NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

School of Science

Information Technologies in Medicine and Biology

Direction: *Bioinformatics*

Algorithms in Structural Bioinformatics

Postgraduate Student: *Begetis Nikolaos* Professor: *Chrysina Evangelia*

Deadline Date: *17/04/2013*

Assignment 2

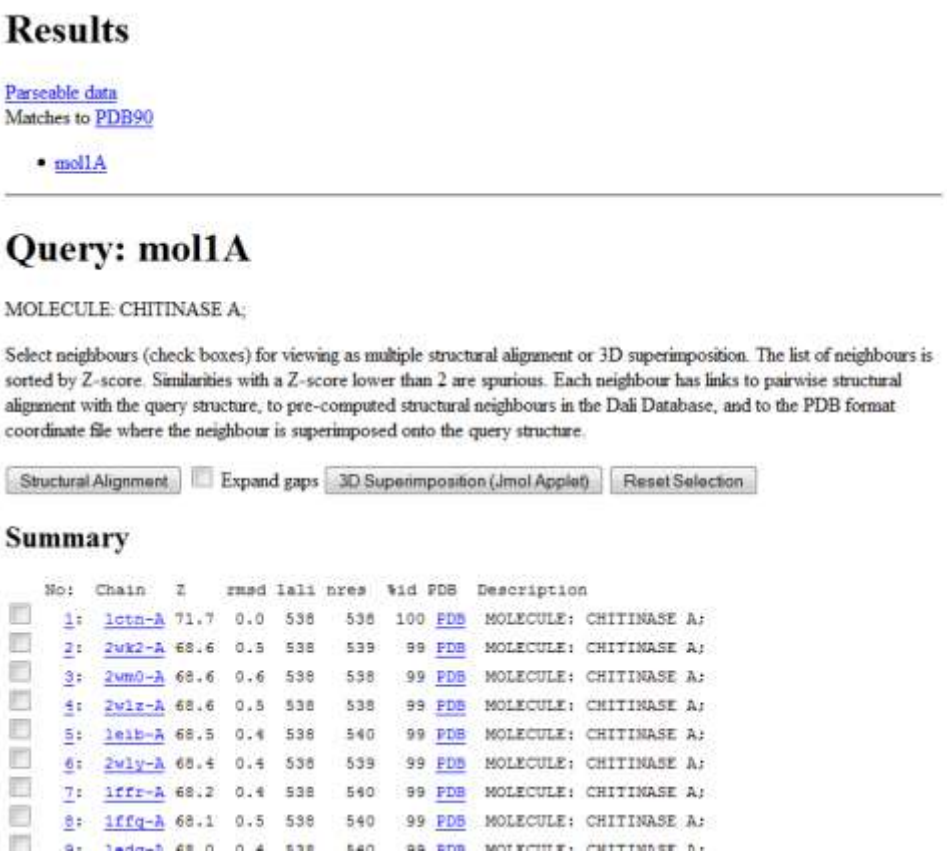
1

In this second assignment we were asked as a first task to continue the previous first assignment and find the structural homologues of the enzyme with PDB code 1ctn, which as we saw corresponds to Chitinase and is found in A chain. In addition, we were asked to use either the PDBeFold tool of EMBL-EBI PDB, or the Dali server, or both of them, and write down the procedure we followed and the search parameters we used.

To begin with, in this task we chose to use the Dali server tool, clearly for personal academic reasons, so as to learn one more tool, as in the previous exercise we used PDBeFold. So, as it is shown in Figure 1, we searched for the enzyme 1ctn in the dali server.

Figure 1

The results returned from this search were similar to those we took in the previous assignment from the PDBeFold tool. Figure 2 shows a screenshot of the results. We also saved them in an HyperText format in file res1.htm provided with the submission.



Results

Parseable data
Matches to PDB90

- mol1A

Query: mol1A

MOLECULE: CHITINASE A;

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimposition. The list of neighbours is sorted by Z-score. Similarities with a Z-score lower than 2 are spurious. Each neighbour has links to pairwise structural alignment with the query structure, to pre-computed structural neighbours in the Dali Database, and to the PDB format coordinate file where the neighbour is superimposed onto the query structure.

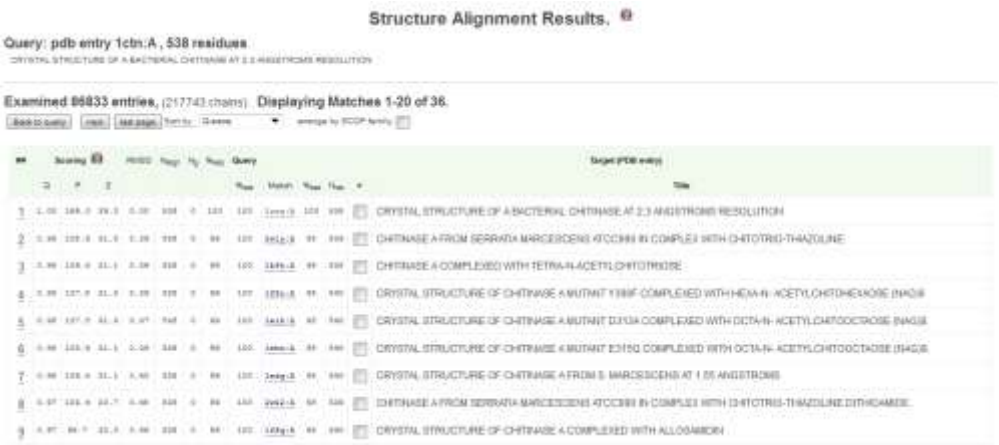
Structural Alignment | Expand gaps | 3D Superimposition (Jmol Applet) | Reset Selection

Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	1ctn-A	71.7	0.0	538	538	100	PDB	MOLECULE: CHITINASE A;
2:	2wk2-A	68.6	0.5	538	539	99	PDB	MOLECULE: CHITINASE A;
3:	2wm0-A	68.6	0.6	538	538	99	PDB	MOLECULE: CHITINASE A;
4:	2wlz-A	68.6	0.5	538	538	99	PDB	MOLECULE: CHITINASE A;
5:	1eib-A	68.5	0.4	538	540	99	PDB	MOLECULE: CHITINASE A;
6:	2wly-A	68.4	0.4	538	539	99	PDB	MOLECULE: CHITINASE A;
7:	1ffr-A	68.2	0.4	538	540	99	PDB	MOLECULE: CHITINASE A;
8:	1ffq-A	68.1	0.5	538	540	99	PDB	MOLECULE: CHITINASE A;

Figure 2

We traced back to the previous assignments results and there we withdrew the file with the **Pairwise** results (Figure 3), with name reslist_pair.dat, so that we could compare the results of both databases and find out if the results are similar to each other (reslist_pair.dat and res1.htm files). We concluded that, indeed, the results were the same.



Structure Alignment Results.

Query: pdb entry 1ctn:A , 538 residues

Examined 86833 entries, (217743 chains) Displaying Matches 1-20 of 36.

Figure 3

The only differences between these two protein data bases for finding structural homologues was that Dali_server had in similar with the PDBeFold database the features: Z-score, rmsd, nres and %id, which respectively in PDBeFold were Z-score, RMSD, N_{align} and $\%_{seq}$. In addition, PDBeFold provides some more features such as the Q-score and P-score and if any of them is hovered with the mouse when used, it is explained in detail what is their role of existence. To complete our study in this task, the results in PDBeFold are sorted using the Q score of proteins while the results In Dali server are sorted with the z-score. Nevertheless, both result set are have the same enzyme proteins included.

2

As a second task we were asked to select the five-first and the five-last protein structure homologues with the highest and lowest scores respectively in the pairwise alignment with the 1ctn chitinase. At first, we were assigned to superpose these structures in the same Cartesian system and describe the procedure and in continuation to build a figure that we present the superimposed 3D structures with a molecular graphics application, in our case the WinCoot 0.7, in which we were referenced in class lecture.

So, to begin with, we gathered the class of the first-five protein structure homologues sorted by the Z-score of Dali server (from now on, Class A) and the last-five protein structure homologues (Class B).

It is worthy saying, that as we emphasized in class lecture, these two classes should not include structure complexes from the same protein, because this would result in finding "*nothing*". For this reason, both classes A and B include the highest and lowest five, respectively, proteins that are different among them.

In order to find where each of the enzyme belong, we made a projection in the feature of "Description" of Dali server results. So the results of class A as taken from Dali server are (we also consider 1ctn as one of the five highest in score, we could also have omitted it and begin from the next in order, but it is not a big deal):

CLASS A

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	1ctn-A	71.7	0.0	538	538	100	PDB	MOLECULE: CHITINASE A;
37:	1itx-A	46.9	1.7	391	419	32	PDB	MOLECULE: GLYCOSYL HYDROLASE;
38:	1lg2-A	44.3	1.6	337	359	26	PDB	MOLECULE: CHITOTRIOSIDASE;
39:	1hkk-A	44.2	1.6	340	364	26	PDB	MOLECULE: CHITOTRIOSIDASE-1;
41:	1ll4-A	44.0	1.5	355	392	32	PDB	MOLECULE: CHITINASE 1;

And the results for class B, respectively are (in here we made the projection in a reverse mode, we kept the worst of every different protein):

CLASS B

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
962:	3faw-A	4.1	3.6	89	775	13	PDB	MOLECULE: RETICULOCYTE BINDING PROTEIN;
963:	3qs9-E	4.0	3.1	61	405	10	PDB	MOLECULE: SL CYTOKINE;
967:	3o4o-C	3.4	3.5	65	316	11	PDB	MOLECULE: INTERLEUKIN-1 BETA;
968:	2vzv-A	3.3	4.7	211	857	8	PDB	MOLECULE: EXO-BETA-D-GLUCOSAMINIDASE;
970:	4amc-A	2.4	4.1	87	1019	5	PDB	MOLECULE: GLUCANSUCRASE;

Now that we defined the members of each class, the next thing we have to do is to superpose these structures in the same Cartesian system so that we can observe the alterations, but before that we have to download the pdb file of each structure homologue. To do that, the only thing needed was to click on the PDB column link for every structure and save the file opened in browser as a PDB file format. All the 10 pdb files are included in the deliverable of this assignment, in separate folders.

