

Visualizing Pseudoknots and Statistics on Pseudoknots

Takumi Otagaki

Master's Student, Asai Laboratory, The University of Tokyo



Benasque 2024 08/01



Contents

1. Pseudoknot visualization
2. Example of Casp16 using pk visualization
3. Statistics on Pseudoknots

About pseudoknots

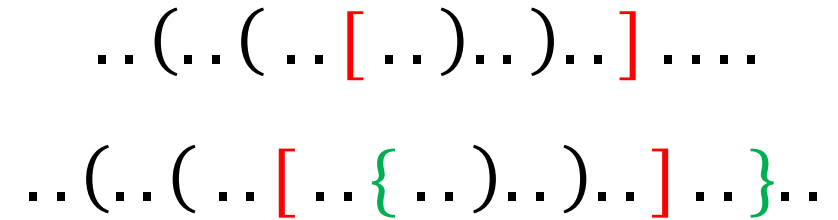
- A pseudoknot is an RNA structure where **bases in a loop pair with bases outside the loop.**

..(..(..[...])..)..
..(..(..[...{...})..)..}

- Pseudoknot Decomposition:
 - Given Base pair set:
{(3, 25), (4, 19), (5,18), (10, 16), (1,14), ... }

About pseudoknots

- A pseudoknot is an RNA structure where bases in a loop pair with bases outside the loop.



- Pseudoknot Decomposition:

- Given Base pair set:
 $\{(3, 25), (4, 19), (5, 18), (10, 16), (1, 14), \dots\}$



decomposition

#bp

Main Layer = Layer1: ..(((.....(.....).....)))..

4

PK1 = Layer2: **[.....]**

2

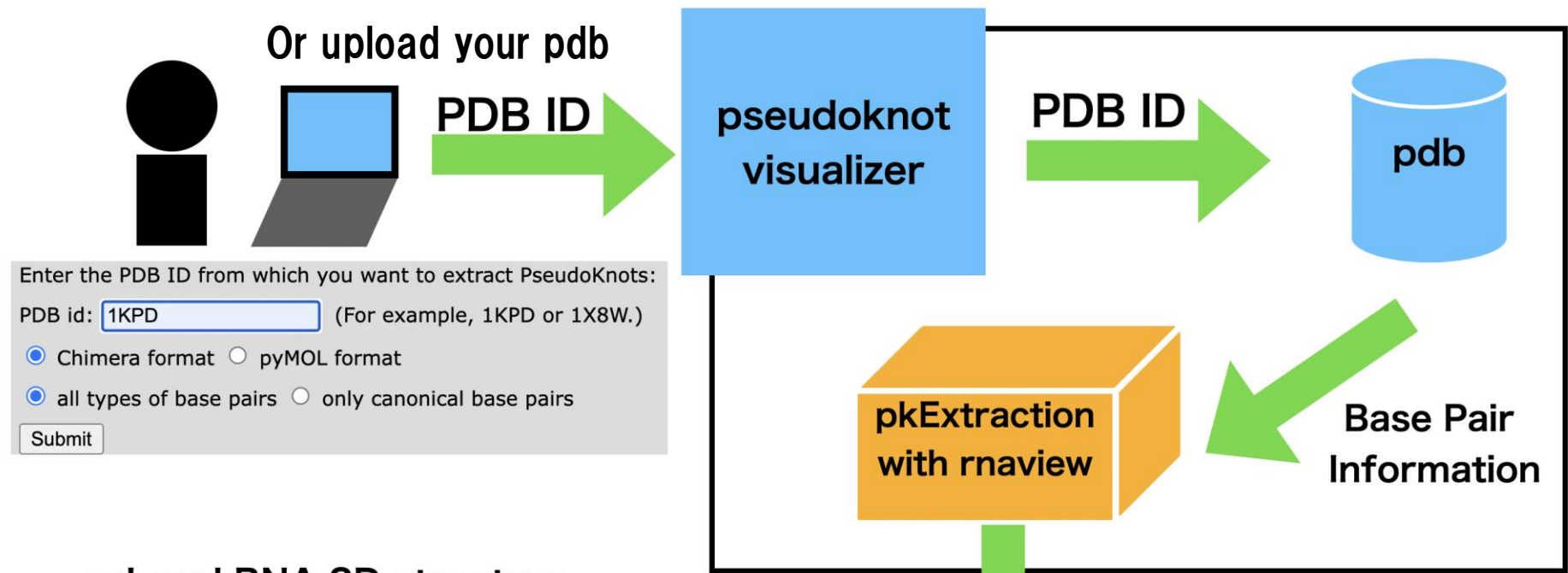
PK2 = Layer3: **{.....}**

1

} Ordered by *#bp*

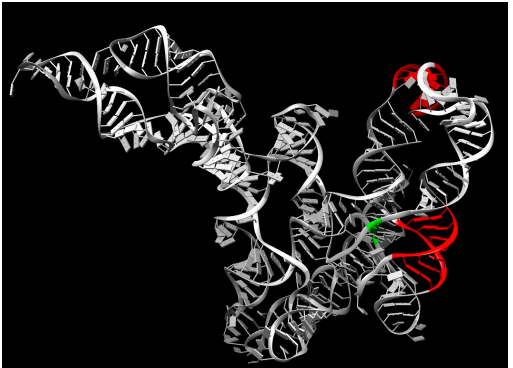
Layer *i*: The layer with the *i*-th most base pairs

Pseudoknot visualization workflow

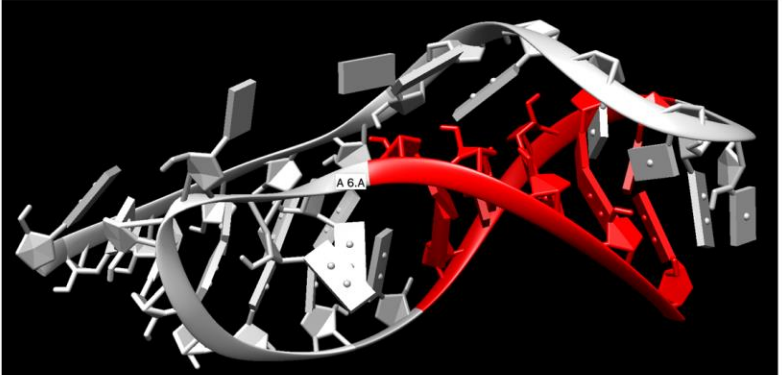


Enter the PDB ID from which you want to extract PseudoKnots:
PDB id: (For example, 1KPD or 1X8W.)
 Chimera format pyMOL format
 all types of base pairs only canonical base pairs

colored RNA 3D structure



Colored by layer ID



pyMOL (Chimera) Command and RNA 2d structure

```
chainA:  
Main Layer:  
.....((((.....)))))  
2nd Layer:  
.((((.....))))).....
```

Copy & Paste to Viewer

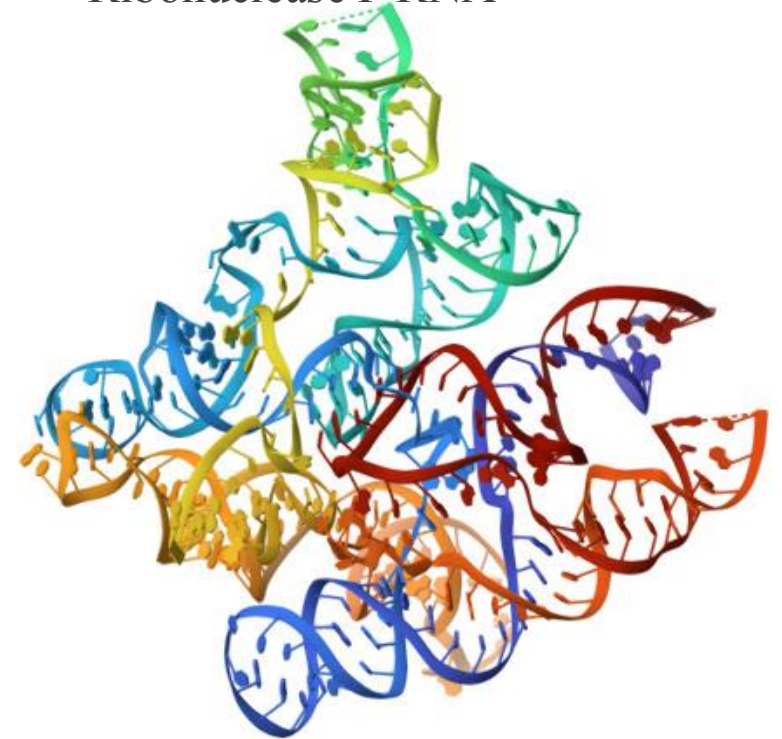
Casp16 prediction using pk visualization

Casp16 prediction using pk visualization

- R1221s2
 - Large RNA (about 400nt.)
 - **The sequence is similar to 2A64 in pdb (Blast search).**
- 2A64
 - Missing regions
 - Pseudoknots
 - Two pseudoknot layer : PK1, PK2

R1221s2 is expected to be similar to 2A64 3d structure.

