DEMOCRITUS UNIVERSITY
OF THRACE
DEPARTMENT OF
MOLECULARBIOLOGY AND
GENETICS

Using AlphaFold2 to predict structures of non-naturally occurring proteins: the Rop case.

Χρησιμοποιώντας το AlphaFold2 για την πρόβλεψη δομών μη φυσικά απαντώμενων πρωτεϊνών: η περίπτωση της πρωτεΐνης Rop.

Alexandros Tafanidis

Supervisor: Dr. Nicholas M. Glykos
Associate Professor of Structural and Computational Biology

Alexandroupolis, June 2024

Acknowledgements

I sincerely thank my supervisor, Dr. Nicholas M. Glykos, for his guidance, willingness to assist me, and unwavering patience throughout the process. I also want to thank him for helping me come up with answers to my questions.

I am also grateful for my friends, especially my family, who encouraged and supported me all this time.

Table of Contents

Abstract

Rop is a protein of significant scientific interest for studying protein folding, which forms a 4-α-helical bundle structure with an antiparallel (syn) topology. This protein has been studied extensively through its mutants carrying amino acid substitutions, deletions, additions, etc., to observe their impact on the three-dimensional structure. The 4-α-helical bundle, forming an anti topology, holds significant importance in structural biology due to its extensive use in numerous studies to explore the effects of different mutations on this specific topology. Unfortunately, our lack of ability to predict the three-dimensional configuration of an artificially synthesized protein prevents us from fully harnessing the capabilities of the Rop protein, as well as any other protein. Here, we show the capabilities of DeepMind's AlphaFold program, which aims to address protein prediction challenges and seeks to evaluate the program's algorithm by testing it on known mutant Rop structures. We found that in most cases, AlphaFold could successfully and correctly predict the structure of an already known Rop mutant, with minor to no differences. Only one out of seven mutants that we will look into in the following pages appear to have significant differences between the structure found in PDB and AlphaFolds' result.

Περίληψη

Μια πρωτεΐνη με σημαντικό επιστημονικό ενδιαφέρον για τη μελέτη της αναδίπλωσης των πρωτεϊνών ως σύνολο είναι η Rop, η οποία διαθέτει μια δομή 4-α-ελικοειδούς δεματίου με αντιπαράλληλη (syn) τοπολογία. Αυτή η πρωτεΐνη έχει μελετηθεί εκτενώς μέσω των μεταλλαγμάτων της που φέρουν υποκαταστάσεις, διαγραφές και προσθήκες αμινοξέων κλπ, για να παρατηρηθεί ο αντίκτυπός αυτών στην τρισδιάστατη δομή της πρωτεΐνης. Η 4-α-ελίκων δέσμη, που σχηματίζει μία anti τοπολογία, έχει σημαντική σημασία στη δομική βιολογία λόγω της εκτεταμένης χρήσης της σε πολυάριθμες μελέτες για τη διερεύνηση των επιδράσεων διαφορετικών μεταλλάξεων σε αυτή τη συγκεκριμένη τοπολογία. Δυστυχώς, η έλλειψη ικανότητας να προβλέψουμε την τρισδιάστατη διαμόρφωση μιας τεχνητά συντιθέμενης πρωτεΐνης μας εμποδίζει να αξιοποιήσουμε πλήρως τις δυνατότητες της πρωτεΐνης Rop, αλλά και οποιασδήποτε άλλης πρωτεΐνης. Εδώ παρουσιάζουμε τις δυνατότητες του προγράμματος AlphaFold της DeepMind, το οποίο επιχειρεί να αντιμετωπίσει τις προκλήσεις πρόβλεψης πρωτεϊνών και επιδιώκει να αξιολογήσει τον αλγόριθμο του προγράμματος δοκιμάζοντας το σε γνωστές μεταλλαγμένες δομές Rop. Βρήκαμε ότι στις περισσότερες περιπτώσεις, το AlphaFold μπορούσε να προβλέψει με επιτυχία και σωστά τη δομή ενός ήδη γνωστού μεταλλάγματος Rop, με μικρές έως καθόλου διαφορές. Μόνο ένα από τα επτά μεταλλάγματα που θα εξετάσουμε στις επόμενες σελίδες, φαίνεται να έχουν σημαντικές διαφορές μεταξύ της δομής που υπάρχει στην PDB και του αποτελέσματος που μας έδωσε το AlphaFold.

1. Introduction

1.1. Coiled-coils

The coiled coils are a topological quaternary structure formed by binding two alpha-helices through various bonds such as hydrophobic interactions, hydrogen bonds, and Van der Waals forces. These interactions hold the two alpha-helices together, resulting in the formation of the coiled-coil structure. In 1951, Linus Pauling described the twisting of two alpha-helices around each other but did not name this type of topological structure [1,2]. The term "coiled coils" was coined by Francis Crick around the same time as Pauling [3]. Pauling is credited with introducing the term [1], while Crick is recognized as the first to propose the structure of coiled coils and develop mathematical methods to predict their structures [3]. Crick also proposed the "knobs-into-holes" structure, where amino acids in the polypeptide chain of a coiled-coil, known as knobs, fit into adjacent holes, influencing the turning of the helices [3,4]. This structure is illustrated in *Figure 1* for better comprehension. Unlike most alpha-helices, the ones forming a coiled-coil are packed somewhat differently, with 3.5 residues per turn rather than the typical 3.6. The coiled-coil structure has distinct features, such as hydrophobic residues occupying the "a" and "d" positions every seven residues, labeled from a to g [2,4,5].

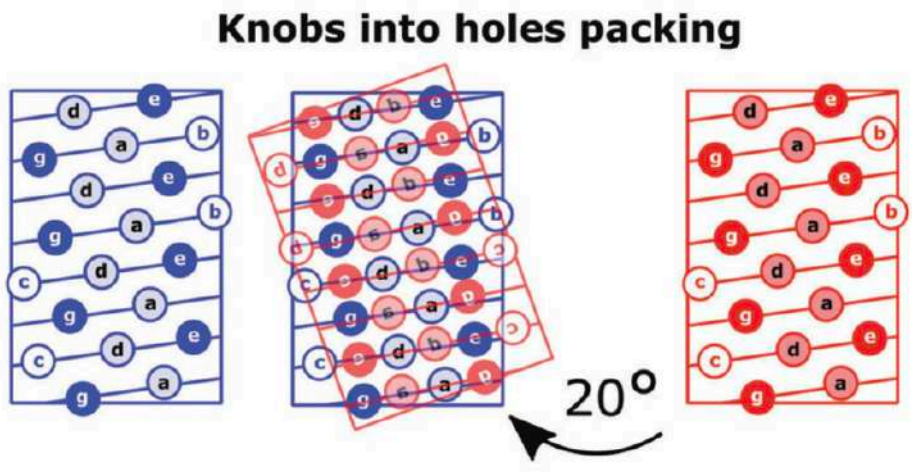

Figure 1 – A representation of how “knobs-into-holes” work. The amino acids shown in red are seen in the second representation to enter, with their direction reversed, into the gaps (in the crevices) between the amino acids shown in blue.

1.2. Four helix bundle

Proteins with a four-helix bundle topology, also known as 4-α-helix bundle topology, consist of 4 helices packed together. These helices twist and turn in such a way that place their hydrophobic residues at the center of the bundle, creating a hydrophobic core. The hydrophilic residues are located on the outer surface of the bundle, making contact with water molecules while surrounding the hydrophobic core [4,6]. The structure of a helical bundle, which is based on ridges-into-grooves arrangement in proteins, involves specific arrangements where the protruding ridges of one protein fit into the grooves of another protein [7]. This structure was first described by Chothia in 1977, who noted a slight twist (about 20°) of all the coiled coils of the 4-α-helical bundle. This twist creates ridges intermittent with grooves, with the residues forming both the ridges and grooves having a distance of 4 amino acids (i – i+4) and about three amino acids (i – i+3) in some cases, as described by Chothia et al. in the same paper [7]. Additionally, the ridges of one helix are packed into the grooves of the other and vice versa [5,7].

1.3. Repressor of Primer (Rop) protein

The Repressor of Primer, also known as Rop, is a protein consisting of a 4-α-helix bundle synthesized in E. coli. Its primary function is maintaining the copy numbers of the ColE1 and similar plasmids at low levels. The protein's structure has been studied using X-ray [8,9] and NMR [10] techniques, revealing that the Rop gene encodes it [11] after being genetically identified in 1980 by Twigg and Sheratt [12].

As mentioned previously, the ROP protein, also known as RNA I (One) modulator (ROM), participates in the mechanism that controls the copy number of plasmid of the ColEl family by increasing the affinity between two complementary RNAs and helps to maintain a constant copy number by counteracting occasional deviations from the steady-state level[13]. To be more precise, ROP binds to the transiently formed RNA I and RNA II kiss complex and reduces the equilibrium dissociation constant of the initial RNA complex[14]. This is achieved by negative control of the frequency of replication

initiation events between RNA I, RNA II and ROP[13]. Furthermore, ROP does not seem to be an essential component of the ColE1 control system. Deleting the rom gene leads to a two-to-threefold increase in *N* (final cell concentration of generations that occurred between the specified period of time) in slowly growing cells, but it has no phenotypic consequences on the *N*-value in fast-growing bacteria[15].

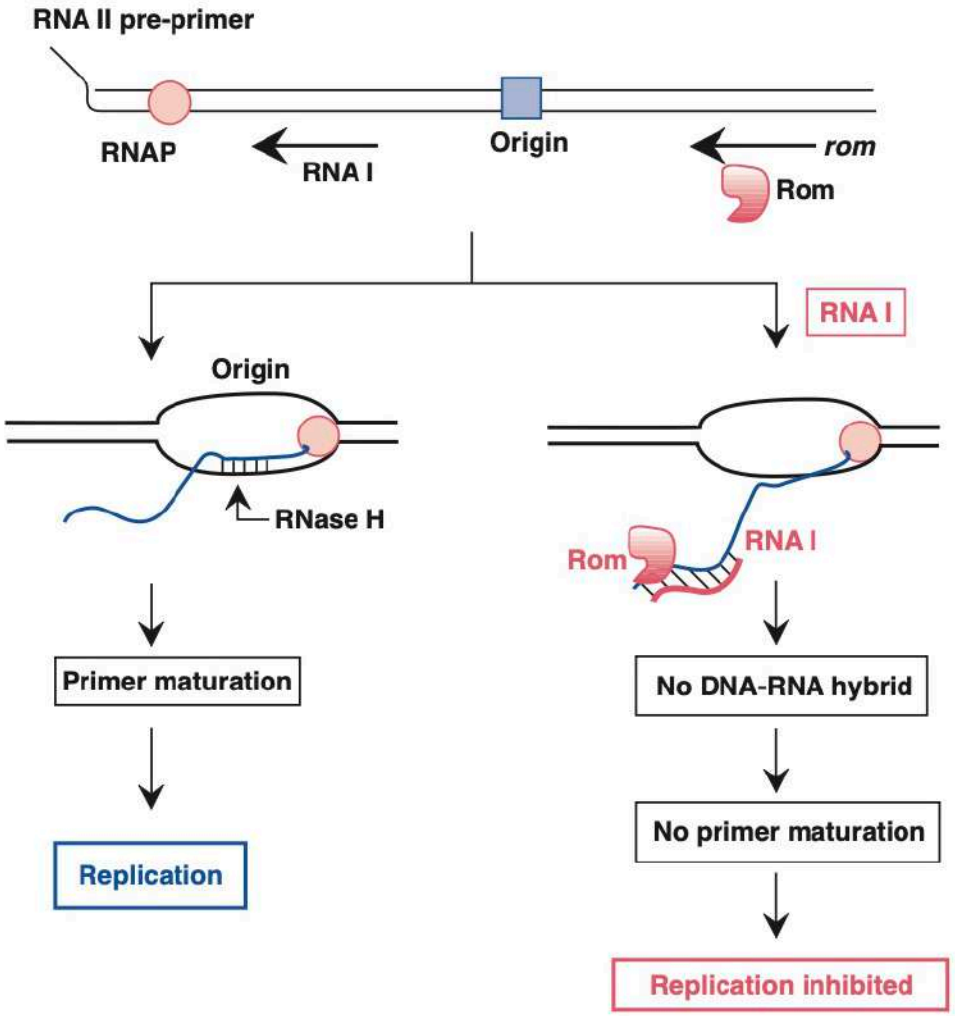

Figure 2 - At the ORC, the RNA II forms a stable hybrid with the template DNA when it does not interact with the RNA I (left). RNase H cleaves this hybrid to produce the RNA primer's 3-OH end, which is where replication begins. The Rom protein facilitates the interaction between the inhibitor RNA I and the complementary region in the RNA II preprimer (right). The maturation of RNA II into the replication primer is prevented by the RNA I-RNA II interaction, which also inhibits the formation of the DNA-RNA II hybrid at the ORC site. Reproduced without permission from Gloria del Solar and Manuel Espinosa, 2000.

The protein folds into a homodimeric four-helix bundle comprised of two helix-turn-helix monomers forming an anti-topology, and each monomer, designated as A and B, consists of 63 residues and is further divided into two chains; the A monomer

into chains 1 and 2 and the B monomer into chains 1' and 2' [10]. These 63 residues form the structure *(Figure 1)*, separating three residues at positions 29-31 and creating the two chains [16]. Furthermore, as we will see with the mutant $\Delta_{30\text{-}34}$, a five-residue loop region serves as a stoppage to the protein heptad pattern[17]. The amino acids are categorized into heptads, and the core structure consists of 8 layers formed by "a" and "d" residues. It is important to note that although the first two residues of each chain are included for convenience, they do not contribute to the helical topology. The binding forces holding the 4-α-helices together are predominantly hydrophobic, with ionic bonds playing no significant role [16].

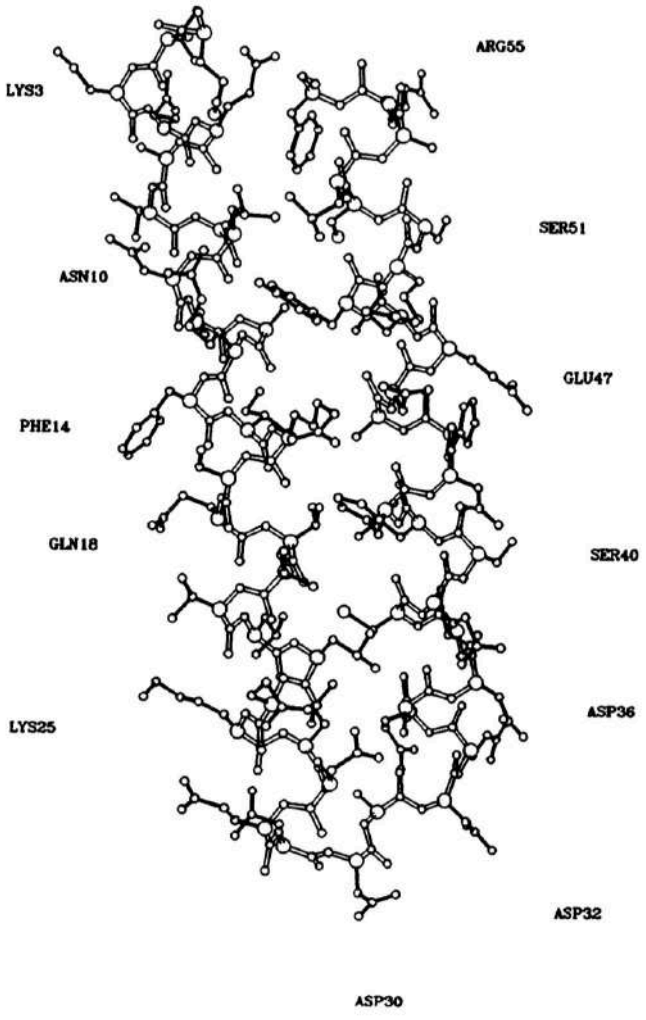

Figure 3 - The structure of the polypeptide of one of the monomers of the Rop protein. It is viewed from such an angle that the 2-fold axis is horizontal and behind the molecule. The number of residues can be seen next to the corresponding amino acid. Reproduced without permission from Banner et al., 1987.

The loop sections of proteins in general, and not only the Rop protein on which this work focuses, have been demonstrated in numerous previous studies to be crucial in determining but also stabilizing the final 3D structure of protein [18–22]. Two loop regions—one in each helix bundle—are present in Rop. Hydrogen bonds are crucial in stabilizing the chains by forming in the loop area and the adjacent residues, specifically residues 28 through 34 [8]. However, only one hydrogen link is seen—the hydrogen bond between the carboxylate and amino groups of residues 32D (aspartic acid) and 33E (glutamic acid). Additionally, the 31st residue of the chain exhibits an uncommon bonding with both helices (26L on the first helix and 35A on the second helix) [8].

We are interested in Rop because it seems to be a perfect natural model system for understanding the thermodynamic and kinetic folding characteristics of 4-α-helix bundle structures[23]. Furthermore, the process by which this specific protein undergoes folding and assumes its distinct structure has been the subject of extensive research for numerous years, resulting in numerous articles that thoroughly investigate it [24–30]. We base our choice on two critical criteria: high-resolution X-ray data [8] and the accessibility of a wide range of strategically mutated proteins [16].

1.4. Rop mutants

Rop mutants come in various forms, each with unique traits. Except for the A31P variant and, to a lesser extent, the 2aa variant, which shows slight differences compared to the original structure of Rop, the tertiary structure of all the other mutant proteins closely resembles the wild-type topological structure. Most mutants adopt an anti-topology (native Rop, A31D, 2aa, Cys-free), except for the A_2L_2 and A_2I_2, which adopt a syn topology. The A31P adopts a topology known as "bisecting-U," which is covered in more detail below. Furthermore, because the $\Delta_{30\text{-}34}$ mutant lacks five residues, it adopts an entirely different configuration and does not appear to fall into the syn-topology, anti, or bisecting-U categories of A31P.

Furthermore, we shall discuss Rop mutants that remain unchanged despite changing their amino acid sequence in the parts that follow this thesis and those that modify their three-dimensional structure even with minute changes. The A31D is one such mutation. Due to a dearth of studies, we do not have much information about this particular mutation. However, its amino acidic sequence is known, and upon loading it into AlphaFold2, we observe that its topological structure does not differ from that of WT Rop.

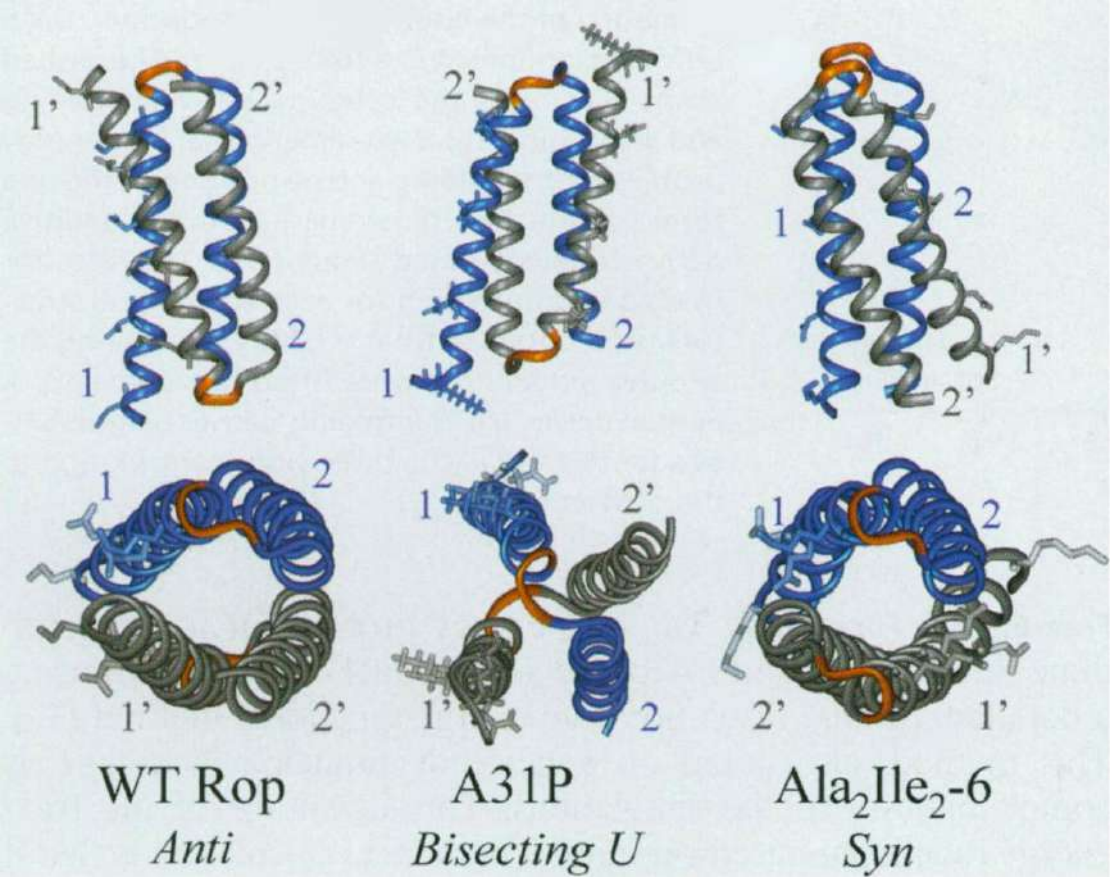

Figure 4 – The three topologies we will examine below are compared side by side in this image. A31P is the only one that takes the bisecting-U topology; A_2L_2 and A_2I_2 are the only ones that take the syn topology and the other mutants take the anti. The loop connecting the two helices is orange, one monomer is grey, and the other is blue. Taken without permission from Y. Levy, Samuel S. Cho et all 2004.

We start with the mutant known as Cys-free when discussing the mutants for which we do have some supporting data and research. Although cysteine residues are often conserved within protein families [31,32], they can cause structural and, specifically, folding complexity in proteins. This is why it is crucial to design a cysteine-free protein [33].

As the name implies, the structure of this specific mutant is identical to that of the wild-type protein due to the removal of cysteine residues at positions 38 and 52 and their replacement by alanine and valine residues, respectively [34]. Designing such a mutant aims to establish a relationship between the thermodynamic modifications and mutations' impact on proteins [34].

Going further, we will now discuss the mutant known as A_2I_2. It forms a syn topology, which is achieved by flipping one monomer 180 degrees around an axis that runs parallel to the dimer interface [35]. Its hydrophobic core, primarily composed of hydrophobic residues [8], has been drastically repacked, redesigning the entire core and losing its capacity to bind RNA. However, at the same time, the protein's thermal stability has been increased [36]. Under typical circumstances, that is, in the protein's wild-type form, the core is made up of residues that occupy positions a and d in the heptad; these residues are typically either leucine, isoleucine, or alanine, cysteine, or threonine, in that order [8]. Because of the little to no polarity in their side chain [37], these residues are hydrophobic and help to stabilize proteins in several ways [36,38,39]. They thus form the central region of this protein. However, on A_2I_2, the core undergoes such a significant redesign that, in addition to being much more densely packed than Rop's core *(Figure 5)*, the "a" sites are almost entirely made up of alanine residues, while the "d" sites are almost entirely made up of isoleucine residues. This is how the mutant gets its name.

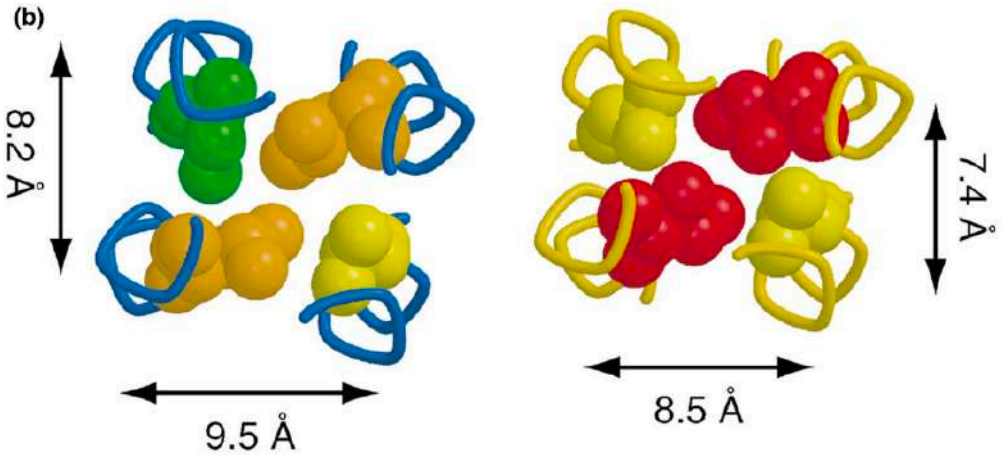

Figure 5 - Side-by-side comparison of the hydrophobic core in Rop and A_2I_2-6 (mentioned below). The closer packing of the helices in A_2I_2-6 is shown (molecule represented in the right image). Reproduced without permission from Mark A. Willis, Barney Bishop, Lynne Regan, and Axel T. Brunger (2000).

