

Identification of different structures in a .pdb file

...

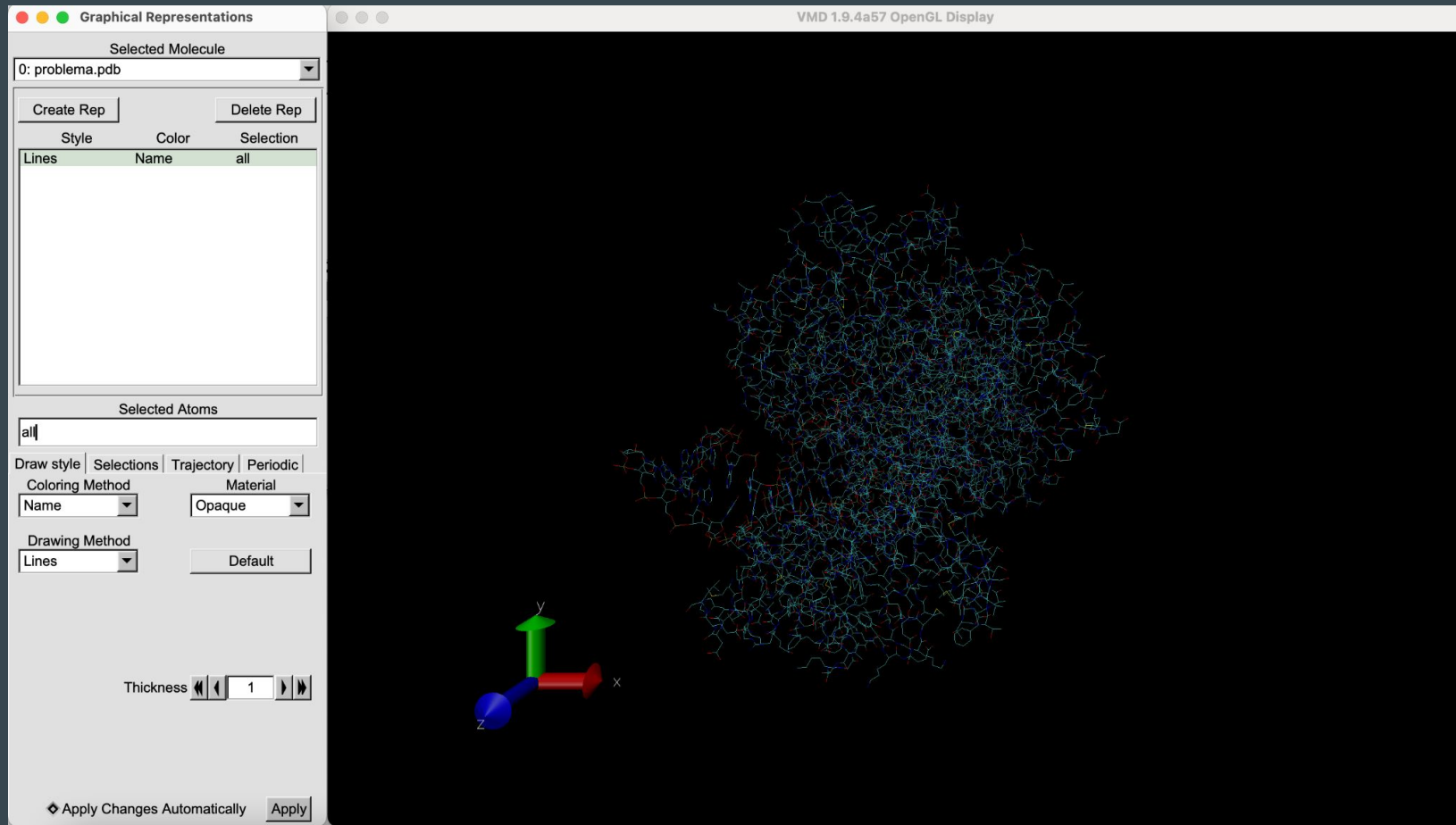
Daniel Bedmar Romero - 1565494
Nanoscience and Nanotechnology 22/23

Justification of the report

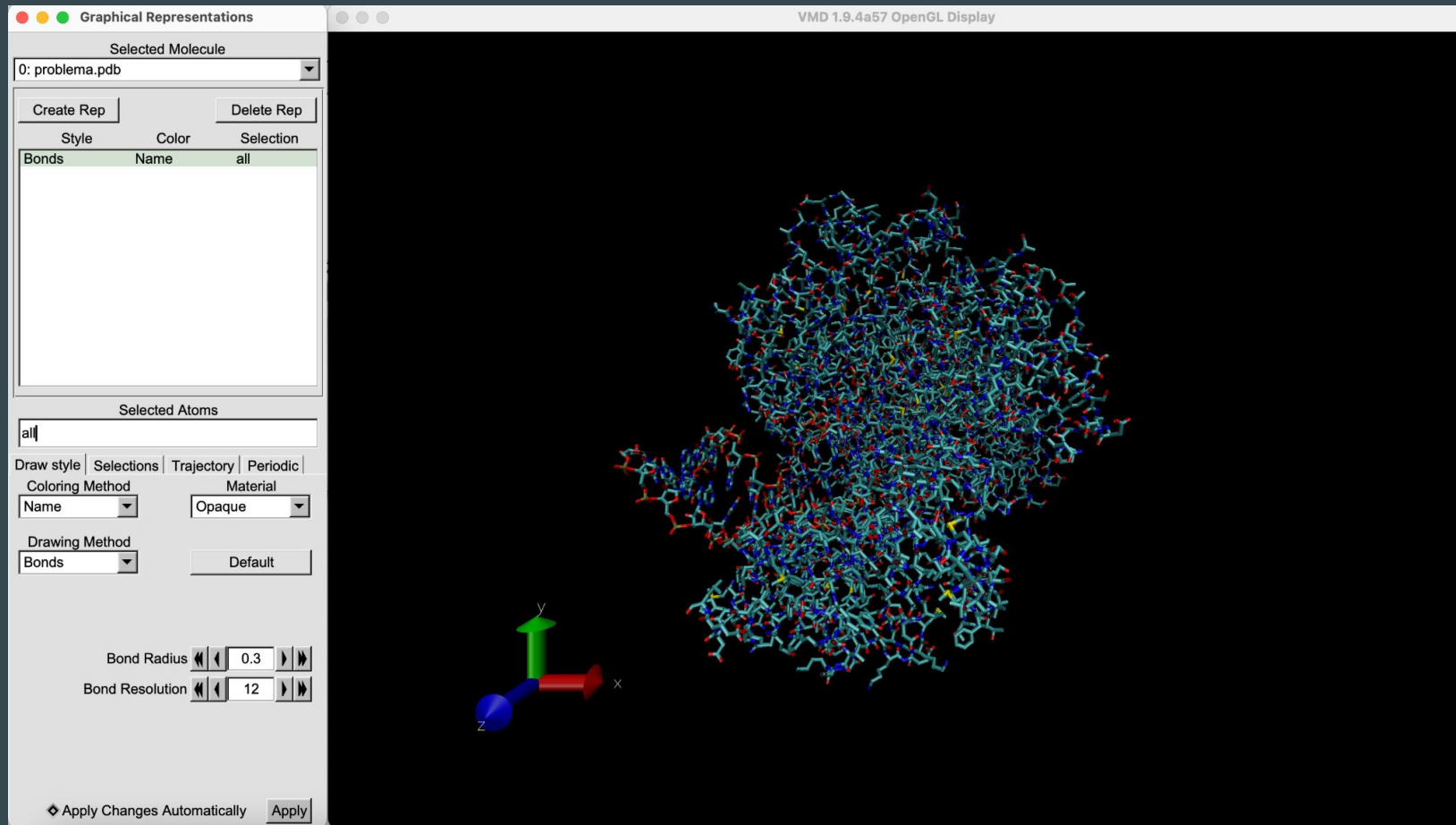
This report is in a power point format because my intention is that it can be followed the process that led me to the identification of the structures. Therefore in this powerpoint we will see this process and at the end the conclusions