We are now ready to superpose these structures in the same Cartesian system and in continuation to build a figure that will present the superimposed 3D structures.

To do that we used the WinCoot molecular graphics application. At first we tried to use the option File→Open→OpenCoordinates, but the .pdb formatted files for an inquisitive reason could not be opened and so we fetched directly from the database the enzyme proteins we needed. Figure 4 shows how we did that.

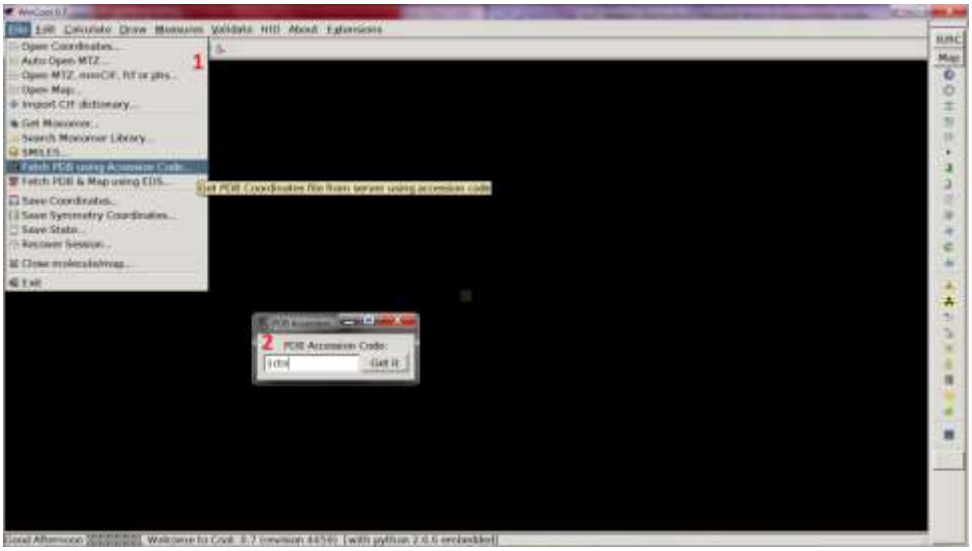


Figure 4

So fetching the pdb file coordinates from server using the accession code of the 1ctn chitinase enzyme protein resulted in Figure's 5 depiction.

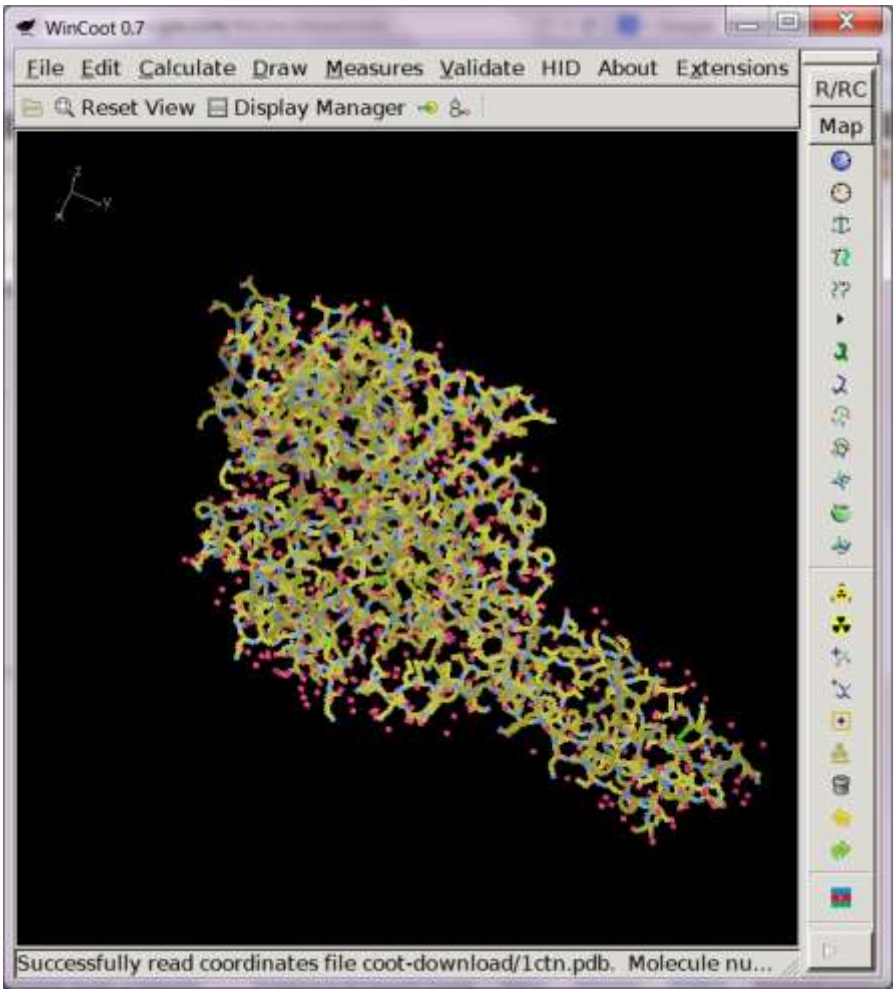


Figure 5

We followed the exact same procedure for the rest four homologue structures highest in score and after that for the 1ctn and the five proteins with the minimum score.