Continuing, there is A_2L_2, analogous to A_2I_2. Just like A_2I_2, it also forms a syn topology [35]. The protein's amino acid sequence indicates that leucine residues are in place of isoleucine residues in d positions in sequence [40]. These d locations are occupied by leucine residues, just like in the earlier mutant we discussed. This is possible due to leucine's hydrophobic properties that shield it energetically inside the protein's core, which helps maintain stability [38,40]. Much like A_2I_2, this results in a more densely packed core and a more stable structure than the wild-type protein. The phenylalanine residue in the polypeptide chain of the mutant was found to be present at position 56 of chains 2 and 2' during the redesign of the mutant's core [40]. Two versions of A_2I_2 were created because it was unclear whether the unique residue for the Rop protein, which consists of 8 layers, had any structural significance. One version, Rop21, had all eight layers repacked, while the other included only six layers—the middle layers—and excluded the outer layers, which contained the phenylalanine residues. The latter one goes by the name Rop13. The two repacked proteins could fold into a stable structure akin to the Rop native state and be more thermodynamically stable [41]. It's also important to note that this specific mutant has been observed to oscillate between the syn and anti structures. The protein is still active and has the same affinity for binding RNA when it is in its anti structure. However, when it is in its syn topology, it becomes inactive and totally loses its ability to bind RNA[42].

We then will go over the mutant known as $\Delta_{30\text{-}34}$. The Rop protein is a homodimer of two 63-residue helix-turn-helix monomers in its normal state. Each monomer is joined to the other by a hairpin, a 5-residue loop region that serves as a stoppage to the protein heptad pattern[17]. The mutant known as $\Delta_{30\text{-}34}$, or RM6, was produced by deleting five residues, specifically from the 30th to the 34th residue, to produce a continuous pattern of heptads. At the same time, its heptad pattern and its sequence are inverted. The formation of a homotetrameric protein, which loses its activity to bound RNA and thus loses its regulatory potency [43], is one of the most prominent and most noticeable distinctions between the natural form of Rop and the modified form. As was previously established, the natural structure of Rop consists of two

homodimers arranged in an anti topology. The heptad pattern of RM6 and the protein's amino acid sequence is reversed in revRM6, another variation of the RM6 mutant. This master's thesis will only focus on and use the RM6 mutant.

Until now, all we had discussed was Rop mutants, which retained their capacity to fold into stable structures and maintain their thermodynamic stability even when one or more of their residues were replaced. Conversely, things either did not change or became better. Not all known Rop mutants, though, are the same as that. We shall then go over the Rop mutation known as 2aa. This specific mutant has been given this name because two alanine residues have been inserted into the protein loop region [44]**.** Except for their bend region, this mutant and the wild-type protein have remarkably similar structures [45]. The insertion of residues in the loop region led to the correction of a discontinuity that the protein had before. This also prevented the smooth continuation of the heptad [44,46], found in past research and may help change the direction of the axis of the Rope-like structure, aka the polypeptide chain [47].

Furthermore, it has been observed that the 2aa mutant of the protein exhibits hydrogen bond formation between loop region residues, specifically between residues that are separated by three positions (i → i + 3) [45]. This contrasts the wild-type form of the protein, where no such formation has been observed. Although no correlation between protein loops and their stability has yet been found [46,48,49], it has been said that the loss of stability that mutant proteins gain is often correlated with increased flexibility [46].

Last but not least, we will discuss the A31P, a Rop mutant with one of the most noticeable differences when comparing its structure with that of native Rop. The A31P mutant is being created similarly to the A31D mutant, with the 31st residue of the amino acid sequence being changed from alanine to proline [50]. This causes the mutant's hydrophobic core to change entirely; it completely changes its topological structure, destabilizes its structure, and changes its surface properties [51]. More precisely, the protein takes on an anti-topology, left-hand orientation [52] in its native form, where the two monomers are "parallel" to one another, and the loop sections

are at opposite ends *(Figure 6*) [53]. On the other hand, the protein adopts a right-handed orientation and a "bisecting U" topological structure when proline is substituted as the 31st residue [50,52–54]. To allow for the turn of the other monomer, the distance and the interhelix angles of the helices had also been increased [30,50]. A sizable, uninterrupted interior cavity forms encircled by the initial pair of hydrophobic layers [50,51].

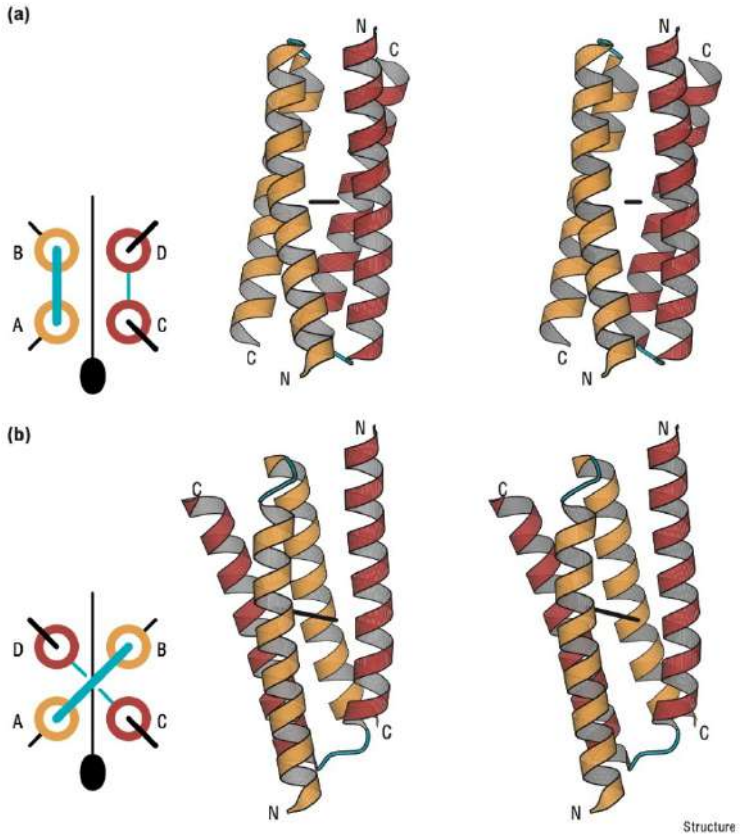

Figure 6 - Comparison between the structure and topology of the native protein next to the A31P mutant. Reproduced without permission from Glykos, N. M., Cesareni, G., & Kokkinidis, M. (1999). Protein plasticity to the extreme: changing the topology of a 4-α-helical bundle with a single amino acid substitution.

Why, therefore, is this occurring following the substitution of alanine with proline? This is because, unlike alanine, proline lacks the cyclic structure necessary to produce the standard dihedral angles (ϕ and ψ) [16].

Furthermore, it has been noted that 31A and 26L establish hydrogen bonds when things are normal. Because proline cannot function as an H^+ donor to create a hydrogen bond, which causes instability and alters the synthesis of Rop, this does not occur in the mutant [8,29].

Finally, the protein's structural shift also impacts the hydrophobic core. A31P disregards the WT core's typical periodicity, which follows the *adad* pattern and instead uses six layers—instead of eight—depending on layer [50].

1.5. Structure prediction: AlphaFold2

The protein folding problem, first presented in the 1960s and now recognized as three distinct problems, is one of the most fundamental issues in structural biology. In summary, it tries to answer how a particular amino acid sequence can reveal a protein's structure and what forces cause it to fold in a particular way [55]. Stated differently, the resolution of the protein folding puzzle would enable precise estimation or ascertainment of the amino acid sequence, furnishing a more comprehensive understanding of the functions of proteins and the processes by which these indispensable macromolecules operate. Nearly a decade later, Anfinsen demonstrated through his now-famous experiment that a peptide may revert to its original three-dimensional shape by relying on amino acid sequence after unfolding by unfolding agents [56]. It was still unclear how exactly knowing a peptide's sequence could help us predict its shape. Subsequently, in 1996, CASP [57] was established as a community where people could predict unknown protein structures in a blind test setting [55]. AlphaFold, a program created by DeepMind that predicts protein structures using artificial intelligence, was presented to the public as the CASP (CASP-XIII) winner. The enhanced AlphaFold version, known as AlphaFold 2, won the round in the upcoming CASP-XIV in 2020 for the second time in a row after submitting protein models that were significantly more accurate than those of the other competitors [57,58].

In more detail, AlphaFold2 is an artificial intelligence program whose primary goal is to predict a protein structure just on its sequence, as was indicated above [59]. It underwent multiple investigations to assess its efficacy in predicting the 3D protein structure. It was compared to other methods of structural prediction, and typically, it

yielded a high rate of success and produced structures that closely resembled the native protein used in each study [60–63].

The primary factor contributing to its effectiveness and ability to accurately forecast the structures of numerous molecules is its ability to accurately map a desired protein sequence to an existing amino acid sequence, regardless of how closely connected the two sequences are to one another in terms of evolution [64]. AlphaFold2 generates multiple sequence alignments (MSAs) using the protein molecule sequence we want to study as a mold. After iteratively runs through the sequence, it generates a complete structure. This structure passes through the Evoformer 48 times (generating a new structure with each cycle), comparing the resultant structure with the one that came before it until the polypeptide's final, complete structure appears [64,65]. The second iteration of AlphaFold outperformed every other participant in CASP XIV, as demonstrated in *Figure 7*, in both the TBM (template-based modeling) and FM (free modeling) categories[61].

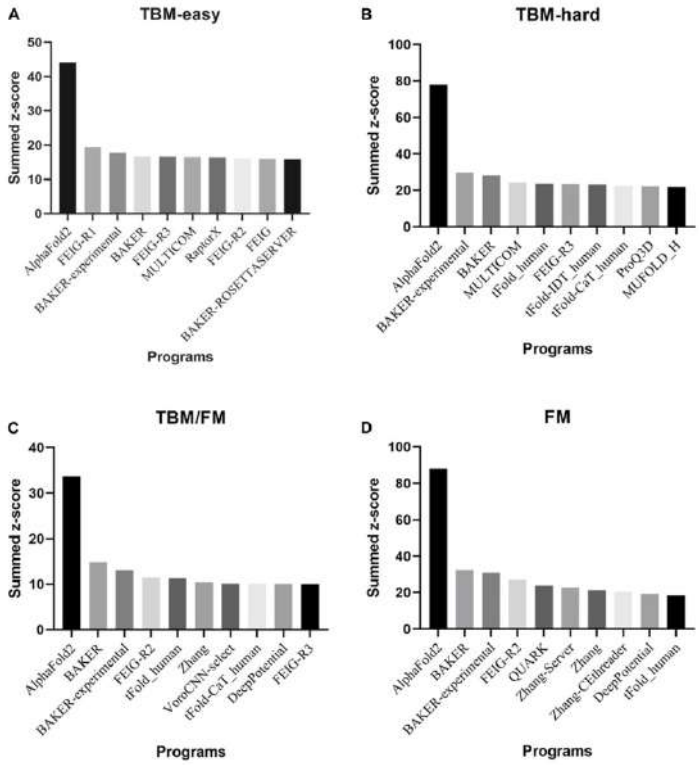

Figure 7 – A four-part graph that contrasts the various competitors and their programs with one another. AlphaFold2 scored higher in both FM (Free Modelling) and TBM (Template-based Modelling) within each category.

DeepMind, the company that founded AlphaFold, made its source code accessible to the public in 2020, allowing anyone to utilize the deep learning program for free [66]. Because of certain limitations with AlphaFold, such as the requirement for large GPU RAM for basic predictions of relatively small proteins and a large number of databases to be searched concurrently with sensitive homology detection techniques [67], scientists had to work around some limitations. This is where ColabFold comes into play. Using Google Collaboratory [68] on a local computer system, ColabFold is software that can be used as a Jupyter notebook. It is approximately 60 times and 90 times quicker than AlphaFold2 [67] for single and batch predictions.

1.6. Main question

We have demonstrated throughout this master's thesis that Rop protein is an excellent choice for studying the characteristics of 4-α-helical bundles. Several mutants were generated to understand this structure more deeply; some changed abruptly, while others retained the same structure as the natural protein. After loading their previously identified sequence from PDB into ColabFold, the outcomes of only a few mutants were contrasted with those from PDB and appeared to take an entirely new structure and topology. On the other hand, the majority of mutants' structures were nearly identical; ColabFold could accurately predict the mutants' protein structure based on their sequence. The program's inability to accurately predict the 3D structure of the mutants under investigation indicates that either missing data was supplied to it, the program processed the data incorrectly, or there are allosteric interactions between amino acids that caused "confusion" in the program and result in an error in the prediction of protein structure. The most probable cause for the three reasons mentioned above is that AlphaFold might lack the necessary tools and computational skills to predict and compare the structures of 2 or more proteins with slight to no evolutionary history [64,69].

2. Methods

2.1. AlphaFold web interface and mode of operation

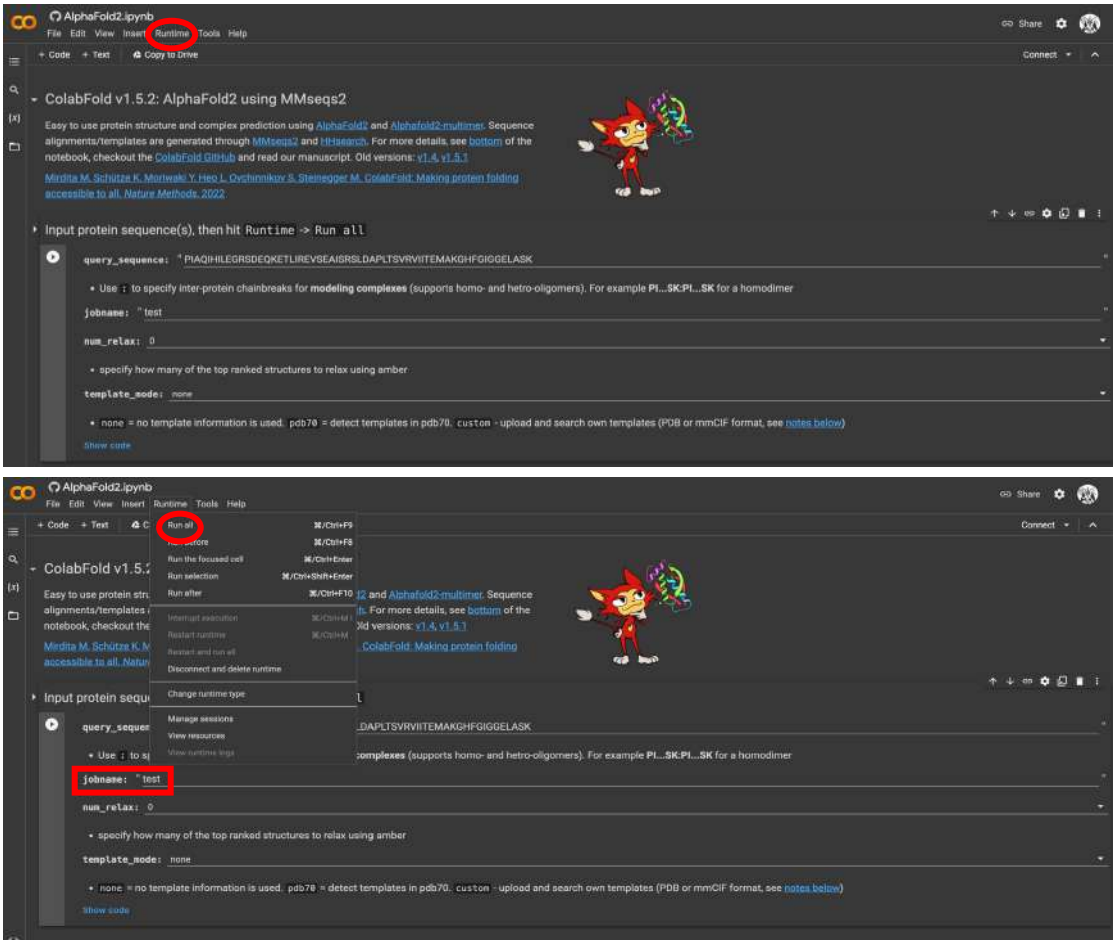

Figure 8 - The main page of AlphaFold, the tool using the algorithm of AlphaFold as its base.
A) The first step (1) is entering the protein sequence that we want AlphaFold to predict its structure. The second step (2) is pressing the option "Runtime".
(B) The third and final step (3) is to run a prediction cycle with the given amino acid sequence by pressing the option "Run all". Optionally, we can name the job cycle in the field named "jobname" right before we hit the "Run all" option (red rectangle).

The first screen we view when we visit the ColabFold website is depicted in the image above. We describe how to load the amino acid sequence into the webpage and what "buttons" we need to hit to activate the program in the caption of *Figure 8*. After the program finishes its calculations, a notification asks if we wish to download the results. The results are automatically saved in the "Downloads" folder and are stored in a folder with a name we have already specified (red rectangle).

2.2. MMalign

We load at least two pdb files from the structure analysis produced by PyMol into MMalign. A computational biology tool called MMalign compares two or more protein structures by superimposing them. Once the program has been installed on our machine via the shell, we can use the following command to compare two structures (which are conveniently located in the same folder):

MMalign 1Rop.pdb Model_WT.pdb

This command compares data from the PDB and AlphaFold concerning the native Rop, providing us with the outcomes shown in Table 1.

The command "MMalign 1.pdb 2.pdb" can run this program from within the Ubuntu terminal. The files designated 1 and 2 contain protein structures; typically, 1 corresponds to the one from the PDB depository and 2 from AlphaFold. Table 1 displays the outcomes of each program computation. The correlation between the wild-type Rop structure and the anticipated WT Rop structure from AlphaFold is displayed in *Table 1*.

We obtain different data for our poorly understood protein structure from every results line. The names of our chains are initially given in the first section of *Table 1* and are indicated by the name of the PDF file and the number of amino acids that make up each polypeptide chain. The second section provides statistics, including the number of successful matches between the two (or more) polypeptide chains and the automatically derived PMSD and TM scores. Subsections that follow will go over the significance of RMSD and TM scores.

Finally, the aligned residues between the two chains, indicated in red as “a” for Chain_1 and “b” for Chain_2, are shown in the third section of the results. These residue pairs are indicated by colons (:) for residue pairs and dots (.) for identical residues. The two chains' sequences are also displayed, with an asterisk (*) denoting the chain's conclusion. The third section of the table will be omitted to condense the amount of data and make the results that are supplied later on more straightforward

to read. The "." symbol indicates other aligned residues that are not shown here, the ":" symbol indicates other residue pairs that are separated by less than 5.0 Å, and the symbol "*" indicates the location of the stop codon. The symbol "-" indicates sequence gaps.

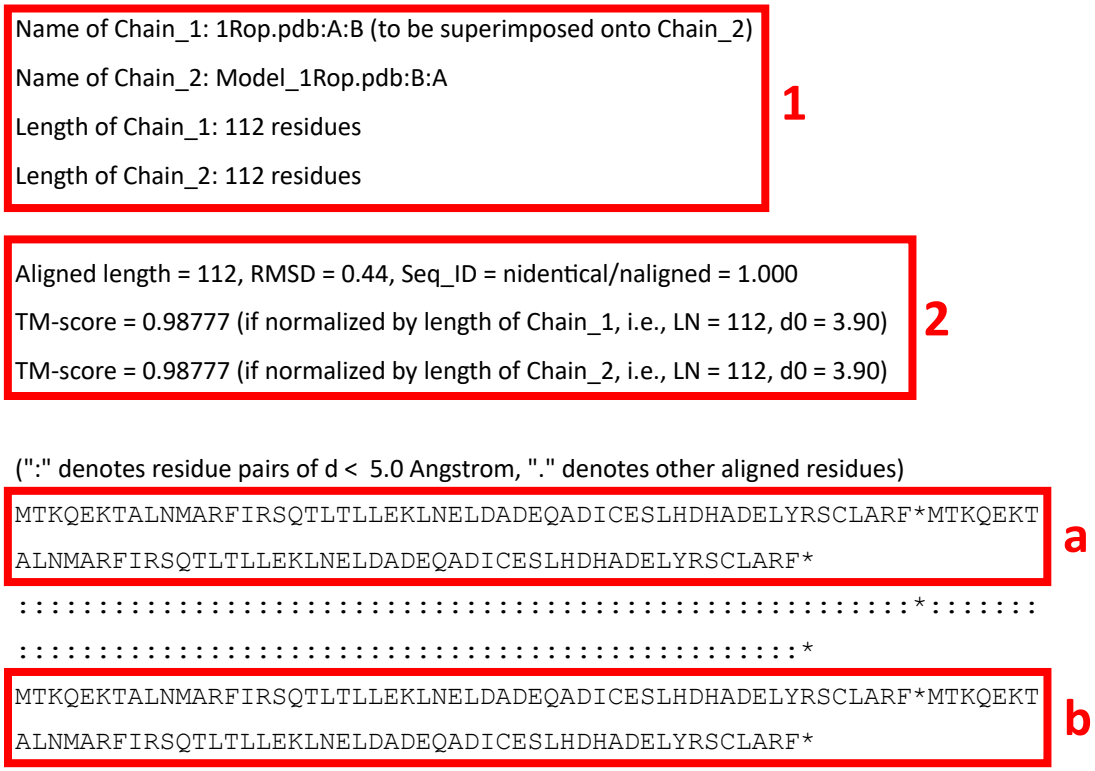

Table 1 – An example of a result after running two .pdb files on MMalign. The results have been divided into three categories for easier explanation. The third part has also been divided into two sections, a and b, each of them representing one of each polypeptide chain.

2.3. Visualization: PyMol

First, to visually compare the structures of the original Rop protein and its mutations in three dimensions, we visualize our data using PyMol, a molecular visualization program. PyMol is an open-source visualization program that shows the molecule's structure we are studying using pdb files [70,71]. We graphically represent the molecule's structure we investigated using PyMol and the pdb file from the PDB database. For instance, we utilized the pdb file for the native form of Rop (PDB ID:

1rop) as the initial step. The outcome was stored for later review and comparison with every other Rop mutant.

The user interface of the program is shown in the following image:

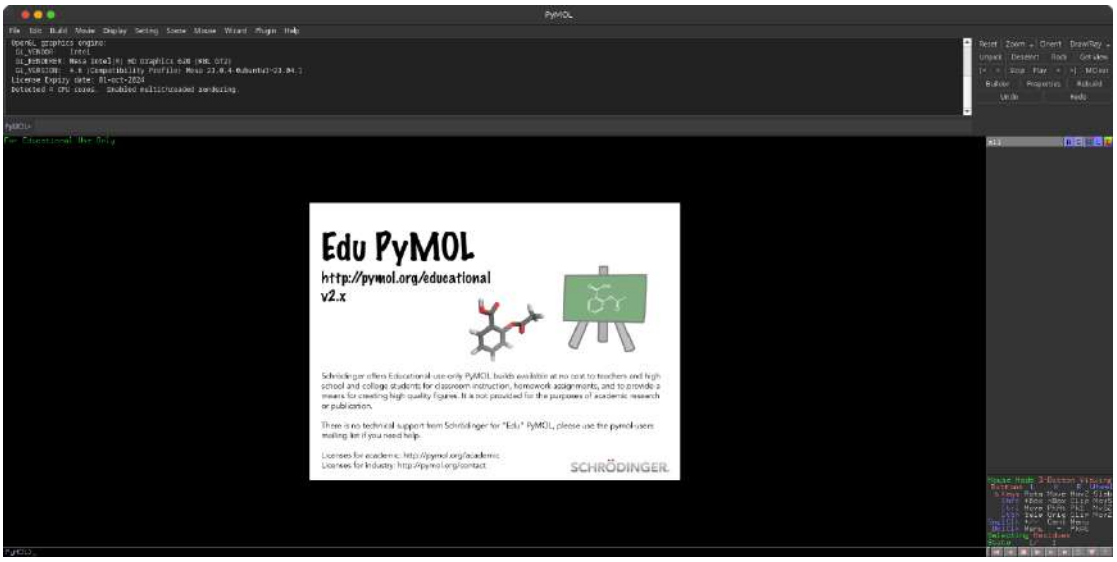

Figure 9 – The UI (User Interface) of PyMol. The complete software version was downloaded at no cost through an institutional email address.