The process

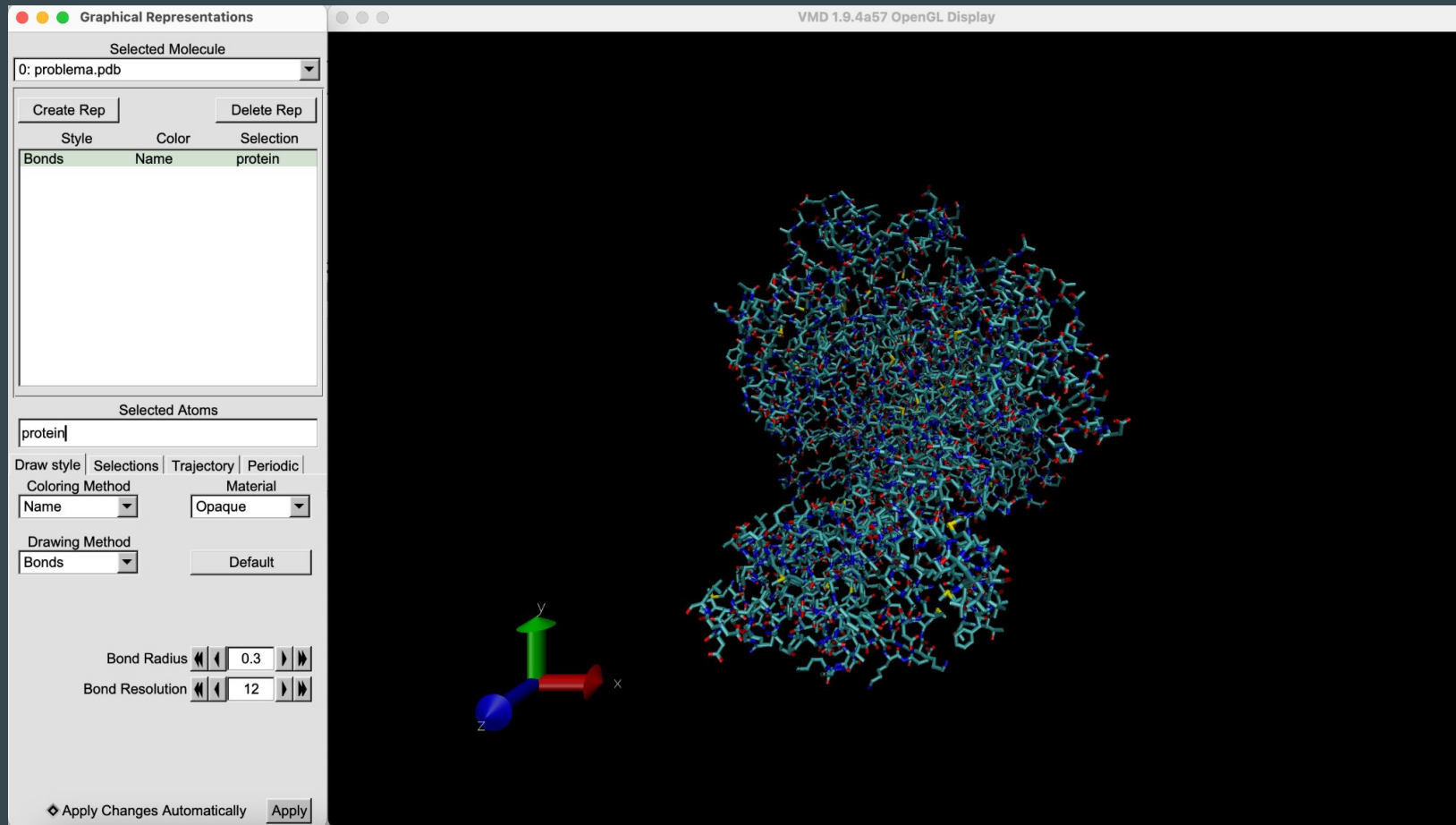
Problema pdb file. This is the problema file opened with VMD



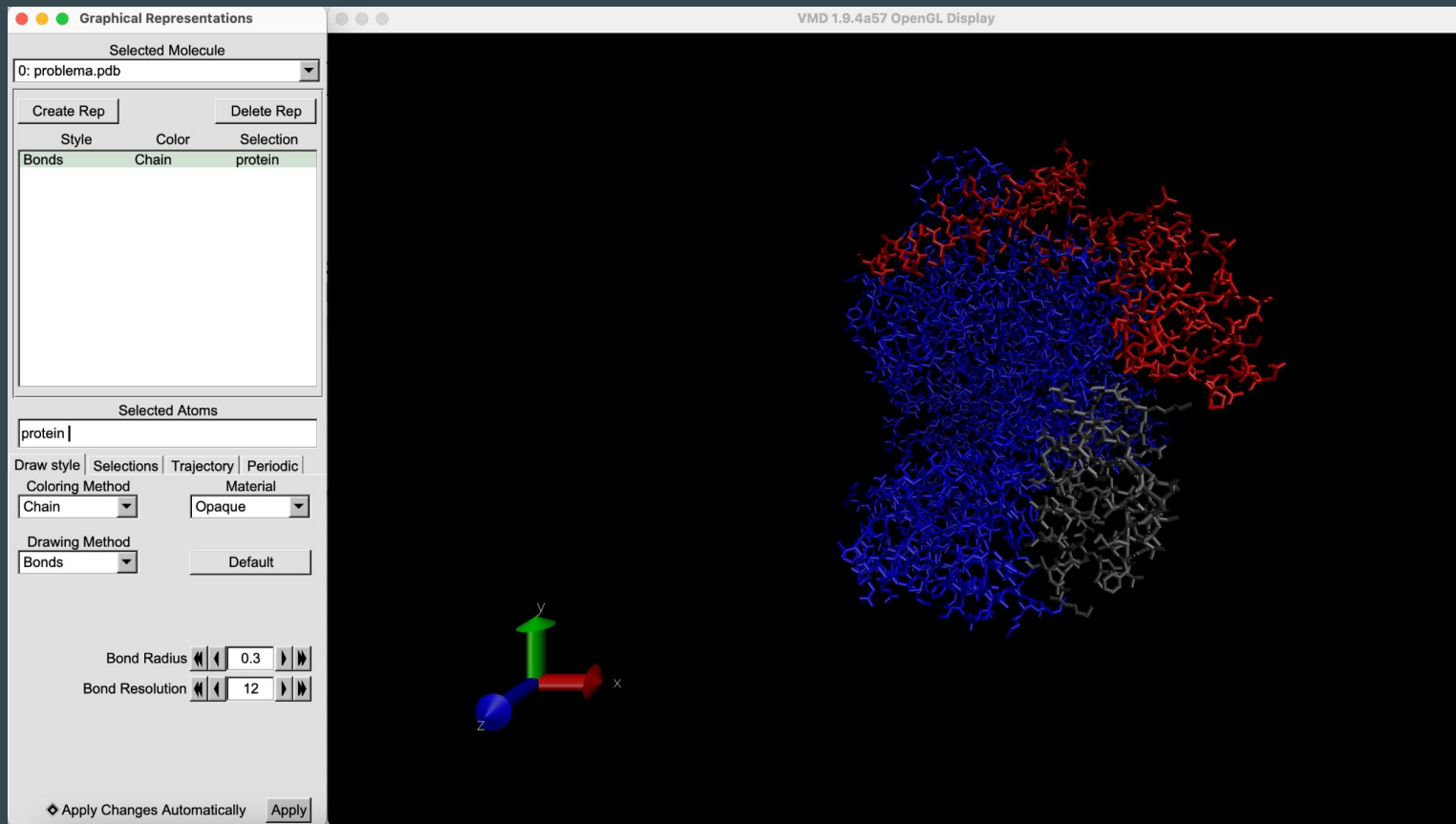
First of all I changed the drawing mode to Bonds so it is more friendly



