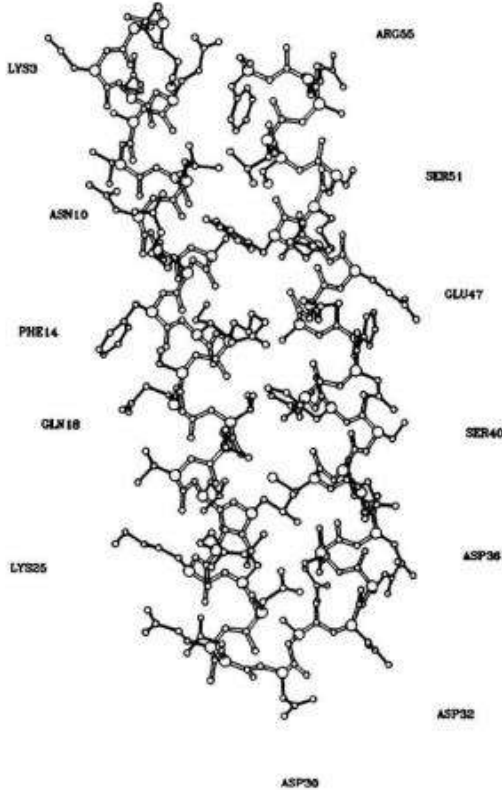DEMOCRITUS UNIVERSITY OF THRACE
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

Using AlphaFold 2 to predict structures of non-naturally occurring proteins: the Rop case
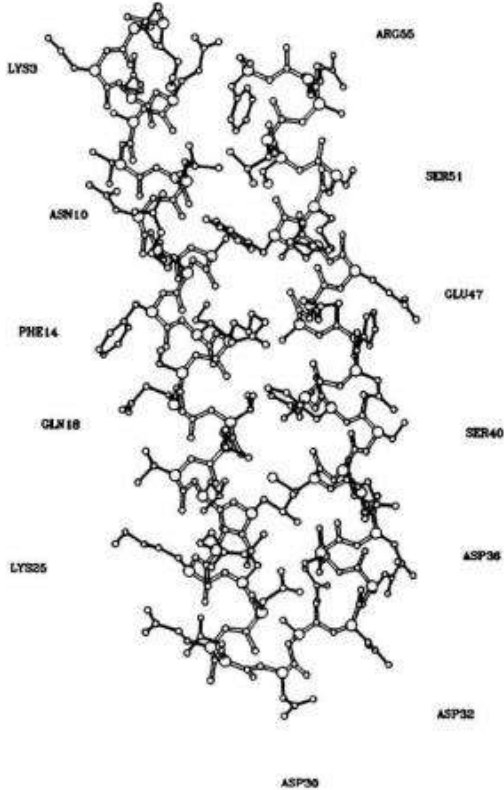
Alexandros Tafanidis

Supervisor: Dr. Nicholas Glykos

The Repressor of Primer (ROP) protein

- A 4-α-helix bundle synthesized in E.coli
- Folds into a homodimeric 4-helix bundle comprised of 2 helix-turn-helix forming an anti-topology.
- Each monomer consists of 63 residues and is further divided into 2 chains.

The Repressor of Primer (ROP) protein

- Residues 29-31 separate each chain.
- Residues into heptads.
- The core structure consists of 8 layers formed by "a" and "d" residues.
- The first 2 residues do not contribute to the helical topology.

Genetic background

- Increases affinity between RNA I – RNA II → limits number of copies of the ColE1 plasmid.

- Achieved by negative control of the frequency of replication initiation events between RNA I, II & ROP.