The results for the Class A are shown in Figure 6 and of 1ctn and Class B are shown in Figure 7. Figure 6 has been rotated 180^{o} degrees keeping the y axon stable, in order to provide a good view of the rest four enzymes. Figure 7 has been rotated -90^{o} degrees keeping the y axon stable , in order to provide a good view of the rest five enzymes.

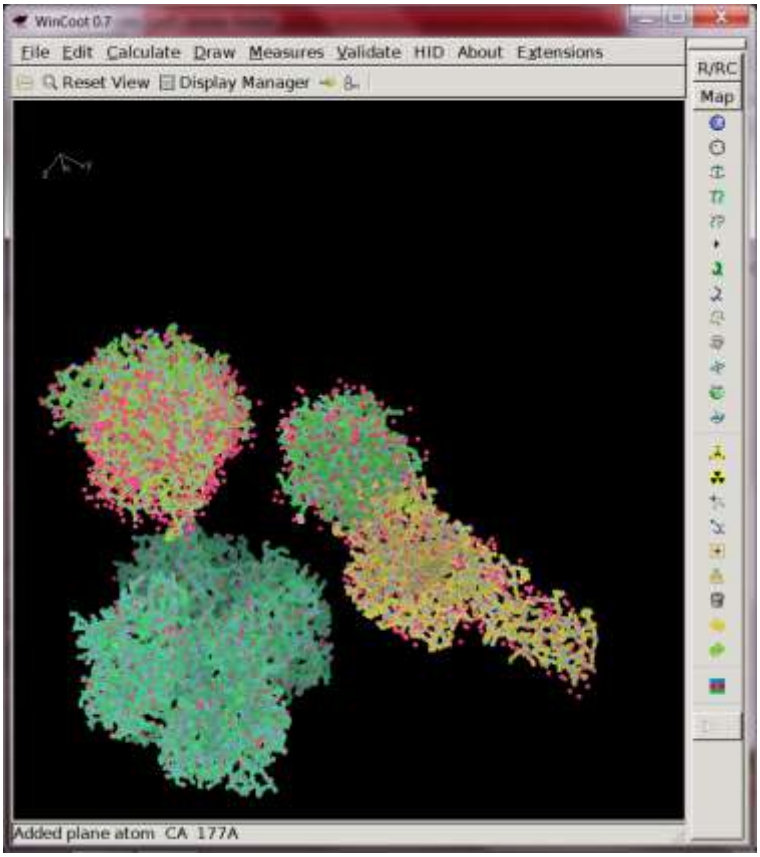


Figure 6

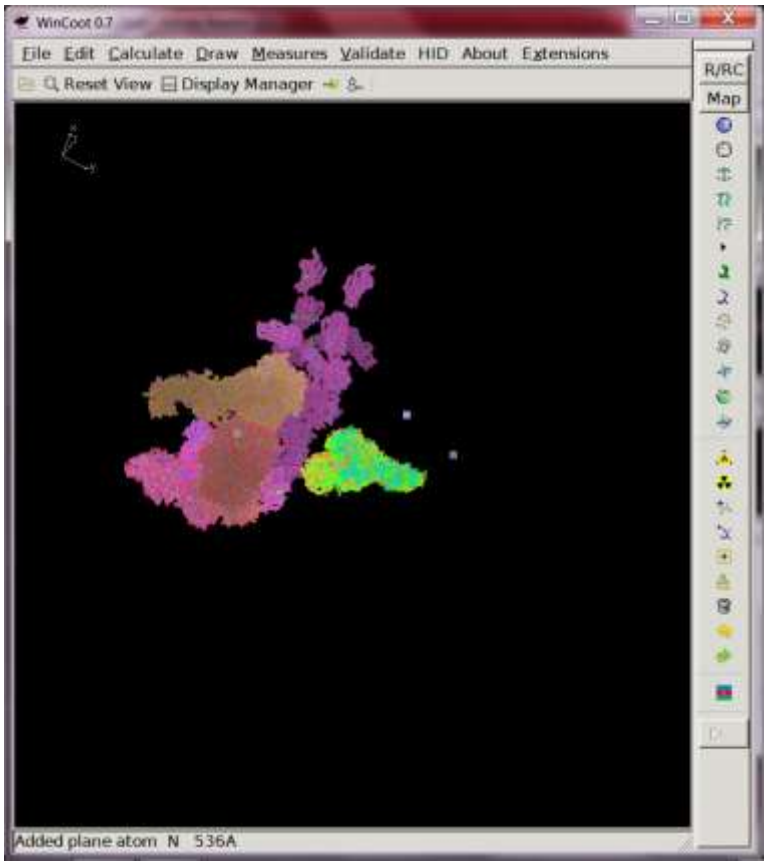


Figure 7

This view (Figure 6 and Figure 7) gives us a lot of details, but for our purpose we do not need such much. To change the representation of the proteins we can hit the button F7 or we can select the option Display Manager at the top menu bar.

