

# Comparative Analysis of Protein, Nucleotide, and Structural Conservation Across Multiple Species

Divyadarshan Soni 22b0387@iitb.ac.in Vikash Kumar 22b0359@iitb.ac.in Rana Das 22b0738@iitb.ac.in

## Introduction

Genetic conservation helps us understand how genes have evolved and how they function across different species. This study focuses on comparing conservation patterns in protein sequences, nucleotide sequences, and 3D structures of genes. We analyzed 50 genes of different lengths (small, medium, large) from five species: human, mouse, zebrafish, chimpanzee, and dog. Using sequence alignment tools and scoring methods like BLOSUM62, we calculated conservation scores for proteins and nucleotides. Structural conservation was measured by analyzing spatial variations. The goal is to see how these domains are related and to understand how evolutionary pressures shape genetic conservation.

## Results 1

- The model was run on 50 different genes with varied gene lengths (Small - GAPDH, Medium - BRCA1, Large - PKD1).
- The model provided conservation scores for each element of the MSA.

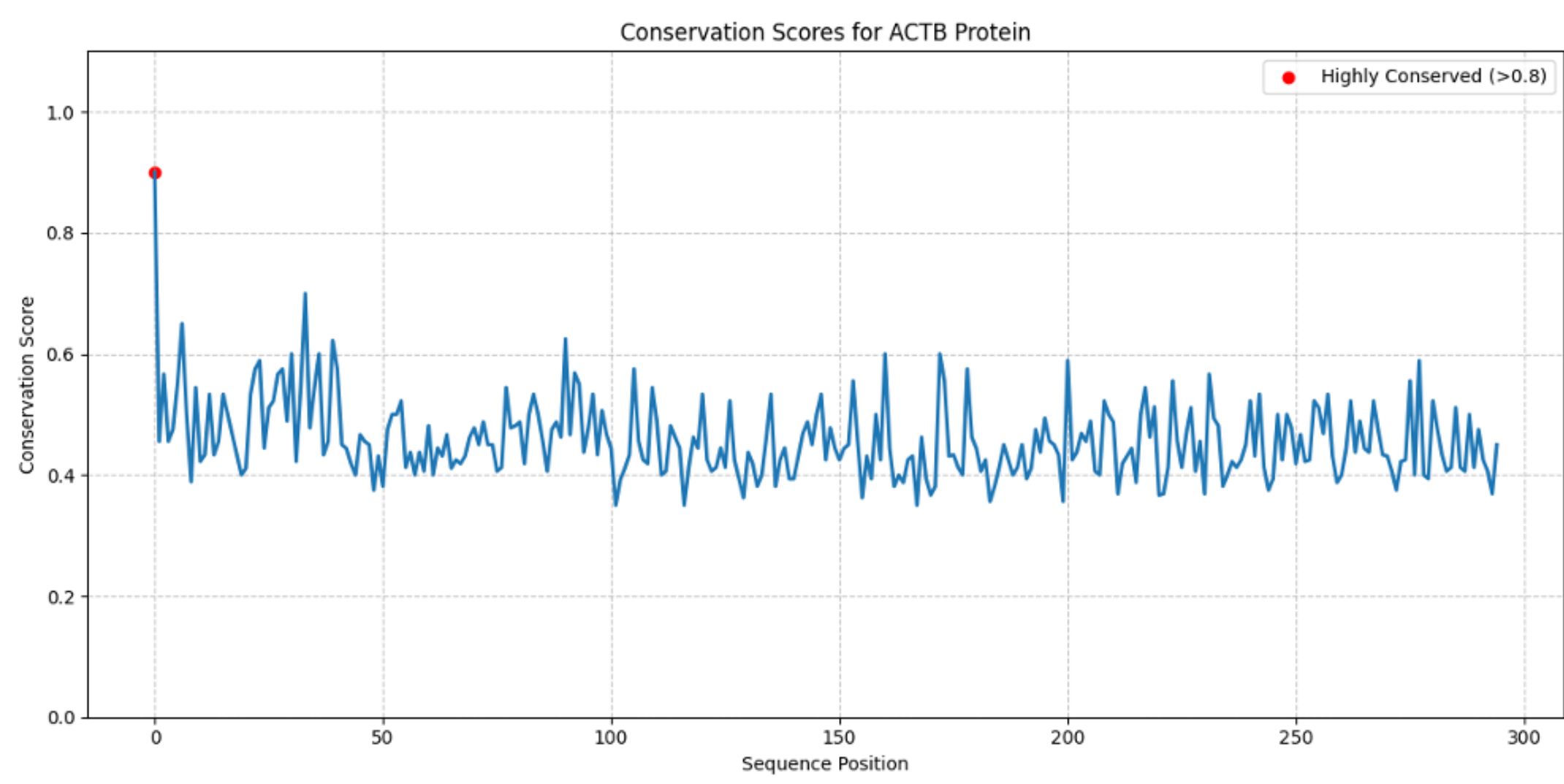


Figure: Conservation scores at all alignment positions of the MSA for protein sequences (ACTB gene).

