Trpcage Folding Simulation Based on Co-Translational Protein Folding

[1]Ankita Mehta, [2]Dr. Navneet Batra, [3]Dr. Rupinderjeet kaur

[1]Research Scholar, [1,2]Head, Associate Professor, [3]Associate Professor
[1]Institute Of Forensic Science, Panjab University, Chandigarh -160014
[2]Department of Biotechnology, GGDSD College, Sec-32, Chandigarh -160030
[3]Department of Biotechnology, DAV College, Sec-10, Chandigarh - 160011

Abstract:
A novel algorithm has been formulated to enable an enhanced and more accurate simulation of Trpcage folding. This computational method primarily draws upon the principles of cotranslational protein folding. In biological systems, polypeptides undergo sequential synthesis with translation rates varying based on codon speed. Studies have demonstrated that helices and sheets typically fold in a timeframe measured within lower milliseconds. As such, certain proteins exhibit faster-folding than elongation capabilities which suggests a high probability for these resulting chains to assume secondary or tertiary structures during cotranslational processes. The examination of the ribosomal exit tunnel likewise demonstrates that peptides can navigate through it in an α-helical structure. Consequently, this study suggests a molecular dynamics simulation algorithm predicated on including one amino acid residue successively during simulations. In biological systems, proteins synthesis occurs at ribosomes by incorporating single amino acids consecutively into sequences. The launch set-up for such simulations need not comprise fully polypeptide chains but could initiate with minimal segments. Commencing the folding simulation with a minimal segment of amino acids, followed by sequential supplementation of individual amino acids during each ensuing simulation cycle strengthen our need to consider interatomic connections between newly incorporated amino acids and their existing counterparts in the protein chain. Essentially, each atom within a protein impacts every other atom. Therefore, during subsequent procedures, we continuously incorporate an amino acid; this allows the newly introduced amino acid to comprehensively examine its interactions with the remaining amino acids in the protein. Such method could be employed for both *ab initio* prediction of protein structures and for designing new proteins. In our study, we analyzed Trpcage simulations and determined that all-atoms root-mean-square deviation is 4.621 Å while backbone root-mean-square deviation is 3.444 Å when compared to original NMR solved structure submitted in Protein Data Bank.

IndexTerms – Algorithm, Trp-Cage, Protein folding, Molecular Dynamics Simulation, RMSD

I. INTRODUCTION

Due to advancements in genome sequencing projects, protein sequencing has become highly accessible resulting in millions of proteins being sequenced with many more still underway[1-4]. However, the sequence does not provide extensive information regarding the functional attributes of a protein [5]. To thoroughly comprehend a protein's function necessitates knowledge about its structure, which progress is considerably slower [6]. The following facts exemplify this reality further. Uniprot stands as a valuable resource of information on protein sequence and function. As per the UniProtKB/TrEMBL Release 2023_03, made available on June 28th, 2023, there are entries for approximately 248272897 types of sequences [7]. Furthermore, the US Research Collaboratory for Structural Bioinformatics Protein Data Bank, supplies experimentally confirmed three-dimensional structures of various biomolecules, with a total count at roughly around 208702 distinct PDB formations. [8]

The comparison between protein sequences and structures reveals that the latter equates to roughly 1/1200 of the former. This punctuates a necessity for further developing computational methodologies with an emphasis on predicting proteins. Additionally, deciphering protein folding is vital when considering innovative custom-tailored protein designs for specific applications. A clear depiction of this challenge was captured by Science Magazine's comprehensive list of top 125 unresolved scientific questions in which they posed - "Can we predict how proteins will fold?" They highlighted its vast complexity emphasizing on the strikingly short time span (ten microseconds) it takes a single protein entity to select one pathway out from nearly limitless potential pathways for folding [9].

Polypeptides are produced in sequence, and the pace of this translation can vary based on codon velocity [10-12]. For instance, a protein translation rate around 0.05 s/codon has been observed in Escherichia coli [10,11]. Meanwhile helices and sheets have demonstrated folding capabilities within the low millisecond range, thereby exhibiting that some proteins may fold at faster rates than their elongation process [13,14]. This would suggest that these resulting chains could take up cotranslational secondary or tertiary structures during formation. Closer examination of the ribosomal exit tunnel revealed peptides traversing it while displaying an α-helical conformation [15-18]. Further scrutiny indicated that as they travel through this channel α-helical arrangements can gain entropic stability[16-21].