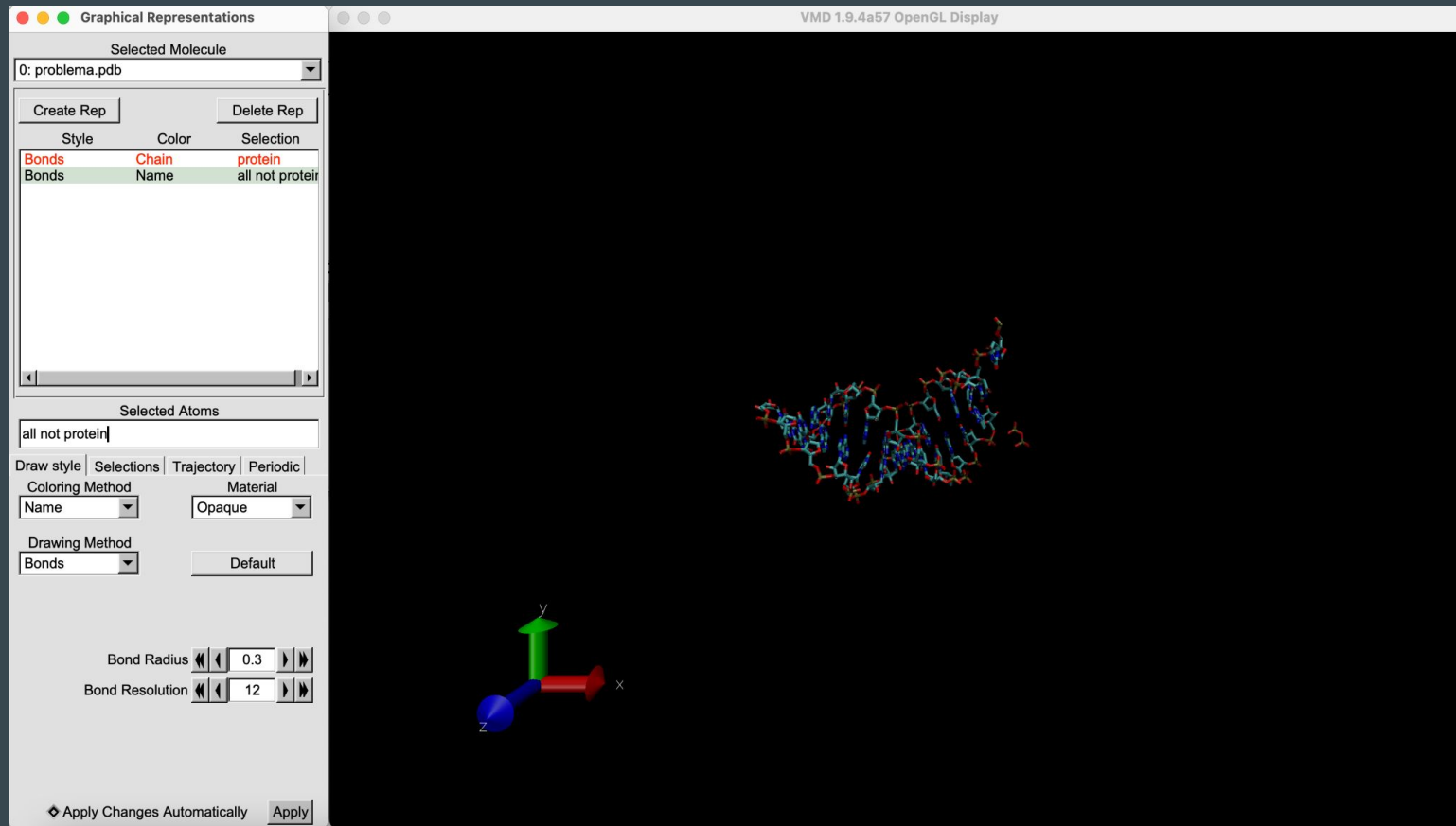
Then I tried to see if there is any protein. Turns out is the most part of it



Then I changed the coloring method to chain in order to see if there are different chains



Then I visualized the rest of the structure by selecting “all not protein”.
It looks like DNA or RNA strands



Because I thought it was DNA or ARN i tried to select the “nucleic”
But by doing this I lost some molecules

Graphical Representations VMD 1.9.4a57 OpenGL Display

Selected Molecule
0: problema.pdb

Create Rep Delete Rep

Style	Color	Selection
Bonds	Chain	protein
Bonds	Name	all not protein
Bonds	Name	nucleic