2A64: Crystal Structure of Bacterial Ribonuclease P RNA

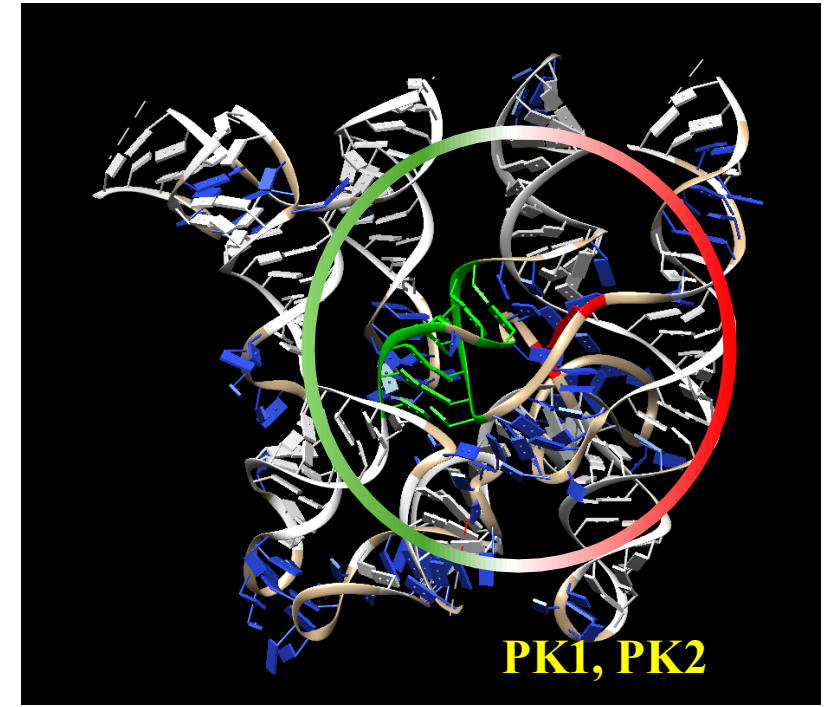


Casp16 prediction using pk visualization

- R1221s2
 - Large RNA (about 400nt.)
 - **The sequence is similar to 2A64 in pdb (Blast search).**
- 2A64
 - Missing regions
 - Pseudoknots
 - **Two pseudoknot layer : PK1, PK2**

R1221s2 is expected to be similar to 2A64 3d structure.

2A64: Crystal Structure of Bacterial Ribonuclease P RNA



Casp16 prediction using pk visualization

R1221s2 should be similar to 2A64 3D structure.

An assumption (that everyone is likely aware of) is

- AlphaFold3 generates a 3D structure similar to that of existing similar sequences in the database.

Casp16 prediction using pk visualization

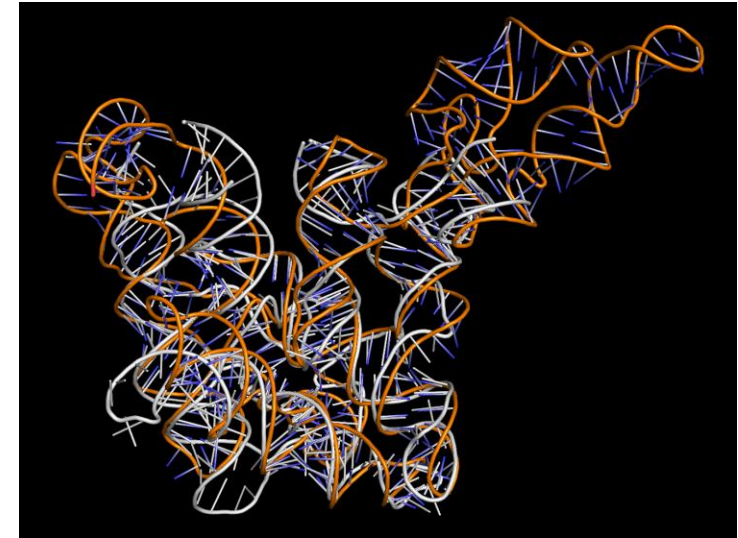
R1221s2 should be similar to 2A64 3D structure.

An assumption (that everyone is likely aware of) is

- AlphaFold3 generates a 3D structure similar to that of existing similar sequences in the database.

In this case, we assumed that **AF3 would predict the structure of R1221s2 to be similar to that of 2A64.**

We can get the structure similar to 2A64!

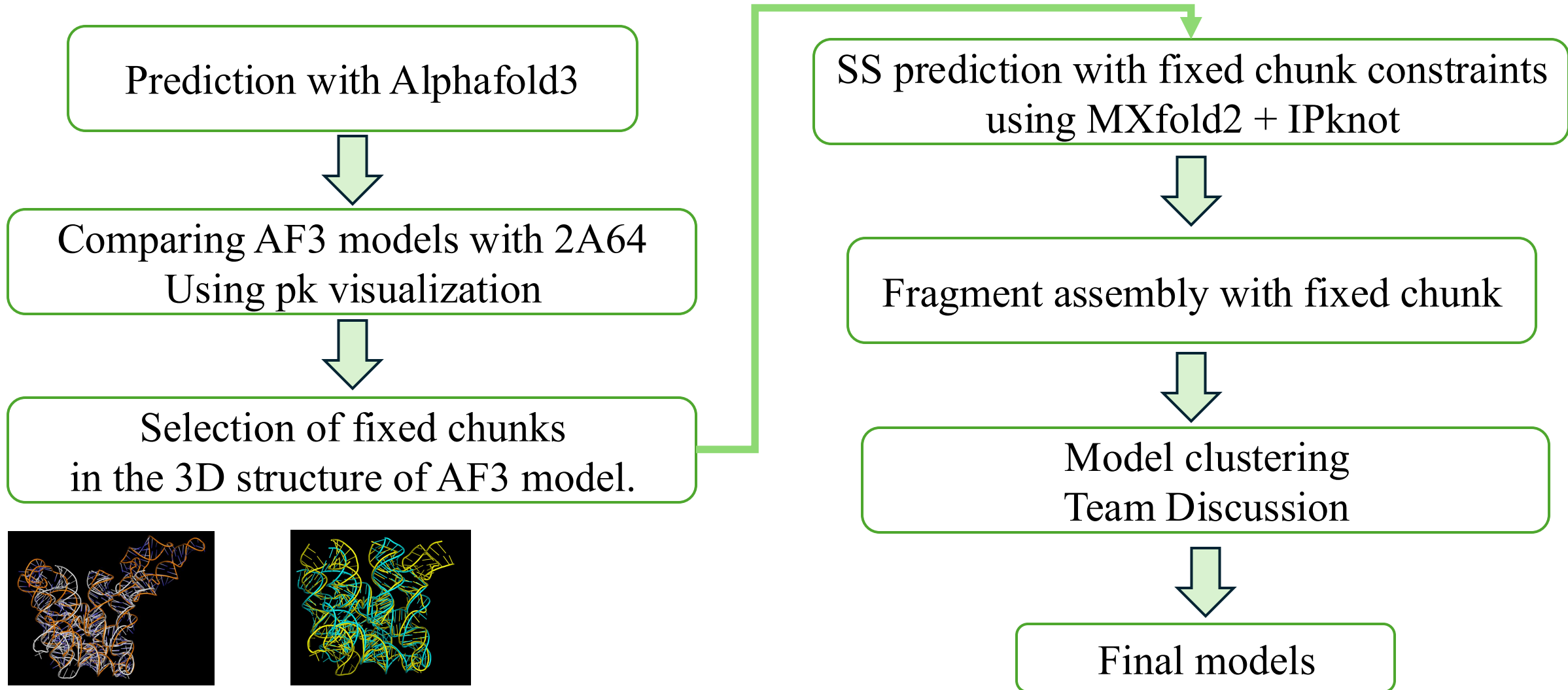


White: 2A64 (in pdb)