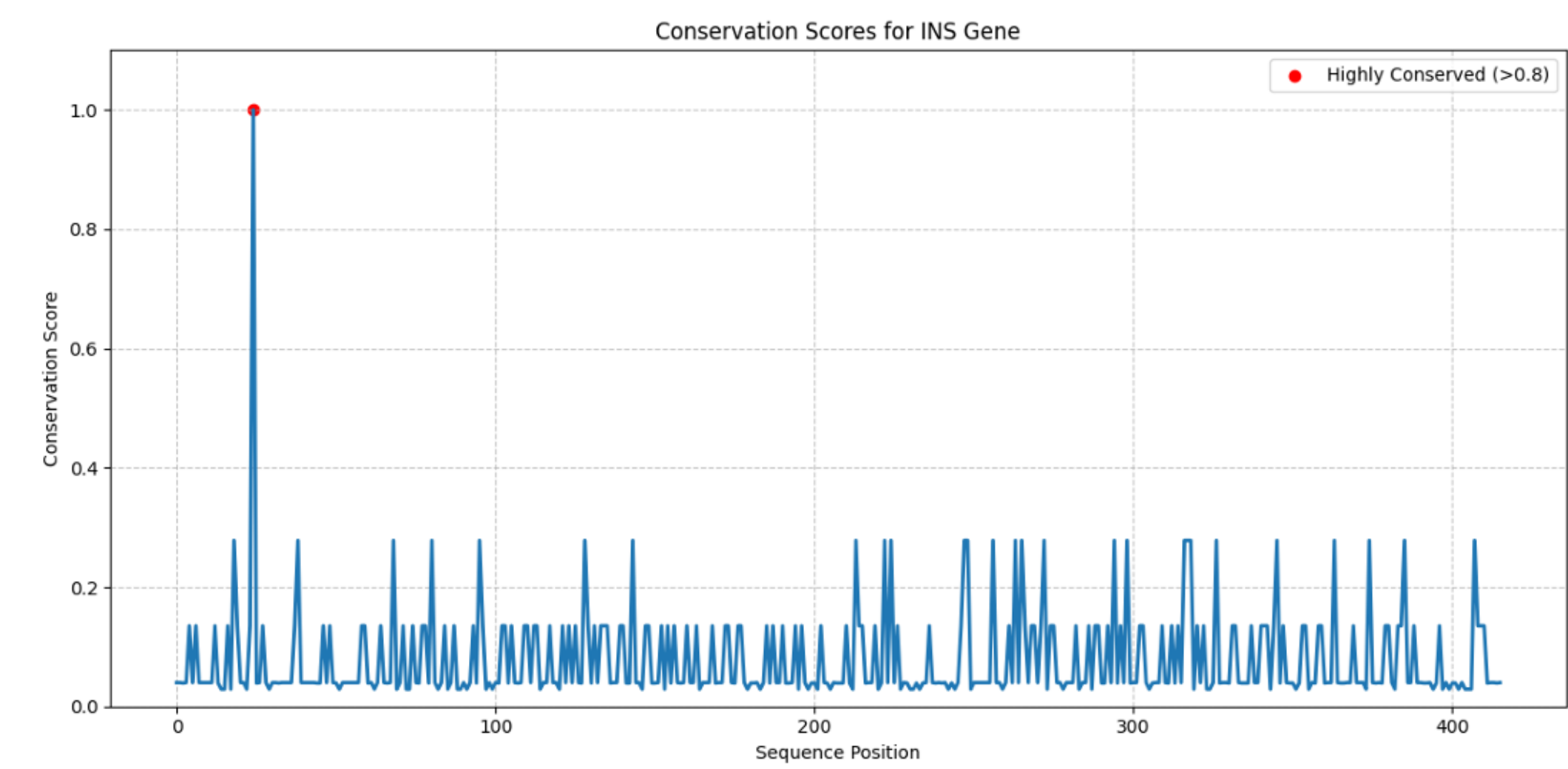


Figure: Conservation scores at all alignment positions of the MSA for nucleotide sequences (INS gene).