Cotranslational folding is a process in which proteins start to fold while still being synthesized on the ribosome. This phenomenon has been observed in various proteins, including Semliki Forest Virus Protein, α-globins, and the bacterial luciferase αβ-heterodimer. For example, SFVP gains its enzyme activity before complete synthesis of the polyprotein precursor [22-26], and α-globins exhibit specific heme-binding activity for truncated nascent chains bound to ribosomes [27]. In the case of the luciferase heterodimer, formation occurs more rapidly when one monomer is translated in the presence of an already folded partner compared to refolding

from a denatured state [28,29] .These examples highlight how cotranslational folding plays a significant role in protein structure formation during translation.

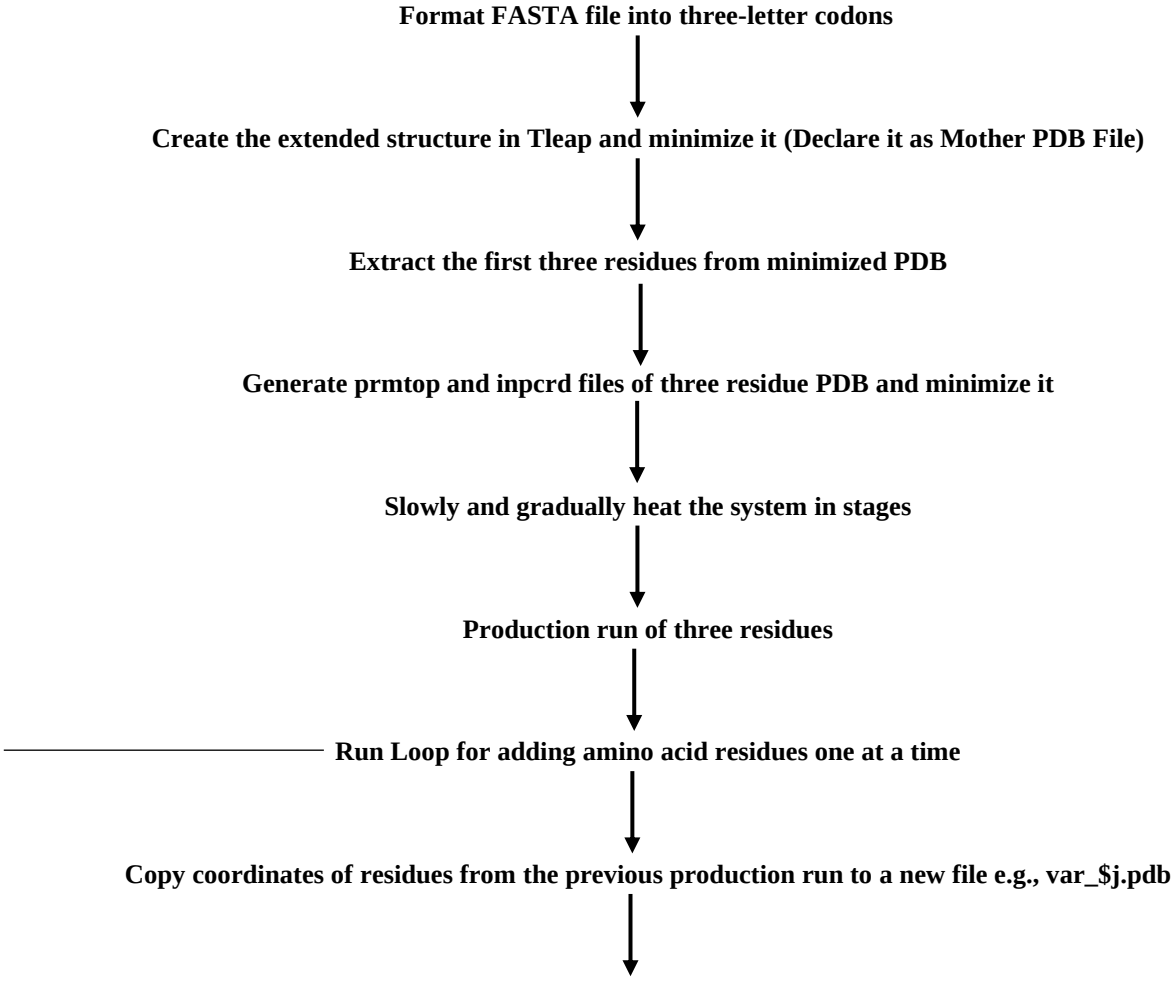
Previous studies by Jonathan J. Ellis et al. have investigated the role of translation directionality in predicting protein folding [30]. In their research, they integrated aspects of cotranslational folding into a protein structure prediction algorithm using the Rosetta program. The results showed that for 94% of the proteins analyzed, tracking translation from N-terminus to C-terminus yielded better predictions compared to tracking the reverse direction. This trend was especially prominent in proteins that exhibited stronger evidence of co-translational folding. Computational models further support this finding, suggesting that such folding favors local contacts in intermediate and terminal folds [31,32].