Selected Atoms
nucleic

Draw style | Selections | Trajectory | Periodic

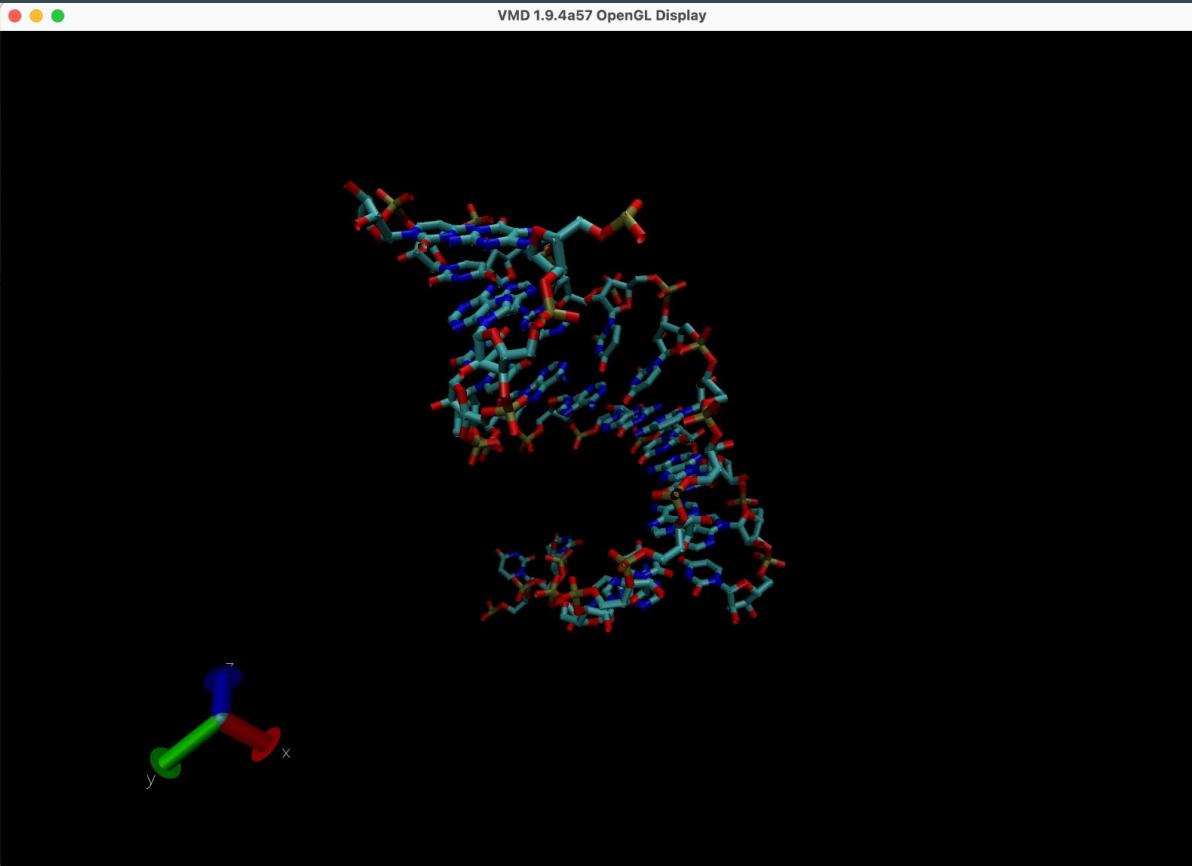
Coloring Method
Name Opaque

Drawing Method
Bonds Default

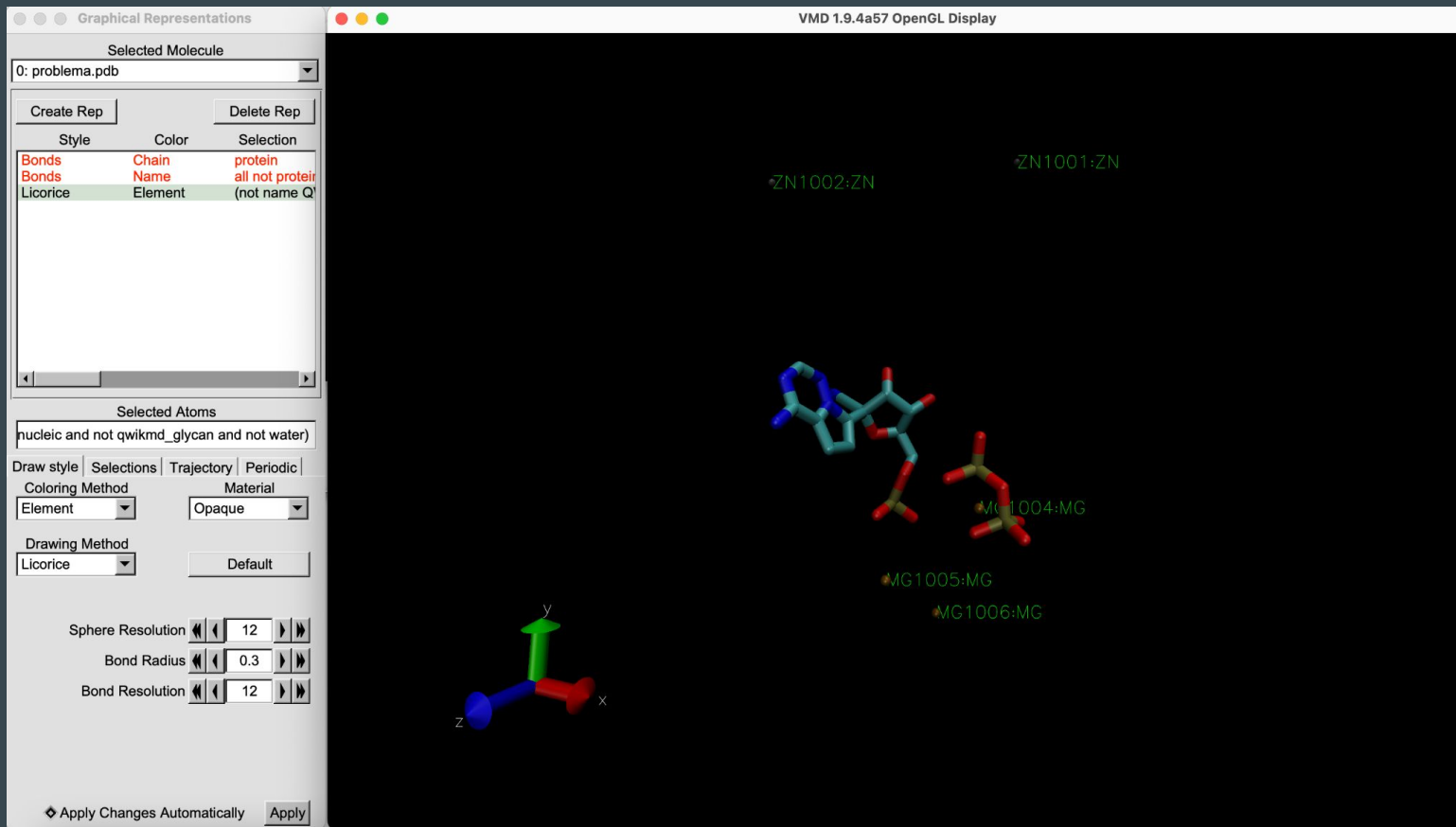
Bond Radius 0.3

Bond Resolution 12

Apply Changes Automatically Apply

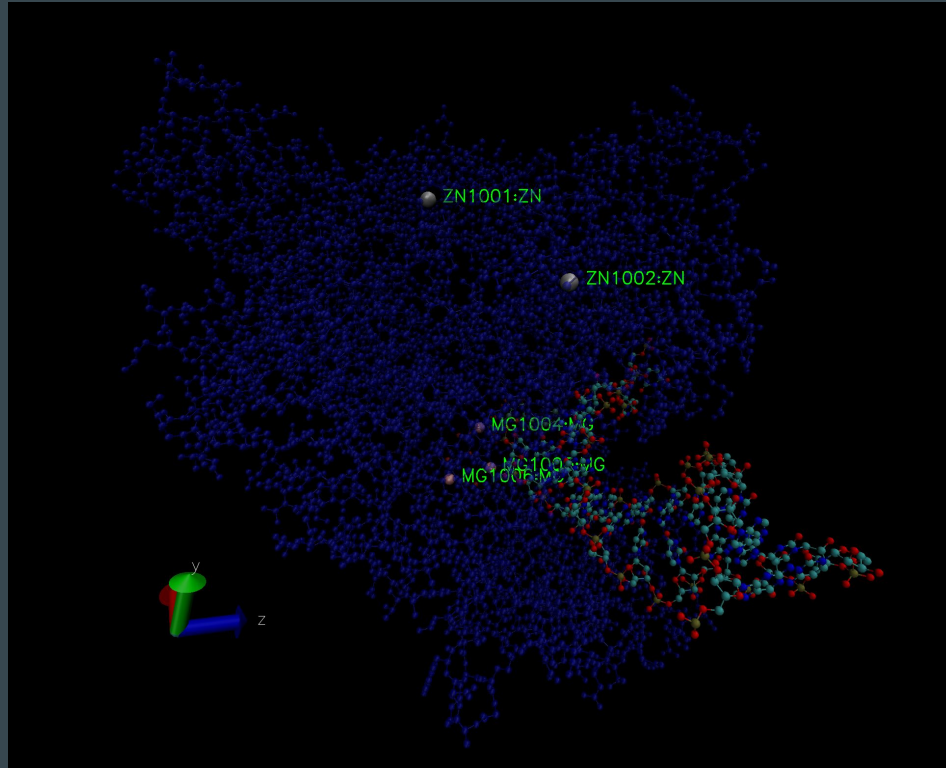


Therefore I went through the different selection options and found one called “qwikmd_hetero” and I found that a part from the nucleotide and the proteins there are 5 atoms and 2 molecules