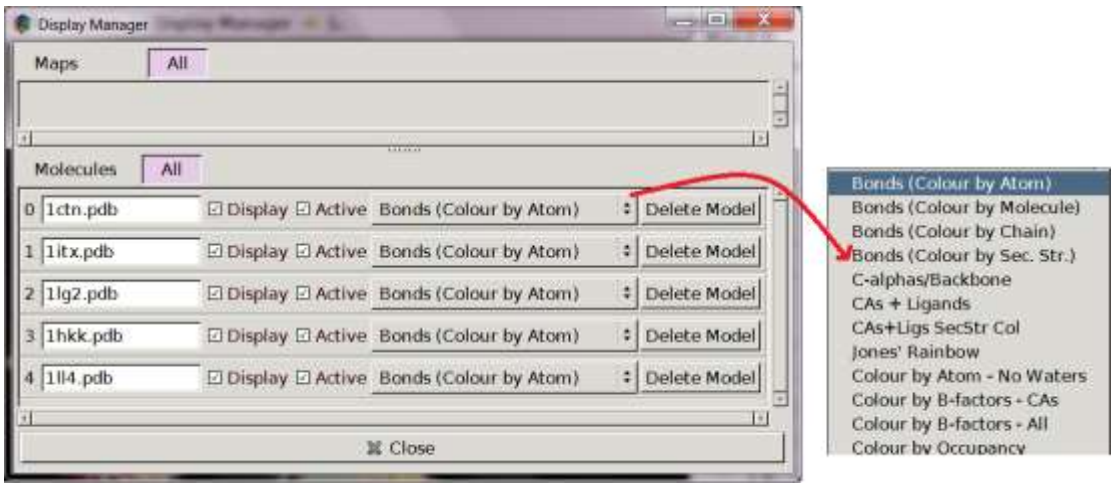


Figure 8

A better representation to distinguish the protein molecules of Class A is for example when we select the option “Color by Molecule”. And an even better to get rid of the information we don’t need at this moment, is to select the option “C-alphas/Backbone”. These two depictions are shown in Figures 9 and 10, respectively.