- Not an essential component of ColE1 control system.
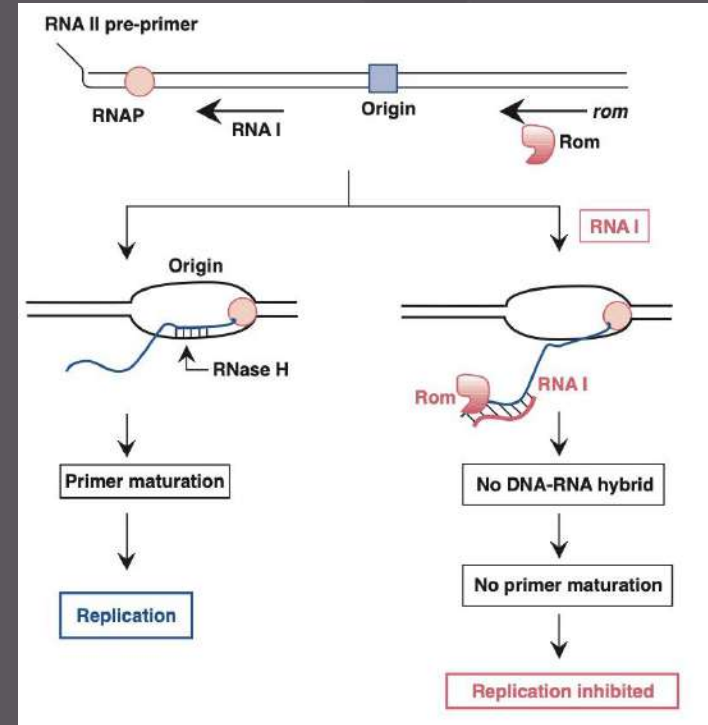
Using AI to predict a protein structure: AlphaFold2

- AI program created by DeepMind.

- High rate of success in predicting a protein structure.

- MSAs use the desired protein sequence as a base.

- The structure runs through Evoformer 48 times.

First Part: The Mutants

Rop mutants

- 7 mutants.

- 4 take the anti topology and 2 the syn.

- Only one takes the "bisecting-U" topology.

Rop mutants: A31D

- Dearth of studies → not much information about it.

- Mentioned only as a name reference to a very few articles.

Rop mutants: A31P

- 31st residue change from alanine to proline.
- Hydrophobic core changes entirely.
- Anti-topology where the 2 monomers are parallel, and the loops are on opposite ends.
- "Bisecting U".
- Distance between helices increased.

WT Rop
Anti

A31P
Bisecting U
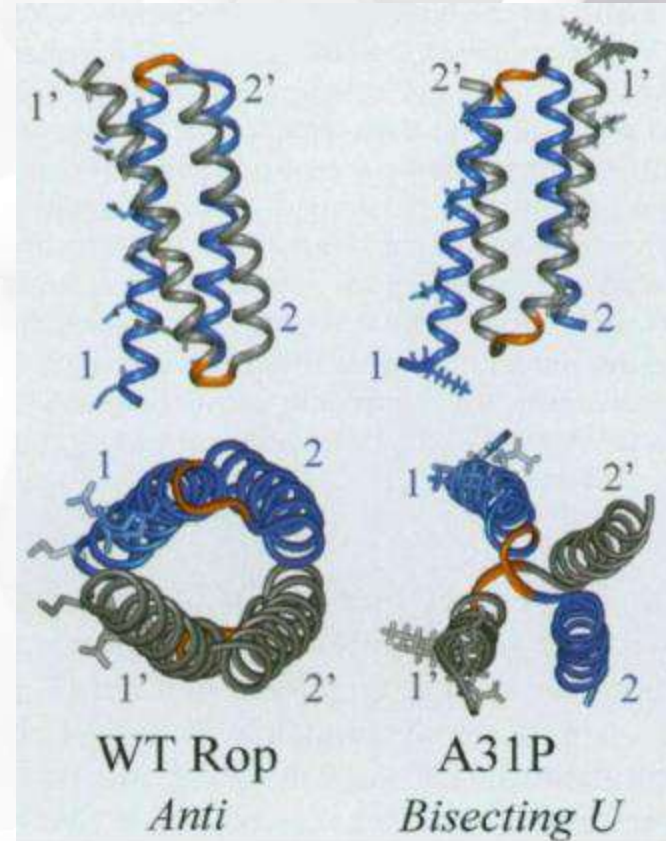
Rop mutants: Cys-free

- Replacement of cysteines 38 and 52 by alanine and valine, respectively.

- Generated to understand the correlation between mutation and protein stability.

Rop mutants: A_2I_2

- Forms a syn topology.

- "a" positions → alanine and "d" → isoleucine.

- The hydrophobic core was drastically repacked, losing the ability to bind RNA.

- Protein's thermal stability increased.