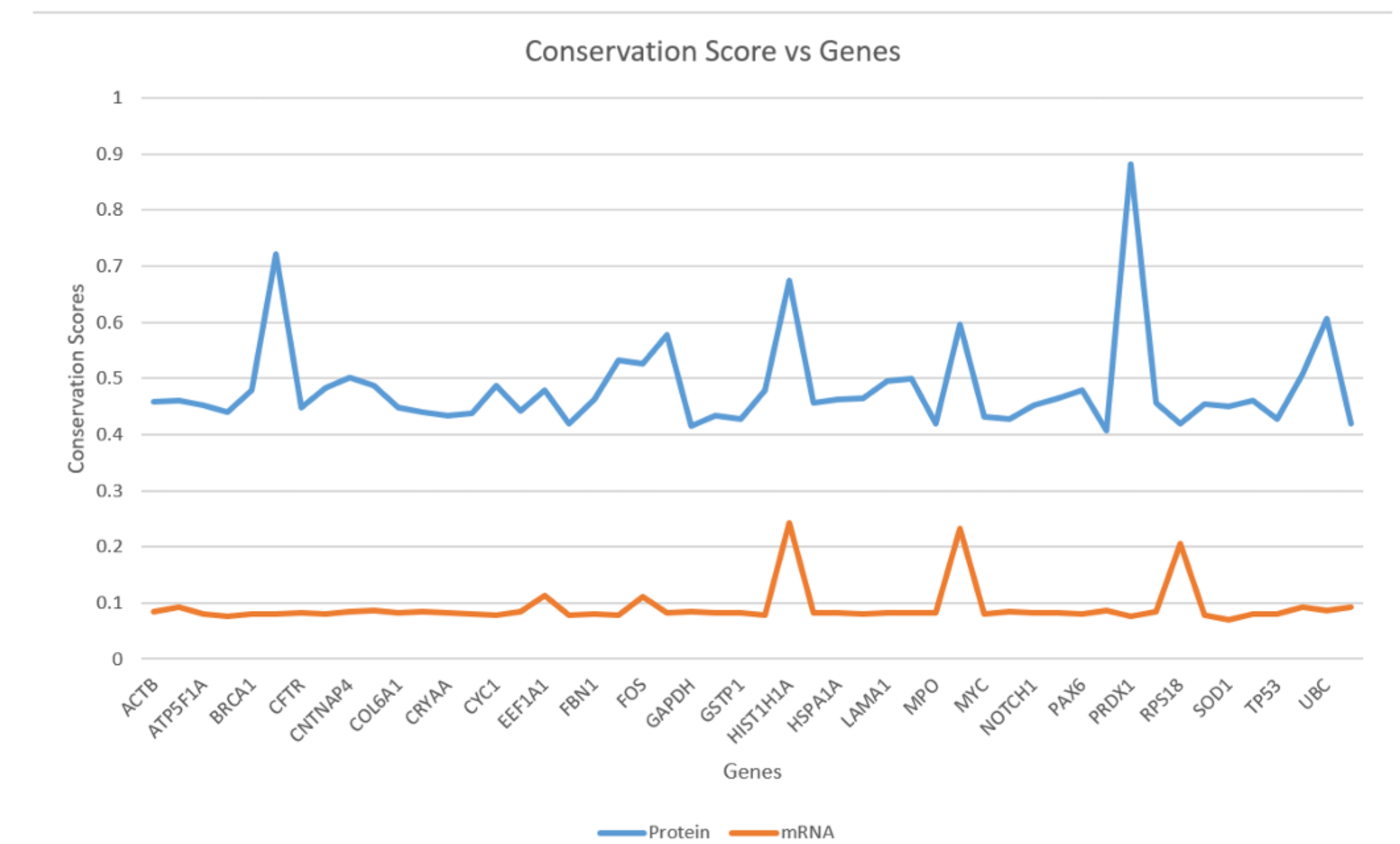


Figure: Mean conservation scores for nucleotide and protein sequences.