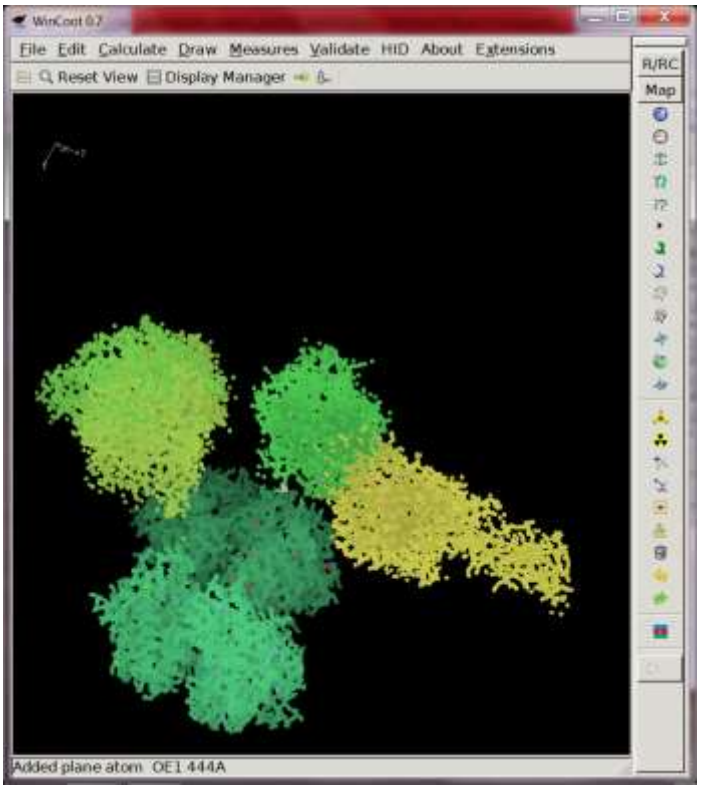


Figure 9

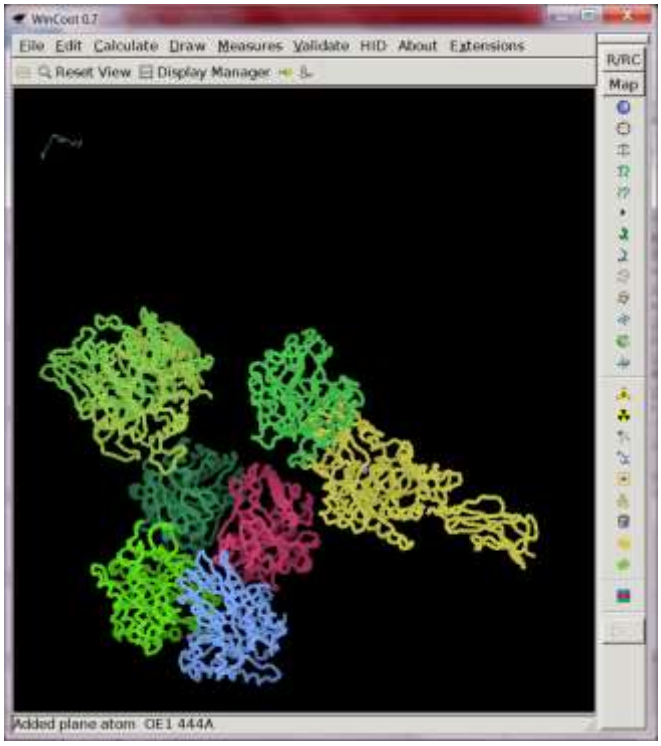


Figure 10

Now that all the homologue structures of both classes are loaded in WinCoot, the only thing remaining for the first subtask of this task is to superpose all the other proteins in the same Cartesian coordinates system with the 1ctn chitinase whose structure we are going to use as the stable structure for these classes A and B superpositions. To do that using WinCoot we have to select the option Calculate→"SSM Superpose" and there for all the homologue structures of both classes we use as the reference structure the 1ctn chitinase and as the moving structure, all of the homologue structures, as it is shown in Figure 11. Then the final superpositions of element structures of each class superposed over 1ctn chitinase are shown respectively in Figures 12 and 14. Figure 13 shows that we must not forget to set the chain of the structure homologue because many of them have more than one chain, in contrast to 1ctn that only has the A-chain.

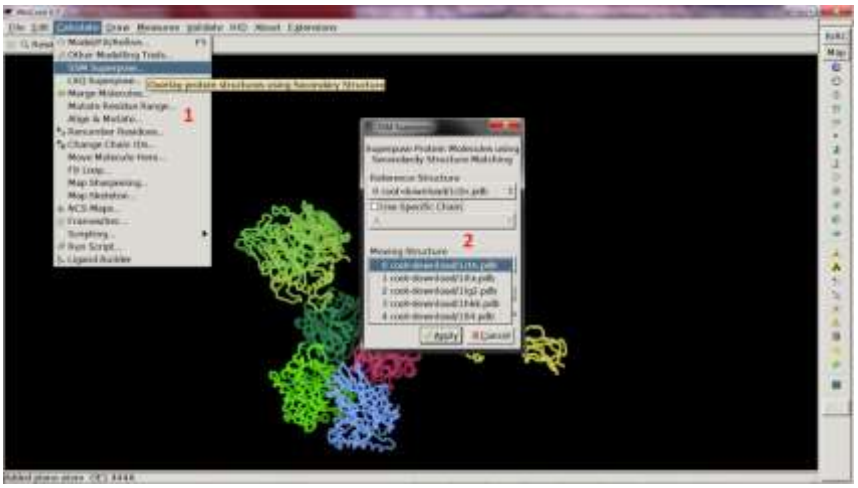


Figure 11

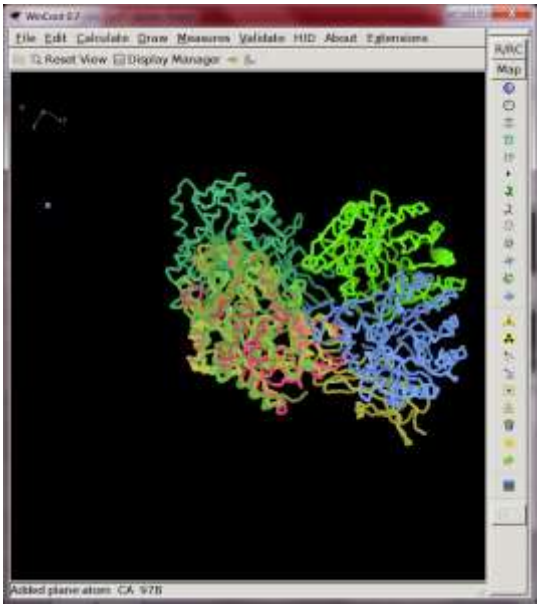


Figure 12

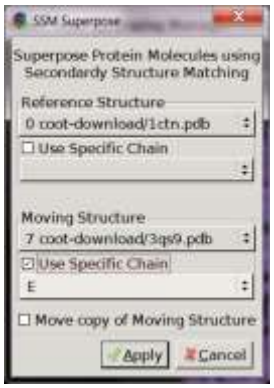


Figure 13

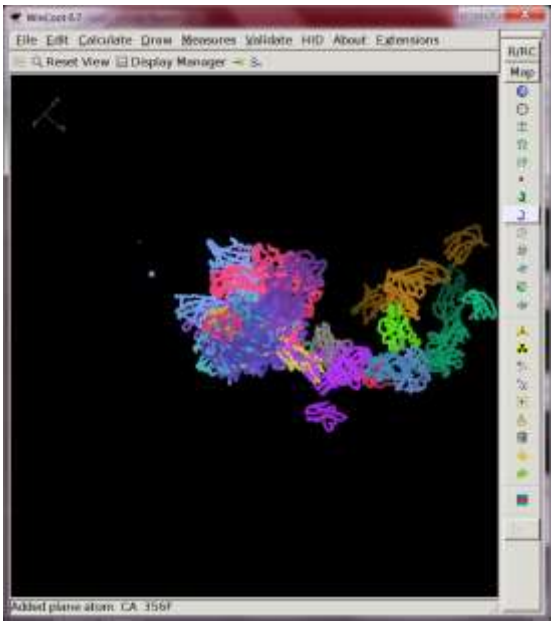


Figure 14

For the final subtask of this assignment we were asked to create an image which will represent all the 3D superimposed structures. To do that the only thing to be done is to change back the option of the Display Manager we depicted in Figure 8 to some other display options such as the “Color by Atom”, “CAs + Ligands”, “CAs + Ligs SecStr Col” and “Color by B-factors - All”. These four alternate 3D superpositions are presented in the next figures in this order for both classes A and B, beginning from Class A (Figures 15-22).