Recent research has focused on the impact of energy barriers in cotranslational models, revealing that the lowest energy state may not always correspond to the sequentially folded protein [33-37]. Computer simulations have shown that nascent chains can adopt partial structures resembling segments of the complete protein [38-41]. However, alternative lattice studies propose a different perspective, suggesting that nascent peptides can remain mostly unstructured until near completion of synthesis [42]. This view considers interactions involving the C terminus and may not be applicable to all proteins. Furthermore, lattice models suggest separate pathways for cotranslational folding and refolding processes [43].

In this study, we introduce a novel molecular dynamics algorithm that can be applied to solve protein structure prediction problems. Our algorithm is based on the concept of cotranslational folding and operates by sequentially adding one amino acid at a time in each iteration of the molecular dynamics simulation. This approach is not exclusively limited to protein folding challenges, as it has broader applications in general molecular dynamics simulations. The algorithm holds potential for accurate ab initio prediction of protein structures and facilitates the design of innovative proteins with various uses such as therapeutics or forensic analysis. The strength of our project lies in its originality and well-founded rationale behind the proposed algorithm.

II. METHODOLOGY

The Trpcage peptide has been widely used as a standard system for protein folding simulations. Starting with the FASTA sequence {NLYIQWLKDGGPSSGRPPPS} of Trpcage, an elongated linear structure was generated using Amber version 18. The calculations were performed on supercomputing resources at IIT Delhi, taking advantage of multiple nodes to run molecular dynamics simulations with the Sander module. Each simulation was conducted under canonical ensemble conditions and included solvation effects through the Generalized Born model. These molecular dynamics simulations aimed to assess the stability and accuracy of this newly developed algorithm to incorporate cotranslational folding pathway. To ensure consistency in results, all simulations used the same parameters throughout. Production runs lasted for 50 ns each, and from these runs, we selected a representative structure based on its minimum potential energy for further analysis purposes. The flowchart of methodology is given in figure 1.