- More densely packed core.
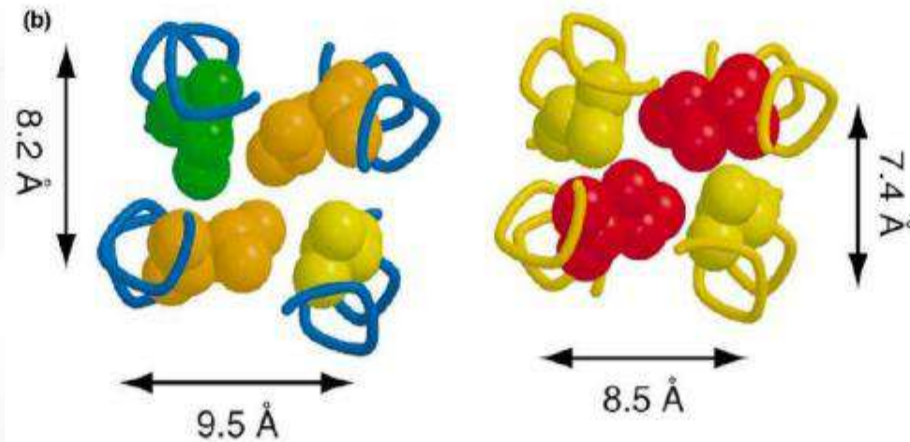
Rop mutants: A_2L_2

- Oscillation between syn & anti topology.
- "a" positions → alanine and "d" → leucines.
- More densely packed core.
- Founding of phenylalanine at position 56 of chains 2 and 2' → reason for designing 2 variations of this mutant.

Rop mutants: 2aa

- 2 extra alanines in the protein loop.

- Prevents the smooth continuation of the heptad.

- Hydrogen bond between the loops (i → i +3).

- No correlation between loops and stability yet.

Rop mutants: Δ_{30-34}

- Deletion of residues 30-34 that form the hairpin so that the heptads are continuous.

- The heptad pattern is inverted.

- Homotetrameric protein without the ability of RNA-binding.
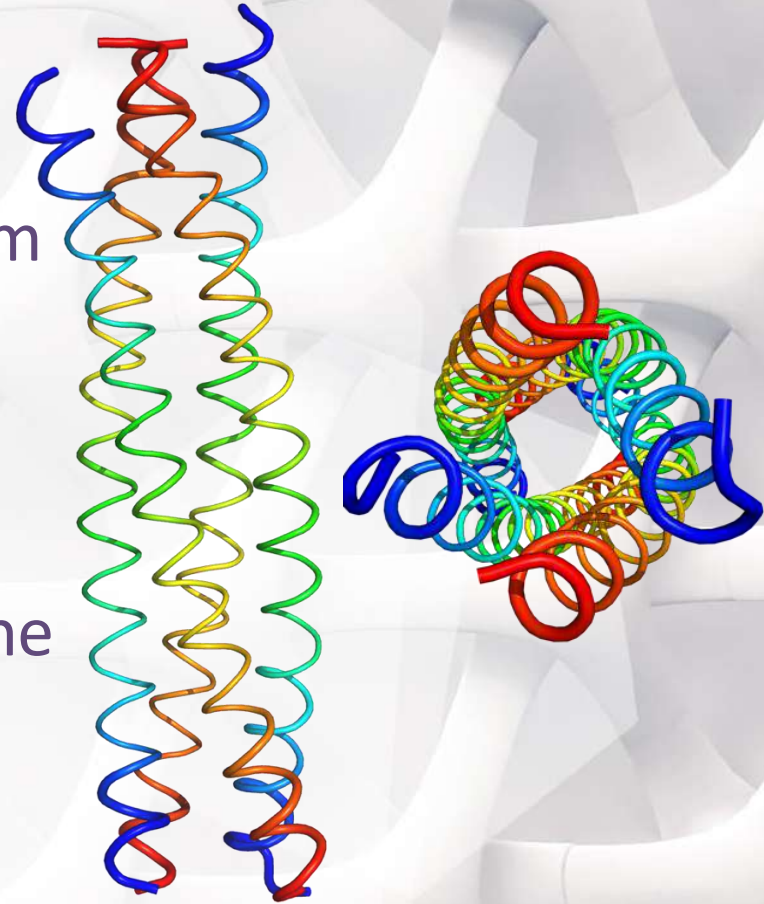
Metrics

RMSD

$\downarrow$

- Statistical similarity assessment of 2 stacked polypeptide chains.
- It relies on the distances between the Cα atoms in these chains.