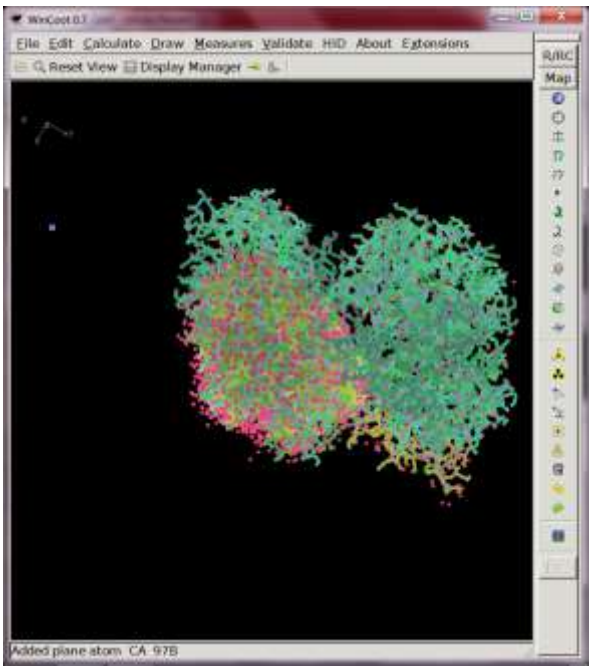


Figure 15

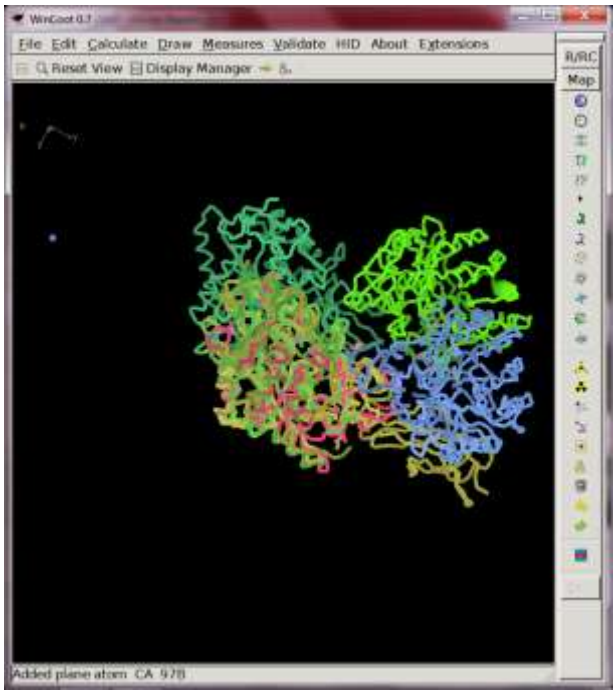


Figure 16

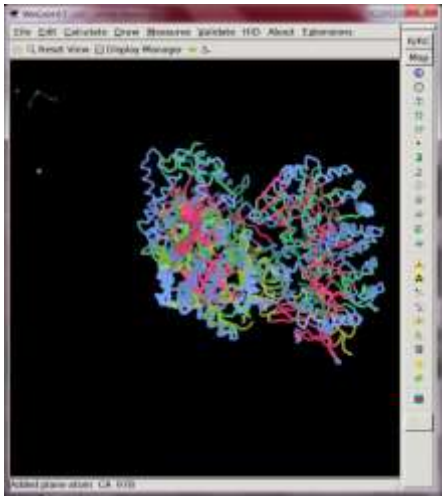


Figure 17

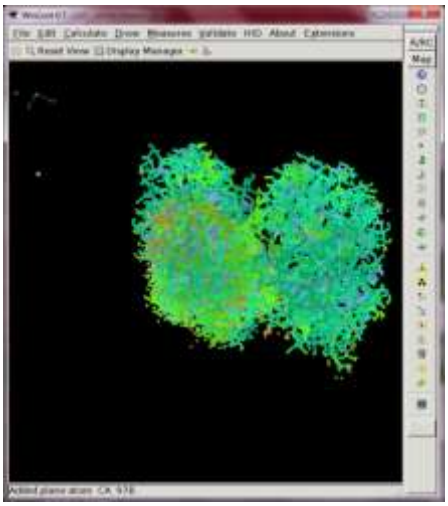


Figure 18

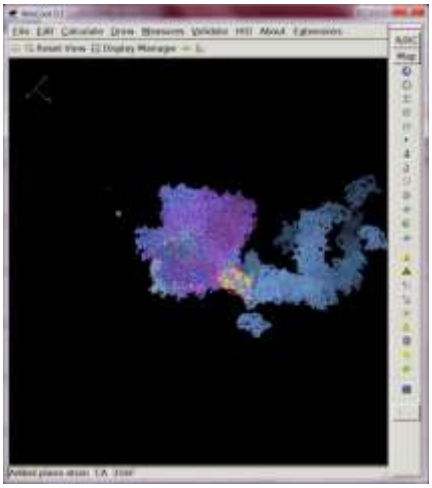


Figure 19

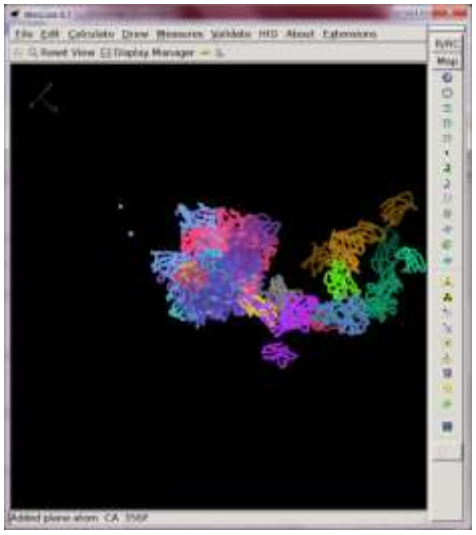


Figure 20