Deduction of the different molecules

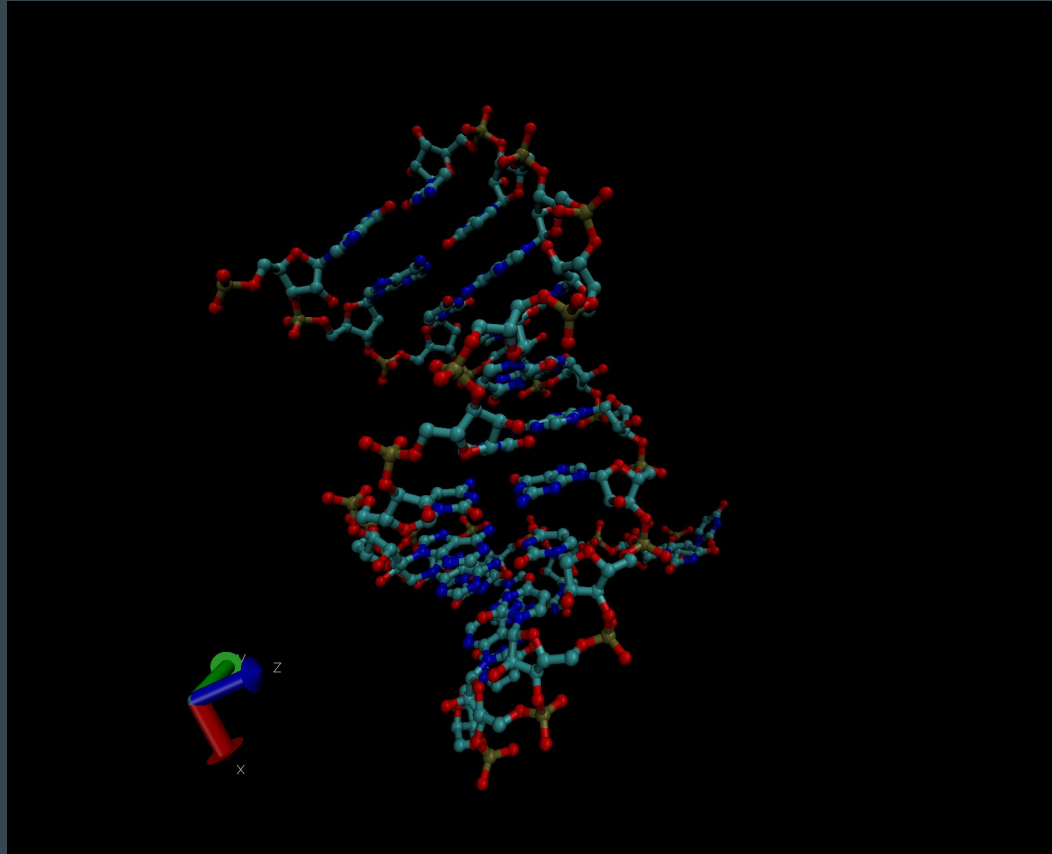
1. The spare atoms: In order to see the atoms I visualized the main chain of the protein and the nucleotide strands and highlighted and labeled the atoms by representing the VDW with a bigger sphere



Based on the location of the atoms in the protein I concluded that the Zinc atoms serve an structural purpose so the chain folds correctly but the Magnesium atoms probably also play a fundamental role in the functionality of the protein. This atoms are called cofactors.

[pdb file of the cofactors](#)

2. The nucleotide strands: In this case the interesting thing to see is if these are RNA or DNA strands. In order to deduce between DNA or RNA I came up with two strategies.



[pdb file of the strands](#)

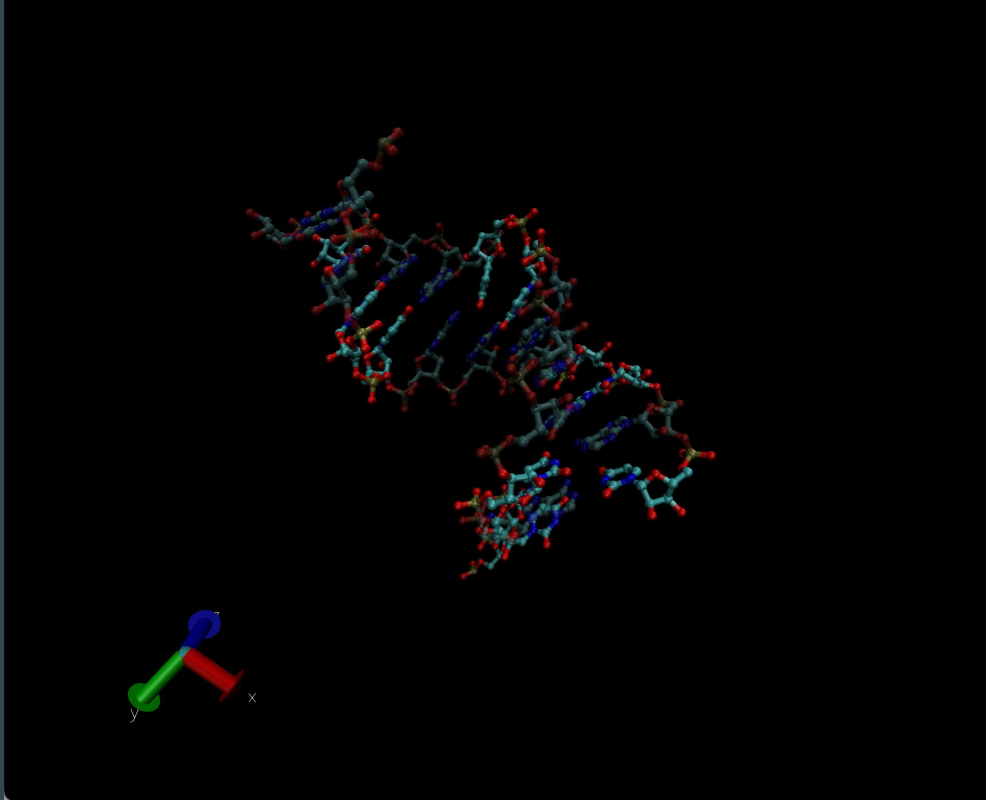
2. The nucleotide strands: In this case the interesting thing to see is if these are RNA or DNA strands.