## Results 2

- Protein sequences were uploaded to AlphaFold for structure prediction.
- Scores were derived from spatial variations in the predicted structures.
- Mean conservation scores were used for further analysis.

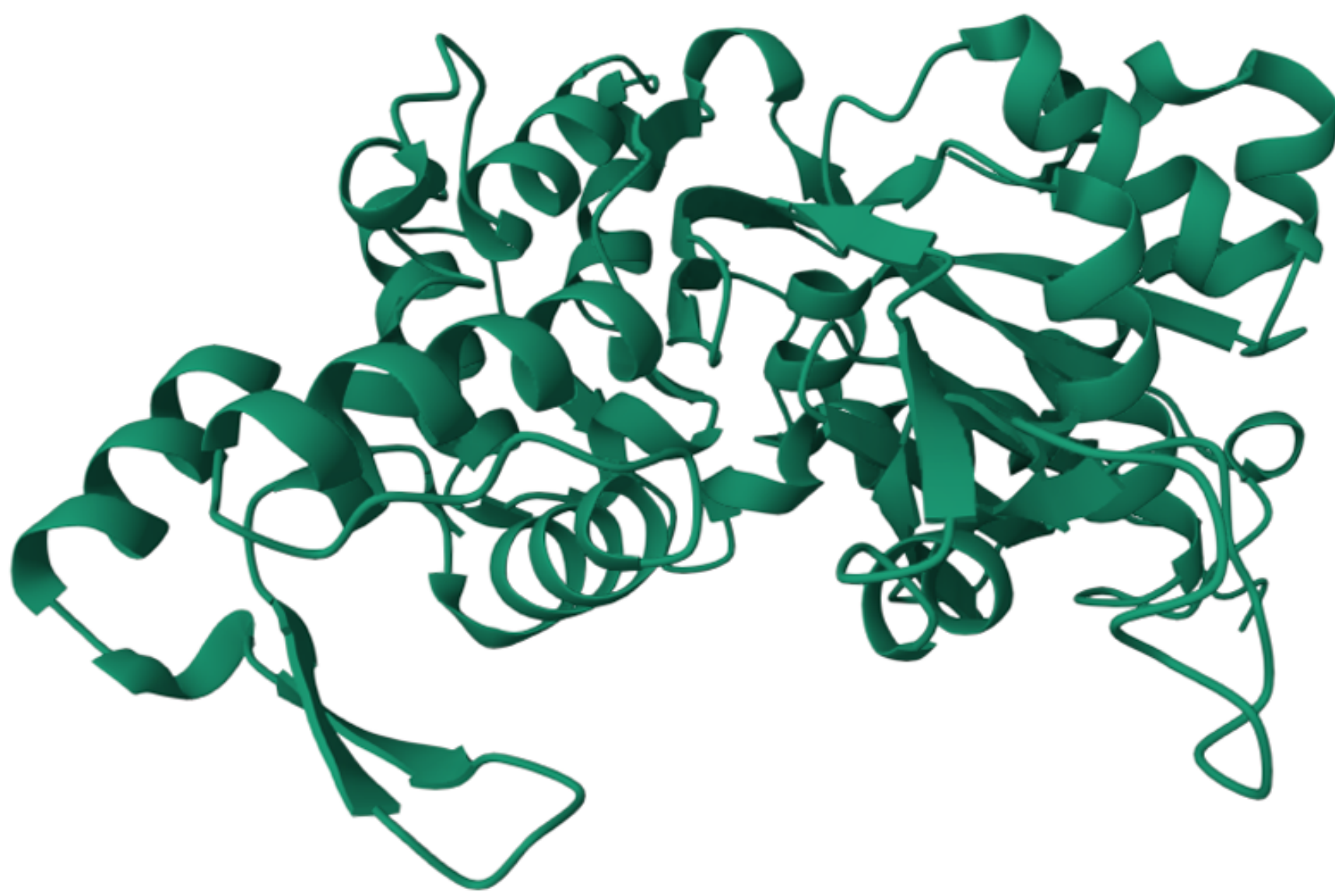


Figure: Predicted structure for *Danio rerio* ACTB protein sequence.