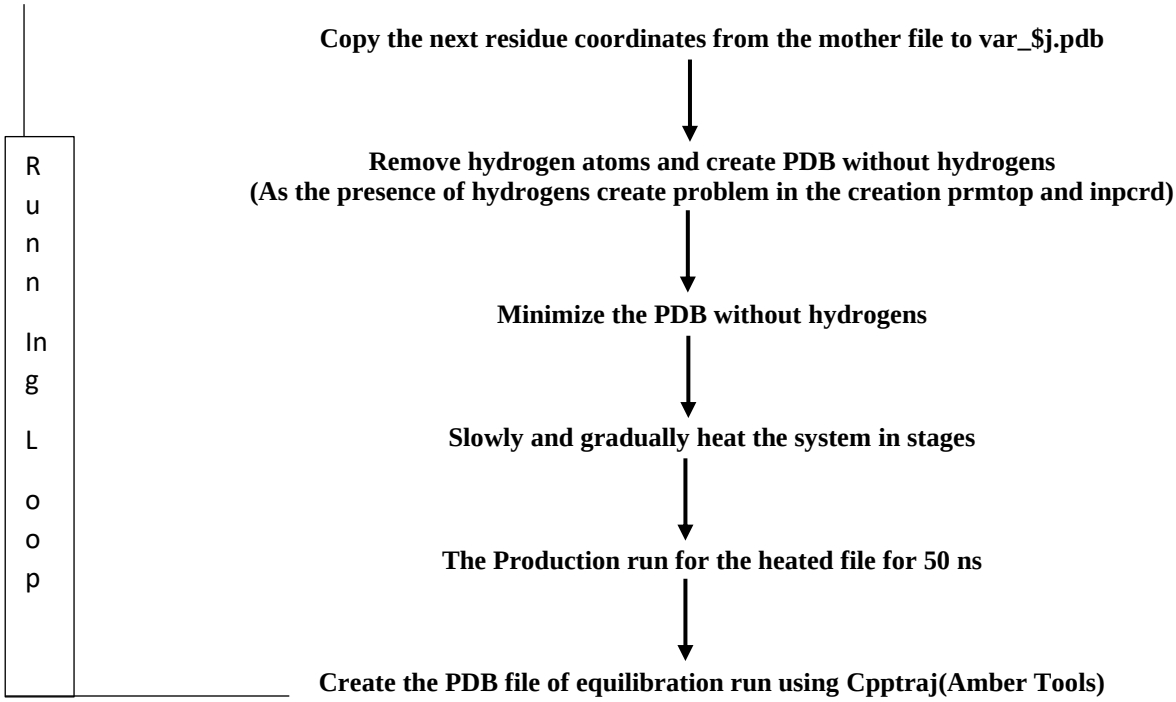


Figure 1. Flow chart

III. RESULT

The obtained pdb file from the Trpcage folding protocol was compared to model 1 of the NMR structure in the PDB database for RMSD calculations. Figure 2 presents a snapshot of the Minimum Potential Energy Structure acquired through one-at-a-time amino acid addition. Whereas figure 3 represents the NMR model of Trpcage which is submitted in Protein Data Bank as 1L2Y.PDB. Minimum Potential energy structure was superimposed on 1 L2Y as shown in figure 4 .When overlaid with 1L2Y, the RMSD values were determined as 4.621 Å for all atoms and 3.444 Å for just the backbone. To ensure temperature consistency throughout simulation, a temperature plot was created displaying constant temperatures across time , which is shown in figure 5. A plot of the energy curve was generated to demonstrate the variations in potential, kinetic, and total energy throughout the simulation, as shown in figure 6.At any particular instance sum of potential and kinetic energy should remain equal to total energy. Additionally, a graph depicting the root-mean-square deviation i.e., RMSD over time was also plotted as shown in figure 7.

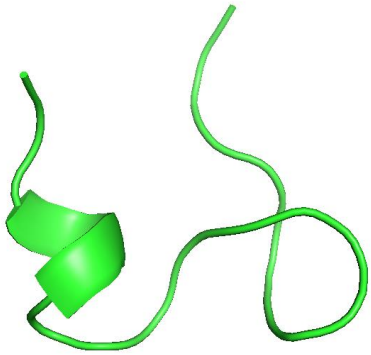


Figure 2. Minimum Potential Energy Structure extracted from Production run.

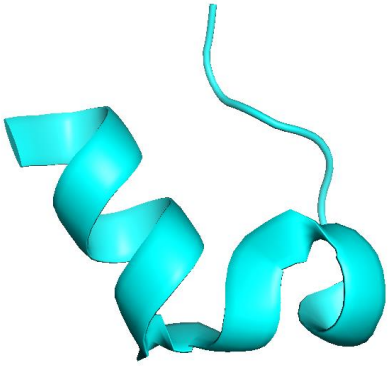


Figure 3. 1L2Y- NMR solved structure of Trpcage.