TM

$\downarrow$

- A more accurate and reliable way of comparing 2 or more structures.
- Does not take into account protein length.

Sequence ID

$\downarrow$

- Measures the percentage of identical residues in the alignment of two protein sequences.

Thesis pipeline

Load sequence in ColabFold and define oligomerization state.

Visual comparison of WT and mutants ROP using PyMol.

Align structures from PDB and AlphaFold using MMalign.

Statistical assessment of stacked polypeptide chains and assessment of the similarity of protein structures.

Main Question

Why do some ROP mutants have a different structure according to AlphaFold in comparison to PDB?

↓

How good is AlphaFold at predicting already-known protein structures?

Second Part: The Results

The control: native Rop

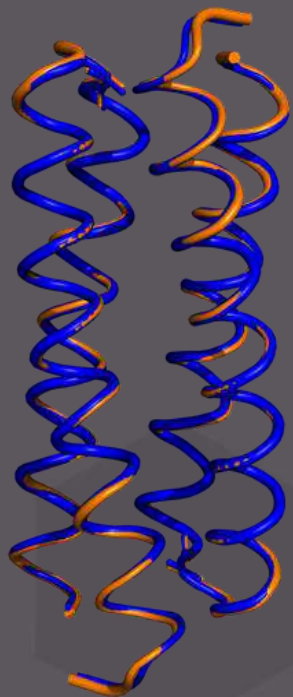|	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop (1Rop) PDB Chains 1 & 2	112	112	0,44	0,98777 (When normalized with 1Rop)	1,000
WT Rop AlphaFold Chains 1' & 2'	126			0,87931 (When normalized with AlphaFold's WT Rop)	

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

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

- Alignment of the structure of WT Rop from PDB and AlphaFold.

- No remarkable changes.

- Orange for WT Rop from PDB and blue for AlphaFold.

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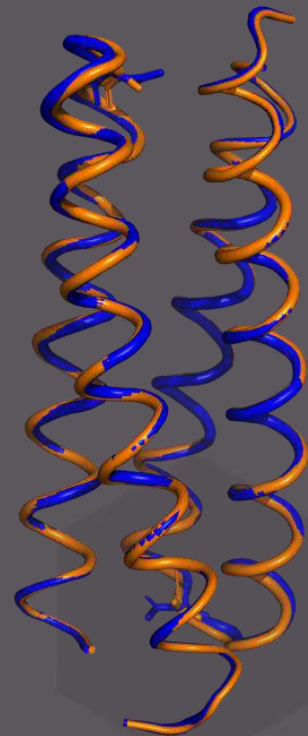
	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop AlphaFold Chains 1' & 2'	126	126	0,88	0,96947	0,984
A31D AlphaFold Chains 1 & 2					

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

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

- A31D mutant has a similar structure as the WT protein.

- AlphaFold's results match PDB.

A31P mutant

AlphaFold WT

AlphaFold A31P

	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				

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

Chain 2:   MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDHADELYRSCLARFGDDGENL*
           ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Chain 2':  MTKQEKTALNMARFIRSQTLTLLEKLNELDPDEQADICESLHDHADELYRSCLARF-------*
```
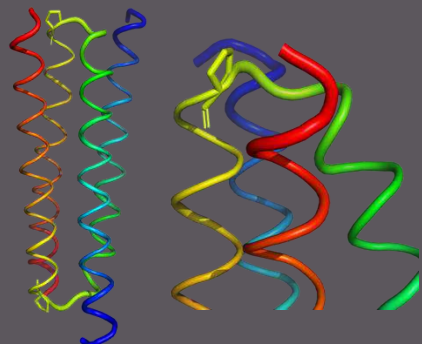
- A31P mutant has the same structure as the WT, according to AlphaFold.

- This is an error as the characteristic "bisecting-U" formation isn't formed.

A31P mutant

PDB A31P

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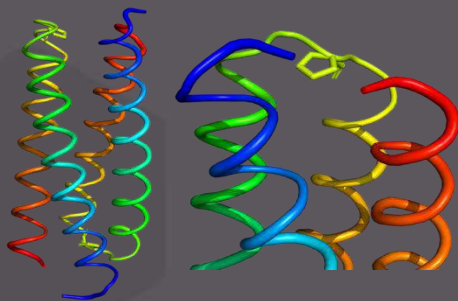
	Chain length	Aligned residues	RMSD	TM	Sequence ID
A31P (1b6q) PDB Chains 1 & 2	112	**80**	3,55	0,44780	0,750
A31P AlphaFold Chains 1' & 2'					