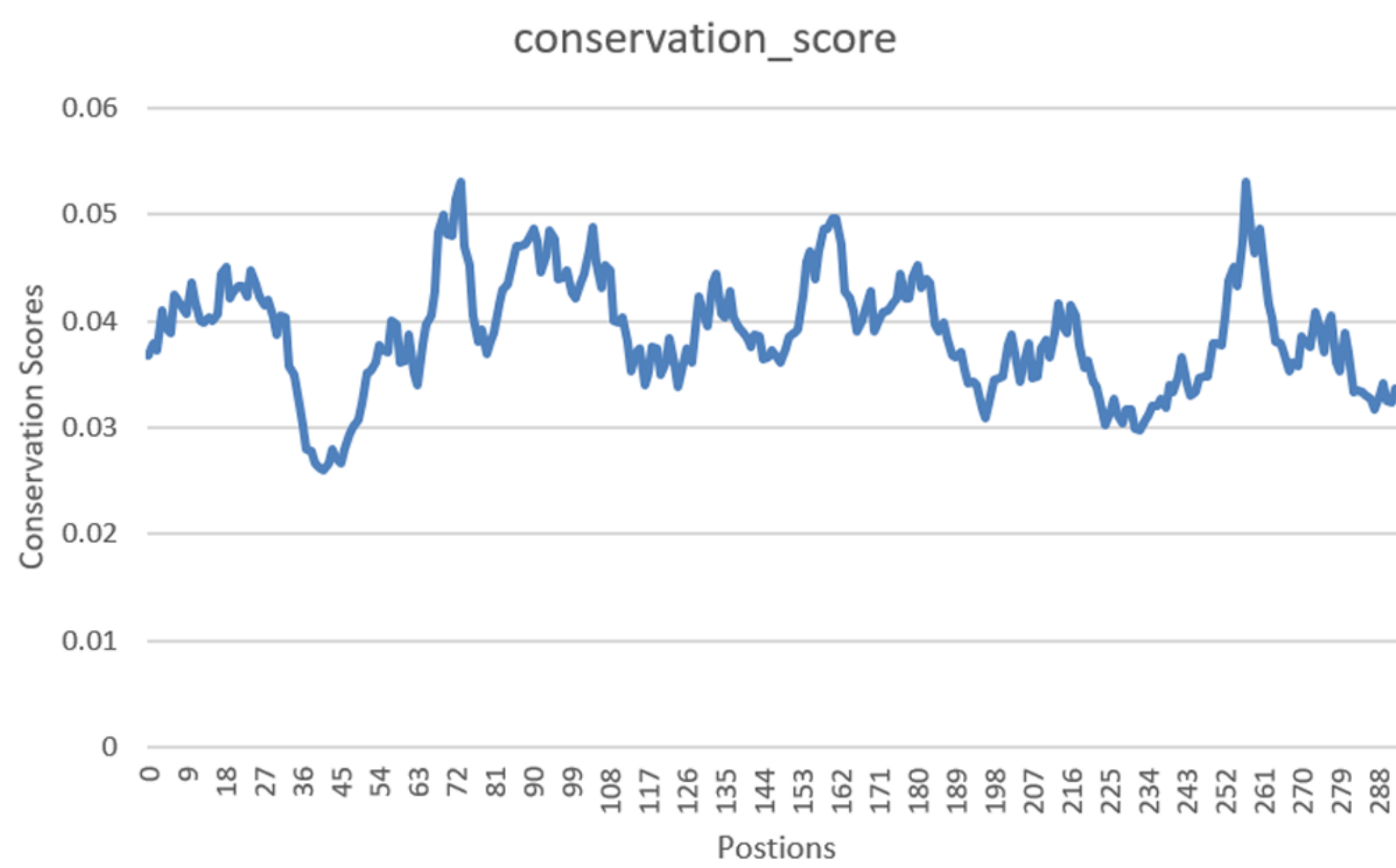


Figure: Conservation scores for structural elements at each position (ACTB gene).

