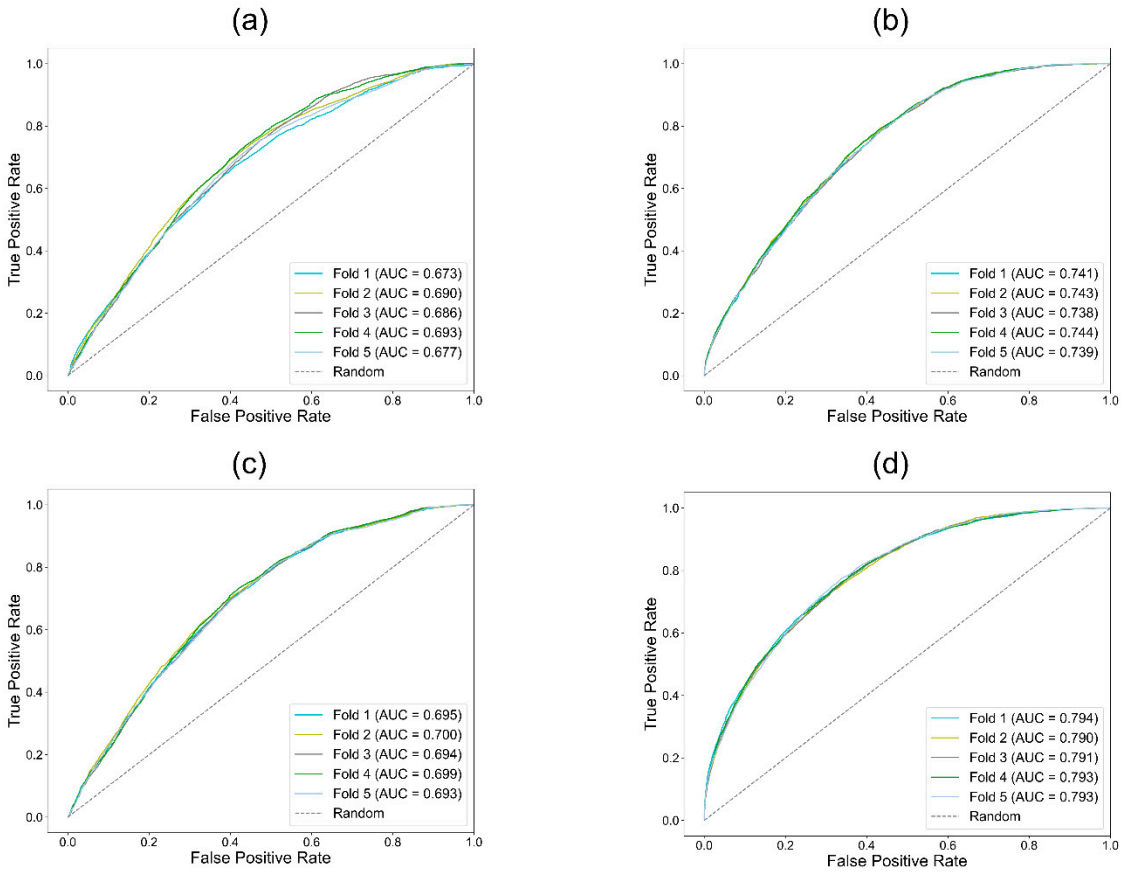


Figure S1. Performance of AUC for machine learning methods through 5-fold cross-validation. (a). Elastic Net(b). Gradient Boosting(c). Logistic Regression (d). XGBoost