```
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*
```

- There are low similarities between the structures according to RMSD and TM scores.
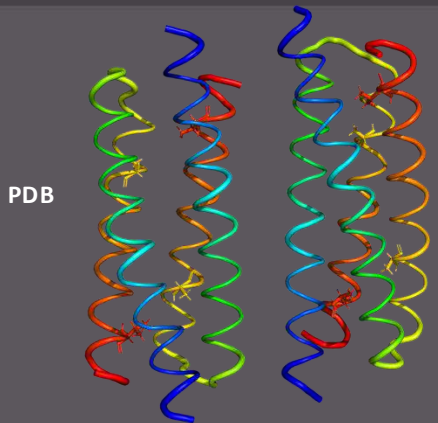
Cys-free mutant
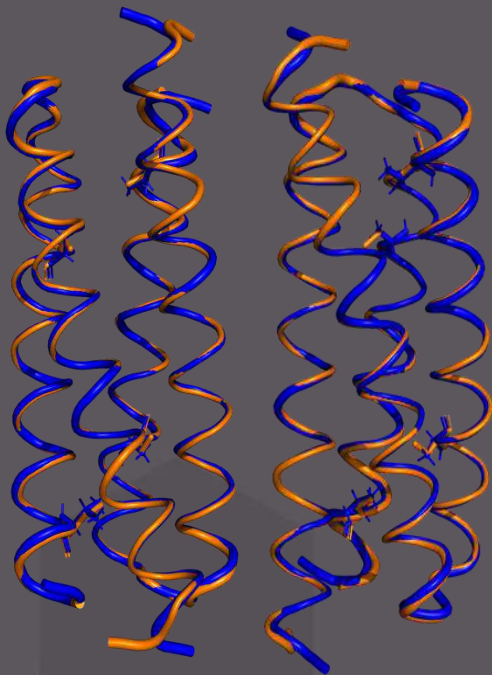
	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop (1Rop) PDB Chains 1 & 2	112	112	1,01	0,97630 (When normalized with 1Rop)	0,946
Cys-free (3k79) PDB Chains 1' & 2'	114			0,95943 (When normalized with 3k79)	

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

Chain 2:   GTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADIAESLHDHADELYRSVLARFG------*
           .:::::::::::::::::::::::::::::::::::::::::::::::::::::::::      *
Chain 2':  GTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADIAESLHDHADELYRSVLARFGDDGENL*
```
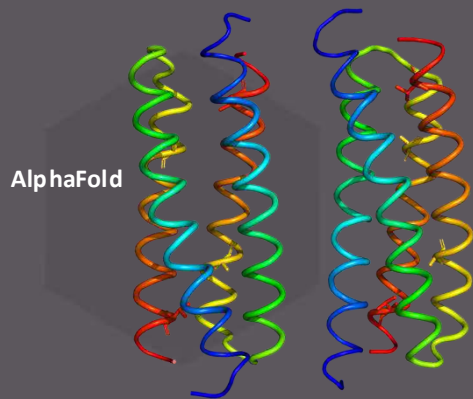
- There are no obvious variations except for the distinct residues at positions 38 and 52.

- Orange for WT Rop and blue for Cys-free, both from PDB.

Cys-free mutant

PDB

AlphaFold

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

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

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

- High score on the RMSD scale and a nearly perfect TM score.
- Almost identical.

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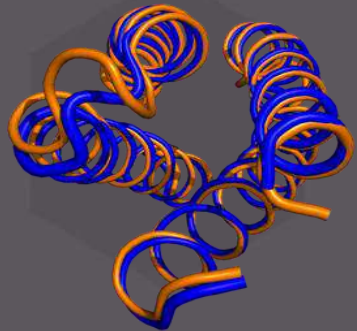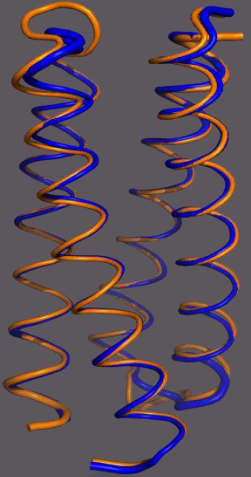
	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop (1Rop) PDB Chains 1 & 2	112	112	0,57	0,97939 (When normalized with 1Rop)	0,982
2aa (1nkd) PDB Chains 1' & 2'	118			0,93065 (When normalized with 1nkd)	


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

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

- Except for the 2 Ala in positions 30 & 32, native Rop and 2aa have high identicality.
- Orange for WT Rop and blue for 2aa mutant, both from PDB.