2.4. RMSD score

The Root Mean Square Deviation (RMSD) can be used to assess how similar two stacked polypeptide chains are statistically. More precisely, it is the mean separation between the atoms of two stacked protein structures, often the backbone or Cα atoms. One important use of this metric in structural biology is comparing the three-dimensional structures of proteins, nucleic acids, and other macromolecules. A smaller RMSD value indicates a higher level of structural similarity.

The process of determining an RMSD score consists of four steps: first, the polypeptide chains under study are superimposed to reduce the distances between atoms; second, the pairs of atoms that make up the chains are identified and compared, mainly using the Cα atoms (e.g., the Cα of the first atom of the first chain with the Cα of the first atom of the second chain); third, the distances (also called

Euclidean distances) between the pairs of atoms are computed; and fourth, the RMSD is determined using the distances calculated in the step before.

An almost identical structure is typically indicated by a low RMSD score of less than 1 Angstrom (1 Å); a moderate score, between 1 and 3 Å, shows slightly different but generally comparable structures; and a high score, above 3 Å, indicates markedly distinct protein structures [72].

2.5. TM score

The TM score assesses the similarity of two or more protein structures comparable to the RMSD previously discussed. However, it offers a more accurate, reliable, and sensitive estimate. The primary distinction between both is that the TM score compares the polypeptide chains we analyze regardless of length. In contrast, the RMSD compares the distances between the Cα atoms in these chains.

Furthermore, there are three steps in the TM score computation instead of four in the RMSD calculation. These steps will only be briefly discussed due to their higher difficulty level than their RMSD counterparts. These are the following steps: firstly, d0 calculation; secondly, distance addition for each aligned amino acid; and thirdly, the final TM score computation. The mathematical equations controlling the TM score are significantly more complex and demanding than their RMSD counterparts despite the score's seeming simplicity of calculation due to its few stages.

A TM-score above 0.5 indicates a high level of similarity to the native structure and a correct folding, suggesting that the predicted model is likely to have a similar structure to the correct native conformation. A TM-score ranging from 0.3 to 0.5 suggests the presence of structural similarity, although the overall fold may not be correct, and the prediction may lack complete accuracy. TM-scores below 0.3 indicate that the predicted structure is likely incorrect or significantly different from the native structure. In this instance, the model is deemed unreliable [73].

3. Results & Discussion

3.1. The control: predicting the structure of native Rop

The aim was to verify that AlphaFold can correctly predict the known structure of native Rop. The sequence given to the program was **MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFGDDGENL,** which is the residue sequence for the native form of the Rop protein. In the following experiments, we are to observe the possible differences and similarities between the native structure of Rop and the structure given by AlphaFold, depicting its prediction for the native form.

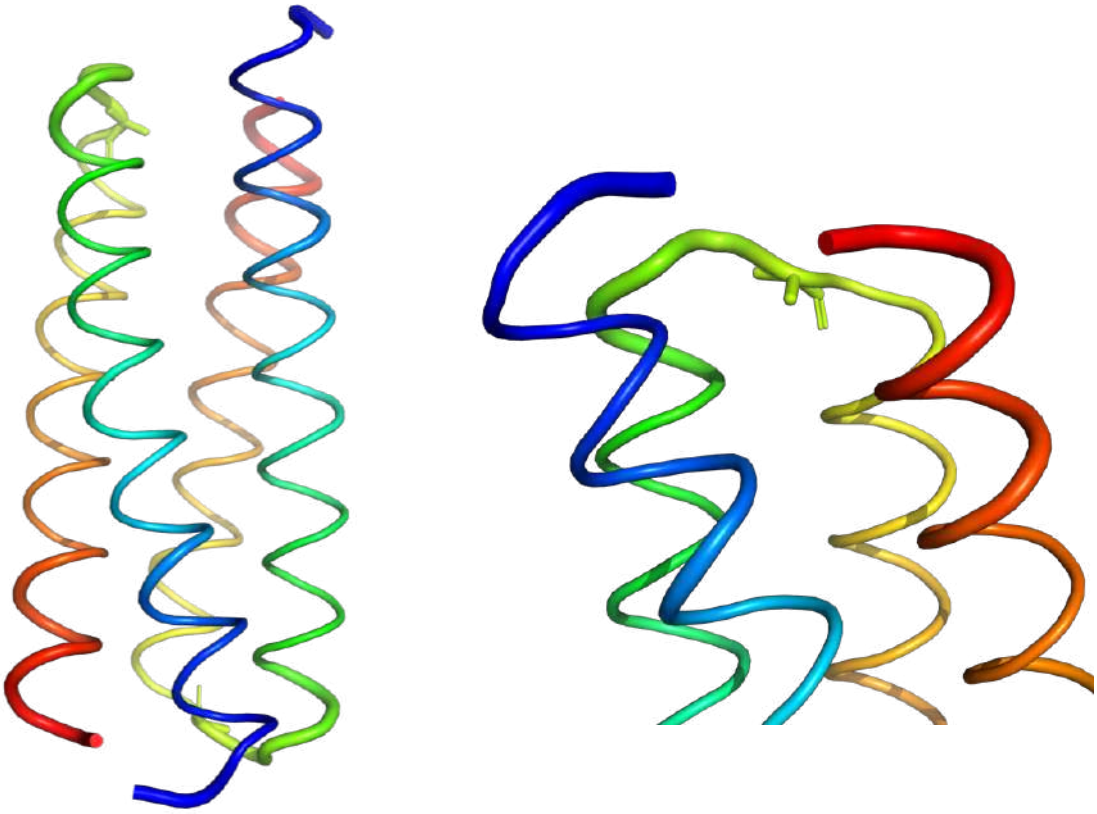

Figure 10 - The 3D structure of the WT Rop, as seen in Pymol, using the file from PDB.
(A) The protein structure comprises two monomers, each folded as a helix-turn-helix motif, forming the complete 4-α-helical bundle.
(B) In the WT form of the protein, the 31st position is taken by an alanine residue. It is depicted as a liquorice stick.

The structure of the Rop protein (PDB: 1rop) is shown in *Figure 10*, using the pdb file from the PDB database and loading it into PyMOL. We use the liquorice stick

style in all structure depictions to minimize unnecessary information. Also, for all the mutants of Rop, we used the same coloring form, from warm to cool colours, as we go from the amino terminus to the carboxy terminus so that we can show any differences that have occurred in the structure of the protein, between the natural structure and the structure resulted from AlphaFold. Finally, in the case of superimposition between two three-dimensional protein structures, we use different colors for each polypeptide chain to make it easier to observe their differences.

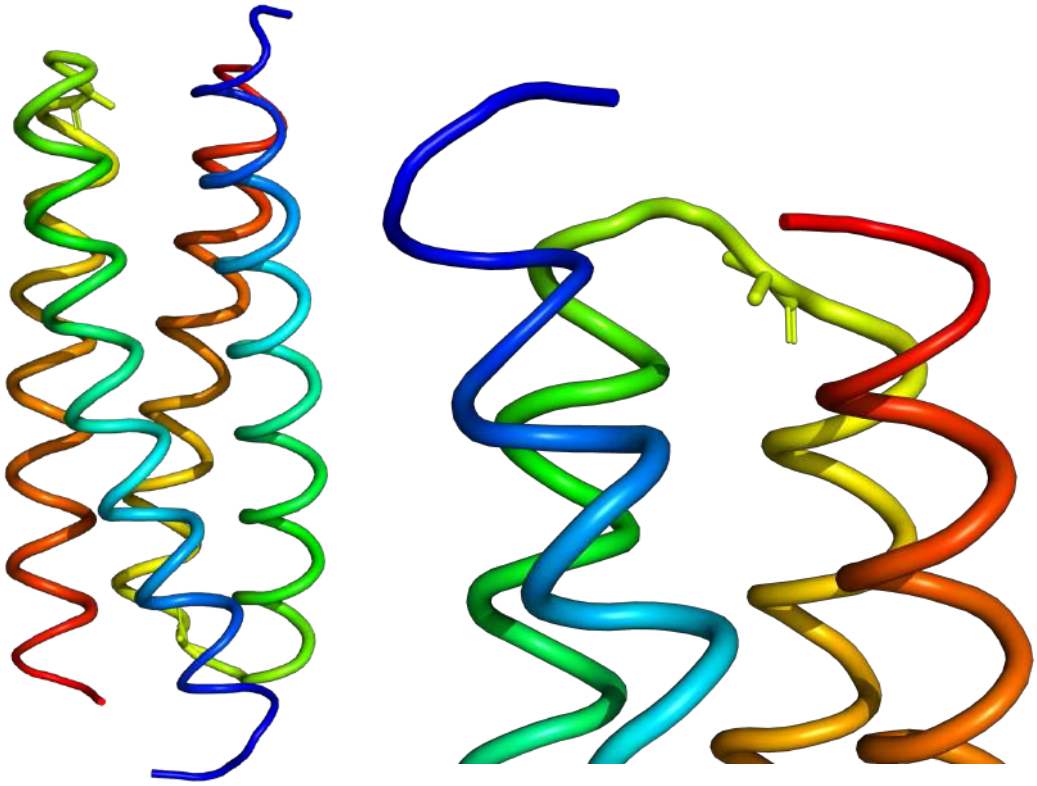

Figure 11 - The structure that depicts WT Rop as predicted by AlphaFold.
(A) No distinct structural or topological differences can be noticed at first glance compared with panel A in Figure 5. We compare the two structures using the MMalign tool and the RMSD and TM-score.
(B) Rotation of the structure depicted in panel B by 90^{o} degrees on the x-axis. The image is zoomed in to better view the 31^{st} amino acid. The amino acid in the 31^{st} position of the WT RP is an alanine residue. It can be distinguished due to its liquorice stick form.

The MMalign is used for all of the calculations. This tool is used to identify similarities and differences between them. If some similarities are noticeable after investigating two protein structures, the tool is also responsible for quantitatively measuring the degree of those structural similarities. Below is a sample of the output

MMalign provides, along with the sequence alignment of the two structures, the WT Rop from PDB and AlphaFold.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop (1Rop) PDB Chains 1 & 2	112	112	0,44	0,98777 (When normalized with 1Rop)	1,000
WT Rop AlphaFold Chains 1' & 2'	126			0,87931 (When normalized with AlphaFold's' WT Rop)	

Table 2 – The data derived from comparing the pdb files of the WT and the AlphaFold's WT Rop.

```
Chain 1:  MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARF-------*
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::       *
Chain 1': MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFGDDGENL*

Chain 2:  MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARF-------*
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::       *
Chain 2': MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFGDDGENL*
```

The MMalign tool's information comes from matching two chains, one representing the PDB WT Rop and the other the AlphaFolds' WT Rop. With a root mean square deviation (RMSD) of 0.44, the alignment resulted in an aligned length of 112 amino acid residues, suggesting a perfect fit between the two chains. Moreover, from the alignment above, we understand that the two structures superimpose each other perfectly, which is evidence of a perfect sequence ID.

These two protein structures are not significantly different based on the results from MMalign and the assumption that we also have the superposition of the two structures, shown in *Figure 12*. They are reasonably close scores, even though their TM score varies based on the structure the program uses as a mold. This also holds true for RMSD. The two polypeptide chains differ in length by 14 amino acids, which is something to note. The sequence -GDDGENL-, which runs from the 58th amino acid to the last, is absent from the file obtained from the PDB. Instead, the WT Rop file from

AlphaFold has this brief sequence. This occurs because these amino acids—the ones that comprise the C-terminus of the chain—could not be found during the initial study efforts to determine the three-dimensional structure of this protein. However, the publication confirmed that these amino acids exist and are a component of proteins [8].

A superposition of the two structures is seen in *Figure 12*. In *Figure 13*, we can easily observe minor to no differences between the resulting protein structure from AlphaFold and its known structure from Protein Data Bank. So, in conclusion, and unsurprisingly, AlphaFold predicts the native Rop structure with complete accuracy.

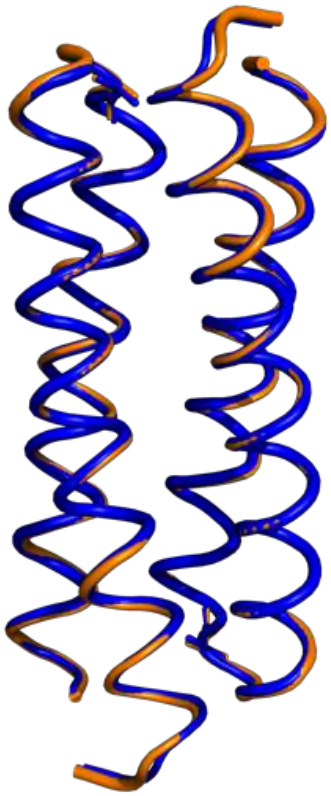

Figure 12 - Alignment of the structures of WT Rop from PDB and AlphaFold. As previously mentioned, and as we can see in the adjacent image, there are no remarkable changes in the structure of the two forms of the protein. The orange colour represents the structure of WT from PDB, and the blue colour the AlphaFold structure.

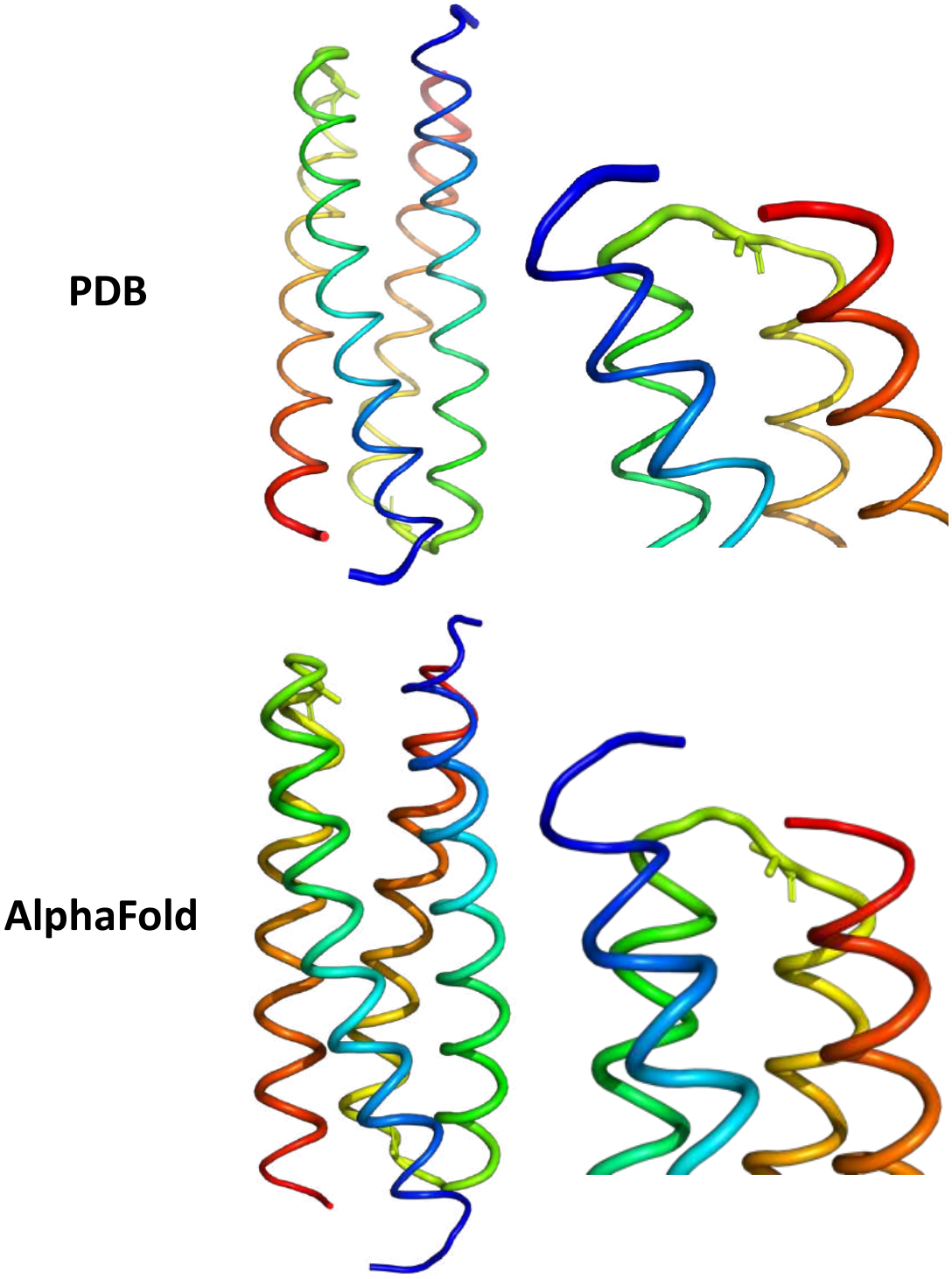

Figure 13 - Side-by-side comparison of both PDB and AlphaFold structures of the WT Rop protein.

3.2. Predicting the structure of the A31D mutant

Before comparing the known structure of A31D with the structure from AlphaFold, we have to graphically depict the structure of the mutant using the pdb file from PDB and AlphaFold separately. As far as we know, no difference must be observed when compared with the native form of the protein. The sequence given to AlphaFold was

MTKQEKTALNMARFIRSQTLTLLEKLNELDDDEQADICESLHDHADELYRSCLARFGDDGENL.

The structure of A31D consists of 2 monomers. Each monomer consists of two helices, thus forming the complete 4-α-helical bundle.

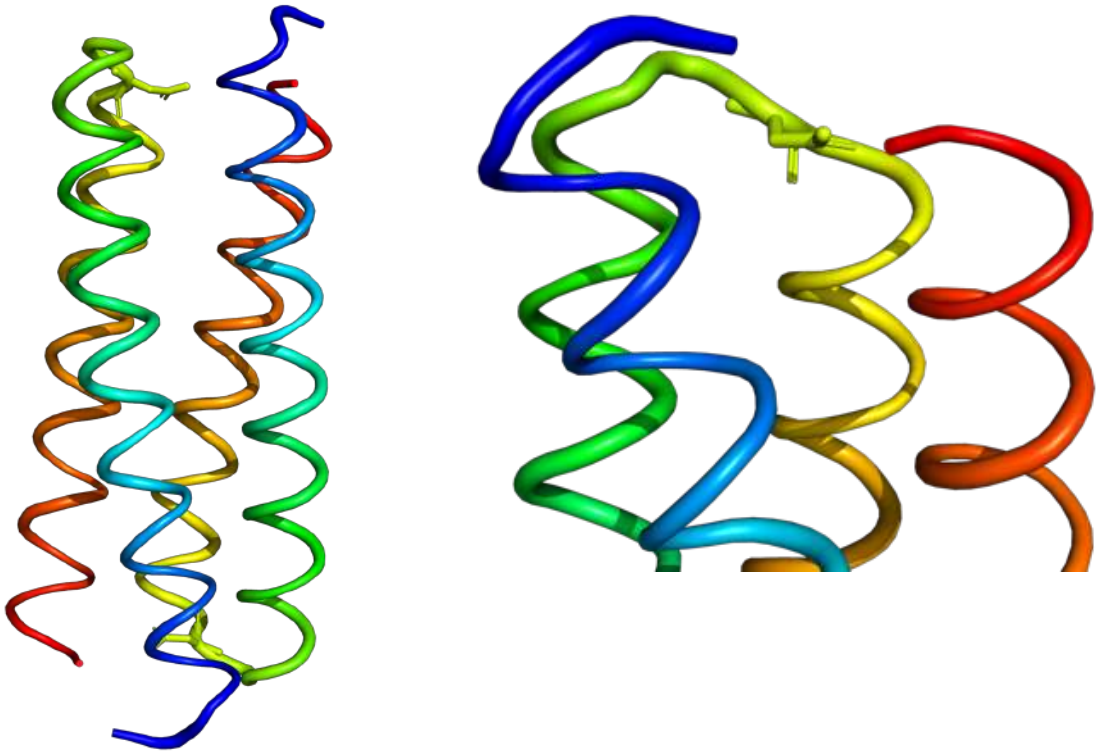

Figure 14 - The 3D structure of the A31D variant, as seen in Pymol. The resulting structure is the one after running the variant sequence in AlphaFold.
(A) The protein structure comprises two monomers/a-helices, which are antiparallel with each other, just like native Rop.
(B) The amino acid in the 31st position. In this mutation, the alanine found in the WT form of Rop is replaced by an aspartic acid residue. It is being displayed as a liquorice stick.

Table 3 displays the RMSD and TM scores for AlphaFolds' A31D and the native Rop, as obtained from MMalign:

Figure 15 - Superposition between the structures of A31D with WT Rop. Both structures resulted after running their sequences in AlphaFold. In all of the above panels (A, B, and C), no structural differences between the two forms of the Rop protein can be seen. The a-helices in both of the monomers are aligned perfectly. In the third (C) panel, we can see the mutation in A31D, where instead of an alanine residue, it has been replaced by an aspartic acid. Due to alanines' simple form, its been covered by the aspartic acid residue. Both residues are depicted as liquorice sticks for better visualization. Note that the orange colour represents the structure of WT Rop, and the blue colour represents the A31D variant.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop AlphaFold Chains 1' & 2'	126	126	0,88	0,96947	0,984
A31D AlphaFold Chains 1 & 2					

Table 3 – The data derived from comparing the pdb files of the WT and A31D. Both files that contained the protein structures came from AlphaFold. Both proteins are identical, with RMSD and TM scores denoting this identicality.

```
Chain 1:  MTKQEKTALNMARFIRSQTLTLLEKLNELDDDEQADICESLHDHADELYRSCLARFGDDGENL*
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::*
Chain 1': MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFGDDGENL*

Chain 2:  MTKQEKTALNMARFIRSQTLTLLEKLNELDDDEQADICESLHDHADELYRSCLARFGDDGENL*
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::*
Chain 2': MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFGDDGENL*
```

With an RMSD of 0.88, the aligned length between the two links is 126. The number of identical residues divided by the number of aligned residues yields a value of 0.984 for the sequence similarity between the two chains, meaning that most residues in both protein structures are identical, and a nearly perfect alignment was achieved. The TM-score value adjusted by either protein is 0.96947. Furthermore, it is evident from the above alignment that the two structures are nearly identical, which contributes to the exceptional alignment.

The A31D mutant is unique not because of its three-dimensional spatial structure but instead because, as previously stated in this subsection and the master's thesis introduction, no studies have been conducted that offer sufficient details for her. However, we can confidently state that aspartic acid (D)'s spatial structure is the same as the natural protein and has no effect whatsoever on the mutant's structure. The data in *Table 3* further supports this, as all amino acids in both structures match precisely, indicating and validating the previously proposed theory. The low RMSD and TM scores further support this.

In conclusion, as expected, AlphaFold successfully gives the same structure for the A31D as the known one found in the Protein Data Bank.

3.3. Predicting the structure of the A31P mutant

So far, we have encountered mutants whose structure is the same as that of the native Rop. Also, on these occasions, AlphaFold created identical structures, with little to no differences, from that of the wild-type protein. We believe a significant divergence between the protein's overall topology and native form must be seen. The sequence given to AlphaFold was **MTKQEKTALNMARFIRSQTLTLLEKLNELDPDEQADICESLHDHADELYRSCLARFGDDGENL**. The A31P structure differs from the WT Rops structure because it has a mutation where an alanine residue has been replaced with a proline. The outcome is a dimer

protein with an inverted U topology, meaning that one of the two monomers that make up the protein enters the cavity of the other monomer, creating an inverted U when viewed from the perspective of the other monomer.

