# Analysis of Ab Initio Protein Structure Prediction Methods

**Maytha Alshammari[1], Jing He[1]**

[1]*Department of Computer Science, Old Dominion University, Norfolk, VA*

## Abstract

Protein structure prediction produces atomic models of three-dimensional structure of a protein from its amino acid sequence. Understanding the function mechanism of proteins requires knowledge of three-dimensional structures. When developing new enzymes and drugs, it's essential to understand the structure of the target protein. In this study, we analyze models predicted using two ab initio protein structure prediction methods, trRosetta and Quark. A set of thirty protein chains was used to evaluate the effectiveness of the methods. The thirty chains were collected from Protein Data Bank (June – November, 2020). The length and the relative position of the predicted secondary structures were examined. We found that the accuracy of models obtained from trRosetta and Quark is good (TM score 0.358 - 0.969). However, in some cases, the methods were not able to accurately predict the relative location of the secondary structures which might affect the overall folding relationship among secondary structures.

## Method

Thirty atomic structures of protein chains were collected from Protein Data Bank (June - November, 2020). The three-dimensional structures of the thirty cases were predicted from the FASTA sequence using two ab initio protein structure prediction methods, trRosetta and Quark. Then, the TM-Score was calculated to measure the similarity between the protein atomic structure and the predicted model. Finally, the length and relative position of the predicted secondary structure were examined.

## Data

| Protein Name | Chain ID | Seq_Len[a]/PDB_Len[b] /SS[c] | Res[d] | TM-Score (trRosetta/Quark) |
|---|---|---|---|---|
| 7JZU | A | 61/55/(3H) | 3.1 Å | 0.50256/0.79155 |
| 7JZM | A | 106/64/(3H) | 3.5 Å | 0.60077/0.88977 |
| 7KDT | A | 500/467/(24H) | 3.5 Å | 0.48621 |
| 7D7R | A,B | 842/579/(15H,12BS) | 4 Å | 0.71887 |
| 7D0I | B, D, F, H | 710/424/(15H, 4BS) | 3 Å | 0.60041 |
| 7ANZ | A,B | 502/410/(14H,13BS) | 3.6 Å | 0.84789 |
| 7ANZ | C | 871/618/(25H, 2BS) | 3.6 Å | 0.68932 |
| 7ANZ | D | 785/613/(22H, 5BS) | 3.6 Å | 0.69662 |
| 7K1W | F | 918/462/(8H, 17BS) | 5.1 Å | 0.79979 |
| 7K22 | L1-L6 | 109/98/(2H, 7BS) | 3.2 Å | 0.86752/0.96950 |
| 7K22 | H1-H5 | 117/99/(6H, 2BS) | 3.2 Å | 0.82242/0.90681 |
| 7K2V | P | 264/264/(6H,10BS) | 6.6 Å | 0.78310 |
| 7K2V | A | 453/451/(15H,11BS) | 6.6 Å | 0.48760 |
| 7K43 | C,F,H | 125/122/(5H,11BS) | 2.6 Å | 0.86828/0.92077 |
| 7K45 | H | 123/122/(4H, 13BS) | 3.7 Å | 0.88867/0.90931 |
| 7K45 | L | 109/103/(2H,11BS) | 3.7 Å | 0.89111/0.93388 |
| 6ZME | CE | 223/ 117/(4H) | 3 Å | 0.3926 |
| 6ZNL | O,o | 186/179/(4H) | 3.8 Å | 0.42899/0.35856 |
| 6ZNL | Y | 467/410/(6H,15BS) | 3.8 Å | 0.41467 |
| 6LVC | A, C | 775/762/(17H,20BS) | 3.0 Å | 0.79235 |
| 6LVC | B, D | 132/124/(5H, 3BS) | 3.0 Å | 0.78033 |
| 6S6B | A,B,C | 155/153/(6H) | 2.75 Å | 0.73359 |
| 6SD4 | A… h | 560/151/(3H, 7BS) | 2.80 Å | 0.69233 |
| 6VQX | A | 100/97/(5H) | 3.15 Å | 0.67602 |
| 6W8U | A,….,o | 142/140/(1H, 8 BS) | 3.8 Å | 0.56965 |
| 6WB9 | 0 | 205/117/(9 BS) | 3.0 Å | 0.46489 |
| 6WB9 | 2 | 292/(13H) | 3.0 Å | 0.65601 |
| 6WR4 | A, B, C | 839/470/(20H, 4BS) | 2.9 Å | 0.59077 |
| 6X8M | A | 314/280/(7H, 15BS) | 2.2 Å | 0.75356 |
| 6XDC | A, B | 284/193/(5H, 8BS) | 2.9 Å | 0.38691 |

**Table 1: The 30 Protein Chains**. [a]The protein sequence length. [b]The protein atomic structure length. [c]The Number of the protein secondary structure. H indicates number of α-helices and BS indicates the number β-Strands. [d]Resolution (electron density).