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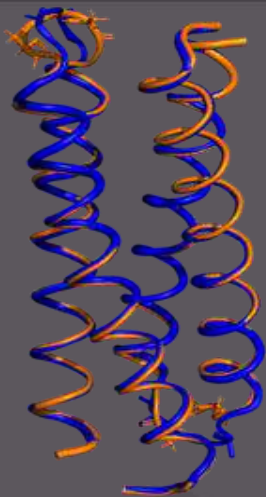
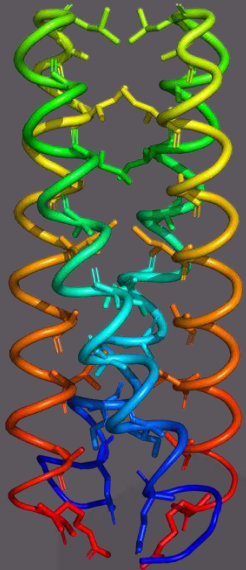|	Chain length	Aligned residues	RMSD	TM	Sequence ID
2aa (1nkd) PDB Chains 1 & 2	118	118	0,89	0,96450 (When normalized with 1nkd)	1,000
2aa AlphaFold Chains 1' & 2'	130			0,87799 (When normalized with AlphaFold's 2aa)	

- High similarity except for a change in the direction of the loop.
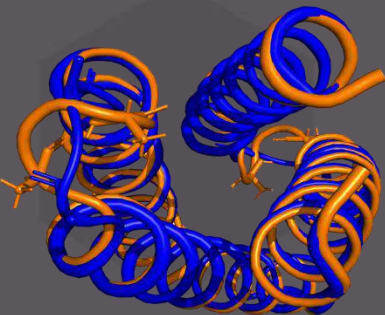
```
Chain 1:   MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFG------*
           :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::     *
Chain 1':  MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFGDDGENL*

Chain 2:   MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFG------*
           :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::     *
Chain 2':  MTKQEKTALNMARFIRSQTLTLLEKLNELADAADEQADICESLHDHADELYRSCLARFGDDGENL*
```
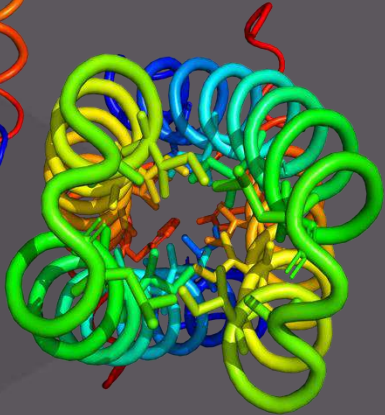
A₂I₂ mutant

	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
A₂I₂ (1f4n) PDB Chains 1' & 2'	109			0,67433 (When normalized with 1f4n)	

- Weak structural similarity between WT and mutant.


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

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

A₂I₂ mutant

PDB

AlphaFold

	Chain length	Aligned residues	RMSD	TM	Sequence ID
A₂I₂ (1f4n) PDB Chains 1 & 2	109			0,94845 (When normalized with 1f4n)	
A₂I₂ AlphaFold Chains 1' & 2'	126	107	0,86	0,82384 (When normalized with AlphaFold's A₂I₂)	0,991

- Strong similarity between PDB and AlphaFold's results.
- AlphaFold accurately predicts the structure of the mutant protein.

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

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

A₂L₂ (Rop21) mutant

	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop AlphaFold Chains 1 & 2	126	92	1,93	0,66477	0,609
A₂L₂ (Rop21) AlphaFold Chains 1' & 2'					

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

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

- The 2 sequences do not align over their whole length because of the mutant's repackaged hydrophobic core.