Firstly, *Figure 16* depicts the known structure of the mutant (PDB: 1b6q). On the contrary, *Figure 17,* which shows the structure that AlphaFold created using the mutant amino-acid sequence, does not produce a protein structure similar to the A31P structure. Instead, the AlphaFold A31P protein's 3D structure is more similar to the WT Rop (again, from the AlphaFold to have an equal measure of comparison). This can be verified by inserting the two structures into the MMalign tool:

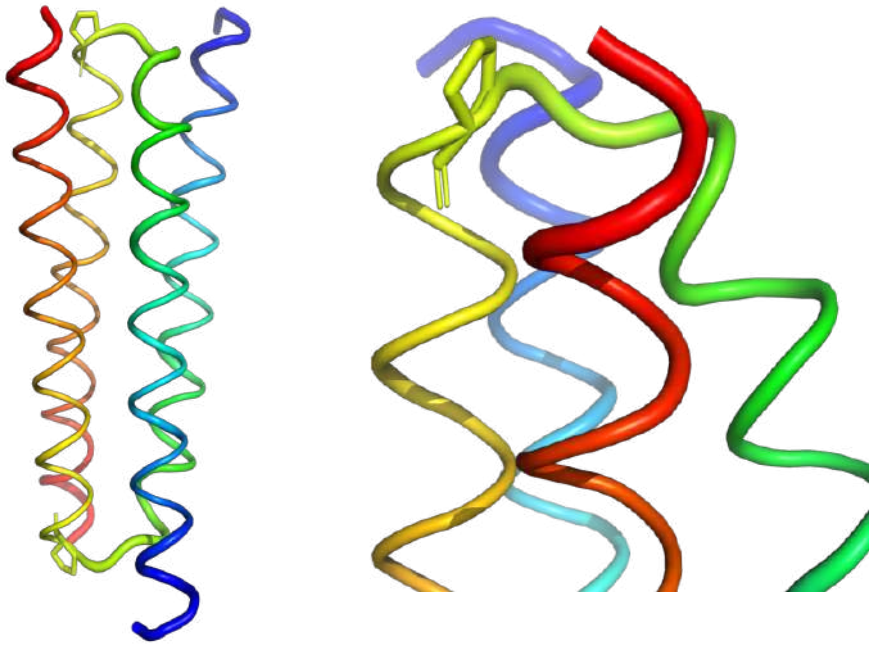

Figure 16 - The known structure of the A31P variant from PDB.
(A) The A31P mutant seems to change the protein structure dramatically compared with the WT form and the A31D variant, where the structure is the same. This structural change is known as “inverted-U” due to the two dimers forming 2 U’s and entering each other’s cavity.
(B) Rotation of the structure depicted in Panel A by 90^{O} degrees on the x-axis. The amino acid proline, located in the 31^{st} position of the variant named A31P and in the image above, is depicted as a liquorice stick.

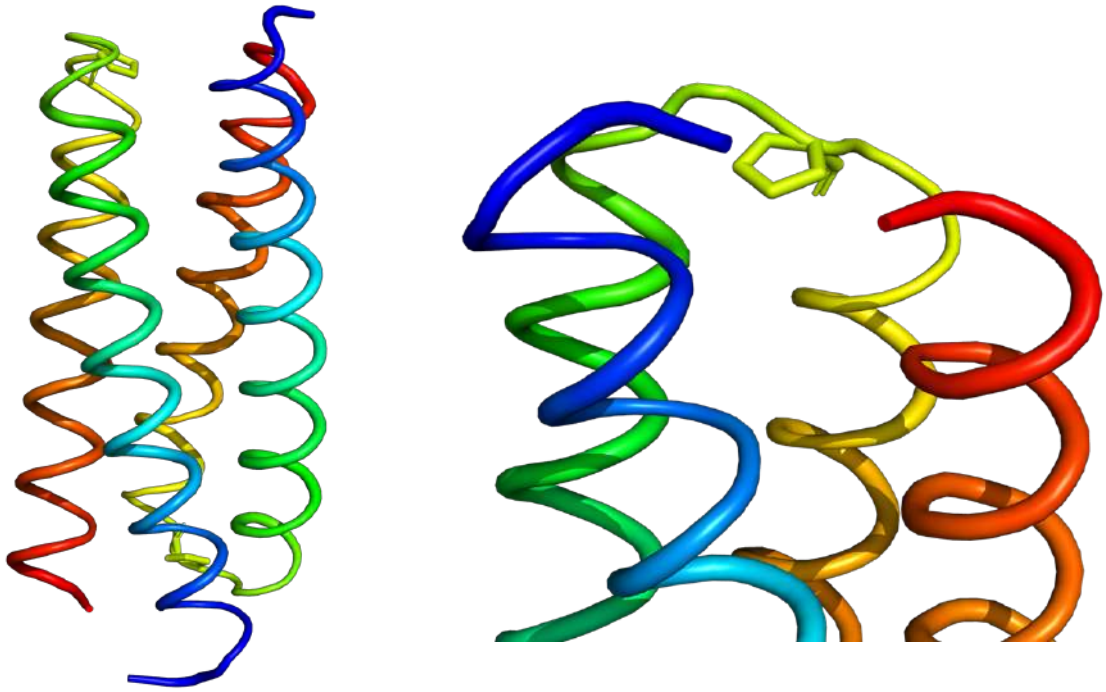

Figure 17 - This structure is a result by AlphaFold for the mutant A31P. The program does not give a result that depicts the known structure of the mutant, as shown in Panel A of Figure 16. Panel B rotates the protein by 90^{o} on the x-axis. The mutation (proline instead of an alanine found in WT) can be seen. We can easily observe that the structure does not form the characteristic “bisecting-U” formation.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop AlphaFold Chains 1 & 2	126	112	0,59	0,982	0,982
A31P AlphaFold Chains 1’ & 2’	112				

Table 3 – The data derived from comparing the pdb files of the WT and the AlphaFold’s WT Rop.

```
Chain 1:  MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFGDDGENL*
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::       *
Chain 1': MTKQEKTALNMARFIRSQTLTLLEKLNELDPDEQADICESLHDHADELYRSCLARF-------*

Chain 2:  MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFGDDGENL*
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::       *
Chain 2': MTKQEKTALNMARFIRSQTLTLLEKLNELDPDEQADICESLHDHADELYRSCLARF-------*
```

We have the WT Rop protein, which comprises chains 1 and 2, and A31P, which comprises the 1' and 2' chains. AlphaFold is the source of the pdb files that contain these polypeptide chains. The first contains 126 amino acids, while the second has 112. After superposition, the total RMSD of the two is 0.59 Å. The TM score further measures the two proteins' structural similarity. The TM-score of 0.92861 indicates a high molecular similarity between the two chains. Given that the bulk of the residues in the two structures are closely aligned, all of the data from AlphaFold, listed in Table 3 above, support the idea that these two structures are comparable. However, there is an issue. As previously stated, the structure of A31P follows a topology known as "bisecting U," which is entirely distinct from WT Rop. This appears to elude the program since, despite the apparent resemblance shown by the MMalign data, when we load the files it provides, we observe that the program has entirely overlooked the topological difference between these two proteins through the amino acid mapping. This will be fully confirmed during the following prediction.

On the other hand, when we ran the A31P variant's two known structures - the one from the PDB database with the code name 1b6q and the one that resulted from AlphaFold - we obtained the following details:

	Chain length	Aligned residues	RMSD	TM	Sequence ID
A31P (1b6q) PDB Chains 1 & 2	112	**80**	3,55	0,44780	0,750
A31P AlphaFold Chains 1’ & 2’					

Table 4 – The data derived from comparing the pdb files of the A31P, with code name 1b6q, and the AlphaFold’s A31P. Data suggest low similarities between the structures.

```
Chain 1:  MTKQEKT-ALNMARFIRSQTLTLLEKLNELDPD-EQADICES-LHDH--ADELYRSCLARF-----------*
           ::.:: :::::::::::::::::::::::: ::: ::.  :.    .                        *
Chain 1': -MTKQEKTALNMARFIRSQTLTLLEKLNELDPDEQAD-ICE-SLH--DHA-----------DELYRSCLARF*

Chain 2:  MTKQEKTALNMARFIRSQTLTLLEKLNEL-DPDEQADICES-LHDH--ADELYRSCLARF------------*
          :::::::::::::::::::::::::::: ::::::  :. ..     .                        *
Chain 2': MTKQEKTALNMARFIRSQTLTLLEKLNELDPDEQAD--IC-ESL--HDH-----------ADELYRSCLARF*
```

We appear to have two protein chains, the first being the known structure of A31P taken directly from the PDB archives and the second one being the A31P's predicted structure by AlphaFold, with each protein consisting of 112 residues. 80 residues make up the matched length between the two chains, and the sequence similarity (Seq_ID) is 0.750 with an RMSD of 3.55. Even though 80 of each chain's 112 total residues were identical, indicating some structural homology between the two chains, their RMSD value of 3.55 and TM-score of 0.4478 indicate that the degree of similarity is insignificant. Also, the topologies of the two structures appear very distinct based on the low TM-score numbers.

Examining the two structures together, we notice a significant discrepancy in their sequence alignment, which validates the conclusion drawn at the end of the previous prediction. Colon, asterisk, period, and other characters, the meaning of which we covered in part 2.2, are characters that indicate the alignment between two sequences. Moreover, there are notable differences between the amino acids comprising the sequences and those extracted from them. All of these data are tied to the RMSD score and the data regarding the overall alignment of the amino acids between the two structures. While the TM score is very low, indicating a structural resemblance, the RMSD of 3.55 indicates a significant difference in the amino acid arrangement between the polypeptide chains of these two proteins.

It is important to note that there is a high degree of alignment between the two structures' total number of amino acids, even though our RMSD score is so high. This indicates a possible low similarity between them. More specifically, only 80 of the 112 total amino acids in both proteins were successfully aligned, based on the alignment shown above (directly below *Table 4*). It also shows that both proteins' amino acid composition varies after a certain point. Additionally, as mentioned above, their structural differences are highlighted by the RMSD, TM score, and sequence ID; this suggests that while their primary structure may be similar, their tertiary structure may not be as much. *Table 4* contains all these data derived from MMalign. All of this supports our theory that AlphaFold entirely ignores the actual structure of A31P. Based

on the data, this appears to result in a protein with the A31P mutation but the WT Rop topology.

To conclude, the experimental data seems to be disregarded by AlphaFold, which instead produces a structure that is the same as the native Rop. This surprising result has the following meaning: either the program's algorithm cannot recognize the mutation or recognizes it but fails to distinguish between alanine and proline. The amino acid alanine (A) is a nonpolar, hydrophobic amino acid. The same applies to the proline (P) residue. It should be possible to differentiate between these two residues due to the prolines side chain characteristics, as it creates steric collisions with neighboring atoms. This might be one of the reasons why such a unique topological structure is formed. Though it is known that the integrity of the bend region plays a critical and significant role in the resulting protein structure, the exact reason why the specific amino acid substitution at the specific position causes this specific topological change in the protein structure has not yet been determined.

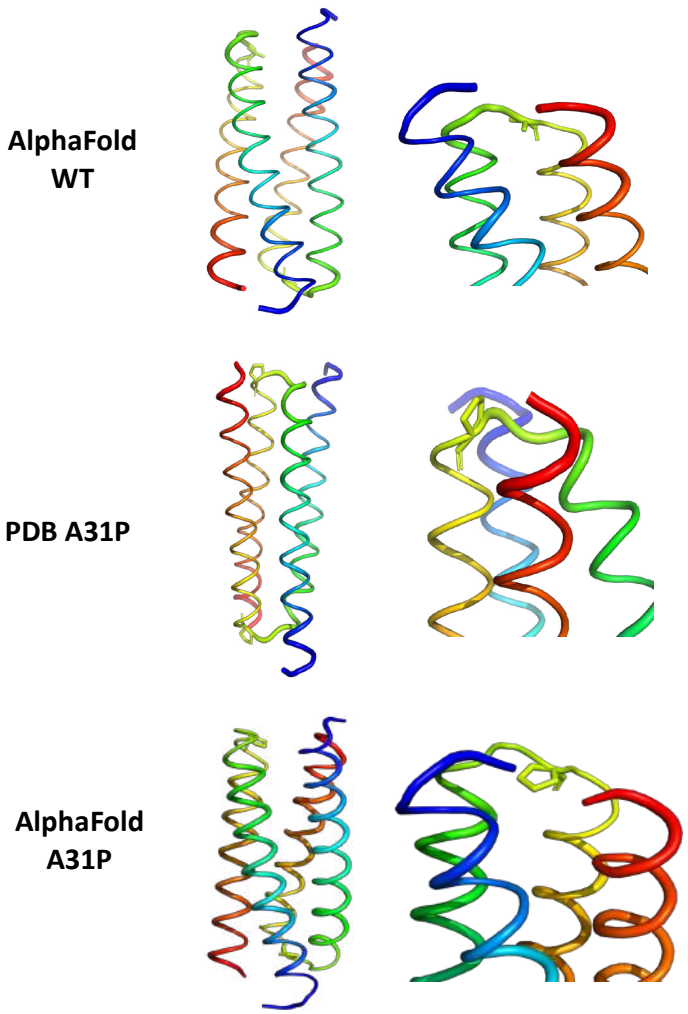

Figure 18 - A closer comparison between WT (AlphaFold) and A31P (PDB & AlphaFold) structures. AlphaFold gives a different structure, which does not represent, at the least, the structure of the known A31P mutant. Instead, the program provides a structure that looks like the WT form of the protein.

3.4. Predicting the structure of the A_2I_2 mutant

Considering the several locations in the amino acid sequence that have changed, it confirms that AlphaFold accurately predicts any potential structural alterations to the stereochemical structure of the A_2I_2 variant (PDB: 1f4n). As far as we know, there have been some significant changes to the protein's 3D structure, particularly a repacking of the hydrophobic core. The sequence that was given to AlphaFold was **GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARFGDDGENL**.

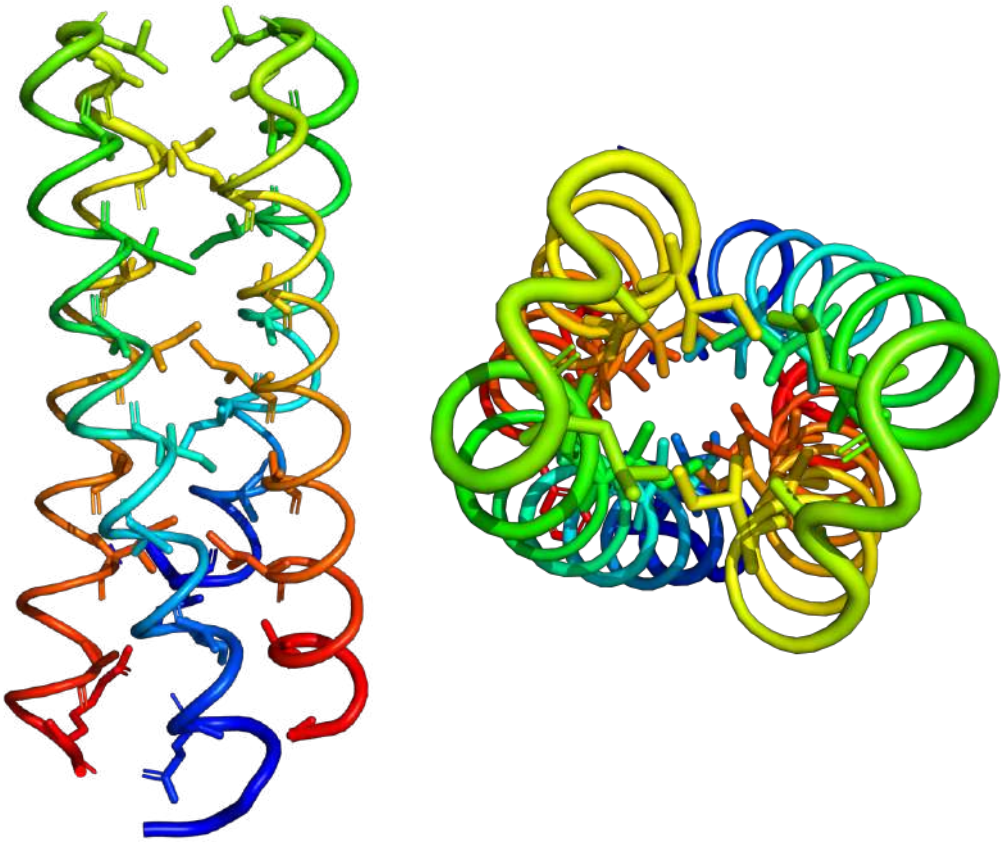

Figure 19 - The Rop A_2I_2 mutant's 3D structure as seen in PyMOL using the PDB file. In Panel A, as was already established, the protein structure consists of two α-helices or monomers, forming a syn topology, unlike native Rop. The repacked core's alanines and isoleucines may be seen as liquorice sticks. In Panel B, the mutant's structure as seen from the side of the turns. We can observe that the modifications brought about by this mutation solely impact the hydrophobic center of the protein.

As the name implies, the hydrophobic core of the variation known as A_2I_2 has been repacked. Specifically, its “a” and “d” locations are primarily altered into hydrophobic alanine and isoleucine residues. The N- and C-termini are now positioned

on the same side of the protein, but the turns are on the other side, drastically altering the protein's overall structure. This is because of the chains' new syn topology (in comparison with the N- and C-termini), as *Figure 19* suggests.

Using MMalign, the native version of Rop and this mutant were compared to determine how much the structure of this variation altered. We obtained the following findings, as seen in *Table 5*.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop (1Rop) PDB Chains 1 & 2	112	83	1,56	0,65824 (When normalized with 1Rop)	0,578
A_2I_2 (1f4n) PDB Chains 1' & 2'	109			0,67433 (When normalized with 1f4n)	

Table 5 – The data derived from comparing the pdb files of the WT Rop with the A_2I_2 protein. Different TM scores suggest differences in their structures when comparing each other structures.

```
Chain 1:  ----------------------MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARF*
                               .:  :::::::::::::::::::::::::                          *
Chain 1': KTILNMARFIRSQALTILEKANELD--ADEIADIAESIHDHADEIYRSALAR---------------------------*

Chain 2:  MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARF---*
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::   *
Chain 2': GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARFGDD*
```

The findings show that the native Rop and A_2I_2 are not structurally identical. The polypeptide chain that corresponds to the native protein has 112 residues, whereas the one that corresponds to the mutant has 109 residues, giving the two chains different lengths. An 83-residue alignment between the two chains has an RMSD of 1.56. The two chains have a sequence identity of 0.578, which indicates that in the alignment, 57.8% of the residues in the two chains are the same. These data show a weak structural similarity between the two proteins under investigation. Also, the TM-score normalized by the WT Rop length is 0.65824. The score increases to 0.67433, suggesting an increased structural similarity if the TM-score is normalized by the length 1f4n.

We will examine the resulting amino acid sequence when A_2I_2 is superimposed on top of the wild-type structure and vice versa. The data MMalign provided us during the superimposition process between these two structures are all listed in *Table 5* (the data are also mentioned in the paragraph above). If we look closely, we can see that the first monomer, which is made up of the WT Rop's chain 1 and the mutant's chain 1', is notably different from the second, which is chain 2 of WT Rop with the 2' chain of 1f4n. The first monomer's two chains lack many amino acids present in only one of them. In addition, the mutant undergoes repackaging of its hydrophobic core, which results in numerous amino acids being replaced and altered compared to the Rop protein's natural form. This could also be explained by the topologically distinct arrangement of amino acids between the two chains.

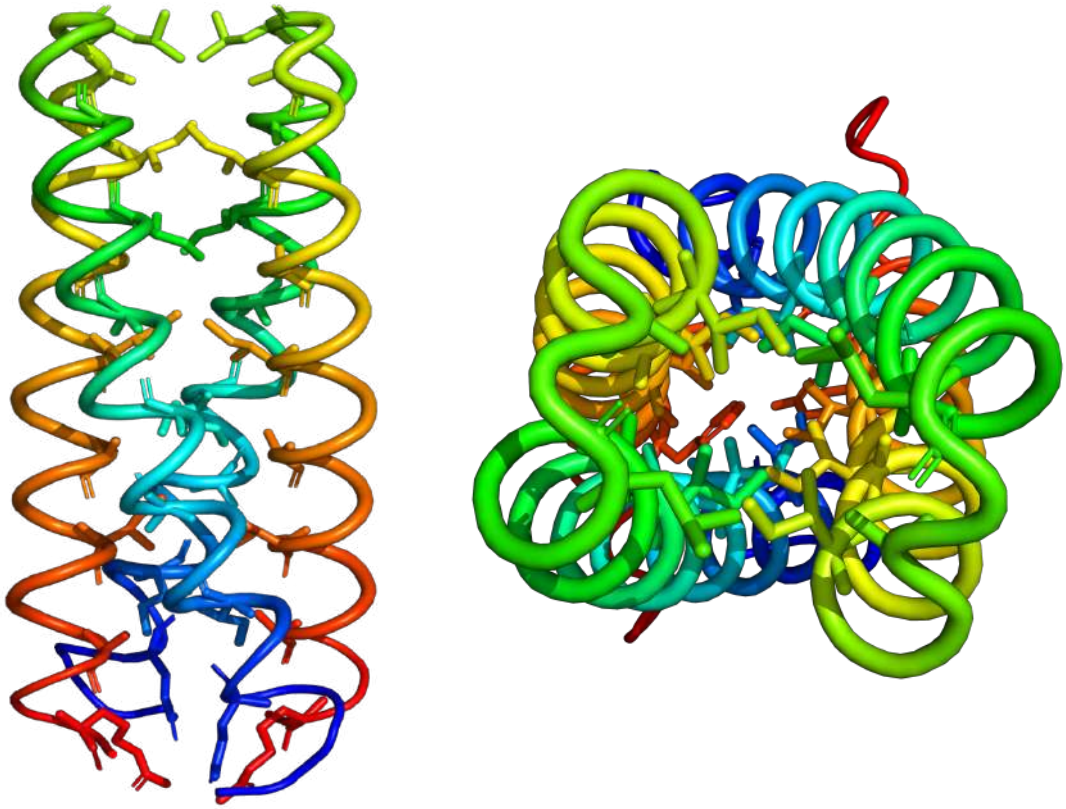

Figure 20 - The outcome of AlphaFold's 3D prediction of the mutant A_2I_2. Panel A's alanine and isoleucine residues are in a liquorice stick form. In Panel B, the repacked hydrophobic core can be seen.

We then enter the variant's amino acid sequence into AlphaFold to produce the protein structure shown in *Figure 20*. When examined closely, the known structure of the mutant and the structure provided by AlphaFold are almost identical.

This is confirmed by using MMalign to align both sequences, as seen in the results below:

	Chain length	Aligned residues	RMSD	TM	Sequence ID
A_2I_2 (1f4n) PDB Chains 1 & 2	109	107	0,86	0,94845 (When normalized with 1f4n)	0,991
A_2I_2 AlphaFold Chains 1' & 2'	126			0,82384 (When normalized with AlphaFold's A_2I_2)	

Table 6 – The data derived from comparing the pdb files of the A_2I_2 protein from PDB and AlphaFold, respectively.

```
Chain 1:  GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARF--GDD----*
          :::::::::::::::::::::::::::::::::::::::::::::::::::  :      *
Chain 1': GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARFGDD--GENL*

Chain 2:   -----KTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALAR--------*
               .::::::::::::::::::::::::::::::::::::::::::::::::        *
Chain 2':  GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARFGDDGENL*
```

These findings come from the alignment of two chains of proteins: the known structure of the mutant from PDB and the structure that AlphaFold predicted. Based on these results, the two protein structures are very similar. With a RMSD of 0.86, the alignment results in 107 out of 109 residues in 1f4n and 126 in AlphaFolds' A_2I_2. Given that the sequence identity is 0.991, 99.1% of the aligned residues in the two sequences are identical. The length of 1f4n is used to adjust the first TM-score, which is 0.94845. Normalized by AlphaFolds' A_2I_2 length, the second TM-score of 0.82384. The better overall structural similarity is suggested when 1f4n is utilized as the reference, as seen by the higher TM-score.

We may determine that these proteins are structurally similar, and AlphaFold correctly predicts the topological structure of A_2I_2 by using *Table 6* and the amino acid alignment of the two structures, listed directly below.

Using the above images, we can combine the findings and compare how the AlphaFold structure differs from the PDB one. In conclusion, AlphaFold accurately predicts the structure of this particular mutant protein with a very high score on the RMSD scale.

