**Features used**

Here I have clearly mentioned why did we choose and how to defend when ma’am ask why did we choose only this particular thing.

**Chou-Fasman**

* **What is it?**  
  A method to predict secondary structure of a peptide based on amino acid propensities to form alpha-helices, beta-sheets, or turns.
* **Why is it included?**  
  Epitopes often lie in specific secondary structures like coils. Chou-Fasman helps in approximating the structure using only sequence.
* **How to defend?**  
  It gives the model structural insight without needing 3D structures. Since loops and turns are common in epitopes, this guides the model towards relevant regions.

**Emini**

* **What is it?**  
  A surface accessibility scale that estimates the likelihood of a residue being on the surface of a protein.
* **Why is it included?**  
  Surface-exposed residues are more likely to interact with immune cells and form B-cell epitopes.
* **How to defend?**  
  This feature directly captures a key epitope characteristic: exposure. Including Emini improves the identification of visible, accessible antigenic regions.

**Kolaskar-Tongaonkar**

* **What is it?**  
  A semi-empirical method that predicts antigenic determinants based on physicochemical properties of amino acids.
* **Why is it included?**  
  It’s tailored for B-cell epitope prediction and considers hydrophilicity, accessibility, and flexibility collectively.
* **How to defend?**  
  It's experimentally validated and widely used. It provides a strong biological baseline for antigenic region identification.

**Parker**

* **What is it?**  
  A hydrophilicity scale that measures how water-loving an amino acid residue is.
* **Why is it included?**  
  Hydrophilic residues are usually on the surface and thus more likely to be recognized as epitopes.
* **How to defend?**  
  Hydrophilicity indicates accessibility and solubility. Parker’s scale complements Emini and Kolaskar to increase surface prediction accuracy.

**Isoelectric Point**

* **What is it?**  
  The pH at which the peptide has a net zero charge.
* **Why is it included?**  
  Charge influences solubility and interaction with immune receptors like MHC or antibodies.
* **How to defend?**  
  It provides an overview of the peptide’s charge behavior, helping estimate its interaction with other biomolecules under physiological conditions.

**Aromaticity**

* **What is it?**  
  Proportion of aromatic amino acids (like Phe, Trp, Tyr) in the peptide.
* **Why is it included?**  
  Aromatic residues contribute to structural stability and specific binding interactions through pi-stacking and hydrophobic packing.
* **How to defend?**  
  Aromatic content affects binding strength and structural conformation. Including it helps in identifying potential hot spots for interaction.

**Hydrophobicity**

* **What is it?**  
  A measure of how water-repellent a residue or region is.
* **Why is it included?**  
  Hydrophobic residues tend to be buried but may form part of binding cores or MHC-anchor residues.
* **How to defend?**  
  Though not usually surface-exposed, some hydrophobic patches contribute to binding affinity. This feature complements accessibility measures.

**Stability**

* **What is it?**  
  A prediction of how structurally stable a peptide is under physiological conditions.
* **Why is it included?**  
  Stable peptides are more likely to survive in vivo and act as effective epitopes or vaccine candidates.
* **How to defend?**  
  An unstable peptide may degrade before immune detection. Including stability filters out poor candidates early in the pipeline.

**Charge**

* **What is it?**  
  Net charge of the peptide at physiological pH.
* **Why is it included?**  
  Charge affects solubility, membrane interaction, and compatibility with MHC or antibody surfaces.
* **How to defend?**  
  It's critical for electrostatic interactions. Along with isoelectric point, it helps capture the peptide's biochemical environment.

**Flexibility**

* **What is it?**  
  A measure of the peptide's mobility, based on empirical or predicted scales like B-factors.
* **Why is it included?**  
  Flexible regions can adopt conformations needed for optimal binding to immune receptors.
* **How to defend?**  
  Antibodies often bind to mobile regions. Flexibility has been experimentally linked with increased epitope potential.

**Solvent Accessibility**

* **What is it?**  
  Degree to which a residue is exposed to the solvent.
* **Why is it included?**  
  Surface-exposed residues are more accessible for immune recognition.
* **How to defend?**  
  It's a core determinant for epitopes, especially B-cell types. Ignoring it can lead to false positives in buried regions.

**BLOSUM Score**

* **What is it?**  
  A score from the BLOSUM substitution matrix indicating how conserved an amino acid is across species.
* **Why is it included?**  
  Conserved regions tend to be functionally important and consistent immune targets.
* **How to defend?**  
  This adds evolutionary robustness. Highly conserved epitopes are ideal for vaccine design due to their universality.

**PTM Sites**

* **What is it?**  
  Potential Post-Translational Modification sites like phosphorylation or glycosylation.
* **Why is it included?**  
  PTMs can mask epitopes or change their conformation, affecting immunogenicity.
* **How to defend?**  
  Including PTM sites ensures the model captures only real-world visible or accessible epitopes, not theoretical ones hidden by sugar or phosphate groups.

**Interaction Energy**

* **What is it?**  
  Computational prediction of binding strength between peptide and MHC or antibody.
* **Why is it included?**  
  Lower (more negative) binding energy suggests stronger interactions, indicating better epitope candidates.
* **How to defend?**  
  This simulates in vitro or in vivo binding in silico. It provides a predictive endpoint that integrates sequence, structure, and biochemical behavior into one measurable score.