25	ATOM	8012	P	A	P	11	60.116	91.177	102.435	1.00110.78	P
26	ATOM	8013	OP1	A	P	11	59.714	92.196	101.432	1.00110.78	O
27	ATOM	8014	OP2	A	P	11	61.066	90.101	102.053	1.00110.78	O
28	ATOM	8015	O5'	A	P	11	60.714	91.941	103.698	1.00110.78	O
29	ATOM	8016	C5'	A	P	11	61.656	92.989	103.540	1.00110.78	C
30	ATOM	8017	C4'	A	P	11	61.330	94.154	104.439	1.00110.78	C
31	ATOM	8018	O4'	A	P	11	60.168	93.843	105.248	1.00110.78	O
32	ATOM	8019	C3'	A	P	11	62.399	94.532	105.451	1.00110.78	C
33	ATOM	8020	O3'	A	P	11	63.416	95.340	104.882	1.00110.78	O
34	ATOM	8021	C2'	A	P	11	61.599	95.236	106.541	1.00110.78	C
35	ATOM	8022	O2'	A	P	11	61.333	96.582	106.177	1.00110.78	O
36	ATOM	8023	C1'	A	P	11	60.281	94.459	106.513	1.00110.78	C
37	ATOM	8024	N9	A	P	11	60.201	93.418	107.557	1.00110.78	N
38	ATOM	8025	C8	A	P	11	60.196	92.058	107.377	1.00110.78	C
39	ATOM	8026	N7	A	P	11	60.098	91.370	108.487	1.00110.78	N
40	ATOM	8027	C5	A	P	11	60.025	92.343	109.471	1.00110.78	C
41	ATOM	8028	C6	A	P	11	59.909	92.263	110.869	1.00110.78	C
42	ATOM	8029	N6	A	P	11	59.844	91.111	111.541	1.00110.78	N
43	ATOM	8030	N1	A	P	11	59.863	93.421	111.564	1.00110.78	N
44	ATOM	8031	C2	A	P	11	59.928	94.575	110.888	1.00110.78	C
45	ATOM	8032	N3	A	P	11	60.038	94.780	109.576	1.00110.78	N
46	ATOM	8033	C4	A	P	11	60.082	93.610	108.914	1.00110.78	C
47	ATOM	8034	P	U	P	12	64.816	95.541	105.642	1.00102.01	P
48	ATOM	8035	OP	U	P	12	65.776	96.158	104.691	1.00102.01	O
49	ATOM	8036	OP	U	P	12	65.163	94.262	106.311	1.00102.01	O
50	ATOM	8037	O5'	U	P	12	64.489	96.618	106.768	1.00102.01	O
51	ATOM	8038	C5'	U	P	12	65.427	96.900	107.793	1.00102.01	C
52	ATOM	8039	C4'	U	P	12	64.774	97.572	108.974	1.00102.01	C
53	ATOM	8040	O4'	U	P	12	63.500	96.938	109.263	1.00102.01	O
54	ATOM	8041	C3'	U	P	12	65.543	97.482	110.282	1.00102.01	C
55	ATOM	8042	O3'	U	P	12	66.566	98.456	110.387	1.00102.01	O
56	ATOM	8043	C2'	U	P	12	64.449	97.632	111.327	1.00102.01	C
57	ATOM	8044	O2'	U	P	12	64.082	98.996	111.475	1.00102.01	O
58	ATOM	8045	C1'	U	P	12	63.289	96.892	110.659	1.00102.01	C
59	ATOM	8046	N1	U	P	12	63.200	95.467	111.064	1.00102.01	N
60	ATOM	8047	C2	U	P	12	63.077	95.110	112.402	1.00102.01	C
61	ATOM	8048	O2	U	P	12	63.054	95.898	113.334	1.00102.01	O
62	ATOM	8049	N3	U	P	12	62.996	93.759	112.632	1.00102.01	N

The first strategy thing I did was to look at the pdb file to see if there were some Uracile nucleotides or not. I found out that it seems like there are Uracile residues in both strands.

pdb file of the strands

2. The nucleotide strands: In this case the interesting thing to see is if these are RNA or DNA strands.

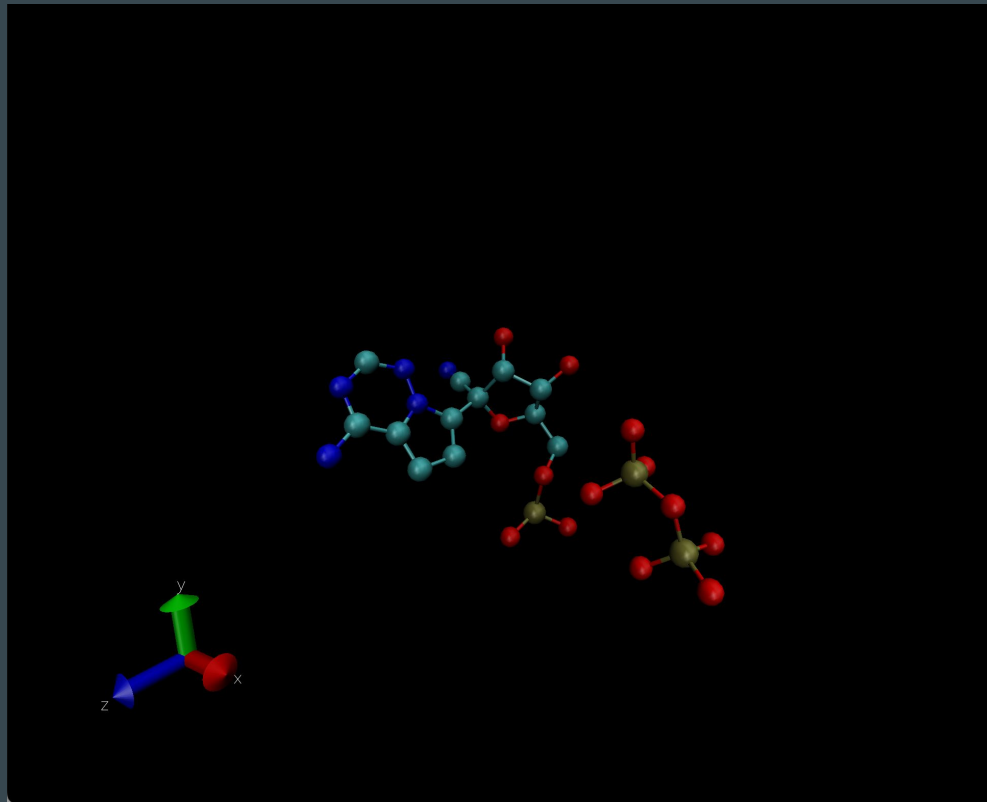


To confirm that the Uracile were in both strands, I created a new pdb file only with the Uracile residues and overlap its representation with the representation of the strands but in transparent “material”.

As it can be seen in the figure besides, there are Uracile residues in both strands, therefore, these are two RNA strands