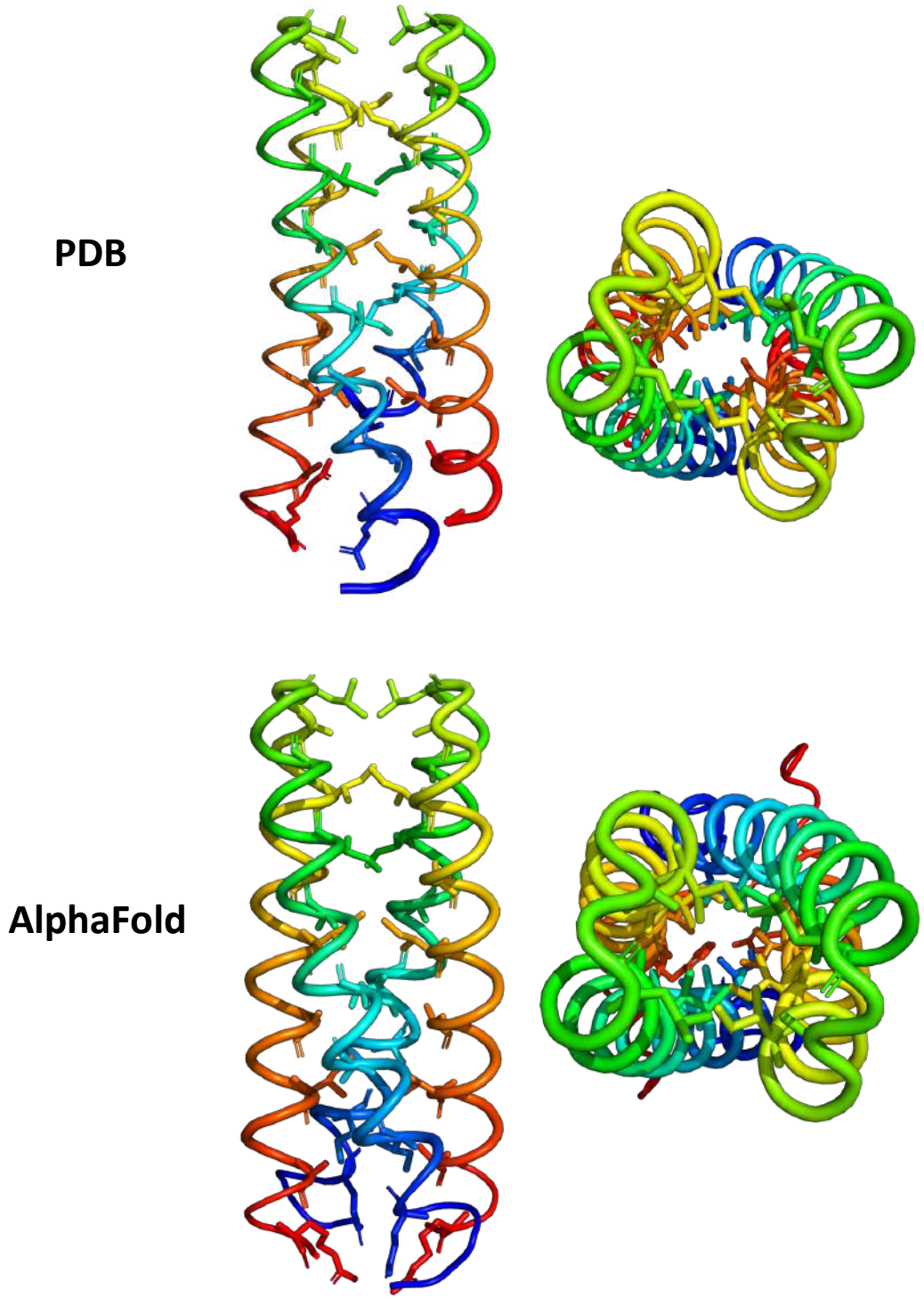

Figure 21 - A more detailed comparison of the A_2I_2 (PDB) and A_2I_2 structures (AlphaFold). Considering the details mentioned above, AlphaFold provides a highly comparable structure.

3.5. Predicting the structure of the A_2L_2 mutant

Like the other in silico experiments and previously mentioned calculations, our primary goal was to accurately foresee any potential structural alterations to the A_2L_2 sequence's stereochemical structure. The hydrophobic core of A_2L_2 has been repacked, and its a and d positions are predominantly changed into hydrophobic alanine and leucine residues, respectively. Like the A_2I_2 variant, the twists are now on the other side of the protein from where they were previously located in the WT form of Rop, significantly affecting its overall structure. Two different versions of this mutant were created so that the stereochemical characteristics of this specific mutant could be studied effectively and accurately. As suggested by the variants' name, the first variety had eight layers completely repacked with alanine and leucine residues in the a and d positions. In contrast, the second variation had its center six layers completely repacked with these residues but had left its outer two layers unchanged. In the associated publication that analyzes these mutations, these Rop protein variations were named Rop21 and Rop13, respectively.

We were anticipating dramatic modifications to the protein's 3D structure, notably a repacking of the hydrophobic core, similar to the repacking of the core of the A_2L_2 mutant. The sequence given to AlphaFold is the one found in PDB, which for the Rop21 is **GTKQAKTLLNMARFLRSQALTLLEKANELDADELADIAESLHDHADELYRSALALFGDDGENL** and for the Rop13 is **GTKQEKTLLNMARFLRSQALTLLEKANELDADELADIAESLHDHADELYRSALARFGDDGENL**.

Even though we know the variant's amino acid sequence and general 3D structure from various studies and publications, the structure has not yet been registered in the PDB. Therefore, we lack any concrete evidence of its 3D structure.

In this present-in-silico experiment, we will look at both variants (Rop21 and Rop13) and any potential structural changes they may have concerning the protein's wild-type form. We start with the Rop21 variant and continue with the Rop13.

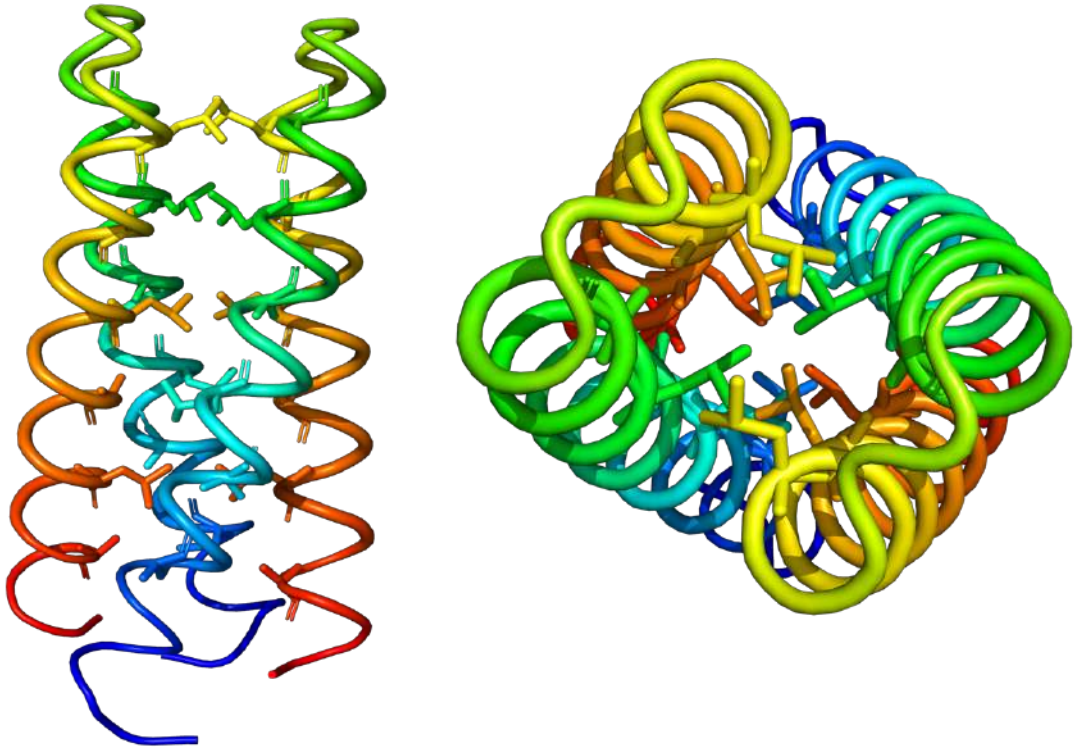

Figure 22 - Running the sequence through AlphaFold produced a more detailed depiction of the Rop21 variant 3D structure of the A_2L_2 mutant. All eight of its layers have been repacked for improved visibility, and the mutations are displayed as liquorice sticks.

In *Figure 22*, the Rop21 variant of A_2L_2 is displayed after we run its sequence to AlphaFold, and *Figure 23* depicts the structure of Rop13 of A_2L_2. The former structure, when superimposed with AlphaFolds' WT Rop, has an RMSD of 1.93 Å across 92 residues and a TM-score of 0.66477, and when superimposed with AlphaFolds A_2I_2 has an RMSD of 1.53 for 126 aligned residues and a TM-score of 0.90777. The latter protein has an RMSD of 1.55 across 89 aligned residues and a TM-score of 0.65565 when superimposed with WT, but when superimposed with A_2I_2, it has a RMSD value of 1.86 for 121 aligned residues and a TM-score of 0.86185.

To ascertain how much the structure of this variant changed, the native form of Rop and Rop21 of A_2L_2 were compared using the MMalign program. Both structures were taken after running their sequences on AlphaFold. The results from MMalign can be seen in *Table 7.*

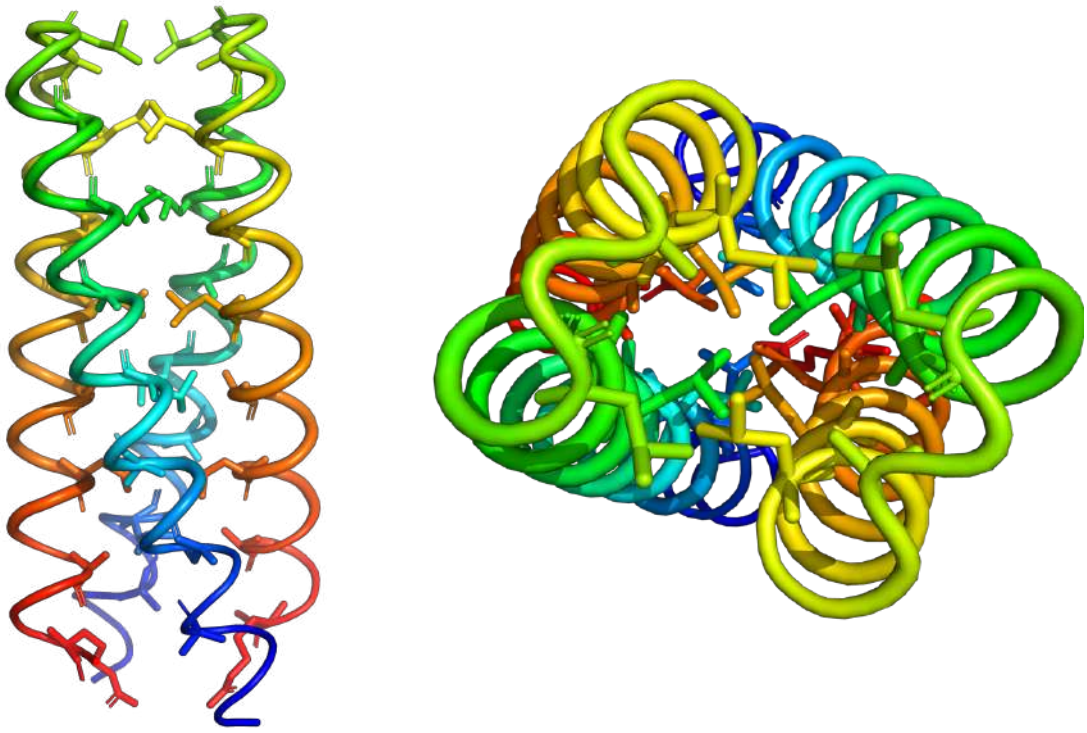

Figure 23 - The Rop13 variation of the A_2L_2 mutant's 3D structure, where the mutations can be seen in the illustrations above and are displayed as liquorice sticks; only 6 of its 8 layers have been repacked. There are no apparent distinctions between these 2 Rop protein structures when compared to one other.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop AlphaFold Chains 1 & 2	126	92	1,93	0,66477	0,609
A_2L_2 (Rop21) AlphaFold Chains 1' & 2'					

Table 7 – The data derived from comparing the pdb files of the WT Rop with the mutant A_2L_2.

```
Chain 1:  ----------------------------MTKQEKTALNMARFIRSQTLTLLEKLNELDA-DEQA---DICESLHDHADELYRSCLARFGDDGENL*
                                       :   :::::::::::::::::::::::::::. .     .                          *
Chain 1': GTKQAKTLLNMARFLRSQALTLLEKANELD---ADELADIAESLHDHADELYRSALARFGDD---GENL--------------------------*

Chain 2:  MTKQEKTALNMARFIRSQTLTLLEKLNELDAD EQADICESLHDHADELYRSCLARFGDDGENL-*
          .:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::.  *
Chain 2': GTKQAKTLLNMARFLRSQALTLLEKANELDADELADIAESLHDHADELYRSALARFGDD GEN-L*
```

First, we have the structure of the native protein as provided by AlphaFold; second, we have the structure of A_2L_2 - Rop21, which is similarly generated from AlphaFold. Since both chains comprise 126 residues, they are similar in size and should have a similar overall structure. 92 residues in both chains may be overlaid with a

distance less than 5.0 Å, according to the aligned length of 92 residues. The aligned residues of the native protein and Rop21 have an average deviation of 1.93 Å according to the RMSD scale. The sequence identity in this instance is 0.609, meaning that 60.9% of the aligned residues are the same in both chains. Based on the TM-score of 0.66477, both proteins have many similarities. However, as the alignment reveals, the two chains, especially the first monomer of both proteins, have noticeable variations.

Table 7, previously discussed in the previous paragraph, is shown directly above the data produced by MMalign A_2L_2, especially Rop21, with WT Rop. Additionally, directly below *Table 7* is the precise amino acid match between these two structures. The two sequences do not align correctly over their whole length, even with the comparatively low RMSD and TM scores. Similar to the A_2I_2 mutant, this is probably the result of the mutant's hydrophobic core repackaging. Because of this, there is a significant difference in the overall number of amino acids correctly aligned between the two structures—only 92 of the 126 amino acids in both proteins have been successfully aligned.

According to some other experiments, due to isoleucine's and leucine's similar properties, the structure of the variants A_2L_2 and A_2I_2 are the same. The structures generated by AlphaFold for the A_2L_2 variant and the A_2I_2 mutant exhibit a significant level of similarity when closely examined. This is supported by the alignment of both sequences using MMalign, as seen in the table below.

Next, we will address the superposition of the A_2L_2 with its corresponding A_2I_2 resulting from AlphaFold. Given that both chains are identical in length at 126 residues, their size and expected overall structure are comparable. Given that both chains have an aligned length of 126, all residues may be overlaid. The aligned residues of Rop21 and A_2I_2 have an average variation of 1.53 Å, according to the RMSD scale. Since the sequence identity in this instance is 0.889, 88.9% of the aligned residues in both chains are the same. The length of the reference structure, in this case Rop21, having a length of 126 residues, is used to normalize the TM-score. With a TM-score of 0.90777, both proteins are very structurally similar.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
A_2L_2 (Rop21) AlphaFold Chains 1 & 2	126	126	1,53	0,90777	0.889
A_2I_2 AlphaFold Chains 1' & 2'					

Table 8 – The data derived from comparing the .pdb files of the A_2L_2, specifically the variant Rop21, with that of A_2I_2. Both structures were derived from AlphaFold.

```
Chain 1:  GTKQAKTLLNMARFLRSQALTLLEKANELDADELADIAESLHDHADELYRSALARFGDDGENL*
          .::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::*
Chain 1': GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARFGDDGENL*

Chain 2:  GTKQAKTLLNMARFLRSQALTLLEKANELDADELADIAESLHDHADELYRSALARFGDDGENL*
          .::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::*
Chain 2': GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARFGDDGENL*
```

We may conclude that these two structures are comparable and that the program accurately predicts the structure of A_2L_2-Rop21 because of the topological similarities between these two mutant proteins and the highly similar amino acid sequence between them. *Table 8* lists the MMalign data, with the amino acid alignment of those data just below.

We obtain the data below using the MMalign software to compare the Rop13 variant of the A_2L_2 mutant to the WT Rop, where only six layers are repacked in its hydrophobic core.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop AlphaFold Chains 1 & 2	126	89	1,55	0,65565	0,584
A_2L_2 (Rop13) AlphaFold Chains 1' & 2'					

Table 9 – The data derived from comparing the pdb files of the A_2L_2, specifically the variant Rop13, with that of WT Rop.

```
Chain 1:  MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFGDDGENL--*
          :::::::::::::::::::::::::::::::::::::::::::::::::::: :: :.  *
Chain 1': GTKQEKTLLNMARFLRSQALTLLEKANELDADELADIAESLHDHADELYRSALARFG-DD-GENL*
Chain 2:  ----------------------------MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFGDDGENL------*
                                     :   ::::::::::::::::::::::::.                                            *
Chain 2': GTKQEKTLLNMARFLRSQALTLLEKANELD---ADELADIAESLHDHADELYRSALARFG--------------------------------DDGENL*
```

The A_2L_2 - Rop13 structure has the native protein's structure superimposed on it. Both structures result from loading each protein's sequences into AlphaFold and running them independently. The length of both proteins is 126 residues. A successful alignment of 89 residues from both chains is shown by the aligned length, which is 89. The RMSD score is 1.55, which indicates that the two structures differ significantly. The number of identical residues divided by the number of aligned residues is used to compute the sequence identity. The sequence identity in this instance is 0.584, which indicates that 58.4% of the aligned residues are the same. When AlphaFold's WT Rop and Rop13 lengths are normalized, the TM-score is 0.65565. This suggests that the two structures have a fair amount in common structurally. However, there is a catch, which we will discuss in the paragraph that follows. When we look at the depiction of the alignment, we can see that there are gaps where the residues are not aligned. The comparatively high RMSD number indicates considerable structural variations between them, as demonstrated by the MMalign findings.

As demonstrated by the comparison of WT Rop vs Rop21, as seen in *Table 7*, our findings indicate a structural similarity, but the alignment disproves this theory. Is this comparable to the aforementioned in purpose? The mutant's topology alters due to the hydrophobic core repackaging, which also affects the alignment of the amino acids. This is seen in how chain 2 of WT and chain 2' of Rop13 align.

The Rop13 variation's structure was then compared to the A_2I_2 variant's structure, as we had done with the Rop21 variant, to identify any potential changes. The information we obtained is shown below:

	Chain length	Aligned residues	RMSD	TM	Sequence ID
A_2L_2 (Rop13) AlphaFold Chains 1 & 2	126	121	1,86	0,86185	0.860
A_2I_2 AlphaFold Chains 1' & 2'					

Table 10 – The data derived from comparing the .pdb files of the A_2L_2, specifically the variant Rop13, with that of A_2I_2. Both structures were derived from AlphaFold.

```
Chain 1:  GTKQEKTLLNMARFLRSQALTLLEKANELDADELADIAESLHDHADELYRSALARFGD-DGENL-*
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::: ...   *
Chain 1': GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARFGDDGEN--L*

Chain 2:  GTKQEKTLLNMARFLRSQALTLLEKANELDADELADIAESLHDHADELYRSALARFGD-DG-ENL-*
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::. .  .   *
Chain 2': GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARFGDDG-EN--L*
```

According to the MMalign results, Rop13 will be superimposed on A_2I_2. Both mutants have a length of 126 residues. The aligned length is 121, indicating that 121 residues from both chains were successfully aligned.

There are noticeable similarities between the two sequences, as can be seen by comparing the data in Table 10 with the sequence alignment directly below it. Because both Rop13 and A_2I_2 are resultant proteins of AlphaFold, their structures are very similar, as evidenced by their nearly perfect alignment of amino acids, low TM score, and relatively low RMSD (remember that structures between 1 and 3 Å on the RMSD scale are considered relatively comparable).

The RMSD value is 1.86 Å, which suggests moderate differences between the structures. The sequence identity is 0.860, meaning that 86.0% of the aligned residues are identical, and the TM-score is 0.86185 when normalized by the length of either mutant. This indicates a moderate level of structural similarity between the two structures. In conclusion, the MMalign results show significant structural differences, as indicated by the relatively high RMSD value.

In conclusion, the structure of the A_2L_2 variant is very similar to that of the A_2I_2 mutant. No noticeable differences were found when comparing the two versions created to research their topological properties and stability. The main distinction between A_2I_2, A_2L_2 (Rop21 and Rop13), and the wild-type version of Rop is that the two latter proteins form a syn topology with their 4-α-helices. In contrast, the former protein forms an anti-topology. If we ignore this one distinction, *Figure 24's* representation of the three variations, A_2I_2, A_2L_2 - Rop21, and A_2L_2 - Rop13, shows that their 3D structures are nearly identical.

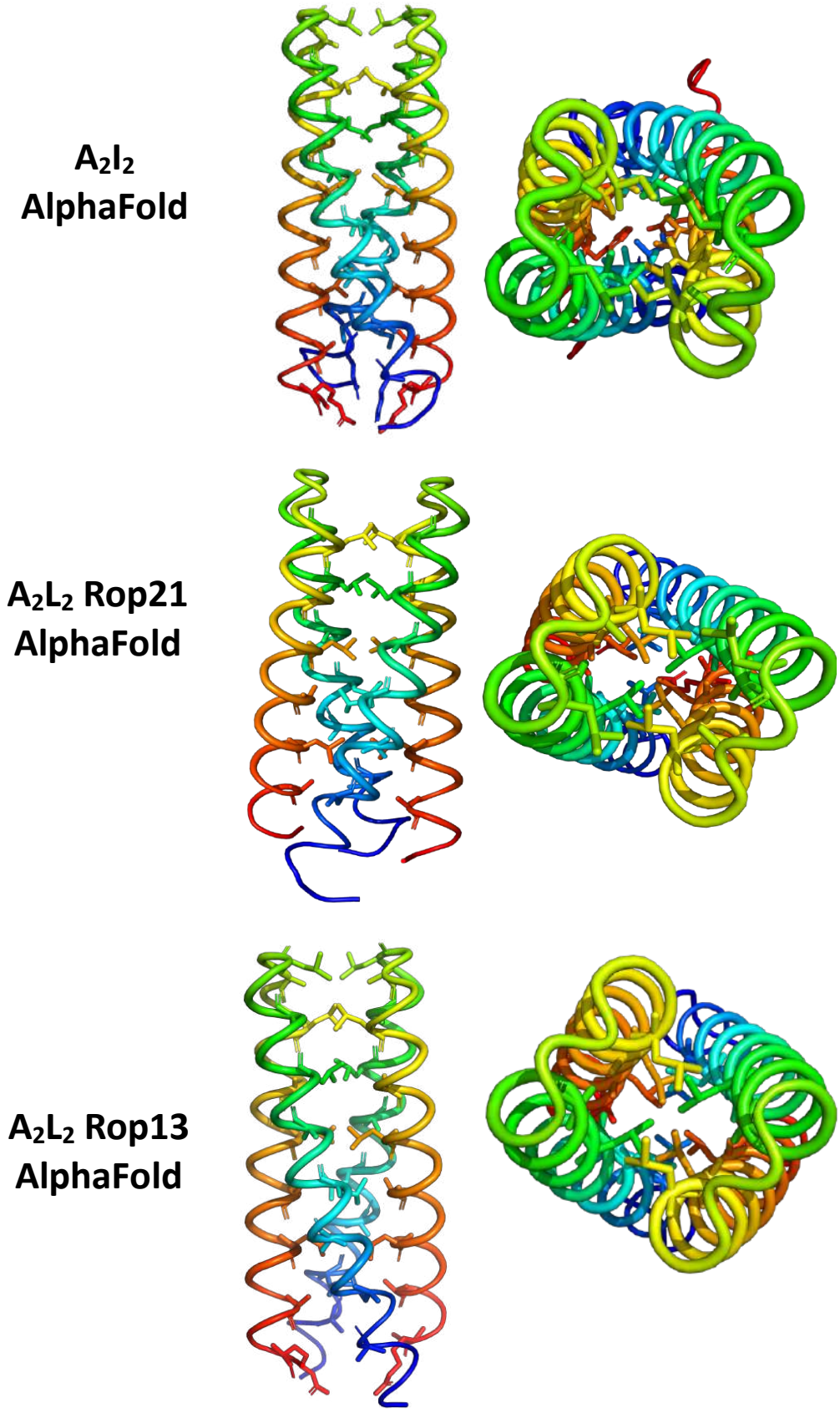

Figure 24 - A comparison between the nearly identical variants named A_2I_2, A_2L_2-Rop21, and A_2L_2-Rop13, respectively (from top to bottom).