## Results 3

- Protein, nucleotide, and structural conservation are positively correlated.
- Correlation between protein and structure is stronger than between structure and nucleotide.
- This aligns with expectations, as protein sequences are the primary basis for predicting structure.

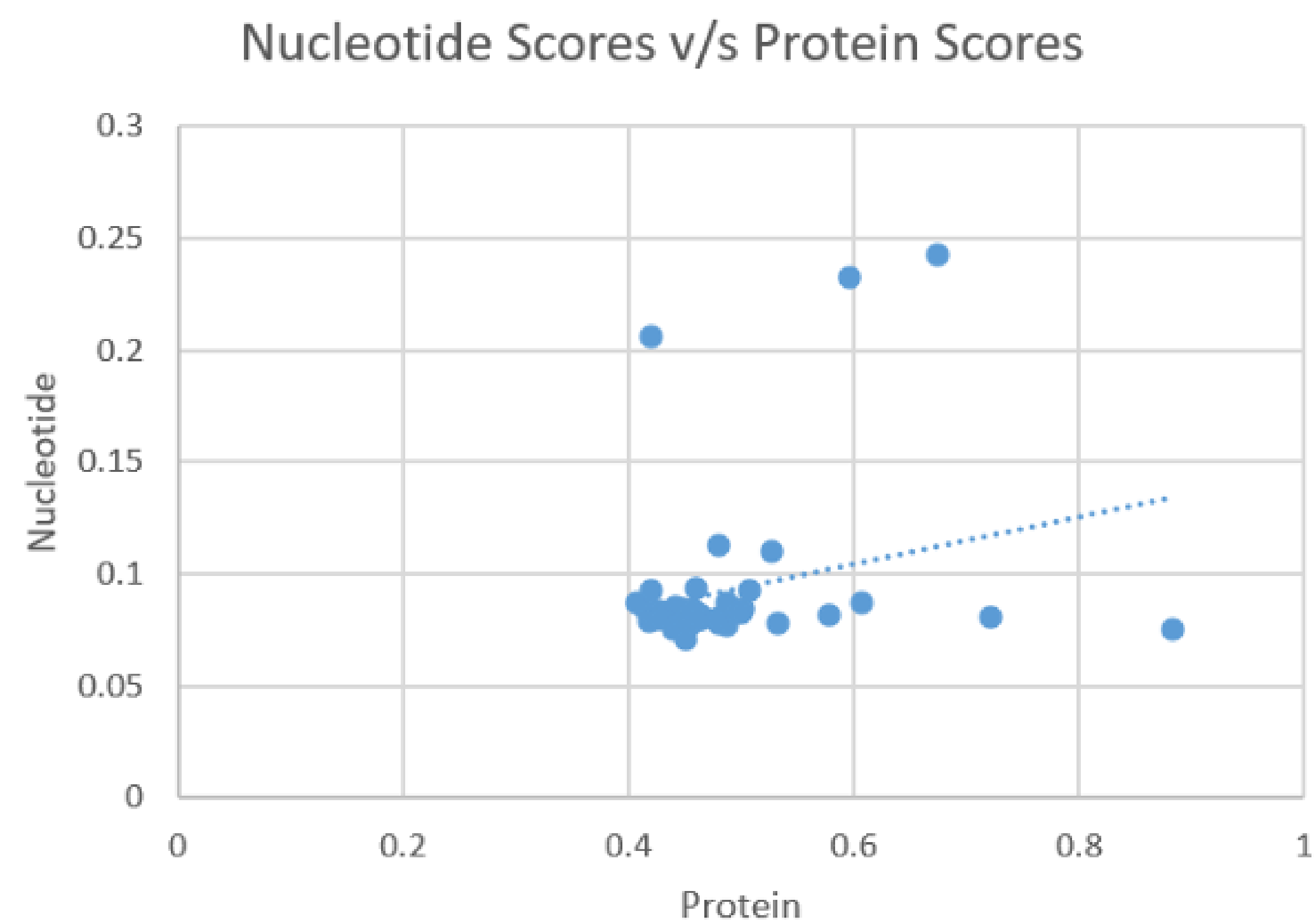


Figure: Correlation between conservation scores of nucleotide and protein.