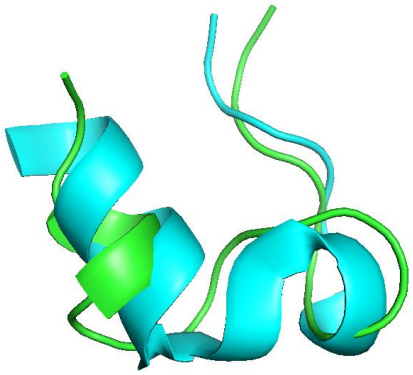


Figure 4. Superimposed 1L2Y with minimum potential energy PDB obtained through one-at-a-time amino acid addition.

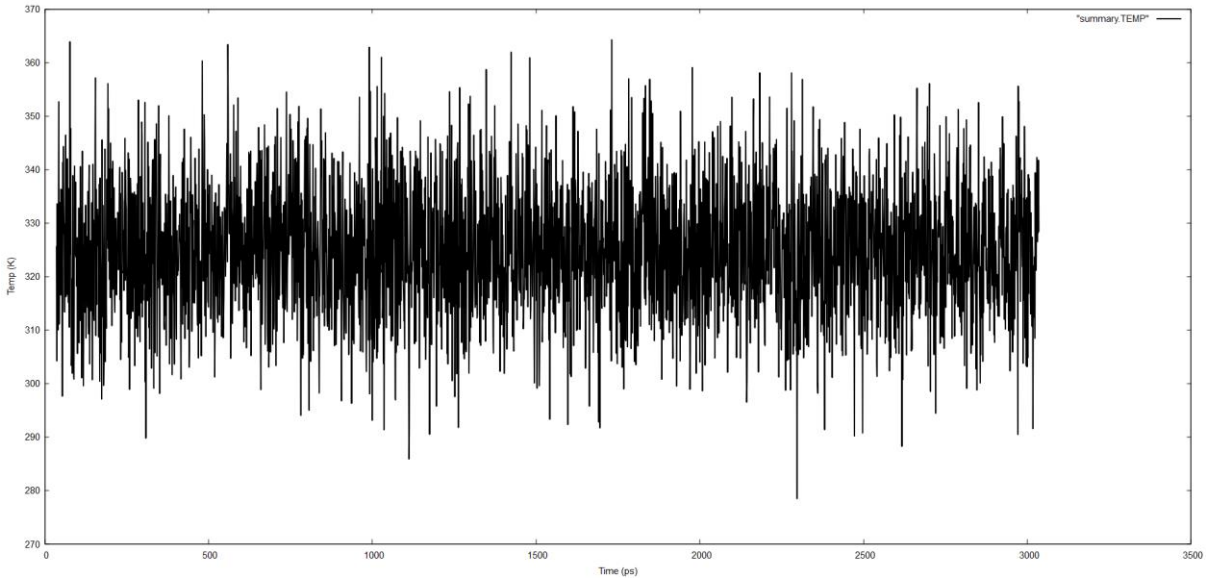


Figure 5. Temperature Plot of Production run.