## Results

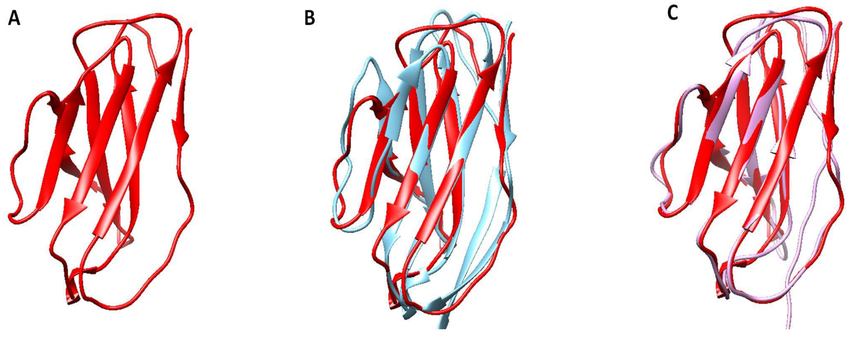


**Figure 1: An example of a well prediction of the relative position of the predicted secondary structure**. (A) The Red ribbon is the atomic structure of 7K22 chain L. (B) The Blue ribbon is the trRosetta Predicted model. TM-Score is 0.86752.(C) The Purple ribbon is the Quark Predicted model. TM-Score is 0.96950.

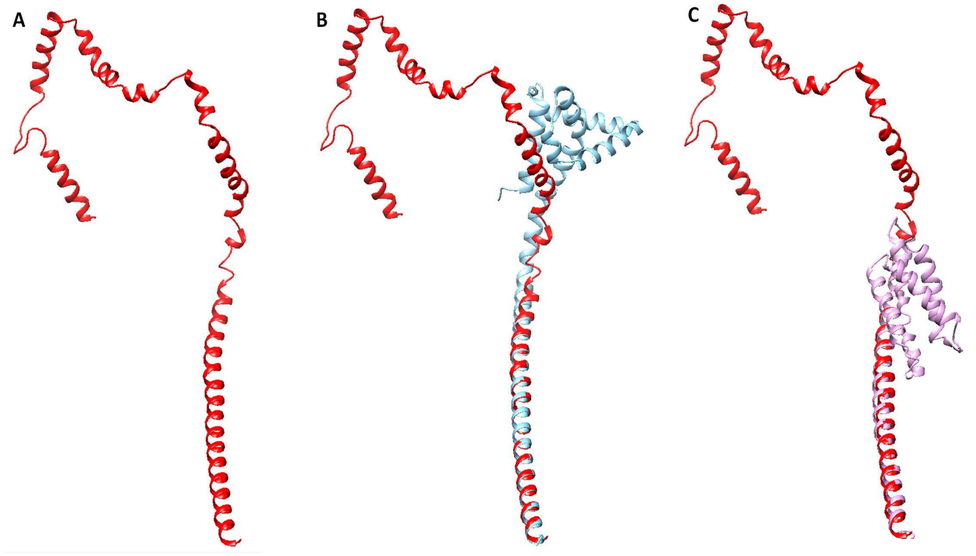


**Figure 2: An example of Inaccurate prediction of the relative position of the predicted secondary structure**. (A) The Red ribbon is the atomic structure of 6ZNL chain O. (B) The Blue ribbon is the trRosetta Predicted model. TM-Score is 0.42899.(C) The Purple ribbon is the Quark Predicted model. TM-Score is 0.3585.

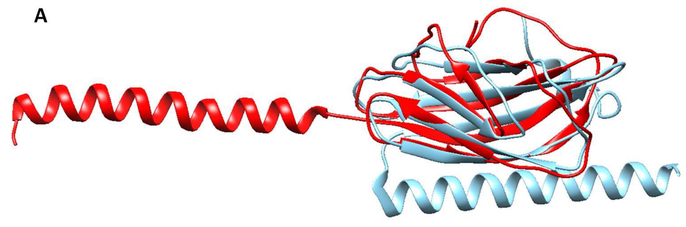


**Figure 3: An example of a Wrong Angel among the predicted Secondary Structures**. (A) The Red ribbon is the atomic structure of 6W8U chain A. The Blue ribbon is the trRosetta Predicted model. TM-Score is 0.56965. The predicted model aligned with the atomic structure at the Beta-Sheet region but not aligned well with the Helix region.

## Conclusion

We examined models predicted using two ab initio protein structure prediction methods, trRosetta and Quark. We found that the methods are performing well in predicting the structure of a protein. However, the methods in some cases were not able to accurately predict the relative position between secondary structures which might affect the overall folding relationship among secondary structures.

## Acknowledgements

The work in this poster was supported by NIH R01-GM062968 and a scholarship to M.A. by the Government of Saudi Arabia.