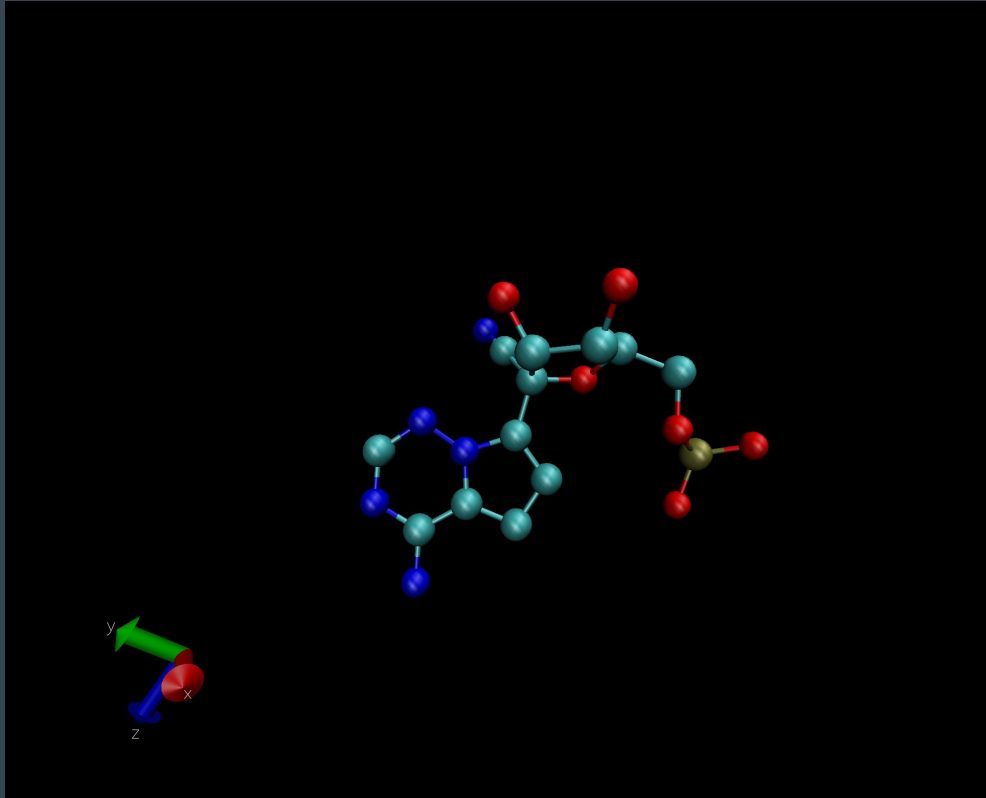
[pdb file of the strands](#)

3. The spare molecules: I wanted to identify both molecules



3. The spare molecules: I wanted to identify both molecules

The first molecule

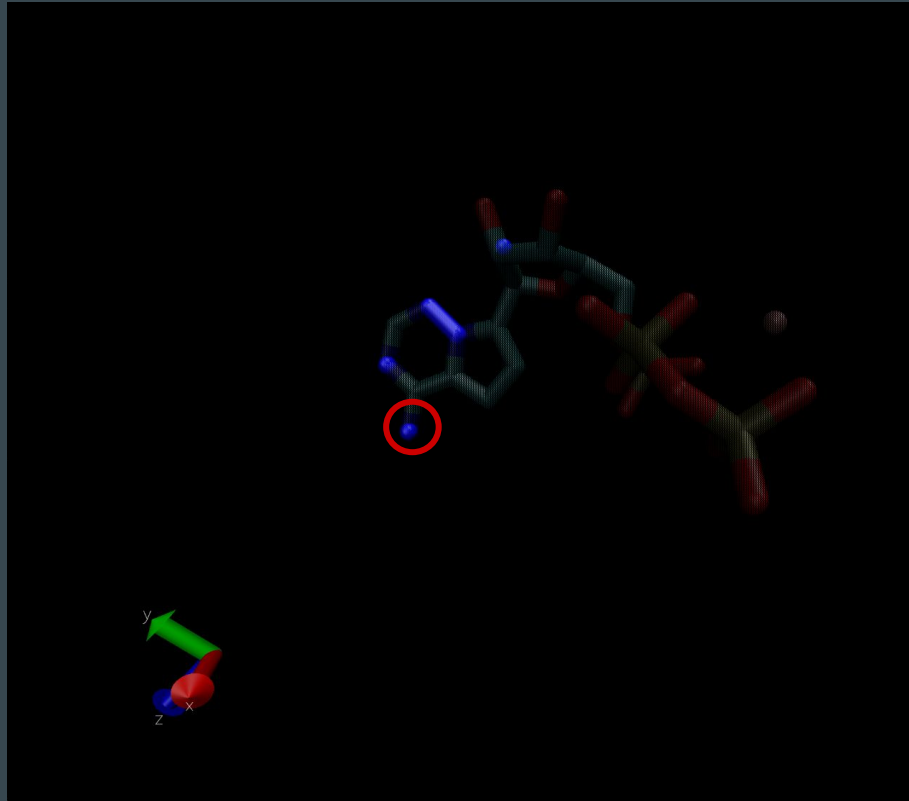


The first molecule I identified was the big one. I identified it as an **Adenine precursor/Adenine rest**. In order to do so I compared it with the rest of the nucleotides because it was similar to them and as it was in the active site of the protein it made sense to me that it had to be related with the nucleotides.

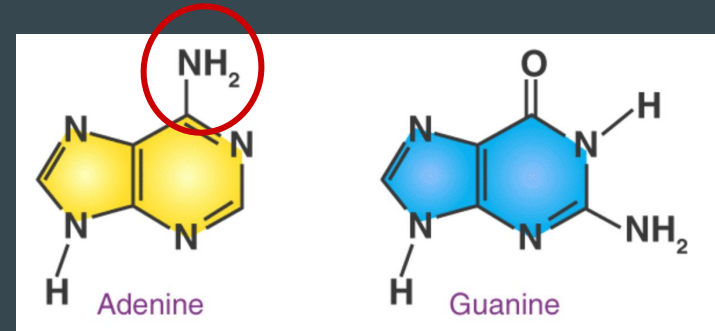
pdb file of the Adenine precursor/rest

3. The spare molecules: I wanted to identify both molecules

The first molecule



I identified it as an Adenine related molecule because from the double ringed nucleotides is the only one that only has a functional group in the external group and it's a NH₂ group

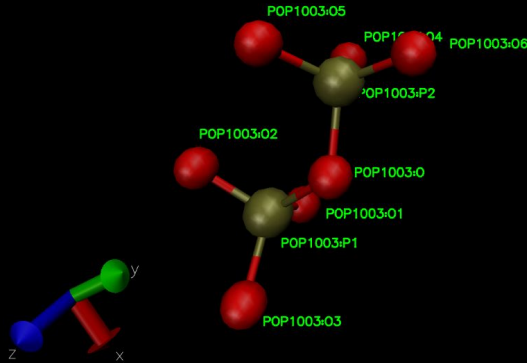


pdb file of the Adenine precursor/rest

3. The spare molecules: I wanted to identify both molecules

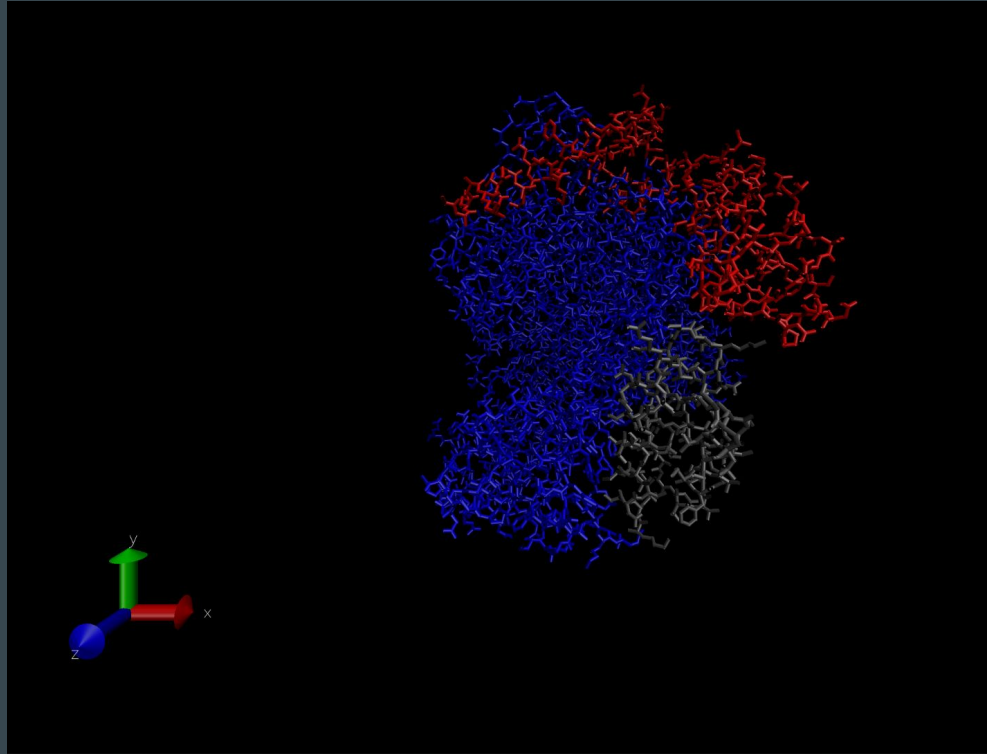
The second molecule

I identified the second molecule as two Phosphate bonded together by identifying the atoms that formed the molecule