3.6. Predicting the structure of the Cys-free mutant

Another mutant that is worth mentioning is that of Cys-free or C38A C52V (PDB: 3k79). "Cys-free variation" refers to a particular form of the Rop protein in which the cysteines in positions 38 and 52 have been replaced by alanine and valine residues, respectively. This variation, like WT Rop, has all of its cysteines in the secondary α-helix of its monomer (2 and 2' α-helix), as seen in *Figure 25*. The 3D structure appears to be the same as the WT one despite the replacement of these residues. The only notable alterations are a bit of curvature and the absence of residues that make up the α-helix in the mutant's N- and C-termini, constituting its monomers. The sequence given to AlphaFold, which can be found in PDB is: **GTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADIAESLHDHADELYRSVLARFGDDGENL**.

The native protein's degree of similarity to the Cys-free variant was calculated using MMalign, just like it was for the other variations. *Figure 25* displays the known structure of the known structure. *Figure 26* depicts the structure of the native protein obtained by AlphaFold. A superposition of the two structures is seen in *Figure 27*. According to MMalign, regarding the superposition between the known structure and the AlphaFold one, the RMSD across 57 residues is 0,86 Å. In more depth detail about the results given by MMalign is as follows:

	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop (1Rop) PDB Chains 1 & 2	112	112	1,01	0,97630 (When normalized with 1Rop)	0,946
Cys-free (3k79) PDB Chains 1' & 2'	114			0,95943 (When normalized with 3k79)	

Table 11 – The data derived from comparing the pdb files of the WT Rop with that of Cys-free. Almost identical alignment and TM scores, indicating high structural similarity.

```
Chain 1:  MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARF-*
          .:::::::::::::::::::::::::::::::::::::::::::::::::::::: *
```

```
Chain 1': GTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADIAESLHDHADELYRSVLARFG*

Chain 2:  MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARF-*
          .:::::::::::::::::::::::::::::::::::::::::::::::::::::: *
Chain 2': GTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADIAESLHDHADELYRSVLARFG*
```

There are 112 and 114 residues in 1Rop and 3k79, which is the known structure of the mutant, respectively. The aligned residues in the two chains have an RMSD of 1.01 Å. Lower RMSD values indicate higher structural similarity between the two chains. The sequence identity between the two chains is 0.946, which means that 94.6% of the aligned residues are identical in the two chains. The TM-score can be normalized by the length of the native's Rop or Cys-free sequence to provide a more meaningful comparison between structures. The TM-score of 0.97630, if normalized by the length of WT, indicates a high degree of similarity between the two chains. Similarly, the TM-score of 0.9943, if normalized by Cys-free length, also indicates a high degree of similarity between the two chains. These results suggest that both structures have a high degree of structural similarity, with a high TM-score, sequence identity, and RMSD. The alignment of the amino acids that comprise these two structures and the data in Table 11 support this hypothesis.

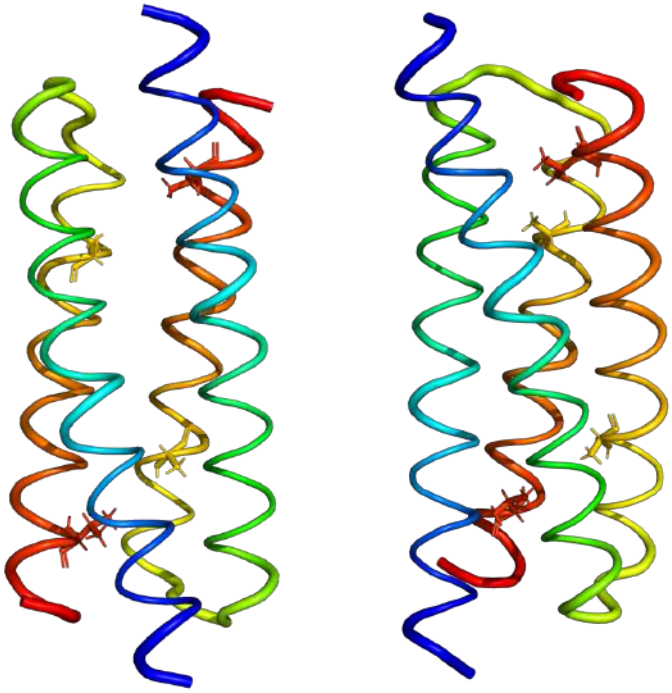

Figure 25 – Illustration of the 3k79 mutant's structure from PDB. The cysteine residues in both monomers' helices 1 and 1' are portrayed as liquorice sticks. Panel B displays the identical protein with an x-axis rotation of 90 degrees. When compared to the native protein, not many differences are immediately apparent (Figure 10).

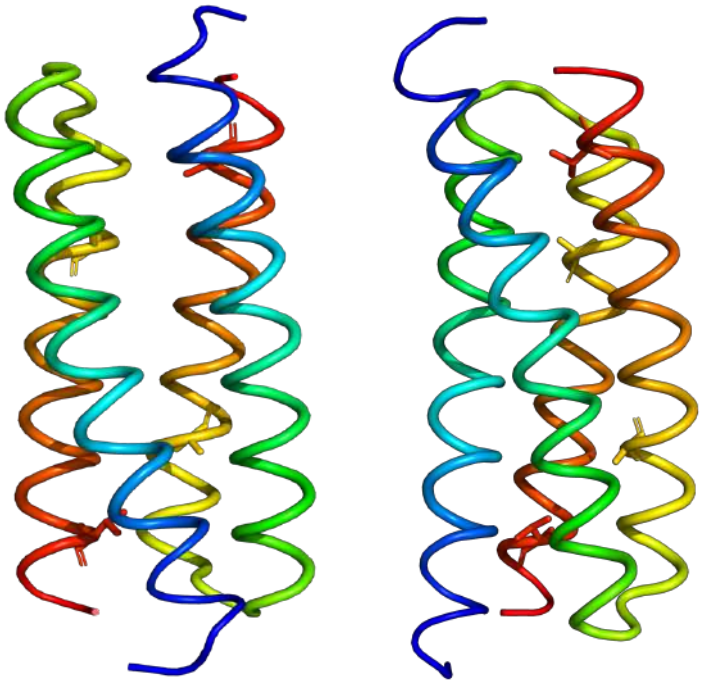

Figure 26 - The 3D structure of the C38A C52V mutant as seen in PyMOL when the AlphaFold file is utilized. The protein structure is represented identically in both panels with no notable differences. This is supported by the results of MMalign, which show that these two proteins are identical in their amino acid sequence with a low RMSD and perfect sequence identity.

We then enter the variant's amino acid sequence into AlphaFold to produce the protein structure shown in *Figure 26*. When examined closely, the known structure of the mutant and the structure provided by Alpha Fold are almost identical.

We utilize both structures' pdb files in the MMalign program to verify this similarity, and the findings are seen in *Table 12*.

The structure generated by AlphaFold and the known structure of the protein are in alignment. There are 114 residues in the former protein and 126 in the latter. It appears to be a successful alignment in 144 of the total amino acids in these two proteins. The positions of the comparable residues in the two structures differ by an average of 0.88 Å, or root mean square deviation (RMSD), between the two. A lower RMSD value indicates a better structural similarity between the two structures. Sequence identity (Seq ID) of 1.000 between the two structures means that all aligned residues are identical. In this case, the TM-score, normalized by PDB's 3k79, is 0.97999, indicating

a high similarity between the two structures. Although slightly lower than the initial TM-score result, AlphaFold's 3k79 TM-score of 0.88771 does not show a significant structural difference.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
Cys-free (3k79) PDB Chains 1 & 2	114	114	0,88	0,97999 (When normalized with 3k79)	1,000
Cys-free AlphaFold Chains 1' & 2'	126			0,88771 (When normalized with AlphaFold's Cys-free)	

Table 12 – The data derived from comparing the Cys-free PDB files from PDB and AlphaFold.

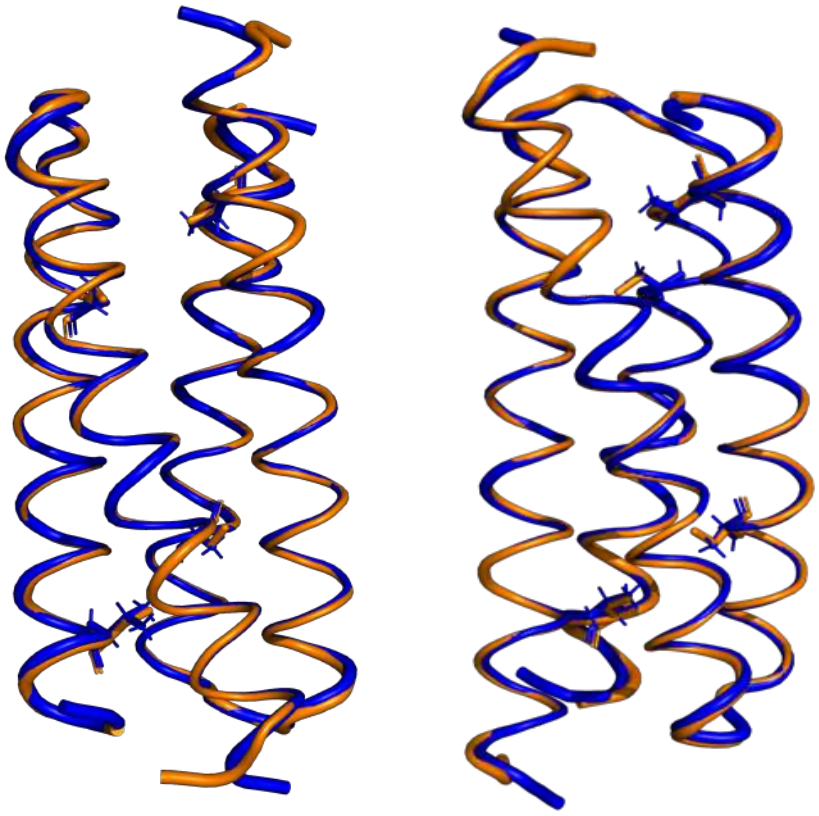

Figure 27 - A superposition of the WT version of the protein and the variation C38A C52V. As far as we can tell, there are no obvious variations between the variant's structure and the native structure. They are oriented in the same direction; however, distinct residues are present at positions 38 and 52.

```
Chain 1:  GTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADIAESLHDHADELYRSVLARFG------*
          .:::::::::::::::::::::::::::::::::::::::::::::::::::::::      *
Chain 1': GTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADIAESLHDHADELYRSVLARFGDDGENL*

Chain 2:  GTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADIAESLHDHADELYRSVLARFG------*
          .:::::::::::::::::::::::::::::::::::::::::::::::::::::::      *
Chain 2': GTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADIAESLHDHADELYRSVLARFGDDGENL*
```

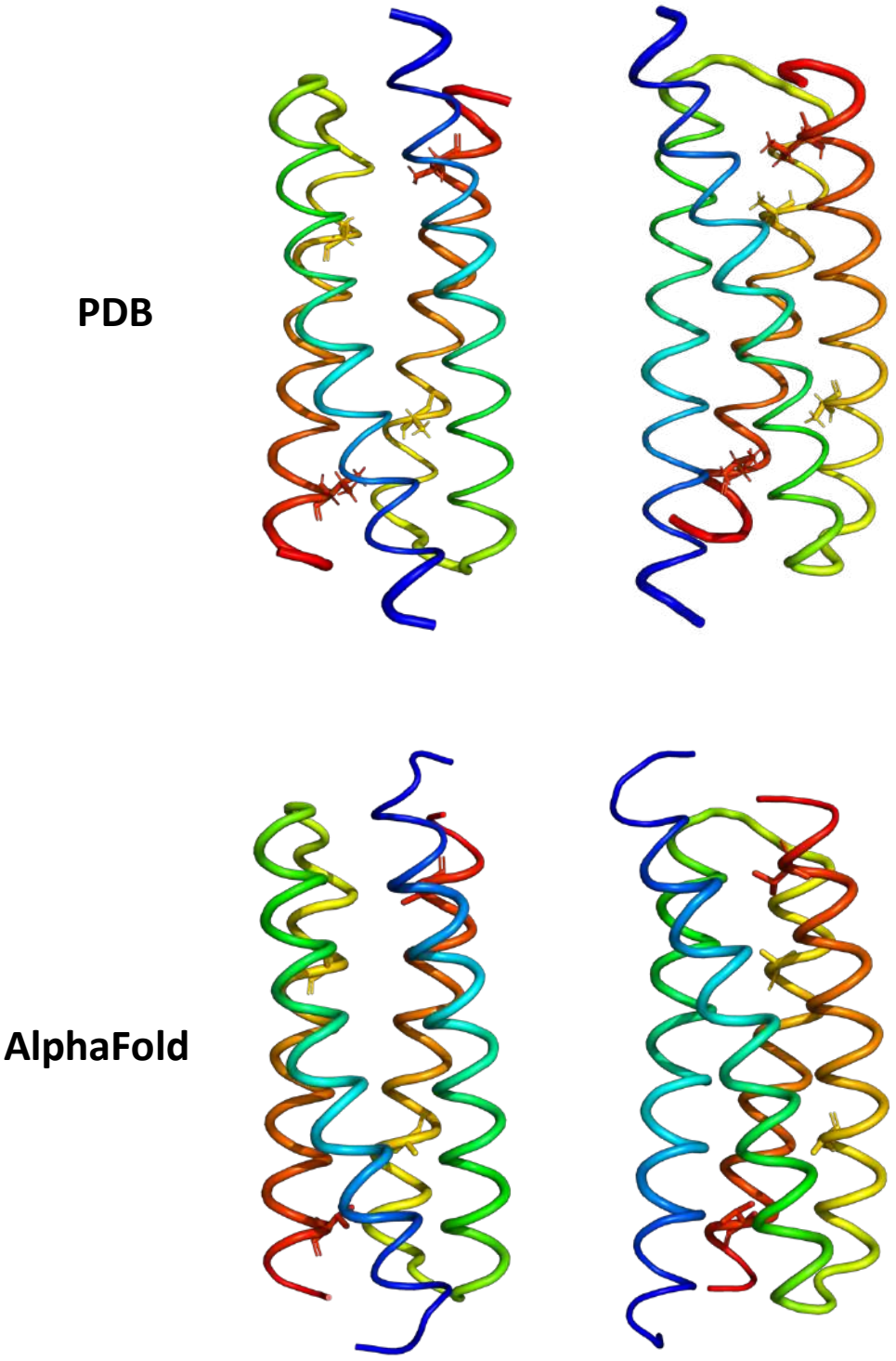

Figure 28 - A closer comparison of the structures resulted by AlphaFold with those stored in PDB.

We may infer notable similarities between the two protein sequences based on the data shown in *Table 12*, which displays an excellent TM and RMSD score. We conclude that these two protein structures are almost identical based on the alignment of the two sequences. Because the final seven amino acids in the Cys-free amino acid sequence are absent, we can observe a slight divergence in the order of chains 1 and 1' and 2 and 2'.

Like most of the mutants we mentioned earlier, AlphaFold accurately predicts the structure of this mutant with a very high score on the RMSD scale and a perfect TM score.

3.7. Predicting the structure of the $\Delta_{30\text{-}34}$ mutant

Another one of the many Rop mutants is $\Delta_{30\text{-}34}$ (PDB: 1qx8), as shown in *Figure 29*. This variation eliminates the turn by deleting five residues from the dimer's turn region, which links the two α-helices. Due to this loss, the mutant now only has four α-helices, each consisting of 58 amino acids. The known sequence given to AlphaFold, representing this protein, is:

MTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARFGDDGENL

Following that, we use AlphaFold to create the protein structure seen in *Figure 30,* which depicts the structure of the native protein obtained by AlphaFold by entering the variant's amino acid sequence. When closely compared, the known structure of the mutant and the structure supplied by AlphaFold are almost similar, as shown in the findings from MMalign below. A superposition of the two structures is seen in *Figure 31*.

According to MMalign, regarding the superposition between the known structure and the AlphaFold one, the RMSD across 98 residues is 0,81 Å.

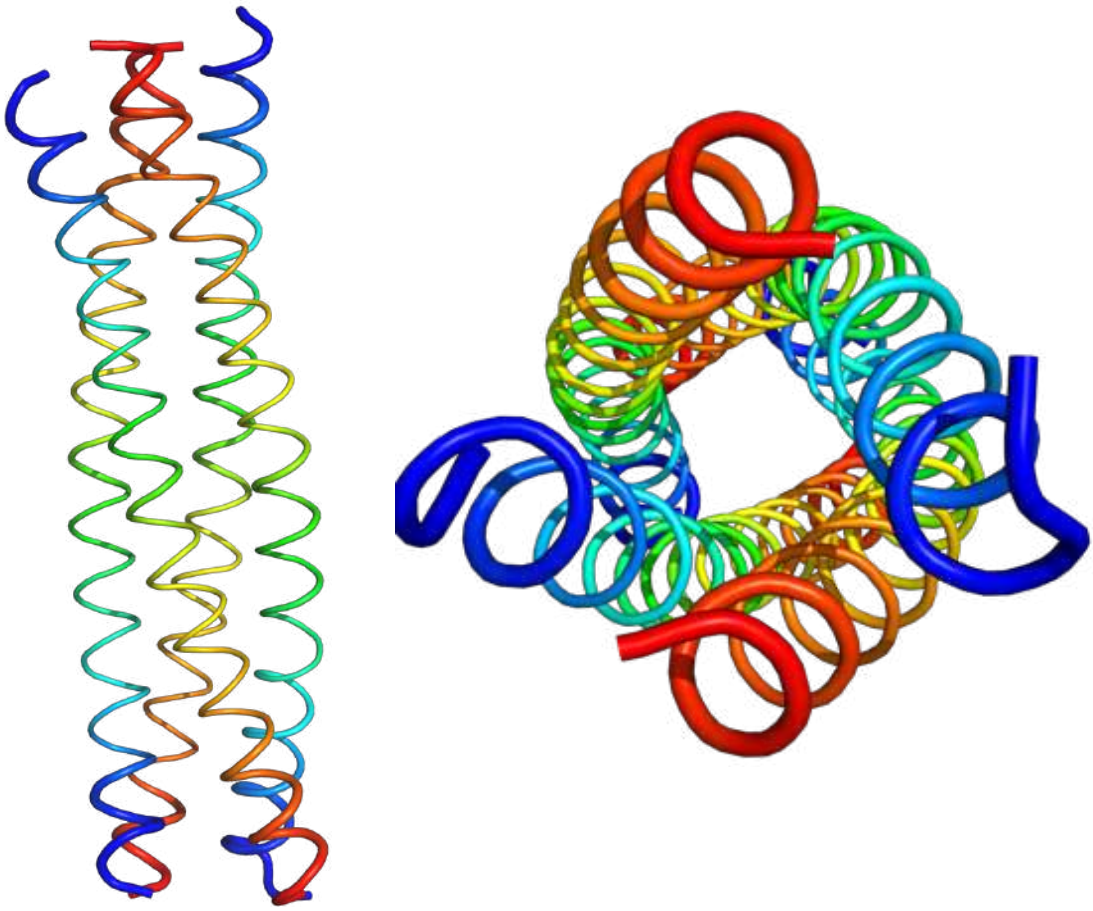

Figure 29 - $\Delta_{30\text{-}34}$ mutant's 3D structure as it appears in PyMOL when the PDB file is used. In Panel A, like all earlier iterations of the Rop protein, the natural form's structure comprises four identical a-helices rather than two monomers with one a-helix each. Panel B has the same design as panel A, but as seen from above (rotated 90^{O} on the y-axis).

	Chain length	Aligned residues	RMSD	TM	Sequence ID
$\Delta_{30\text{-}34}$ (1qx8) PDB Chains 1 & 2	196	196	0,91	0,97172 (When normalized with 1qx8)	1,000
$\Delta_{30\text{-}34}$ AlphaFold Chains 1' & 2'	232			0,82431 (When normalized with AlphaFold's $\Delta_{30\text{-}34}$)	

Table 13 – The data comparing the pdb files of $\Delta_{30\text{-}34}$ from PDB and AlphaFold.

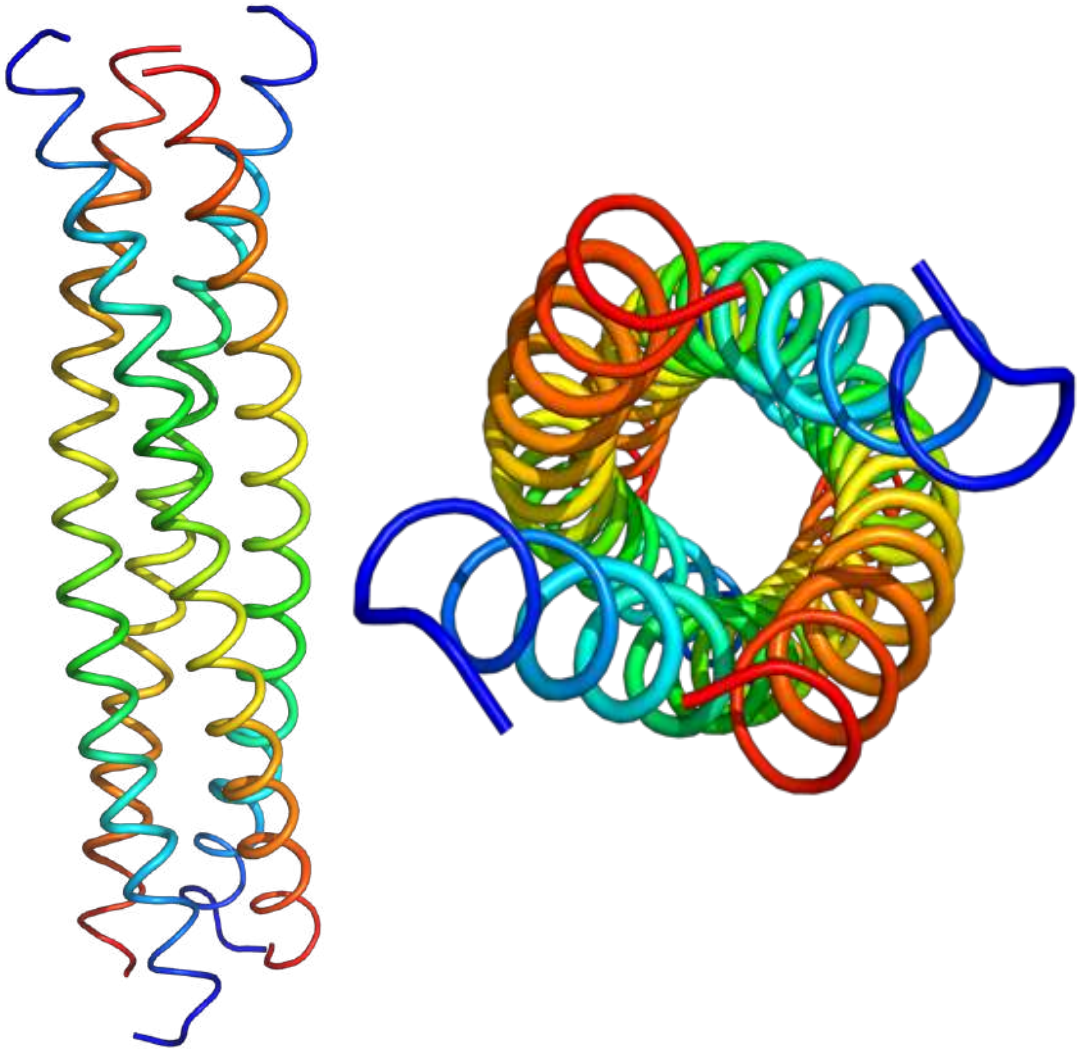

Figure 30 - The variation's structure, as provided by AlphaFold. In Panel A, AlphaFold gives four identical α-helices, just like the known structure of the mutant. The same structure is depicted in Panel B, but it is rotated by 90^{O} on the y-axis.