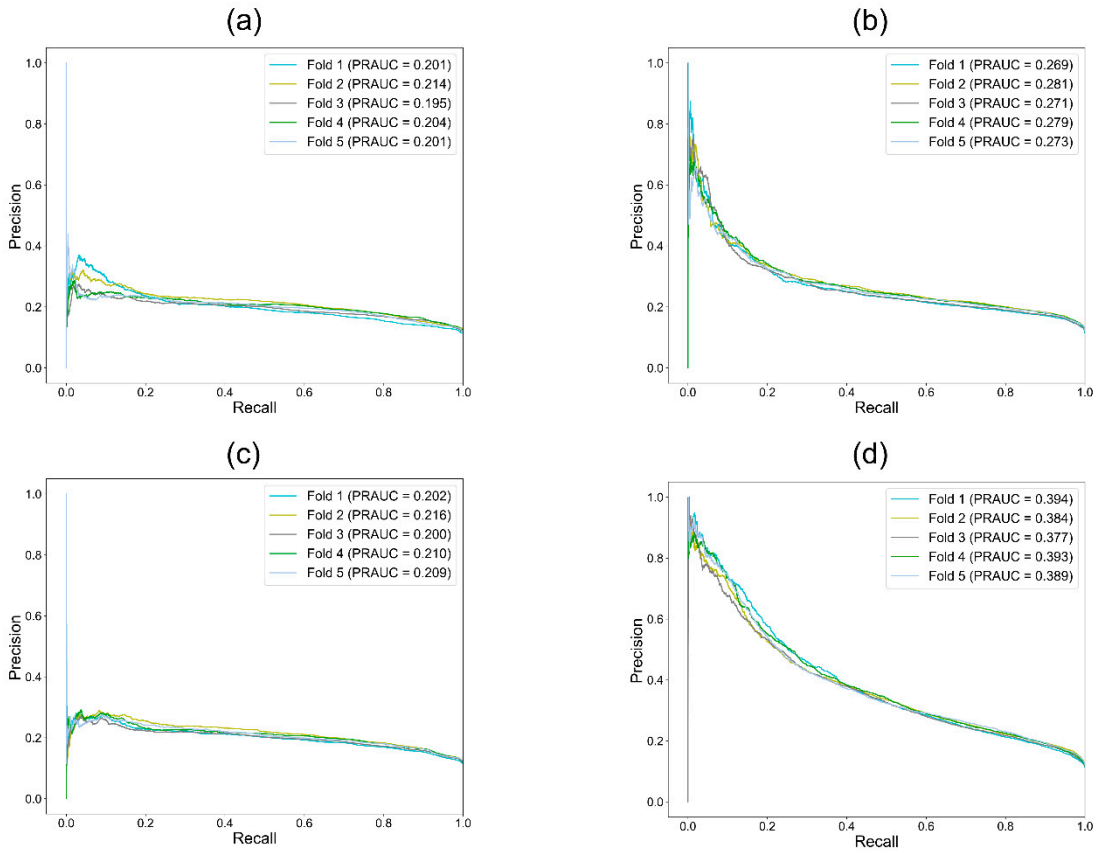


Figure S2. Performance of AUPRC for machine learning methods through 5-fold cross-validation. (a). Elastic Net(b). Gradient Boosting(c). Logistic Regression (d). XGBoost

