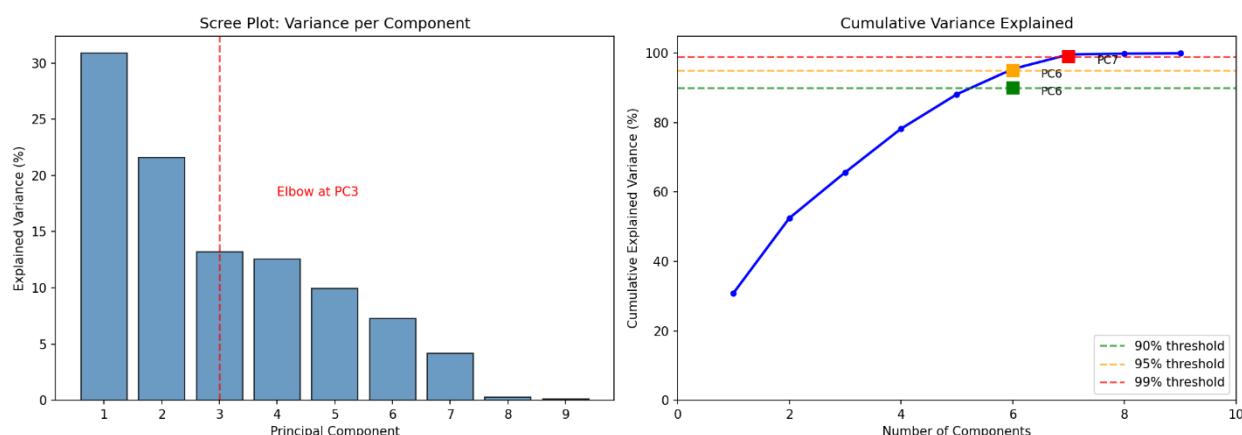
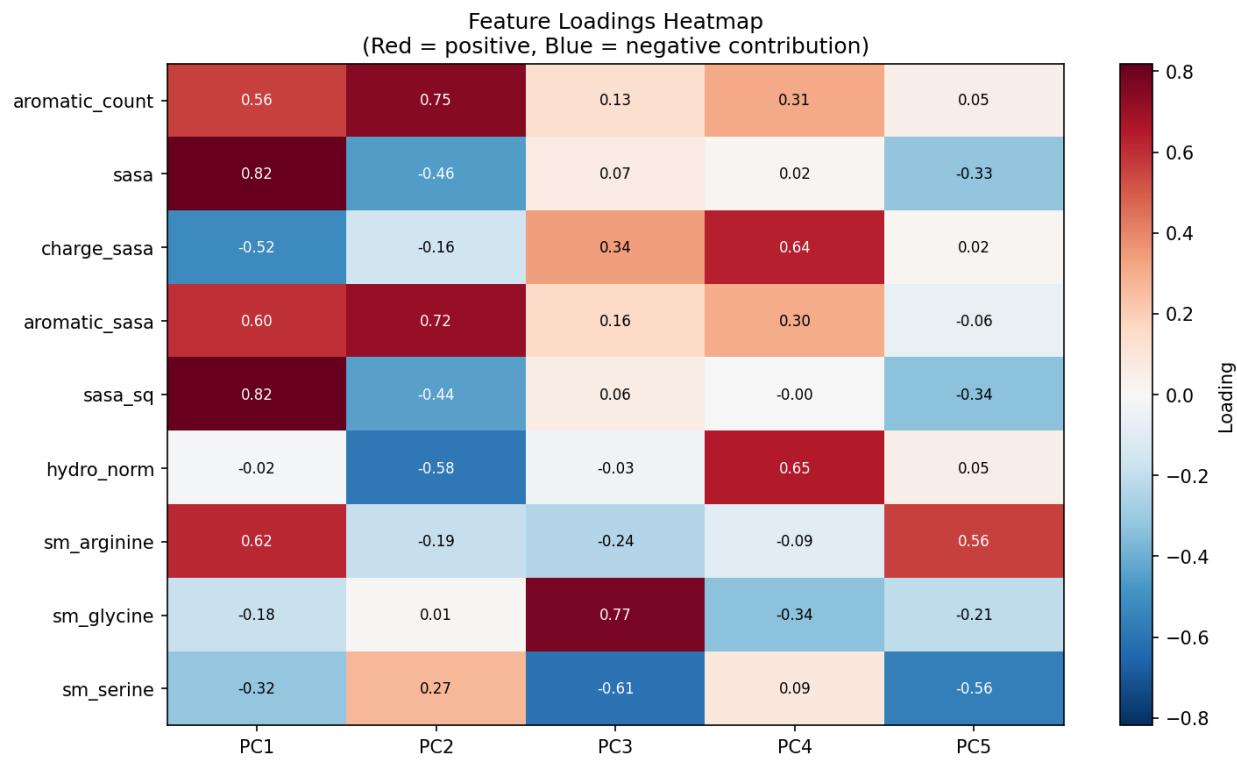


Table 1. PCA Feature analysis; to understand the interpretation column, take a look at the PCA-Loadings-Heatmap plot

Component	Variance	Cumulative	Interpretation
PC1	30.88%	30.88%	Surface area & geometry - solvent-accessible surface dominates
PC2	21.60%	52.47%	Aromaticity & hydrophobicity - π-cation-stacking and hydrophobic interactions
PC3	13.20%	65.67%	Small molecule identity - amino acid type discrimination
PC4	12.59%	78.26%	Charge solvent-accessible surface area and normalized hydrophobic interaction
PC5	9.93%	88.19%	Competitive effect of small molecules
PC6	7.28%	95.47%	Small but structured electrostatic and chemical effects
PC7	4.16%	99.63%	-----
PC8-9	<0.5%	100%	-----





Features KEPT (coefficient above threshold):

Kept Feature	Type	Why Important
sasa	Basic	Binding pocket size
sasa_sq	Polynomial	Non-linear size effects
aromatic_count	Basic	π -stacking potential
aromatic_sasa	Interaction	Exposed aromatic surface
hydro_norm	Normalized	Hydrophobicity (scaled)
charge_sasa	Interaction	Electrostatic surface
sm_arginine	One-hot	Charged amino acid binding
sm_glycine	One-hot	Small amino acid binding
sm_serine	One-hot	Polar amino acid binding

Features REMOVED (coefficient too small):

Removed Feature	Likely Reason
hydrophobicity	Redundant with hydro_norm (normalized version)
charge	Captured by charge_sasa interaction
hydro_charge	Weak interaction, not predictive
hydro_sasa	Weak interaction, not predictive
hydro_sq	Redundant with hydro_norm
prot_HSA	Protein identity not predictive (key insight!)
prot_Lysozyme	Protein identity not predictive
prot_BSA	Protein identity not predictive
prot_* (others)	Protein identity not predictive
sm_glucose	Low binding affinity variation
sm_ATP	Low binding affinity variation