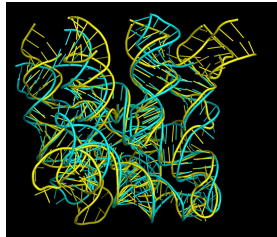
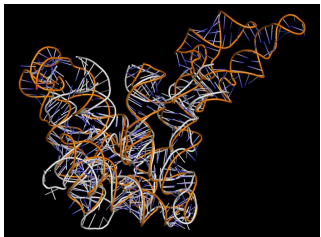
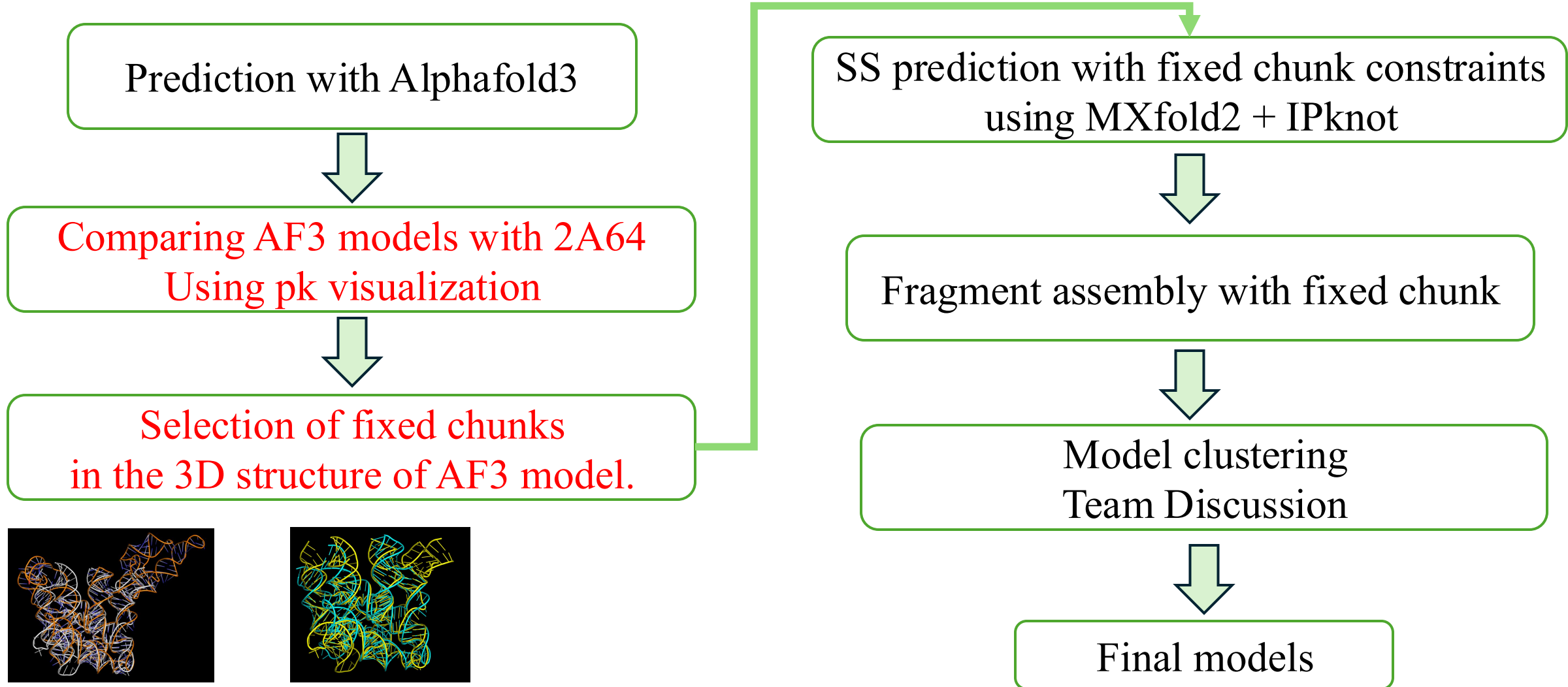
Orange: af3 model

↑ We can see they are similar.

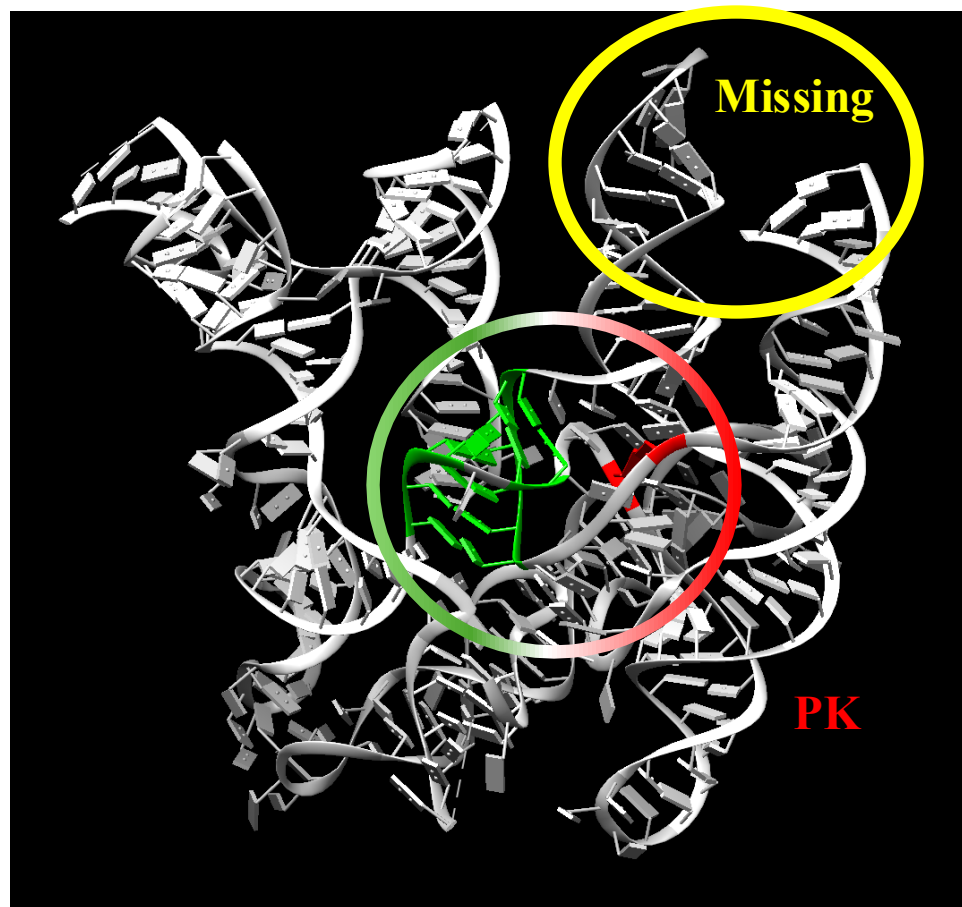
Our specific strategy for Casp16 R1221s2 using pk visualization and AF3