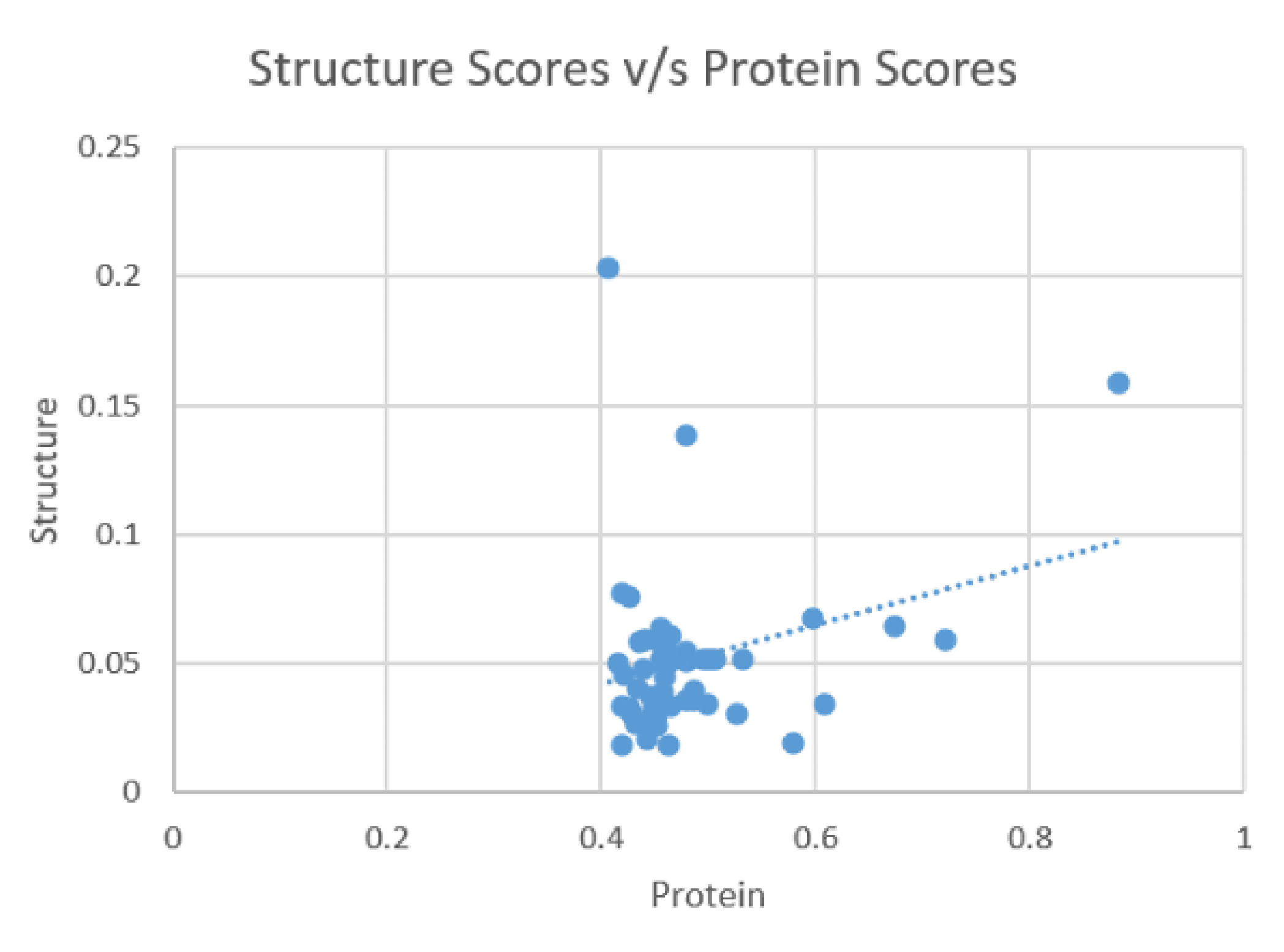


Figure: Correlation between conservation scores of structure and protein.