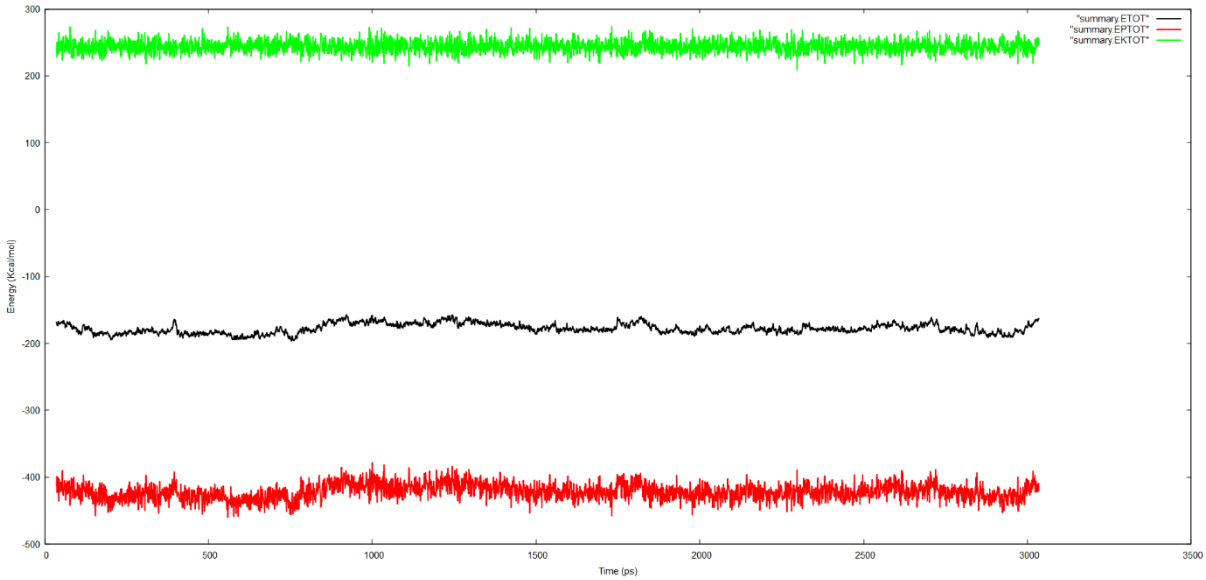


Figure 6. Energy Plot of Production run.

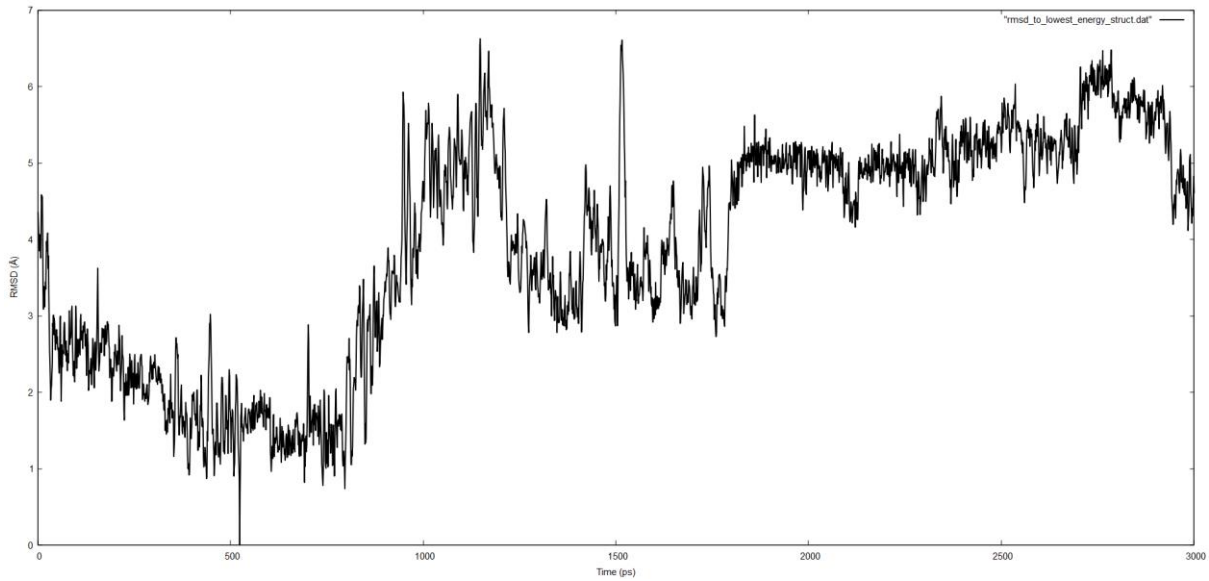


Figure 7. RMSD vs Time(ps).

IV. DISCUSSION

This computational approach is based on the principles of cotranslational protein folding. The method begins with a minimal segment of amino acids and gradually adds individual amino acids in each subsequent iteration, considering the interatomic connections between the newly incorporated amino acids and those already present in the protein chain. This allows for a thorough exploration of interactions between each atom within the protein throughout the procedure.

In summary, the current study explores the idea of adding amino acids one-at-a-time in protein folding simulations, using the Trpcage fold as a test case. The results show that this approach could partially reproduce the native structure of the Trpcage protein with a moderate RMSD value. Additionally, it is suggested that further validation of this approach should be conducted on other small proteins, considering factors such as protein length and secondary structure. Furthermore, future improvements can include testing the effectiveness of this one-at-a-time approach in energy minimization protocols. Further testing and validation are necessary to fully understand the effectiveness of this approach in protein folding simulations.