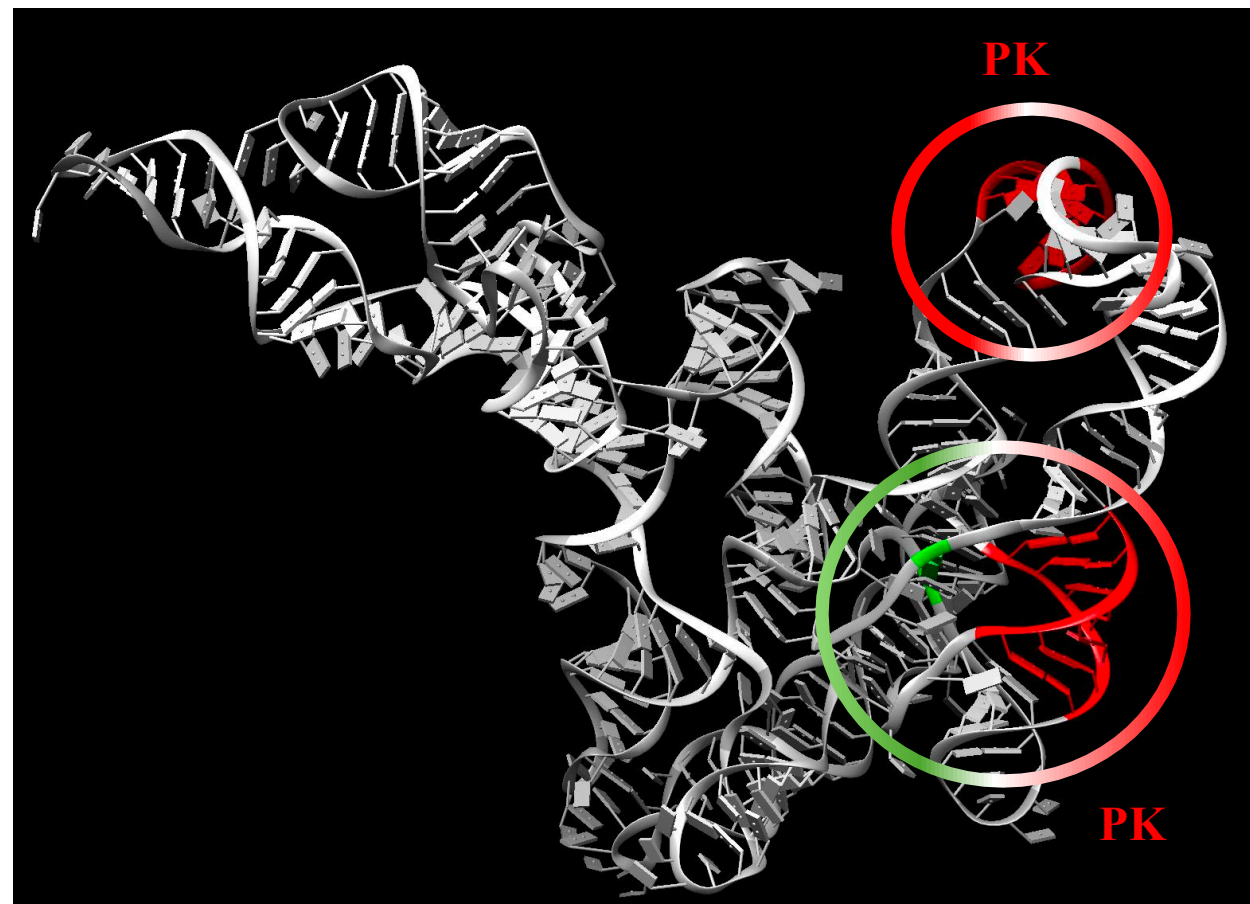
Our specific strategy for Casp16 R1221s2 using pk visualization and AF3



Comparing AF3 models with 2A64 using pk visualization



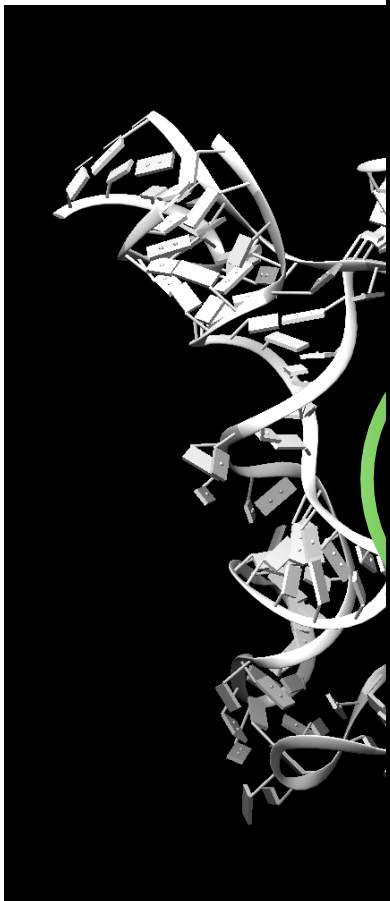
2A64 (pdb)



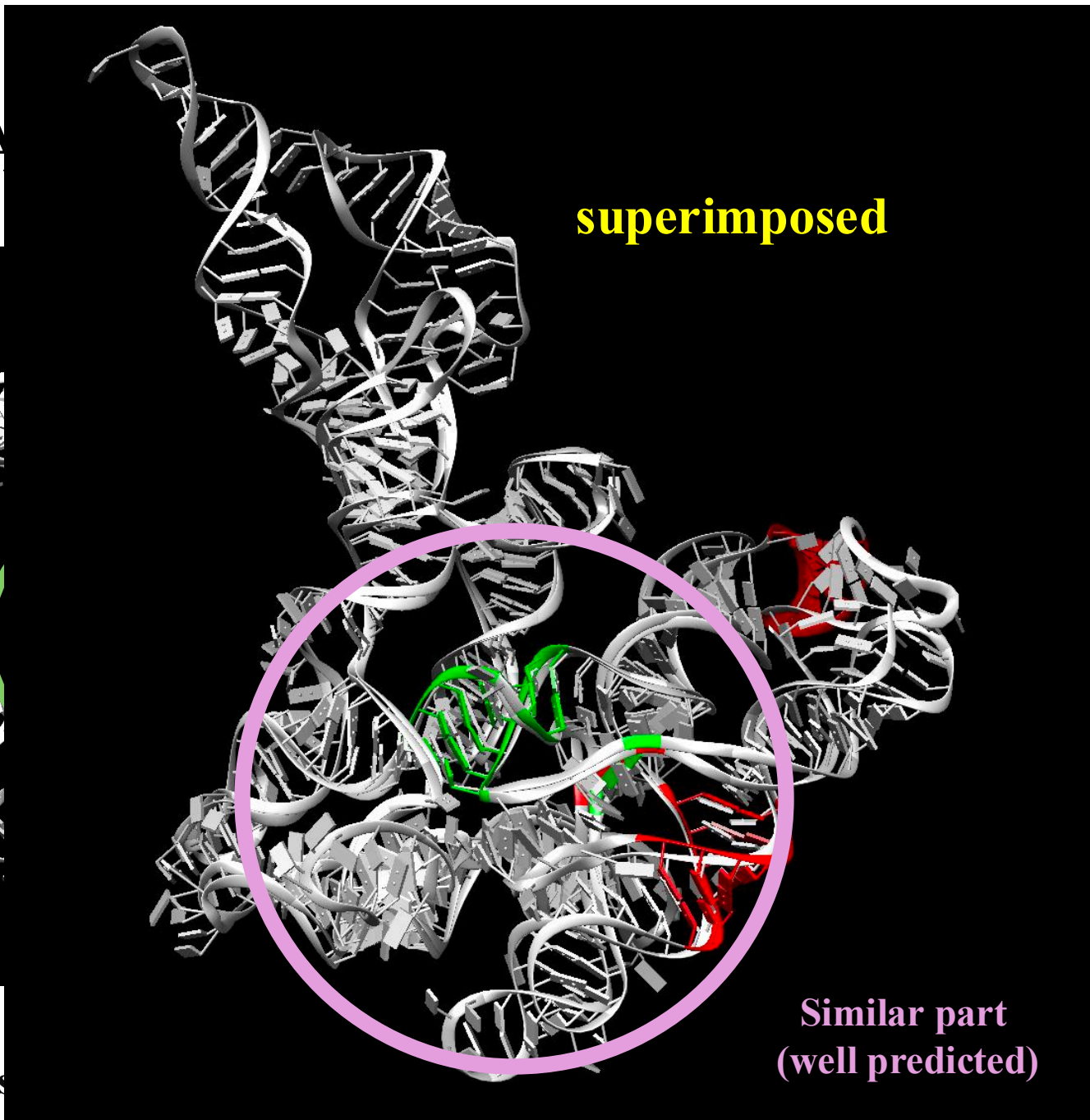
R1221s2 AF3 model

The region near the missing area in 2A64 forms a pseudoknot. Our team assume that the **pseudoknot structures output by AF3 to be unreliable** due to their complexity.