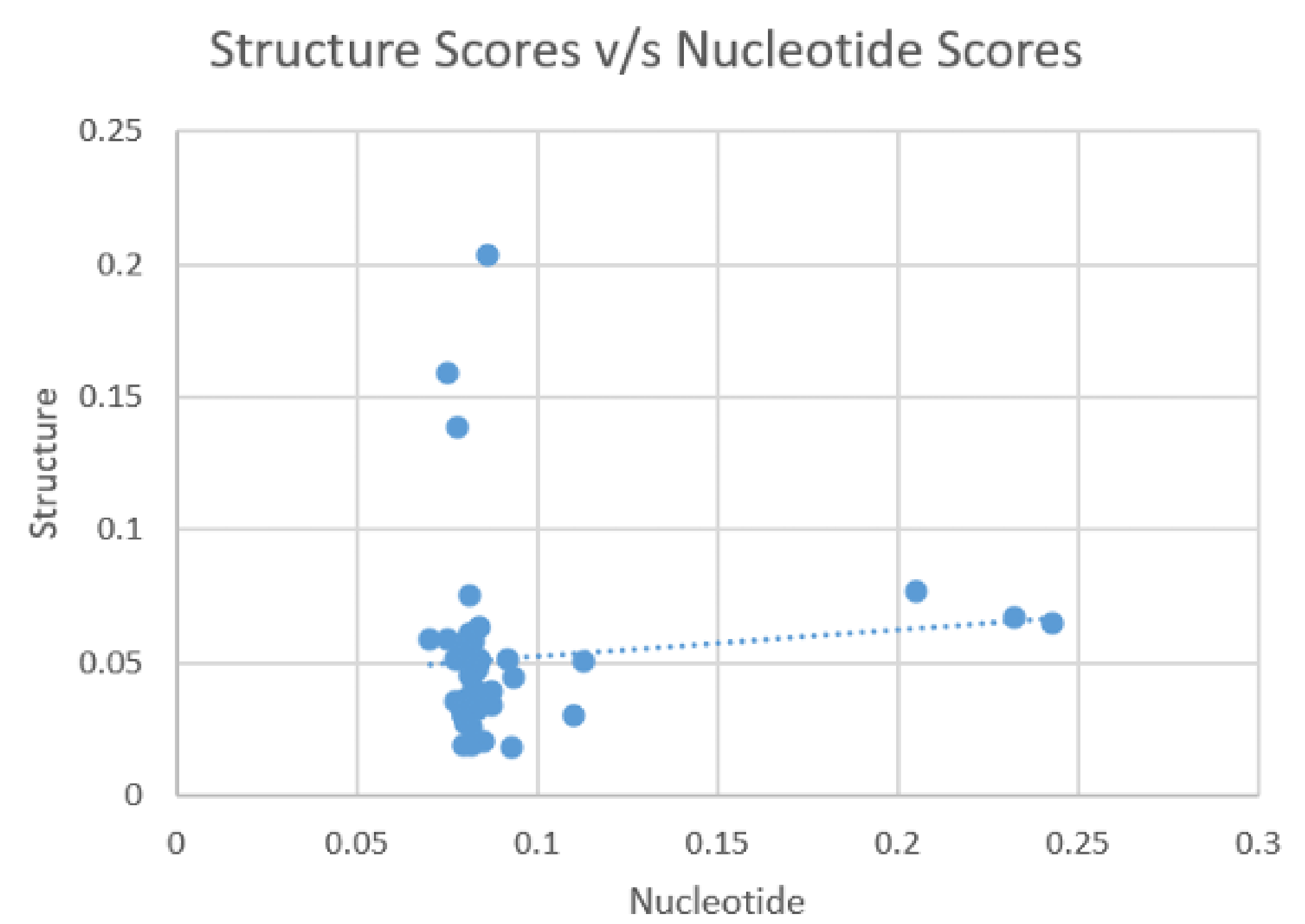


Figure: Correlation between conservation scores of structure and nucleotide.

	Protein	mRNA	Structure
Protein	1		
mRNA	0.25816	1	
Structure	0.2902	0.105	1

Figure: Correlation summary: protein, nucleotide, and structure.

## Discussion

- Blosum62 should be used for sequences with similarities less than 62% which is followed in almost all the cases.
- The model prediction comes with an error of around 2 percent which is minor and may result in false results.
- Structure is more related to protein sequence than mRNA (Nucleotide).
- During structure prediction some sequences could not predict a structure so that case the conservation score is taken to be mean of the rest.

