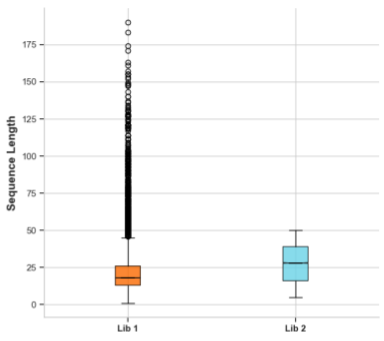
6. Comparative Descriptor Analysis Between Libraries

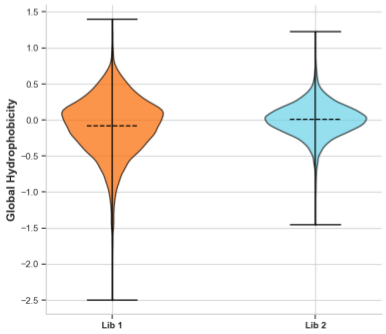
This summary of descriptor comparisons between two peptide libraries shows variations in amino acid composition, global charge, hydrophobicity, and hydrophobic moment.

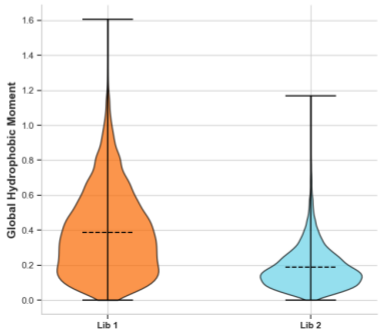
Summary

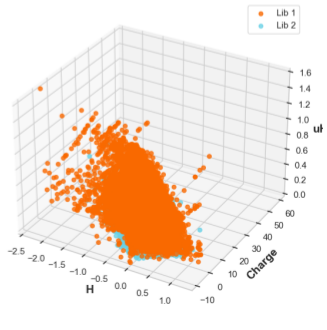

Lib 1
Lib 2
Fraction
Amino Acids
A C D E F G H I K L M N P Q R S T V W Y


pH: 7.4 , amide: true
Lib 1
Lib 2
Fraction
Global Charge


Sequence Length
Lib 1
Lib 2


Global Hydrophobicity
Lib 1
Lib 2


Global Hydrophobic Moment
Lib 1
Lib 2


Lib 1
Lib 2
H
Charge