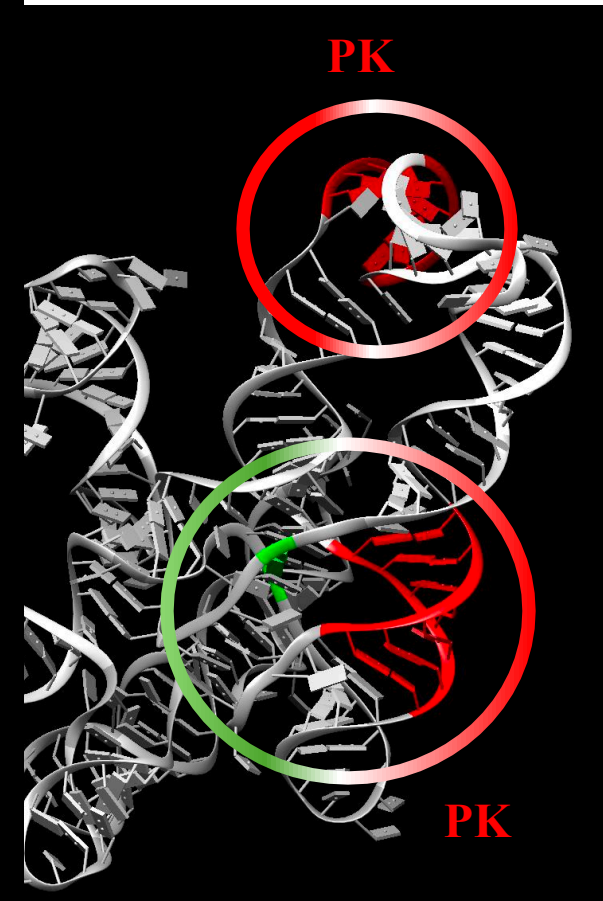
Comparing AF3



The region near the miss...
AF3 to be unreliable due to their complexity.



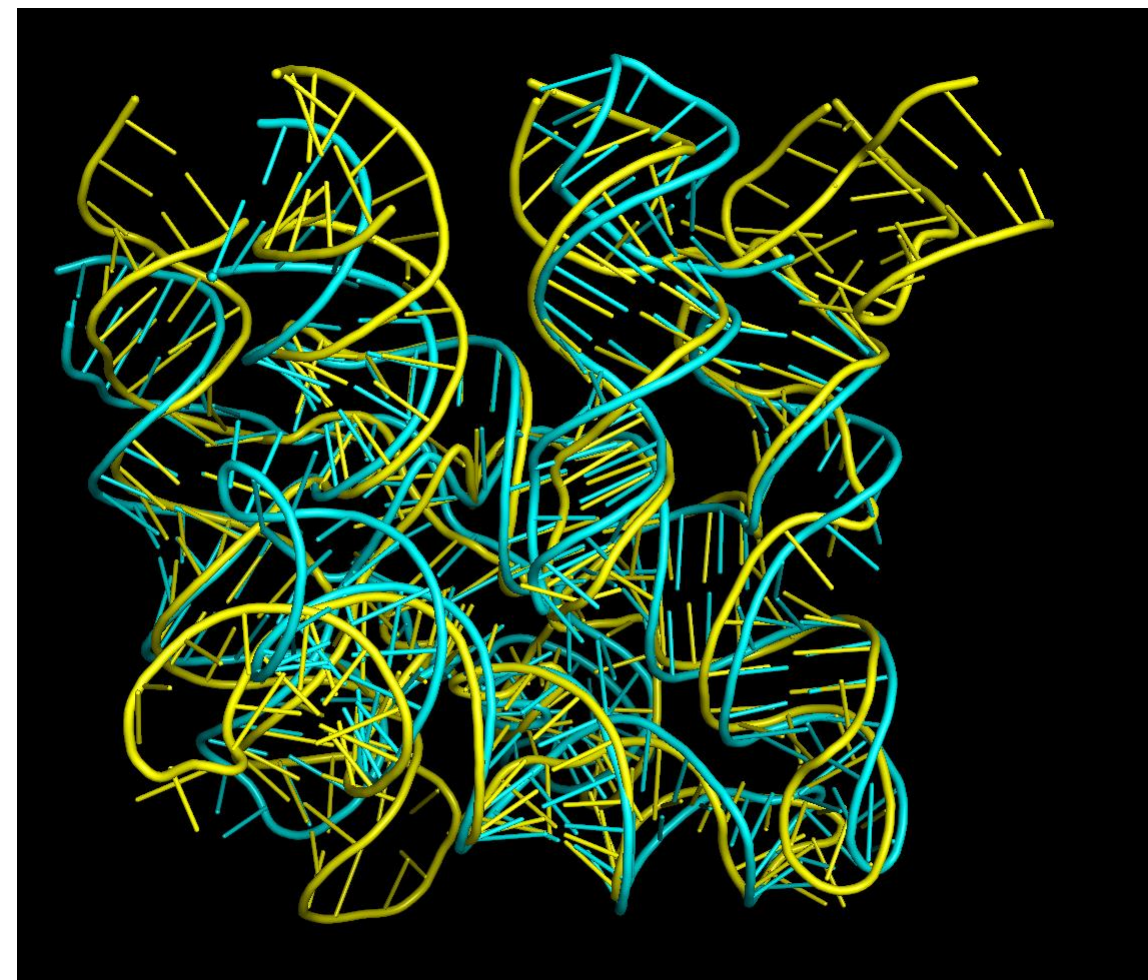
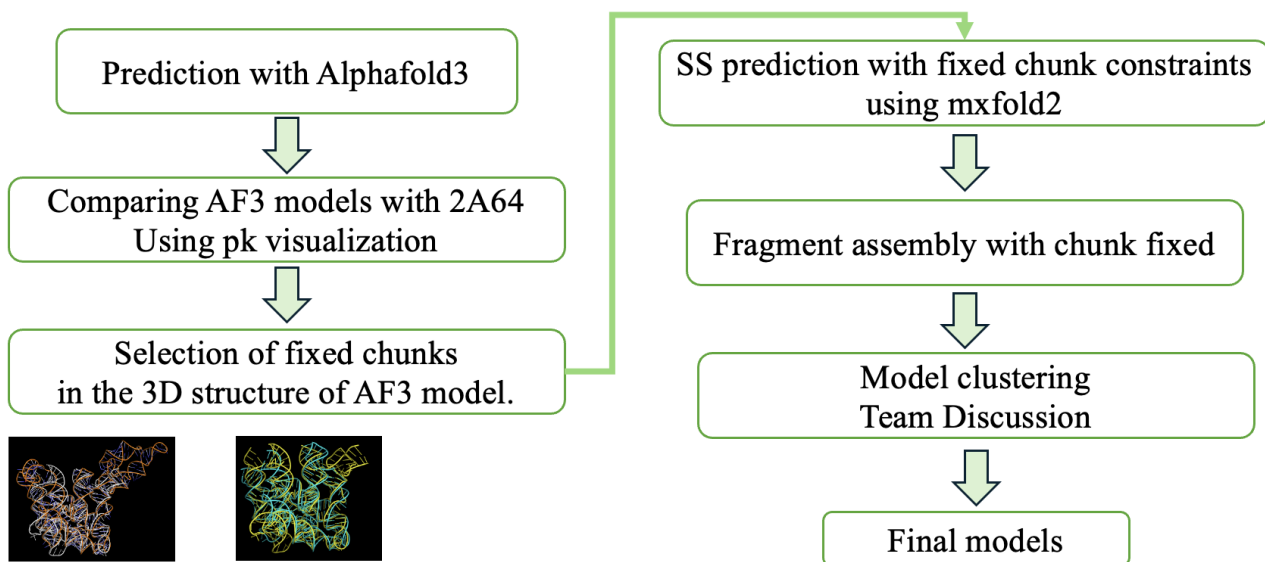
Visualization