Furthermore, to verify the similarities and/or the differences between these two structures of the $\Delta_{30\text{-}34}$, we insert the residue sequence into MMalign, and we get the following results:

```
Chain 1:  ----EKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARF-------*
              ::::::::::::::::::::::::::::::::::::::::::::::::       *
Chain 1': MTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARFGDDGENL*

Chain 2:  MTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARF-------*
          ::::::::::::::::::::::::::::::::::::::::::::::::::::       *
Chain 2': MTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARFGDDGENL*

Chain 3:  ----EKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARF-------*
              ::::::::::::::::::::::::::::::::::::::::::::::::       *
Chain 3': MTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARFGDDGENL*
```

```
Chain 4:  MTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARF-------*
          :::::::::::::::::::::::::::::::::::::::::::::::::::       *
Chain 4': MTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARFGDDGENL*
```

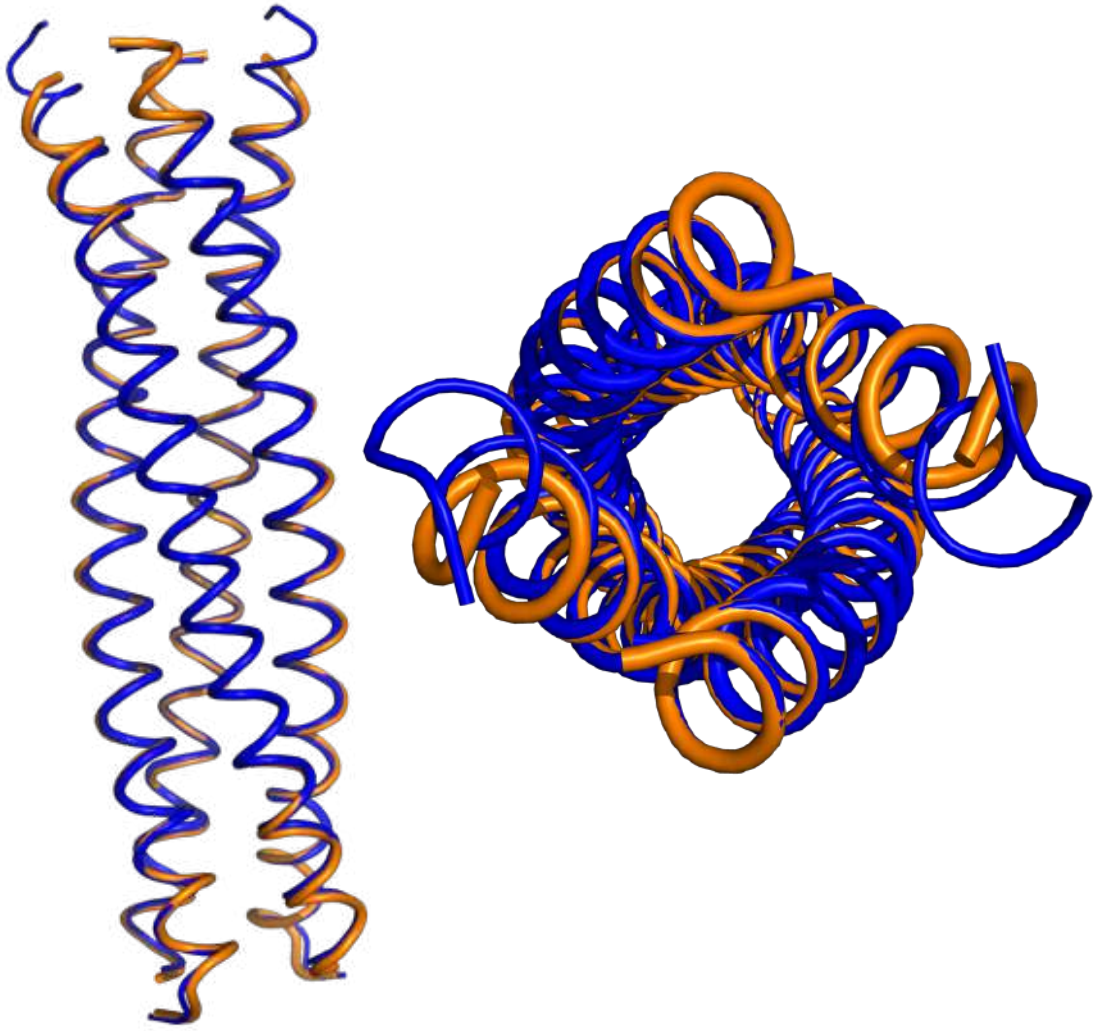

Figure 31 - A superposition between the PDB and AlphaFolds structure of the $\Delta_{30\text{-}34}$ variant. Along its axis, we can see that both proteins are almost entirely aligned.

The structural alignment between AlphaFold's $\Delta_{30\text{-}34}$, which represents AlphaFold's structural prediction of the protein, and PDB's $\Delta_{30\text{-}34}$, with the code name 1qx8, can be seen in the above image. Additionally, the program produces a protein with 232 amino acids, whereas the original form of this mutant has 196. Every residue in 1qx8 could line up with the structure that AlphaFolds predicted. In this case, the RMSD is 0.91, indicating a close structural similarity between the two chains. Every residue in the aligned portions is identical, as indicated by the Seq ID, which in this case is 1.000 and assesses similarities at the amino acid sequence level. When

normalized by the length of 1qx8, the TM-score is 0.97172, indicating a substantial degree of structural similarity between the two chains. When normalized by AlphaFold's protein, the TM-score is 0.82431; a lower value. The alignment shows that most of the residues in the aligned sections of the two chains are the same, with very few dissimilarities.

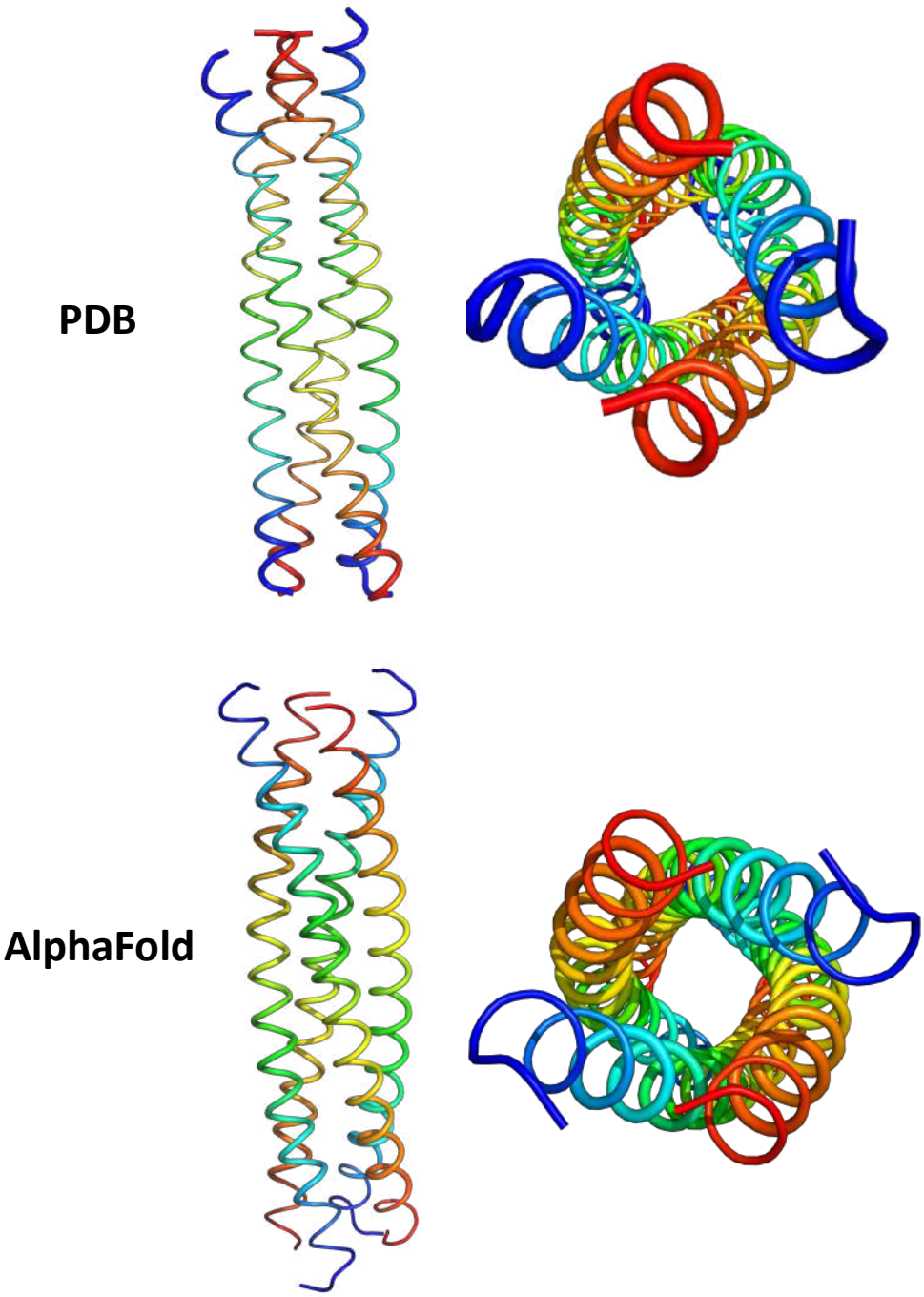

Figure 32 - A comparison of the variant's known structure, which we have been investigating in our eighth computation and the structure we obtain from running its sequence through AlphaFold.

As previously indicated, the $\Delta_{30\text{-}34}$ mutant is one of the most unique mutants we have examined and employed for this master's thesis. This is because, unlike WT Rop and most other mutants, it comprises four single helices rather than two monomers. Their alignment also makes this evident. In every previous instance, we contrasted monomer A's helices 1 and 2 with monomer B's helices 1' and 2', respectively. Four distinct helices are superimposed here and compared to the equivalent of the other protein structure. Despite this, the program appears to have accurately predicted the structure of $\Delta_{30\text{-}34}$, as evidenced by the data obtained from MMalign, as seen in *Table 13*, and the alignment of the amino acids comprising the two almost identical protein structures. Some of the data that support this are the low TM and RMSD scores and the high alignment of residues between the two sequences.

Overall, the findings and the data imply that the known structure of this variation and the one obtained using AlphaFold have extremely comparable overall structures. At the same time, there may be discrepancies in some locations because of the two chains' different lengths.

3.8. 2aa mutant

Last but not least is the mutant 2aa (PDB: 1nkd), in which two alanine residues were inserted at the 30th and 32nd positions. The known sequence for this polypeptide is **MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFGDDGE NL**. Two alanine residues, colored red in the sequence mentioned above, are positioned in the loop region of the protein at positions 30 and 32 of the amino acid sequence. The mutant has the same 4-α-helix bundle as the WT structure of Rop and the same anti-topology, so there are no immediately noticeable differences between it and the WT structure.

Figure 33 displays the known structure of the 2aa mutant. The topology of the variant produced by AlphaFold is shown in *Figure 34*. *Figure 35* displays the

superposition of PDB's WT and 2aa proteins, whereas *Figure 36* displays the superposition of PDB's 2aa and AlphaFold. According to Mmalign, regarding the superposition between the known structure and the AlphaFold one, the RMSD across 112 residues is 0,57 Å.

When comparing the known structures of the protein's WT and 2aa forms using the MMalign tool, the following information is returned by the program:

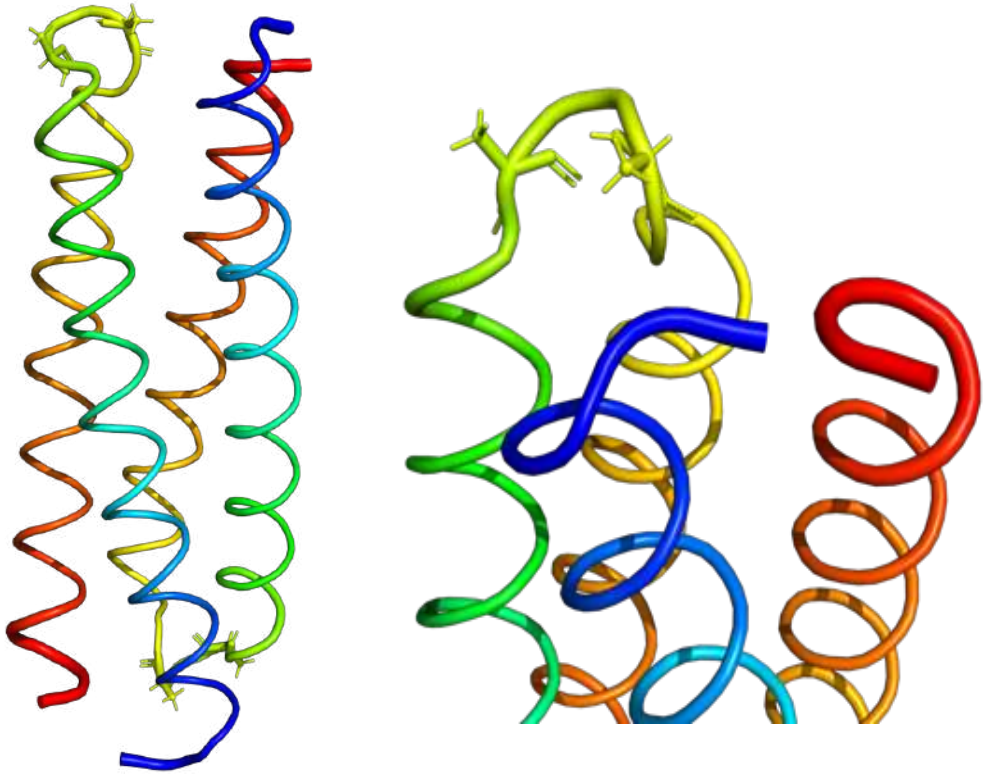

Figure 33 - An illustration of the mutant name 2aa's known structure's 3D topology. The two alanine residues are represented as liquorice sticks, as is better seen in Panel B.

These results result from two protein chains' structural alignment, with 1Rop superimposed over 1nkd. The aligned regions of the two chains have 112 residues in common. The two chains' structural divergence is relatively small, as the RMSD value of 0.57 shows. Most aligned residues are the same between the two chains, with a high sequence identity of 0.982. The two chains exhibit high structural similarity, as indicated by the TM-score of 0.97939, normalized by the length of one Rop. When normalized by the length of 1nkd, the TM-score of 0.93065 is lower, suggesting that the native protein rather than Chain_2 provides a more accurate reference structure

for this structural alignment. The 0.57 RMSD score indicates a slight difference between the two topologies. Furthermore, as can be seen from the image above, the high sequence identity of 0.982 indicates that the two protein sequences are almost the same, differing only in a few residues.

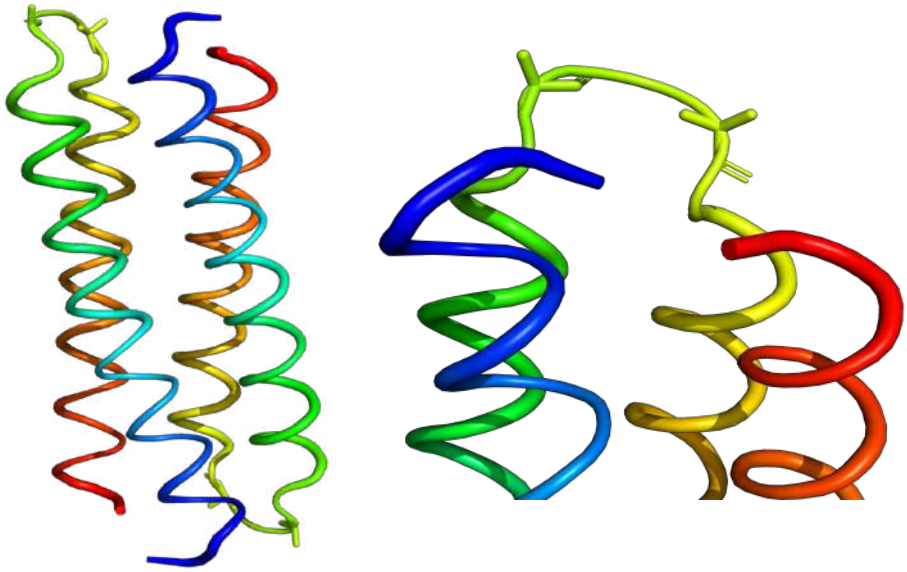

Figure 34 - The outcome of AlphaFold's 3D prediction of the mutant 2aa. Panel B shows a closer look at the mutation in the loop region.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop (1Rop) PDB Chains 1 & 2	112	112	0,57	0,97939 (When normalized with 1Rop)	0,982
2aa (1nkd) PDB Chains 1' & 2'	118			0,93065 (When normalized with 1nkd)	

Table 14 – The data derived from comparing the pdb files between WT Rop and the mutant named 2aa.

```
Chain 1:  MTKQEKTALNMARFIRSQTLTLLEKLNELD--ADEQADICESLHDHADELYRSCLARF-*
          ::::::::::::::::::::::::::::::  :::::::::::::::::::::::::: *
Chain 1': MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFG*


Chain 2:  MTKQEKTALNMARFIRSQTLTLLEKLNELD--ADEQADICESLHDHADELYRSCLARF-*
          ::::::::::::::::::::::::::::::  :::::::::::::::::::::::::: *
Chain 2': MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFG*
```

Based on our current understanding and the studies conducted to compare these two structures, the topological structures of 2aa and WT must be nearly identical. One exception is the alteration caused by inserting these two alanine residues, which will be noticeable in the loop area and illustrated in *Figure 35*, which shows the superposition of the known structures of WT Rop and 2aa. Despite this, the stability and overall similarity of the two proteins remain unchanged, as indicated by the MMalign results mentioned above.

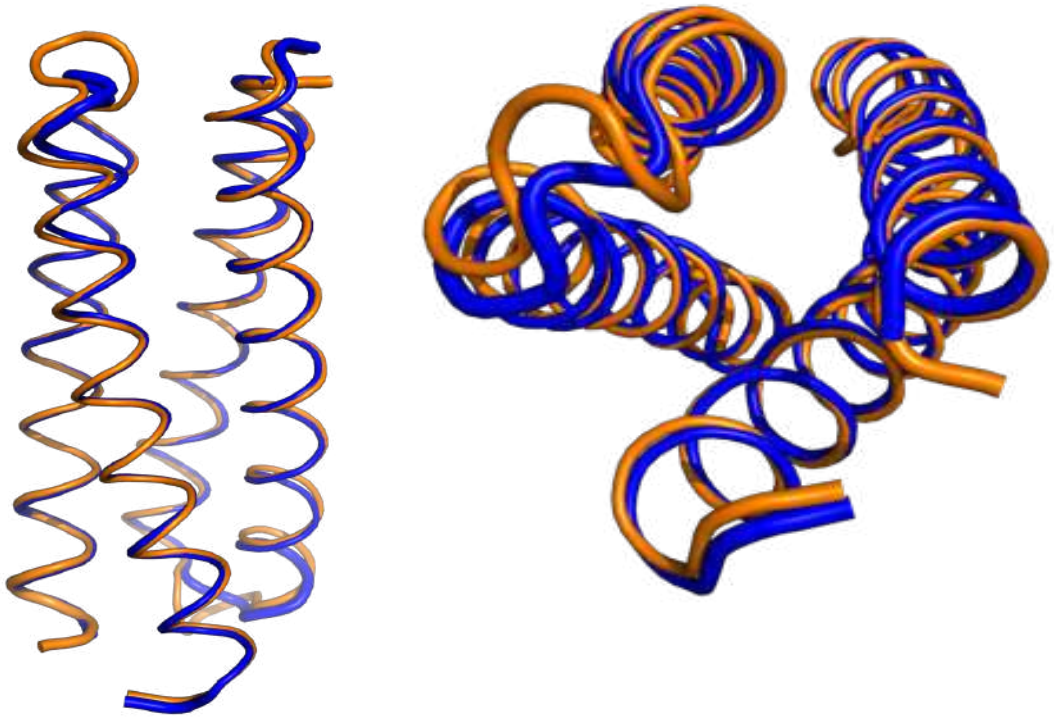

Figure 35 - Superposition of the 2aa and WT Rop structures. PDB was used to get both topological structures. Apart from the loop area, they are precisely aligned along their sequence for the reasons we discussed.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
2aa (1nkd) PDB Chains 1 & 2	118	118	0,89	0,96450 (When normalized with 1nkd)	1,000
2aa AlphaFold Chains 1' & 2'	130			0,87799 (When normalized with AlphaFold's 2aa)	

Table 15 - The data derived from comparing the .pdb files between the two variations of 2aa, the one resulting from AlphaFold and the other one is the known structure taken from PDB.

```
Chain 1:  MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFG------*
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::::      *
Chain 1': MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFGDDGENL*

Chain 2:  MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFG------*
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::::      *
Chain 2': MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFGDDGENL*
```

Next, we combined the structure predicted by AlphaFold's algorithm with the structure of the known 2aa. These outcomes come from superimposing the structural alignment of two protein chains, 1nkd, over AlphaFold's 2aa. While AlphaFolds' 2aa has 130 residues, 1nkd has only 118. Based on the aligned length of 118, all of the residues in 1nkd were aligned with those in AlphaFold's 2aa. The RMSD value of 0.89 indicates a significant difference between the two structures, which is to be expected. The alignment of all the residues between the two structures was identical, as evidenced by the sequence identity of the aligned residues, which is 1.000. This result indicates that conformational changes or model errors are more likely to cause the differences between the two structures than sequence modifications. Normalized by PDB's 2aa, the TM-score is 0.96450; when normalized by AlphaFold's, it is 0.87799.

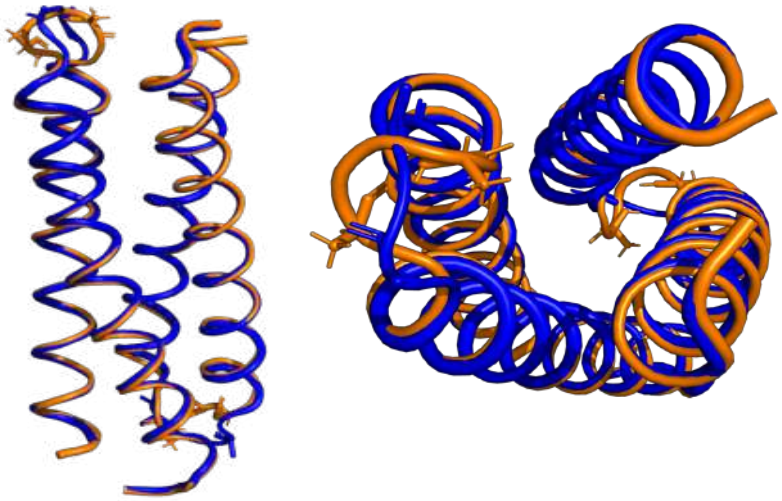

Figure 36 - A superposition of the 2aa from the PDB with the resulting structure from AlphaFold and are represented by the orange and blue colors, respectively. The two extra alanine residues inserted into positions 30 and 32 are displayed as liquorice sticks for better visualization. We can see the direct contrast between the two structures' identical α-helices as opposed to the loop area, where we can see that they are entirely out of alignment.

As mentioned above, along the axis of the chains, we can see that both structures are perfectly aligned. The only difference in the overall superposition is the

variant's different orientation compared with the known structure of 2aa due to the insertion of these two alanine residues.