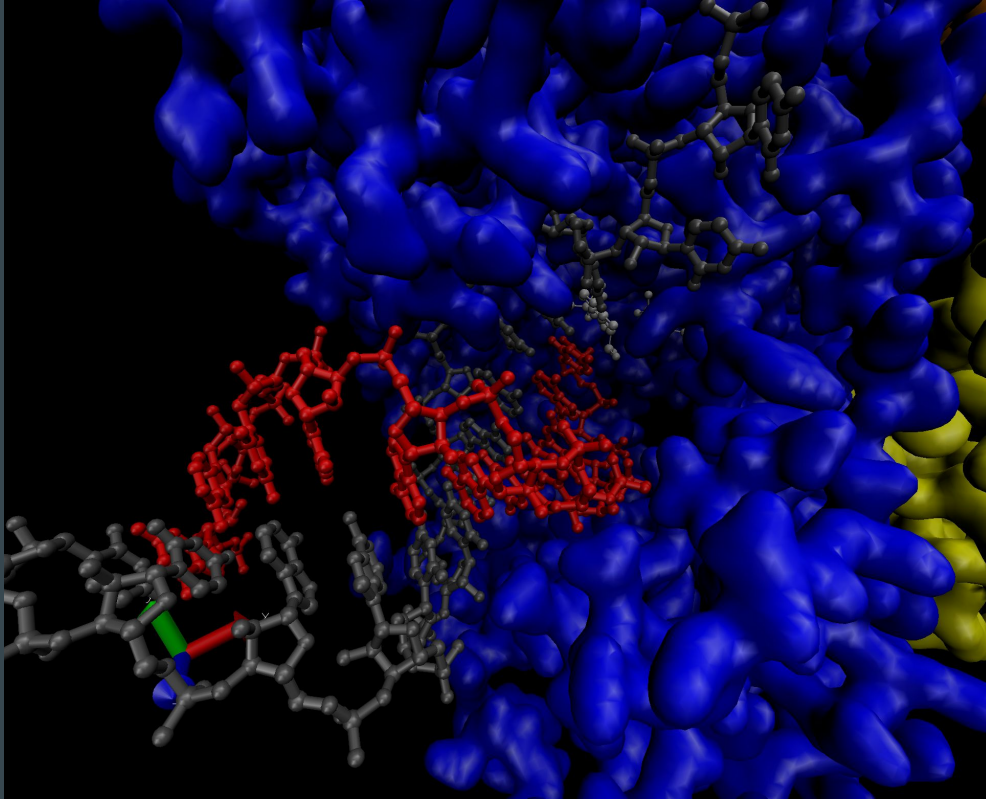
pdb file of the 2 Phosphate group

4. The protein. In this case I did not intended to identify the exact protein but to make a valid guess of which type of protein was and what was its biological role



pdb file of the protein

4. The protein. In this case I did not intended to identify the exact protein but to make a valid guess of which type of protein was and what was its biological role



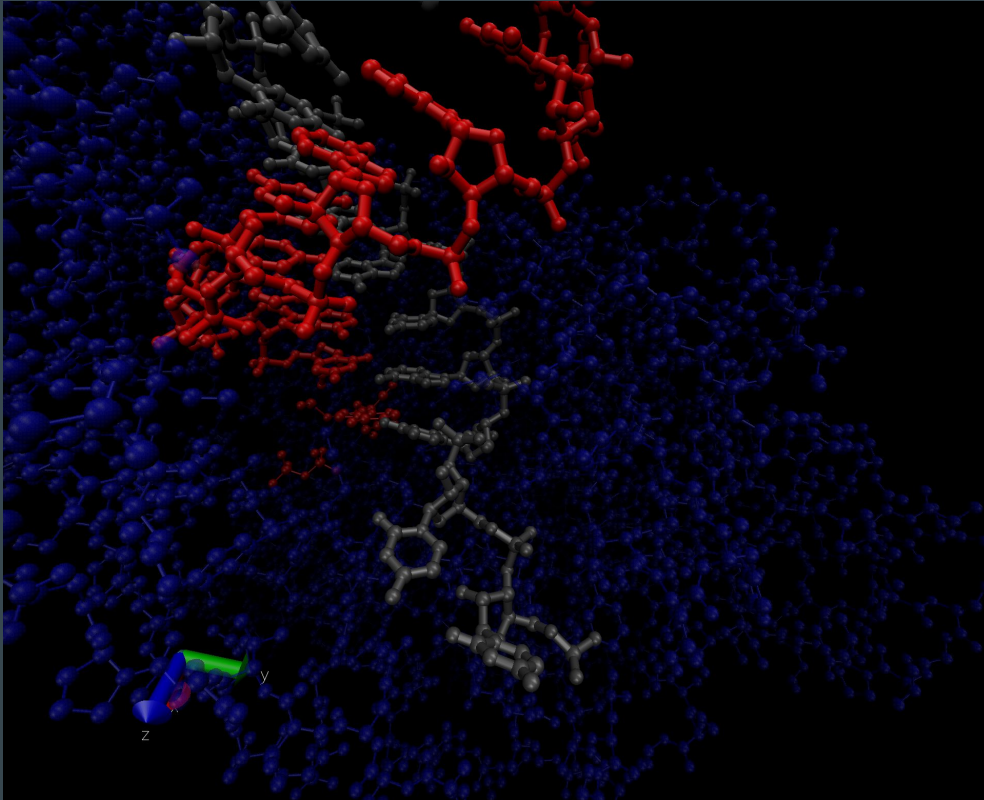
In order to make a guess about this protein I had different things I have deduced during this whole exercise:

- The protein clearly has an active site where a double strand of RNA is located
- Only one of the strands is inside the active site and the other one seems to be separating from the strand located in the protein
- Near the active site there is an Adenine precursor (or the rests of an Adenine) and two bonded phosphate groups
- It seems to be a hole from where molecules enter/exit the protein in the opposite side from which the protein enters

pdb file of the Adenine precursor

Conclusion

Conclusion



I have deduced that this protein could be **copying an ARN strand**, because the 2 bonded phosphate could be the rest of an ATP and it is being used to use the energy of cleaving the Phosphate groups to add one of them to link the nucleotides and the nucleotides could be entering to the active site by the hole in the main chain (maybe if there are not enough Zn this hole could remain closed and the protein turned off). In this case, the model strand would be the gray one.

This protein could be found in virus or near a celular nuclei and it would be used in the transcription of genetic material.

[Github repository](#)