F3 model
seudoknot structures output by

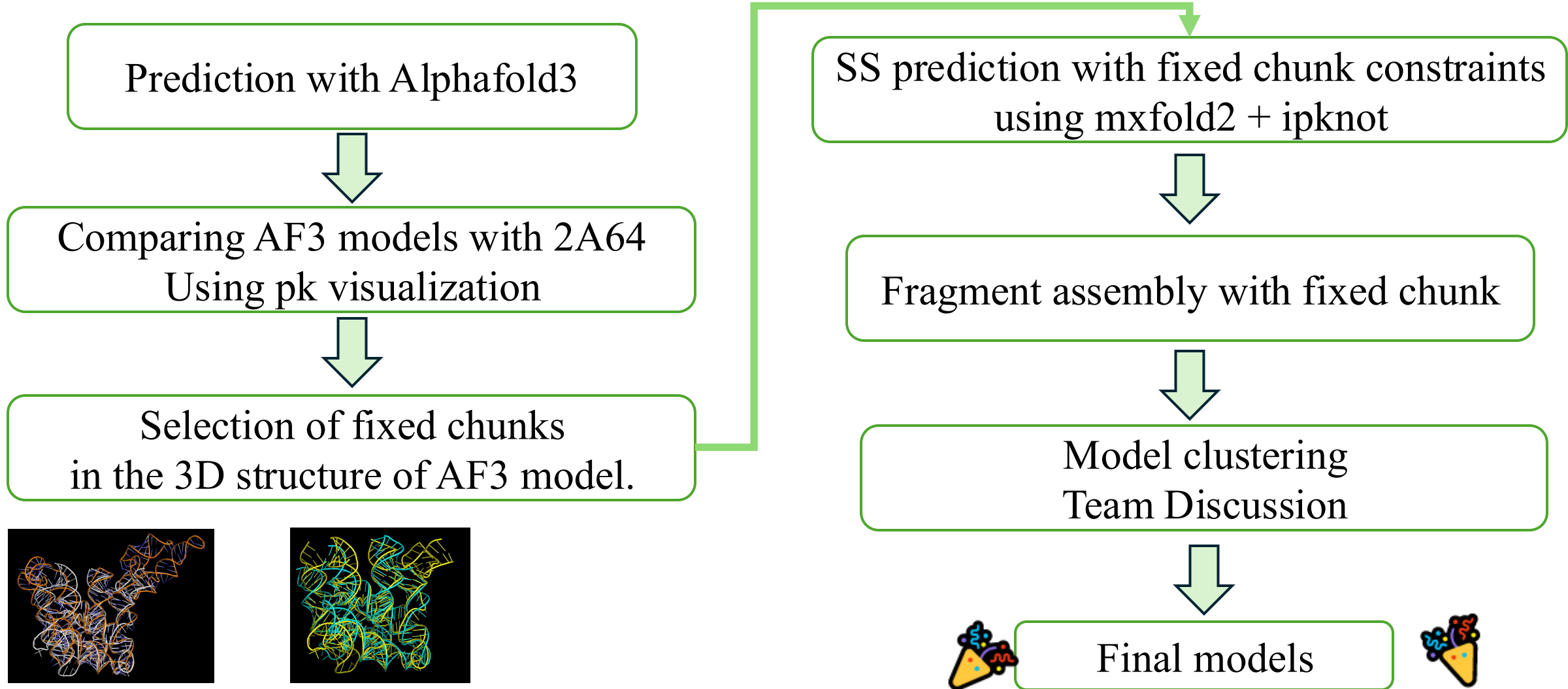
Comparing AF3 models with 2A64 using pk visualization

We fixed the substructures shown in this figure as chunks and used them for downstream predictions.



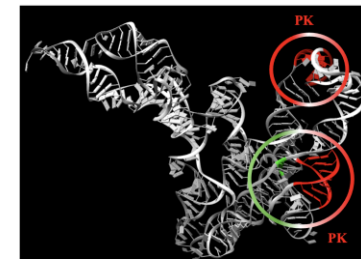
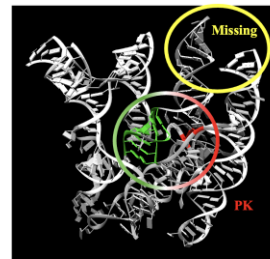
Blue: part of the model predicted by alphafold3
Yellow: 2A64 (pdb entry)

Our specific strategy for Casp16 R1221s2 using pk visualization and AF3



Minor Observations on alphafold3 usage.

- AF3 may help to predict the 3D structures of sequences that are expected to similar to already known structures
- To determine the 3D structure of a sequence x, we can refer to a **similar sequence y from existing databases.**
- **AlphaFold3** is expected to predict x's structure based on its similarity to y.
- However, AlphaFold3 is **unreliable for regions unique to x.**
- We identify these different regions and re-predict them using FARFAR2.
- **By leveraging AlphaFold3's memorization, we can efficiently predict RNA 3D structures.**



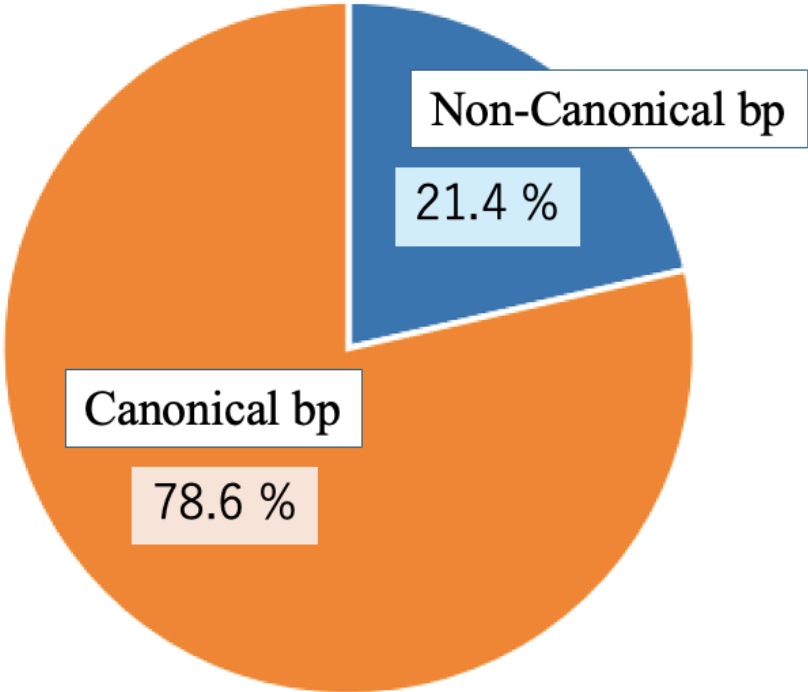
Statistics on Pseudoknots in PDB

Base pair classification in this presentation

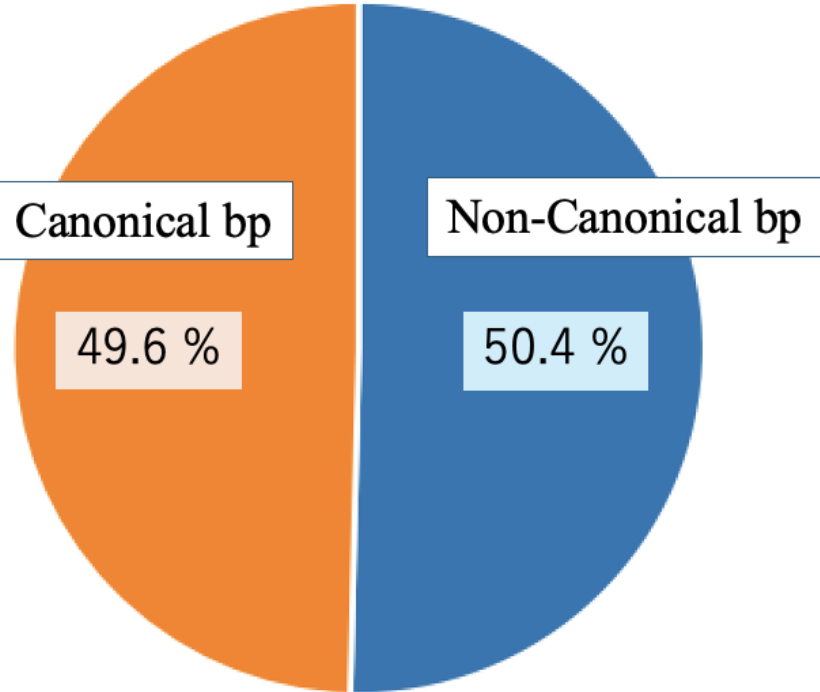
Base pair	Classification
A-U, G-C	Canonical bp
G-U	Canonical bp
The others	Non-Canonical bp

Statistics on pseudoknots

Main Layer
(First Dimension)



Pseudoknot Layer
(Higher Dimension)