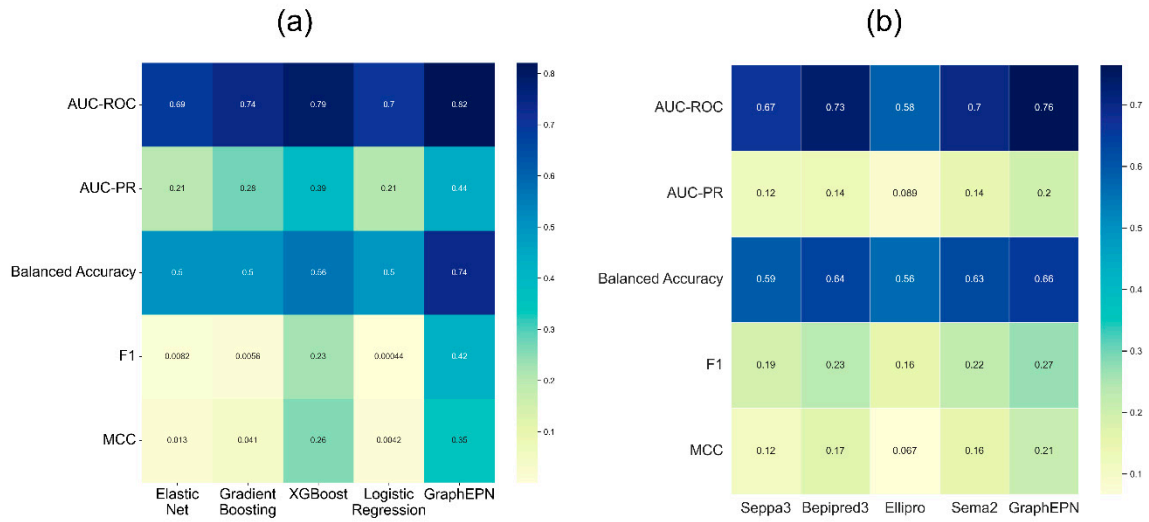


Figure S3. Metrics Heatmap (a). Comparison with machine learning methods(b). Comparison with peer methods

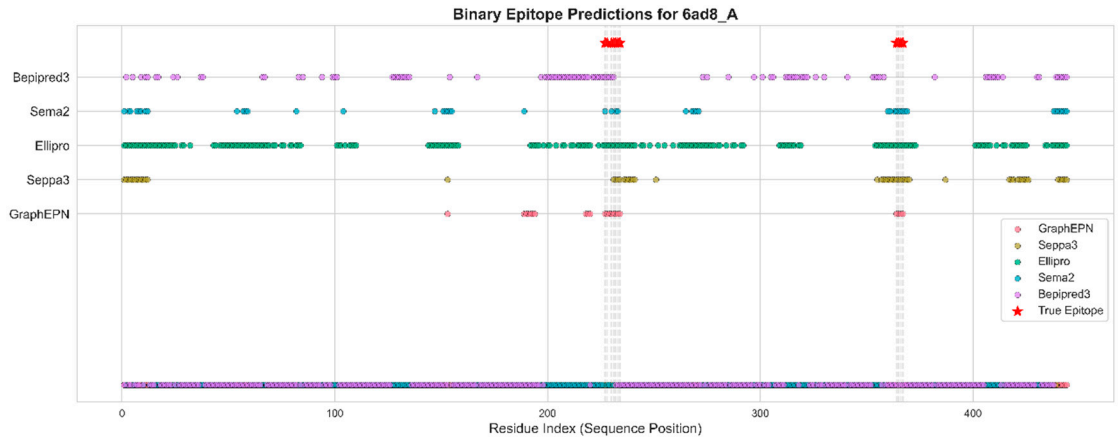


Figure S4. Sequence-based comparison of epitope predictions for PDB 6ad8_A. The x-axis represents the residue index, while the y-axis displays the predicted epitope positions from different methods. Red stars indicate the true epitope residues.

 # GraphEPN Pseudocode:

 # 1. Pretrain VQ-VAE model

Function PretrainVAE(parameters):
    Load pretraining dataset
    Initialize VQ-VAE model and optimizer
    For each epoch:
        For each batch in the dataset:
            Encode node features using the Encoder
            Quantize features via the Vector Quantizer
            Decode features with the Decoder
            Compute total loss: reconstruction loss + quantization loss + masked loss
            Backpropagation and optimizer update
        Save trained VQ-VAE model

 # 2. Train GNN for epitope prediction

Function TrainGNN(VAE_model, GNN_model, training_data, parameters):
    Set VAE_model to evaluation mode
    Set GNN_model to training mode
    Initialize optimizer
    For each epoch:
        For each batch in training dataset:
            Extract node features from VAE_model
            Concatenate encoded (z) and quantized (e) features
            Predict epitope probabilities using GNN model
            Compute total loss: BCE + Dice
            Backpropagation and optimizer update
        Validate model performance and save best checkpoint if improved
        If early stopping condition met, break