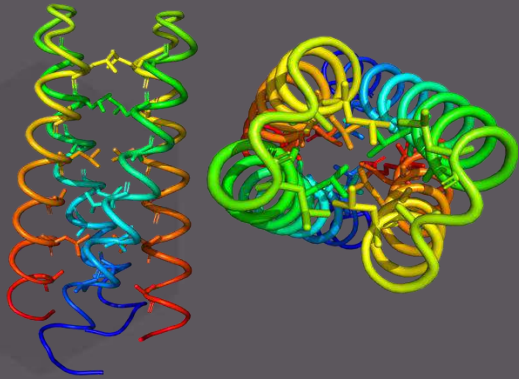- Especially the 1st monomer has noticeable variations.

A₂L₂ (Rop 21) mutant

A₂I₂ AlphaFold

A₂L₂ Rop21 AlphaFold
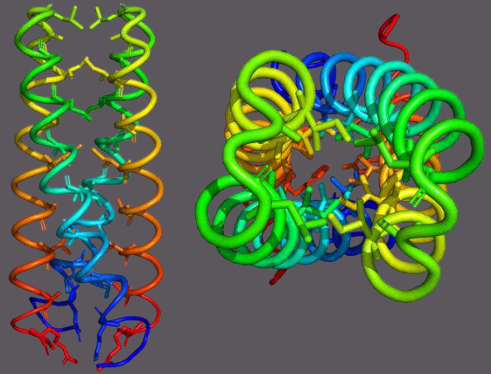
	Chain length	Aligned residues	RMSD	TM	Sequence ID
A₂I₂ (Rop21) AlphaFold Chains 1 & 2	126	126	1,53	0,90777	0.889
A₂I₂ AlphaFold Chains 1' & 2'					