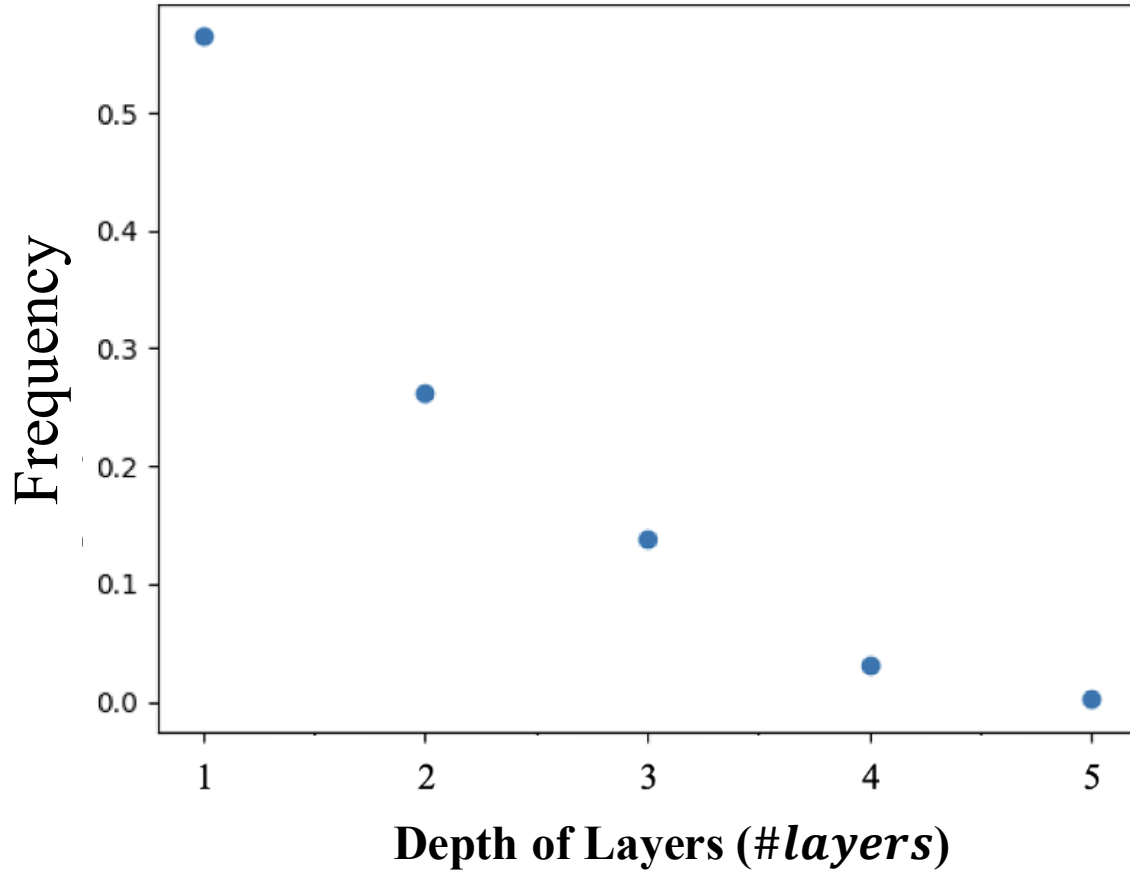


Main Layer Layer1: ..(((.....(.....).....))).....
 PK1 Layer2:[[.....]].....
 PK2 Layer3: {.....}

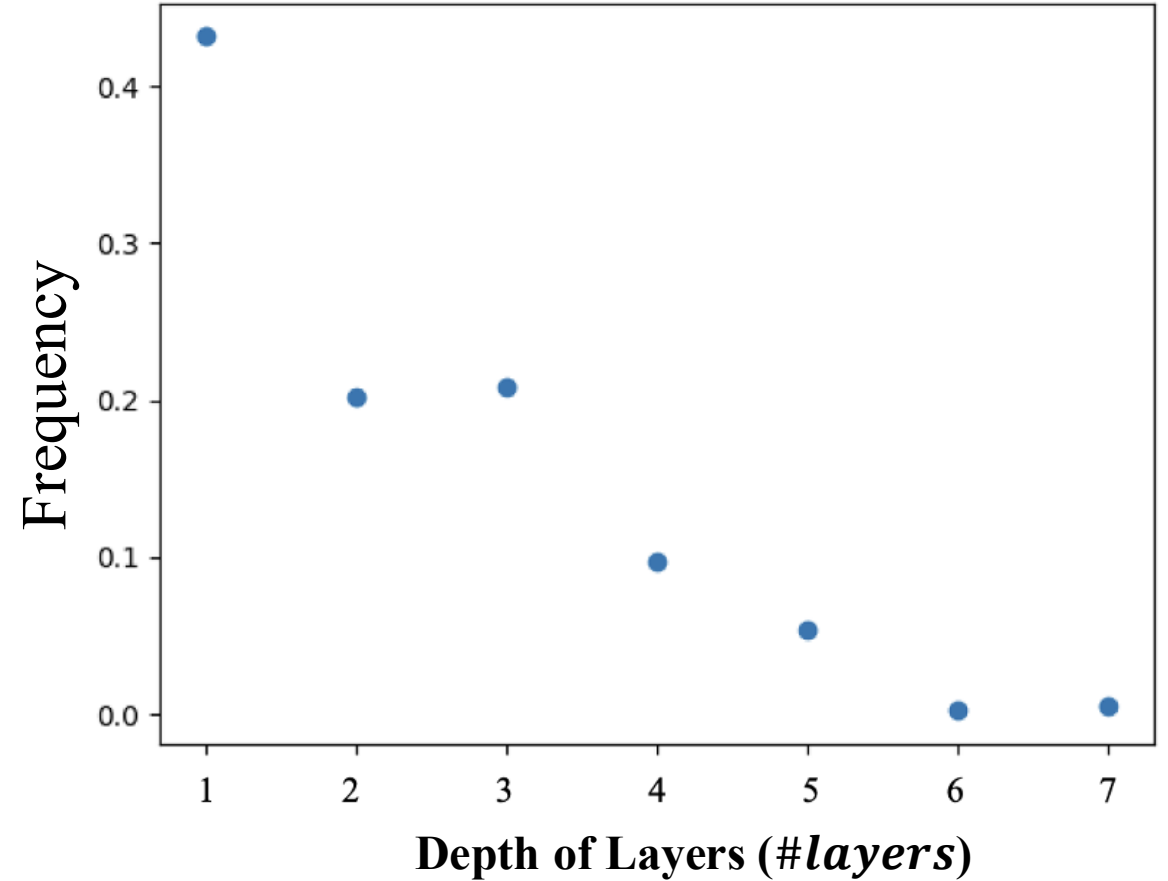
Note:
PK layers contain more non-canonical bp than Main Layer contains.

Statistics on pseudoknots

Only Canonical bp



All bp including non-canonical



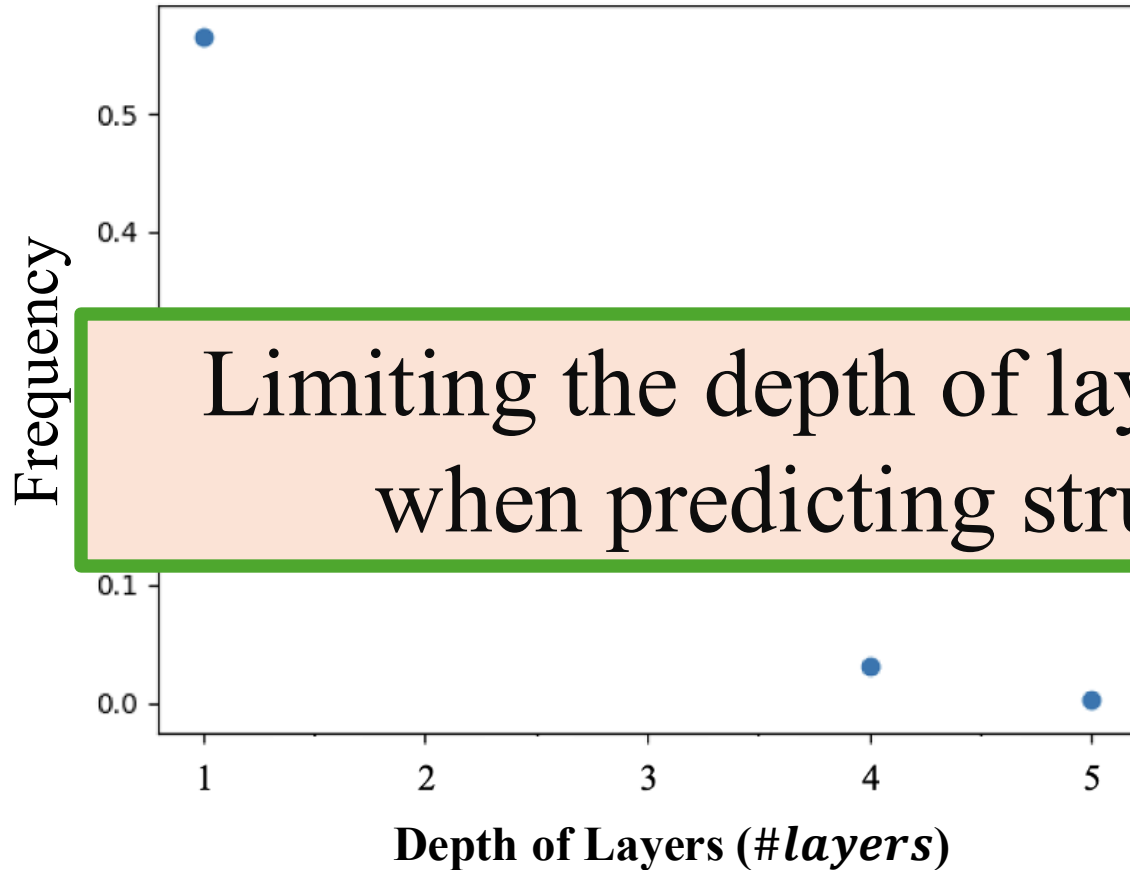
Main Layer Layer1: ..(((.....(.....).....))).....
 PK1 Layer2:[[.....]].....
 PK2 Layer3: {.....}