V. REFERENCES

1. 1.Pareek, Chandra Shekhar, Rafal Smoczynski, and Andrzej Tretyn. "Sequencing technologies and genome sequencing." Journal of applied genetics 52 (2011): 413-435.
2. 2.Lewin, Harris A., Denis M. Larkin, Joan Pontius, and Stephen J. O'Brien. "Every genome sequence needs a good map." Genome research 19, no. 11 (2009): 1925-1928.
3. 3.Ng, Pauline C., and Ewen F. Kirkness. "Whole genome sequencing." Genetic variation: Methods and protocols (2010): 215-226.
4. 4.Venter, J. Craig, Hamilton O. Smith, and Leroy Hood. "A new strategy for genome sequencing." Nature 381, no. 6581 (1996): 364-366.
5. Galperin, Michael Y., and Eugene V. Koonin. "From complete genome sequence to 'complete'understanding?." Trends in biotechnology 28, no. 8 (2010): 398-406.
6. 6.Rigden, Daniel J., ed. From Protein Structure to Function with Bioinformatics. Springer, 2017.
7. 7.https://www.ebi.ac.uk/uniprot/TrEMBLstats
8. 8.https://www.rcsb.org/
9. 9.So much more to know. . . Science. 2005;309:78–102. DOI: 10.1126/science.309.5731.78b
10. Pedersen, Steen. "Escherichia coli ribosomes translate in vivo with variable rate." The EMBO journal 3, no. 12 (1984): 2895-2898.
11. Wilson, Kevin S., and Harry F. Noller. "Molecular movement inside the translational engine." Cell 92, no. 3 (1998): 337-349.
12. 12 Abel, Kenton, and Frances Jurnak. "A complex profile of protein elongation: translating chemical energy into molecular movement." Structure 4, no. 3 (1996): 229-238.
13. Briggs, Martha S., and Heinrich Roder. "Early hydrogen-bonding events in the folding reaction of ubiquitin." Proceedings of the National Academy of Sciences 89, no. 6 (1992): 2017-2021.
14. Lu, Jirong, and Frederick W. Dahlquist. "Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR." Biochemistry 31, no. 20 (1992): 4749-4756.
15. Lim, V. I., and A. S. Spirin. "Stereochemical analysis of ribosomal transpeptidation conformation of nascent peptide." Journal of molecular biology 188, no. 4 (1986): 565-574.
16. 16.Ziv, Guy, Gilad Haran, and D. Thirumalai. "Ribosome exit tunnel can entropically stabilize α-helices." Proceedings of the National Academy of Sciences 102, no. 52 (2005): 18956-18961.
17. Nissen, Poul, Jeffrey Hansen, Nenad Ban, Peter B. Moore, and Thomas A. Steitz. "The structural basis of ribosome activity in peptide bond synthesis." Science 289, no. 5481 (2000): 920-930.