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

Chain 2:   GTKQAKTLLNMARFLRSQALTLLEKANELDADELADIAESLHDHADELYRSALARFGDDGENL*
           .:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::*
Chain 2':  GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARFGDDGENL*
```
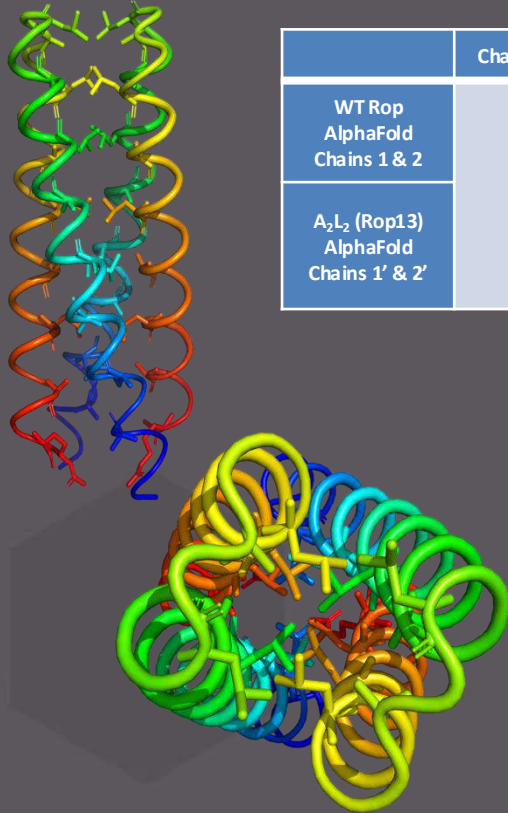
- Both proteins are very structurally similar.

- AlphaFold made the right prediction for Rop21's structure.

A₂L₂ (Rop 13) mutant

	Chain length	Aligned residues	RMSD	TM	Sequence ID
WT Rop AlphaFold Chains 1 & 2	126	89	1,55	0,65565	0,584
A₂L₂ (Rop13) AlphaFold Chains 1' & 2'					

- They might look similar, but there is a catch.

- The mutant's topology changes due to the hydrophobic core repackaging, which also affects the alignment of the rest of the amino acids.

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

A₂L₂ (Rop 13) mutant

A₂L₂ Rop13 AlphaFold
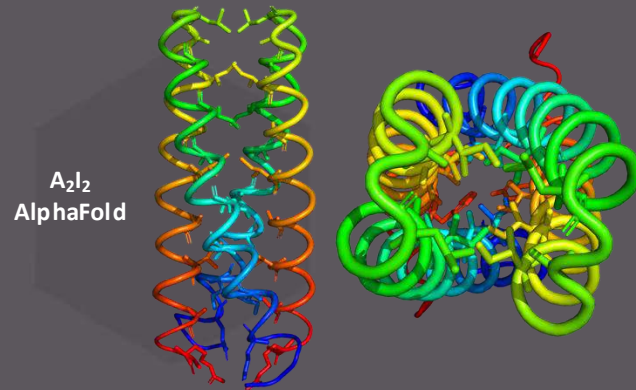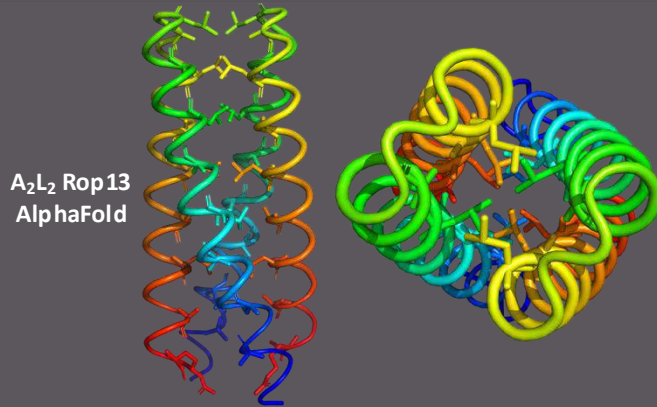
A₂I₂ AlphaFold

- Moderate level of similarity between the 2 mutants.

	Chain length	Aligned residues	RMSD	TM	Sequence ID
A₂L₂ (Rop13) AlphaFold Chains 1 & 2	126	121	1,86	0,86185	0.860
A₂I₂ AlphaFold Chains 1' & 2'					

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

Chain 2:  GTKQEKTLLNMARFLRSQALTLLEKANELDADELADIAESLHDHADELYRSALARFGD-DG-ENL-*
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::.  .  *
Chain 2': GTKQEKTILNMARFIRSQALTILEKANELDADEIADIAESIHDHADEIYRSALARFGDDG-EN--L*
```
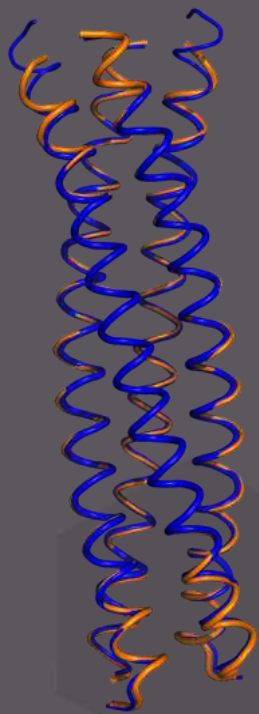
Δ$_{30-34}$ mutant


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

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

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

Chain 4:  MTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARF-------*
          ::::::::::::::::::::::::::::::::::::::::::::::::::::      *
Chain 4': MTKQEKTALNMARFIRSQTLTLLEKLNELADICESLHDHADELYRSCLARFGDDGENL*
```
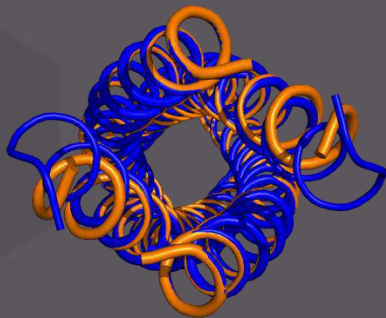
- Both PDB and AlphaFold form 4 single helices instead of 2 monomers consisting of 2 helices.
- AlphaFold results match those from PDB.
- Orange for PDBs' Δ$_{30-34}$ and blue for AlphaFolds'.

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

Conclusions

- Only A31P has different topology and 3D structure.

- AlphaFold failed to predict the right structure.

- The mutation was either missed by the algorithm or it could not distinguish P from A.

- Mutation 2aa had a minor change in the turn region.

- AlphaFold predicted the right structure for 2aa.

- Different topologies occur when correcting heptad discontinuity by insertion (2aa) and deletion (Δ_{30-34}).

Conclusions

- Considering the rapid advancement of technology, it is only logical that one day- soon- we will be able to predict protein structures using AI.

Thank you!