Note:

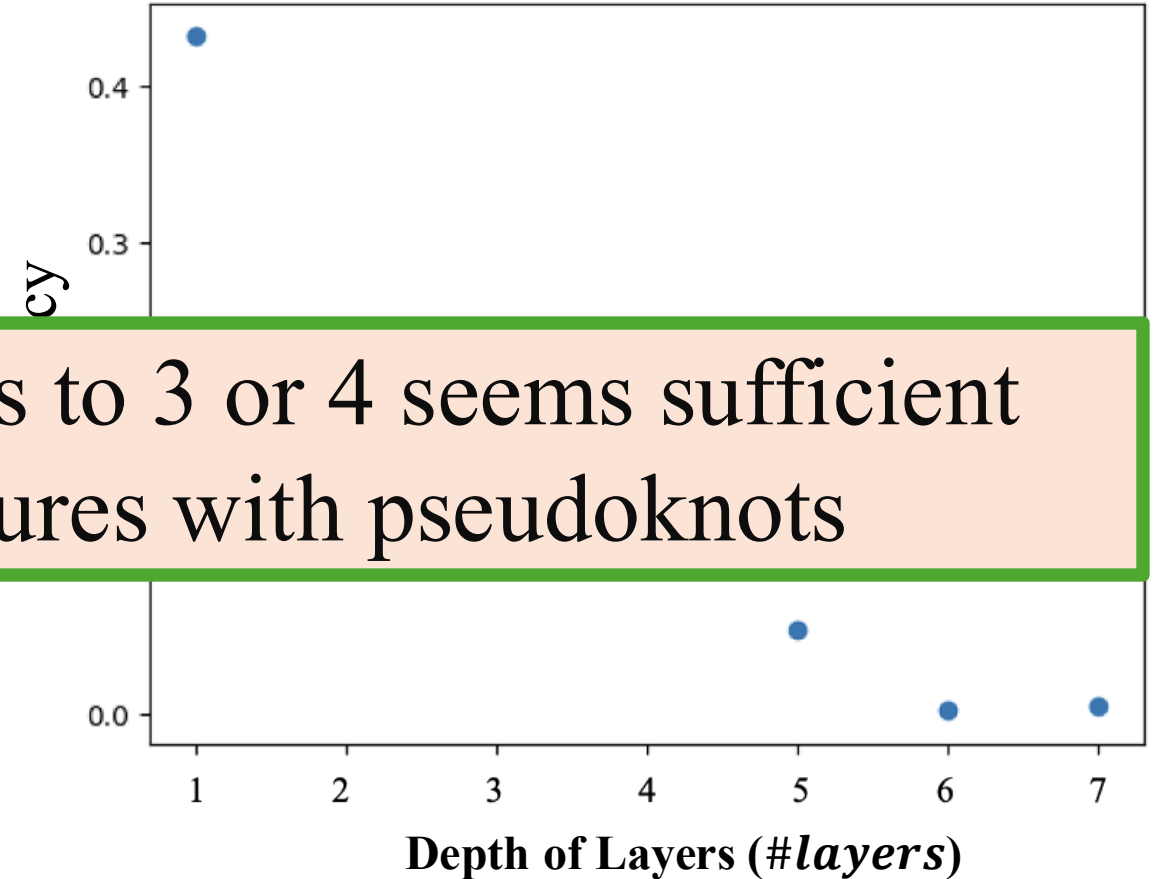
- **Canonical Only: #layers ≤ 4 for 90%**
- **All bp: #layers ≤ 5 for 90%**

Statistics on pseudoknots

Only Canonical bp



All bp including non-canonical



Limiting the depth of layers to 3 or 4 seems sufficient when predicting structures with pseudoknots

Main Layer Layer1: ..(((.....(.....).....))).....
 PK1 Layer2:[[.....]].....
 PK2 Layer3: {.....}

Note:

- **Canonical Only: #layers ≤ 4 for 90%**
- **All bp: #layers ≤ 5 for 90%**

Summary

- PK Visualization

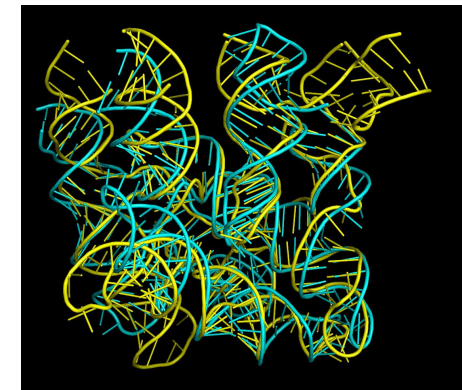
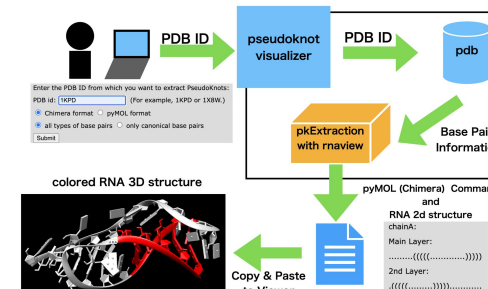
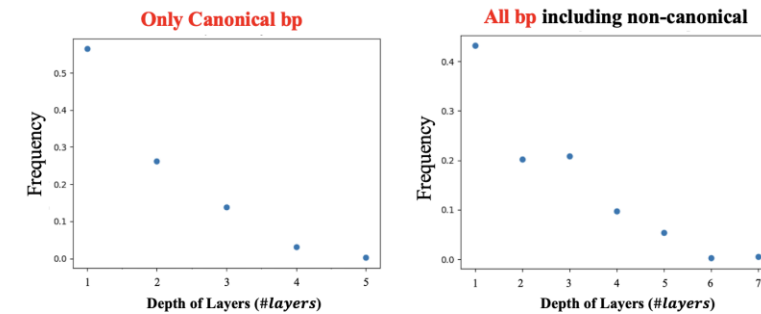
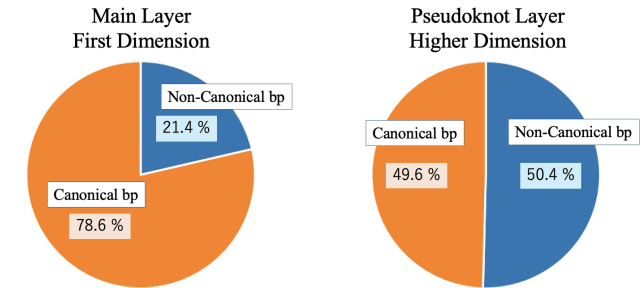
- Canonical bp only or All bp including non-canonical

- Statistics on PK

- Non-canonical bp may be crucial for pseudoknot
- Limiting the depth of layers to 3 or 4 may be sufficient when predicting structures with pseudoknots

- Visualization of PK is useful.

- R1221s2, Casp16 prediction.



Acknowledgements

- Kiyoshi Asai (Asai lab)
- Kengo Sato (casp16)
- Junnichi Iwakiri (casp16)
- Ikuo Kurisaki (casp16)
- Jiro Kondo (casp16)
- Shunsuke Sumi (casp16)
- Kazuteru Yamamura (casp16)

Casp16
RNA_Dojo