18. Ban, Nenad, Poul Nissen, Jeffrey Hansen, Peter B. Moore, and Thomas A. Steitz. "The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution." Science 289, no. 5481 (2000): 905-920.
19. 19.Harms, Joerg, Frank Schluenzen, Raz Zarivach, Anat Bashan, Sharon Gat, Ilana Agmon, Heike Bartels, François Franceschi, and Ada Yonath. "High resolution structure of the large ribosomal subunit from a mesophilic eubacterium." Cell 107, no. 5 (2001): 679-688.
20. Milligan, R. A., and P. N. T. Unwin. "Location of exit channel for nascent protein in 80S ribosome." Nature 319, no. 6055 (1986): 693-695.
21. Yonath, Ada, K. R. Leonard, and H. G. Wittmann. "A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction." Science 236, no. 4803 (1987): 813-816.
22. 22.Nicola, Anthony V., Wei Chen, and Ari Helenius. "Co-translational folding of an alphavirus capsid protein in the cytosol of living cells." Nature cell biology 1, no. 6 (1999): 341-345.
23. Strauss, James H., and Ellen G. Strauss. "The alphaviruses: gene expression, replication, and evolution." Microbiological reviews 58, no. 3 (1994): 491-562.
24. 24.Choi, Hok-Kin, Guoguang Lu, Sukyeong Lee, Gerd Wengler, and Michael G. Rossmann. "Structure of Semliki Forest virus core protein." Proteins: Structure, Function, and Bioinformatics 27, no. 3 (1997): 345-359.
25. Melancon, P. A. U. L., and H. E. N. R. I. K. Garoff. "Processing of the Semliki Forest virus structural polyprotein: role of the capsid protease." Journal of virology 61, no. 5 (1987): 1301-1309.
26. Aliperti, Giuseppe, and Milton J. Schlesinger. "Evidence for an autoprotease activity of Sindbis virus capsid protein." Virology 90, no. 2 (1978): 366-369.
27. Komar, Anton A., Aigar Kommer, Igor A. Krasheninnikov, and Alexander S. Spirin. "Cotranslational folding of globin." Journal of Biological Chemistry 272, no. 16 (1997): 10646-10651.
28. Fedorov, Alexey N., and Thomas O. Baldwin. "Process of biosynthetic protein folding determines the rapid formation of native structure." Journal of molecular biology 294, no. 2 (1999): 579-586.
29. Evans, Michael S., Thomas F. Clarke Iv, and Patricia L. Clark. "Conformations of co-translational folding intermediates." Protein and peptide letters 12, no. 2 (2005): 189-195.
30. Ellis, Jonathan J., Fabien PE Huard, Charlotte M. Deane, Sheenal Srivastava, and Graham R. Wood. "Directionality in protein fold prediction." BMC bioinformatics 11 (2010): 1-16.
31. Morrissey, M. P., Z. Ahmed, and E. I. Shakhnovich. "The role of cotranslation in protein folding: a lattice model study." Polymer 45, no. 2 (2004): 557-571.
32. Levinthal, Cyrus. "Are there pathways for protein folding?." Journal de chimie physique 65 (1968): 44-45.
33. Huard, Fabien PE, Charlotte M. Deane, and Graham R. Wood. "Modelling sequential protein folding under kinetic control." Bioinformatics 22, no. 14 (2006): e203-e210.
34. Anfinsen, Christian B. "Principles that govern the folding of protein chains." Science 181, no. 4096 (1973): 223-230.
35. Basharov, M. A. "Protein folding." Journal of cellular and molecular medicine 7, no. 3 (2003): 223-237.
36. Baskakov, Ilia V., Giuseppe Legname, Stanley B. Prusiner, and Fred E. Cohen. "Folding of prion protein to its native α-helical conformation is under kinetic control." Journal of Biological Chemistry 276, no. 23 (2001): 19687-19690.
37. Bujnicki, Janusz M. "Protein-structure prediction by recombination of fragments." Chembiochem 7, no. 1 (2006): 19-27.
38. Lu, Hsiao-Mei, and Jie Liang. "A model study of protein nascent chain and cotranslational folding using hydrophobic-polar residues." Proteins: Structure, Function, and Bioinformatics 70, no. 2 (2008): 442-449.
39. Frydman, Judith. "Folding of newly translated proteins in vivo: the role of molecular chaperones." Annual review of biochemistry 70, no. 1 (2001): 603-647.
40. Thulasiraman, Vanitha, Chin-Fen Yang, and Judith Frydman. "In vivo newly translated polypeptides are sequestered in a protected folding environment." The EMBO journal 18, no. 1 (1999): 85-95.
41. Bukau, Bernd, Elke Deuerling, Christine Pfund, and Elizabeth A. Craig. "Getting newly synthesized proteins into shape." Cell 101, no. 2 (2000): 119-122.
42. 42 . Wang, Peiyu, and Dmitri K. Klimov. "Lattice simulations of cotranslational folding of single domain proteins." Proteins: Structure, Function, and Bioinformatics 70, no. 3 (2008): 925-937.
43. Wang, Peiyu, and Dmitri K. Klimov. "Lattice simulations of cotranslational folding of single domain proteins." Proteins: Structure, Function, and Bioinformatics 70, no. 3 (2008): 925-937.