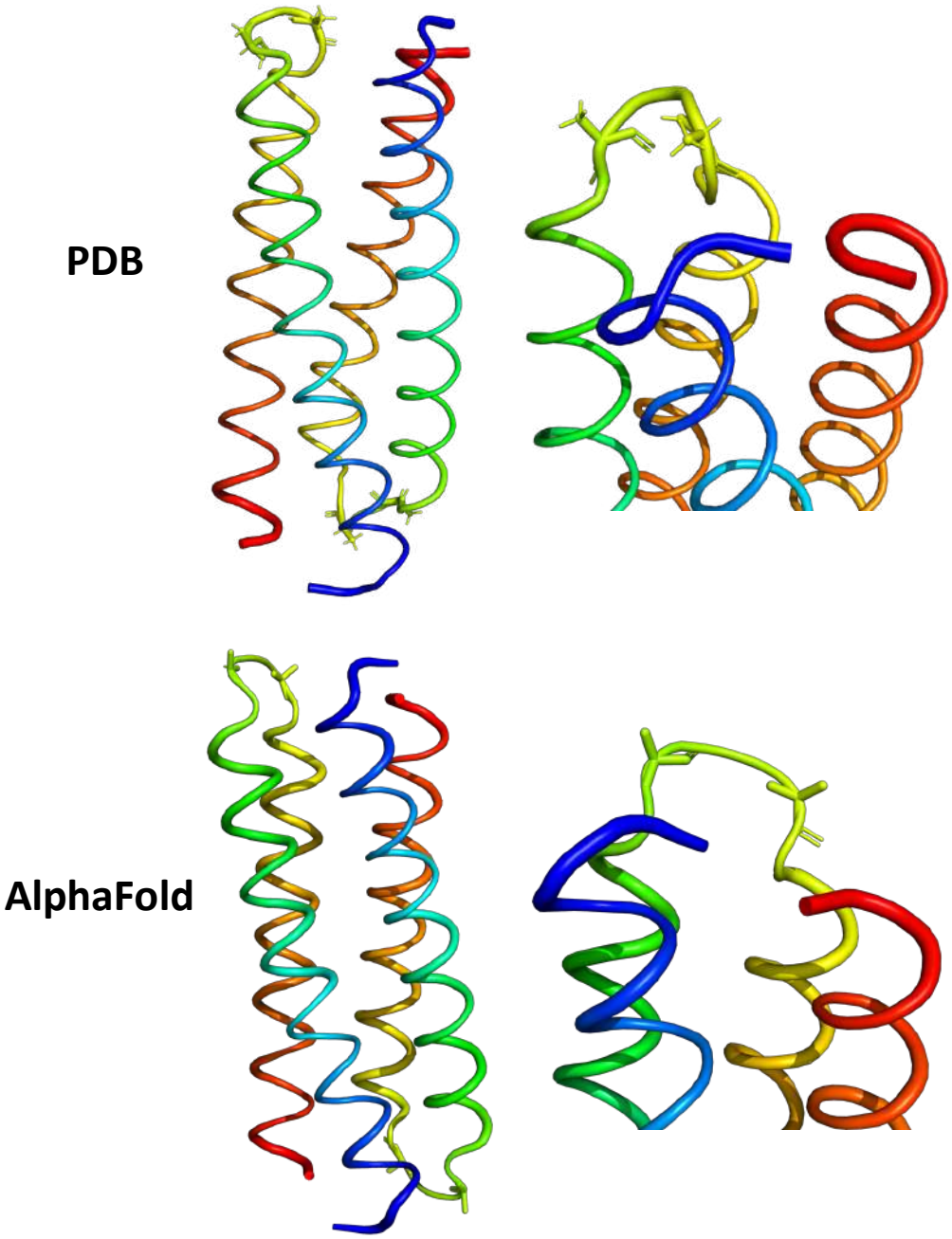

Figure 37 - A side-by-side comparison of the topological structure of 2aa, where Panel A represents the mutant's known structure and Panel B represents the structure predicted by AlphaFold. To better visualize and observe the curvature of the loop region and the two alanine residues, the second photos of both panels are not identical.

The program findings show structural differences between the crystal structure and the model, indicating that these differences are most likely due to conformational changes or model errors despite the aligned residues having perfect sequence identity, according to the MMalign results. These conformational changes appear to be caused by the loop's altered orientation in the structure produced by AlphaFold, which closely resembles the orientation that develops in most variant structures, including WT. As we previously saw with the A31P variation, when the protein's overall structure is drastically altered, AlphaFold produces a result that is more like the structure of the WT version of Rop than the mutant that we are studying.

4. Conclusions

Since the mutants of the Rop protein were created by humans rather than by natural selection, as was previously noted, this thesis aimed to test AlphaFold's ability to forecast the architectures of these altered proteins. This allows us to assess the "power" of the program's algorithm and the degree of accuracy of its findings, as we are already aware of the protein structures of the mutant proteins we are looking at. Only one of the seven mutants, A31P, has a different topology and three-dimensional structure than the normal Rop protein. 2aa, on the other hand, exhibits some minor differences in its bend region but is nothing too noticeable. Even if some of the hydrophobic cores of the remaining five mutants have been repacked, they still exhibit a relatively high similarity index (RMSD and TM scores, total aligned amino acids, and sequence ID) to that of native Rop.

The conformational changes that appear to occur in A31P once the 31st residue is swapped from alanine to proline are most likely caused by the unique structure of proline. As a consequence, the two residues (Leu29 & Asp30) right before proline (31P) have their ϕ and ψ angles altered, thus creating this peculiar "bisecting U" structure [50]. It also strengthens the notion and belief that, although the turns in a protein's structure are crucial and fundamental to how the protein folds in three dimensions, a

single alteration in the amino acid sequence might have unexpected effects [50,74,75]. As may be seen from the results, AlphaFold incorrectly guesses its structure. Consequently, the mutation appears to be missed by the program's algorithm, or it cannot distinguish between proline and alanine.

Furthermore, the 2aa mutant is the other of the two mutant proteins with localized and minor changes in the turn region in relation to the wild-type Rop. The comparison suggests that the protein structure results from AlphaFold are pretty accurate and closely resemble those known from the PDB. The slight differences could be attributed to the mutant's changed angle, potential model errors, or conformational changes. It may be necessary to modify the program's algorithm or conduct additional testing to uncover the reasons behind these discrepancies. Also important to note is the fact that 2aa maintains a native-like topology even after correcting its heptad discontinuity with two insertions in the loop region. On the contrary, $\Delta_{30\text{-}34}$ topology completely changes and adopts the known homotetrameric structure when the discontinuity is corrected, but this time through deletions in the bend region. It’s possible that it takes into consideration evolutionary information alongside information provided by PDB to predict right its unique structure rather than predicting a native anti-topology, just like it did with 2aa.

In summary, considering the rapid advancement of technology, especially in the field of protein structure prediction—especially given the latest data acquired by the AlphaFold program [76]—it is logical to believe that the day when we will be able to predict protein structures with accuracy and precision is not too far off.

References

1. Pauling, L.; Corey, R. B.; Branson, H. R. The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. Proceedings of the National Academy of Sciences 1951, 37 (4), 205–211. https://doi.org/10.1073/pnas.37.4.205.
2. Lupas, A. N.; Gruber, M. The Structure of α-Helical Coiled Coils. Fibrous Proteins: Coiled-Coils, Collagen and Elastomers 2005, 37–38. https://doi.org/10.1016/s0065-3233(05)70003-6.
3. Crick, F. H. C. Is α-Keratin a Coiled Coil? Nature 1952, 170 (4334), 882–883. https://doi.org/10.1038/170882b0.
4. Crick, F. H. C. The Packing of α-Helices: Simple Coiled-Coils. Acta Crystallographica 1953, 6 (8),689–697. https://doi.org/10.1107/s0365110x53001964.
5. Branden, C. I.; Tooze, J. Introduction to Protein Structure; Garland Science: Boca Raton, FL, 2012.
6. Fadouloglou, V. E.; Glykos, N. M.; Kokkinidis, M. Side-Chain Conformations in 4-Alpha-Helical Bundles. Protein Engineering 2001, 14 (5), 321–328. https://doi.org/10.1093/protein/14.5.321.
7. Chothia, C.; Levitt, M.; Richardson, D. Helix to Helix Packing in Proteins. Journal of Molecular Biology 1981, 145 (1), 215–250. https://doi.org/10.1016/0022-2836(81)90341-7.
8. Banner, D. W.; Kokkinidis, M.; Tsernoglou, D. Structure of the ColE1 Rop Protein at 1.7 Å Resolution. Journal of Molecular Biology 1987, 196 (3), 657–675. https://doi.org/10.1016/0022-2836(87)90039-8.
9. Struble, E. B.; Ladner, J. E.; Brabazon, D. M.; Marino, J. P. New Crystal Structures of ColE1 Rom and Variants Resulting from Mutation of a Surface Exposed Residue: Implications for RNA-Recognition. Proteins 2008, 72 (2), 761–768. https://doi.org/10.1002/prot.21965.
10. Predki, P. F.; Mike Nayak, L.; Gottlieb, M. B. C.; Regan, L. Dissecting RNA-Protein Interactions: RNA-RNA Recognition by Rop. Cell 1995, 80 (1), 41–50. https://doi.org/10.1016/0092-8674(95)90449-2.

11. Lacatena, R. M.; Banner, D. W.; Castagnoli, L.; Cesareni, G. Control of Initiation of PMB1 Replication: Purified Rop Protein and RNA I Affect Primer Formation in Vitro. Cell 1984, 37 (3),1009–1014,https://doi.org/10.1016/0092-8674(84)90435-5.
12. Twigg, A. J.; Sherratt, D. Trans-Complementable Copy-Number Mutants of Plasmid ColE1. Nature 1980, 283 (5743),216–218. https://doi.org/10.1038/283216a0.
13. Castagnoli, L.; Scarpa, M.; Kokkinidis, M.; Banner, D. W.; Tsernoglou, D.; Cesareni, G. Genetic and Structural Analysis of the ColE1 Rop (Rom) Protein. The EMBO Journal 1989, 8 (2), 621–629. https://doi.org/10.1002/j.1460-2075.1989.tb03417.x.
14. Lee, A. J.; Crothers, D. M. The Solution Structure of an RNA Loop–Loop Complex: The ColE1 Inverted Loop Sequence. Structure 1998, 6 (8), 993–1007. https://doi.org/10.1016/s0969-2126(98)00101-4.
15. Del Solar, G.; Espinosa, M. Plasmid Copy Number Control: An Ever-Growing Story. Molecular Microbiology 2002, 37(3),492–500. https://doi.org/10.1046/j.1365-2958.2000.02005.x.
16. Glykos, N. M.; Papanikolau, Y.; Vlassi, M.; Kotsifaki, D.; Cesareni, G.; Kokkinidis, M. Loopless Rop: Structure and Dynamics of an Engineered Homotetrameric Variant of the Repressor of Primer Protein. Biochemistry 2006, 45 (36),10905–10919. https://doi.org/10.1021/bi060833n.
17. Brunet, A. P.; Huang, E. S.; Huffine, M. E.; Loeb, J. E.; Weltman, R. J.; Hecht, M. H. The Role of Turns in the Structure of an Alpha-Helical Protein. Nature 1993, 364 (6435), 355–358. https://doi.org/10.1038/364355a0.
18. Nagi, A. D.; Regan, L. An Inverse Correlation between Loop Length and Stability in a Four-Helix-Bundle Protein. Folding and Design 1997, 2 (1), 67–75. https://doi.org/10.1016/s1359-0278(97)00007-2.
19. Predki, P. F.; Regan, L. Redesigning the Topology of a Four-Helix-Bundle Protein: Monomeric Rop. Biochemistry 1995, 34 (31), 9834–9839. https://doi.org/10.1021/bi00031a003.

20. REGAN, L. Protein Redesign. Current Opinion in Structural Biology 1999, 9 (4), 494–499. https://doi.org/10.1016/s0959-440x(99)80070-0.
21. Kresse, H. P.; Czubayko, M.; Nyakatura, G.; Vriend, G.; Sander, C.; Bloecker, H. Four-Helix Bundle Topology Re-Engineered: Monomeric Rop Protein Variants with Different Loop Arrangements. Protein Engineering, Design and Selection 2001, 14 (11), 897–901. https://doi.org/10.1093/protein/14.11.897.
22. Steif, C.; Weber, P.; Hinz, H. J.; Flossdorf, J.; Cesareni, G.; Kokkinidis, M. Subunit Interactions Provide a Significant Contribution to the Stability of the Dimeric Four-Alpha-Helical-Bundle Protein ROP. Biochemistry 1993, 32 (15), 3867–3876. https://doi.org/10.1021/bi00066a005.
23. Kamtekar, S.; Hecht, M. H. The Four-Helix Bundle: What Determines a Fold? The FASEB Journal 1995, 9 (11),1013–1022. https://doi.org/10.1096/fasebj.9.11.7649401.
24. Lassalle, M. W.; Hinz, H. J. Unfolding of the Tetrameric Loop Deletion Mutant of ROP Protein Is a Second-Order Reaction. Biochemistry 1998, 37 (23), 8465–8472. https://doi.org/10.1021/bi9730691.
25. Lassalle, M. W.; Hinz, H.-J.; Wenzel, H.; Vlassi, M.; Kokkinidis, M.; Cesareni, G. Dimer-To-Tetramer Transformation: Loop Excision Dramatically Alters Structure and Stability of the ROP Four α-Helix Bundle Protein. Journal of Molecular Biology 1998, 279 (4), 987–1000. https://doi.org/10.1006/jmbi.1998.1776.
26. Lassalle, M. W.; Hinz, H.-J. . Refolding Studies on the Tetrameric Loop Deletion Mutant RM6 of ROP Protein. Biological Chemistry 1999, 380 (4). https://doi.org/10.1515/bc.1999.060.
27. Van Nuland, N. A. J.; Dobson, C. M.; Regan, L. Characterization of Folding the Four-Helix Bundle Protein Rop by Real-Time NMR. Protein Engineering Design and Selection 2008, 21 (3), 165–170. https://doi.org/10.1093/protein/gzm081.
28. Glykos, N. M.; Kokkinidis, M. Structural Polymorphism of a Marginally Stable 4-α-Helical Bundle. Images of a Trapped Molten Globule? Proteins 2004, 56 (3), 420–425. https://doi.org/10.1002/prot.20167.
29. Schug, A.; Whitford, P. C.; Levy, Y.; Onuchic, J. N. Mutations as Trapdoors to Two Competing Native Conformations of the Rop-Dimer. Proceedings of the National

Academy of Sciences of the United States of America 2007, 104 (45), 17674–17679. https://doi.org/10.1073/pnas.0706077104.

30. Bashford, D.; Chothia, C.; Lesk, A. M. Determinants of a Protein Fold. Journal of Molecular Biology 1987, 196 (1), 199–216. https://doi.org/10.1016/0022-2836(87)90521-3.
31. Bowie, J. U.; Reidhaar-Olson, J. F.; Lim, W. A.; Sauer, R. T. Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions. Science 1990, 247 (4948), 1306–1310. https://doi.org/10.1126/science.2315699.
32. Matsumura, M.; Matthews, B. W. Control of Enzyme Activity by an Engineered Disulfide Bond. Science 1989, 243(4892), 792–794. https://doi.org/10.1126/science.2916125.
33. Hari, S. B.; Byeon, C.; Lavinder, J. J.; Magliery, T. J. Cysteine-Free Rop: A Four-Helix Bundle Core Mutant Has Wild-Type Stability and Structure but Dramatically Different Unfolding Kinetics. Protein Science 2010, 19 (4), 670–679. https://doi.org/10.1002/pro.342.
34. Levy, Y.; Cho, S. S.; Shen, T.; Onuchic, J. N.; Wolynes, P. G. Symmetry and Frustration in Protein Energy Landscapes: A near Degeneracy Resolves the Rop Dimer-Folding Mystery. Proceedings of the National Academy of Sciences 2005, 102(7), 2373–2378. https://doi.org/10.1073/pnas.0409572102.
35. Willis, M. A.; Bishop, B.; Regan, L.; Brunger, A. T. Dramatic Structural and Thermodynamic Consequences of Repacking a Protein's Hydrophobic Core. Structure 2000, 8 (12), 1319–1328. https://doi.org/10.1016/s0969-2126(00)00544-x.
36. Garrett, R. H. Biochemistry; Brooks/Cole, Cengage Learning: Belmont, Ca, 2010.
37. Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. 1959, 1–63. https://doi.org/10.1016/s0065-3233(08)60608-7.
38. Makhatadze, G. I.; Privalov, P. L. Energetics of Protein Structure. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0065323308605483 (accessed 2024-07-24).

39. Munson, M.; Regan, L.; O'Brien, R.; Sturtevant, J. M. Redesigning the Hydrophobic Core of a Four-Helix-Bundle Protein. Protein Science 1994, 3 (11), 2015–2022. https://doi.org/10.1002/pro.5560031114.

40. Munson, M.; Balasubramanian, S.; Fleming, K. G.; Nagi, A. D.; O'Brien, R.; Sturtevant, J. M.; Regan, L. What Makes a Protein a Protein? Hydrophobic Core Designs That Specify Stability and Structural Properties. Protein Science 1996, 5(8), 1584–1593. https://doi.org/10.1002/pro.5560050813.

41. Gambin, Y.; Schug, A.; Lemke, E. A.; Lavinder, J. J.; Ferreon, A. C. M.; Magliery, T. J.; Onuchic, J. N.; Deniz, A. A. Direct Single-Molecule Observation of a Protein Living in Two Opposed Native Structures. Proceedings of the National Academy of Sciences 2009, 106 (25), 10153–10158. https://doi.org/10.1073/pnas.0904461106.

42. Kefala, A.; Kotsifaki, D.; Providaki, M.; Amprazi, M.; Kokkinidis, M. Expression, Purification and Crystallization of a Protein Resulting from the Inversion of the Amino-Acid Sequence of a Helical Bundle. Acta Crystallographica Section F Structural Biology Communications 2017, 73 (1), 51–53. https://doi.org/10.1107/s2053230x16020173.

43. Vlassi, M.; Dauter, Z.; Wilson, K. S.; Kokkinidis, M. Structural Parameters for Proteins Derived from the Atomic Resolution (1.09 Å) Structure of a Designed Variant of the ColE1 ROP Protein. Acta Crystallographica Section D Biological Crystallography 1998, 54 (6),1245–1260. https://doi.org/10.1107/s0907444998002492.

44. Vlassi, M.; Steif, C.; Weber, P.; Tsernoglou, D.; Wilson, K. S.; Hinz, H. J.; Kokkinidis, M. Restored Heptad Pattern Continuity Does Not Alter the Folding of a 4-α-Helix Bundle. Nature Structural Biology 1994, 1 (10), 706–716. https://doi.org/10.1038/nsb1094-706.

45. Kokkinidis, M.; Vlassi, M.; Papanikolaou, Y.; Kotsifaki, D.; Kingswell, A.; Tsernoglou, D.; Hinz, H. J. Correlation between Protein Stability and Crystal Properties of Designed ROP Variants. Proteins 1993, 16 (2), 214–216. https://doi.org/10.1002/prot.340160208.

46. Cohen, C.; Parry, D. A. D. α-Helical Coiled Coils - a Widespread Motif in Proteins. Trends in Biochemical Sciences 1986, 11 (6), 245–248. https://doi.org/10.1016/0968-0004(86)90186-6.

47. Paliakasis, C. D.; Kokkinidis, M. The Stability of the Four-α-Helix Bundle Motif in Proteins. Protein Engineering 1991, 4(7), 849–850. https://doi.org/10.1093/protein/4.7.849.

48. Jaenicke, R. Folding and Association of Proteins. Progress in Biophysics and Molecular Biology 1987, 49 (2-3), 117–237. https://doi.org/10.1016/0079-6107(87)90011-3.

49. Glykos, N. M.; Cesareni, G.; Kokkinidis, M. Protein Plasticity to the Extreme: Changing the Topology of a 4-α-Helical Bundle with a Single Amino Acid Substitution. Structure 1999, 7 (6), 597–603. https://doi.org/10.1016/s0969-2126(99)80081-1.

50. Peters, K.; Hinz, H. J.; G Cesareni. Introduction of a Proline Residue into Position 31 of the Loop of the Dimeric 4-α-Helical Protein ROP Causes a Drastic Destabilization. Biological chemistry 1997, 378 (10). https://doi.org/10.1515/bchm.1997.378.10.1141.

51. Presnell, S. R.; Cohen, F. E. Topological Distribution of Four-α-Helix Bundles. Proceedings of the National Academy of Sciences 1989, 86 (17), 6592–6596. https://doi.org/10.1073/pnas.86.17.6592.

52. Glykos, N. M.; Kokkinidis, M. Meaningful Refinement of Polyalanine Models Using Rigid-Body Simulated Annealing: Application to the Structure Determination of the A31P Rop Mutant. Acta Crystallographica. Section D, Biological Crystallography 1999, 55 (Pt 7), 1301–1308. https://doi.org/10.1107/s0907444999004989.

53. Amprazi, M.; Kotsifaki, D.; Providaki, M.; Kapetaniou, E. G.; Fellas, G.; Kyriazidis, I.; Pérez, J.; Kokkinidis, M. Structural Plasticity of 4-α-Helical Bundles Exemplified by the Puzzle-like Molecular Assembly of the Rop Protein. Proceedings of the National Academy of Sciences of the United States of America 2014, 111 (30), 11049–11054. https://doi.org/10.1073/pnas.1322065111.

54. Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The Protein Folding Problem. Annual Review of Biophysics 2008, 37(1), 289–316. https://doi.org/10.1146/annurev.biophys.37.092707.153558.

55. Anfinsen, C. B. Principles That Govern the Folding of Protein Chains. Science 1973, 181 (4096),223–230. https://doi.org/10.1126/science.181.4096.223.

56. Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIV. Proteins: Structure, Function, and Bioinformatics 2021, 89 (12), 1607–1617. https://doi.org/10.1002/prot.26237.

57. Callaway, E. "It Will Change Everything": DeepMind's AI Makes Gigantic Leap in Solving Protein Structures. Nature2020, 588. https://doi.org/10.1038/d41586-020-03348-4.

58. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T. Highly Accurate Protein Structure Prediction with Alphafold. Nature 2021, 596 (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

59. Yin, R.; Feng, B. Y.; Varshney, A.; Pierce, B. G. Benchmarking AlphaFold for Protein Complex Modeling Reveals Accuracy Determinants. Protein Science 2022, 31 (8). https://doi.org/10.1002/pro.4379.

60. Bertoline, L. M. F.; Lima, A. N.; Krieger, J. E.; Teixeira, S. K. Before and after AlphaFold2: An Overview of Protein Structure Prediction. Frontiers in Bioinformatics 2023, 3. https://doi.org/10.3389/fbinf.2023.1120370.

61. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. Science 2021, 373 (6557), 871–876. https://doi.org/10.1126/science.abj8754.

62. Pereira, J.; Simpkin, A. J.; Hartmann, M. D.; Rigden, D. J.; Keegan, R. M.; Lupas, A. N. High-Accuracy Protein Structure Prediction in CASP14. Proteins: Structure, Function, and Bioinformatics 2021, 89 (12). https://doi.org/10.1002/prot.26171.
63. Skolnick, J.; Gao, M.; Zhou, H.; Singh, S. AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. Journal of Chemical Information and Modeling 2021, 61(10), 4827–4831. https://doi.org/10.1021/acs.jcim.1c01114.
64. Rubiera, C. O. AlphaFold 2 Is here: What's behind the Structure Prediction Miracle | Oxford Protein Informatics Group. Oxford Protein Informatics Group. https://www.blopig.com/blog/2021/07/alphafold-2-is-here-whats-behind-the-structure-prediction-miracle/.
65. Al-Janabi, A. Has DeepMind's AlphaFold Solved the Protein Folding Problem? BioTechniques 2022, 72 (3), 73–76. https://doi.org/10.2144/btn-2022-0007.
66. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. Nature Methods 2022, 19, 1–4. https://doi.org/10.1038/s41592-022-01488-1.
67. Bisong, E. Building Machine Learning and Deep Learning Models on Google Cloud Platform; Apress: Berkeley, CA, 2019. https://doi.org/10.1007/978-1-4842-4470-8.
68. Vouzina, O.-D.; Tafanidis, A.; Glykos, N. M. The Curious Case of A31P, a topology-switching Mutant of the Repressor of Primer Protein: a Molecular Dynamics Study of Its Folding and Misfolding. arXiv.org. https://doi.org/10.48550/arXiv.2404.01405.
69. CCP4 NEWSLETTER on PROTEIN CRYSTALLOGRAPHY. https://legacy.ccp4.ac.uk/newsletters/newsletter41.pdf (accessed 2024-07-24).
70. The PyMOL Molecular Graphics System. PyMOL by Schrödinger. https://pymol.org.
71. Kufareva, I.; Abagyan, R. Methods of Protein Structure Comparison. Methods in molecular biology (Clifton, N.J.) 2012, 857, 231–257. https://doi.org/10.1007/978-1-61779-588-6_10.

72. Zhang, Y.; Skolnick, J. Scoring Function for Automated Assessment of Protein Structure Template Quality. Proteins: Structure, Function, and Bioinformatics 2004, 57 (4), 702–710. https://doi.org/10.1002/prot.20264.
73. Mukherjee, S.; Zhang, Y. MM-Align: A Quick Algorithm for Aligning Multiple-Chain Protein Complex Structures Using Iterative Dynamic Programming. Nucleic Acids Research 2009, 37 (11), e83–e83. https://doi.org/10.1093/nar/gkp318.
74. Rose, G. D. Protein Folding and the Paracelsus Challenge. Nature Structural Biology 1997, 4 (7), 512–514. https://doi.org/10.1038/nsb0797-512.
75. Callaway, E. Major AlphaFold Upgrade Offers Boost for Drug Discovery. Nature 2024. https://doi.org/10.1038/d41586-024-01383-z.