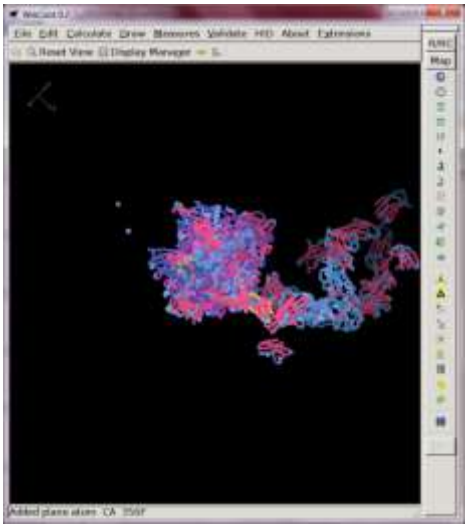


Figure 21

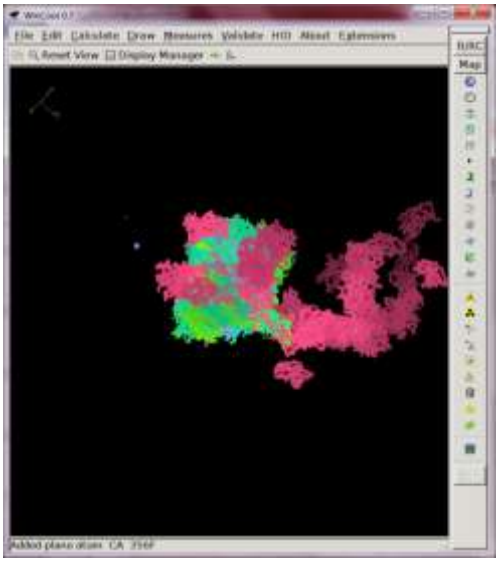


Figure 22

Figures 23-26 shows both of Class A and B 3D transpositions together, in the above four display options and Figure27 shows the transposition in display option “C-alphas/Backbone”.

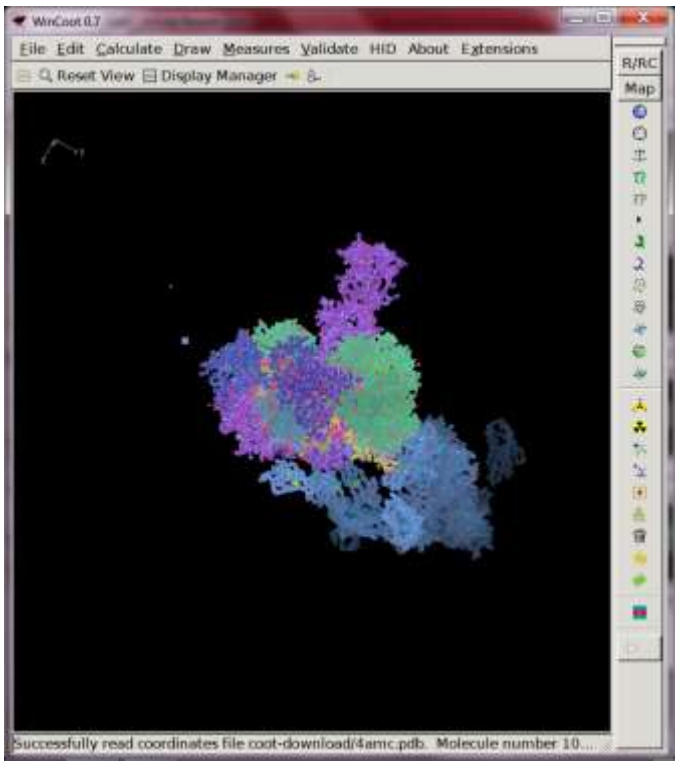


Figure 23

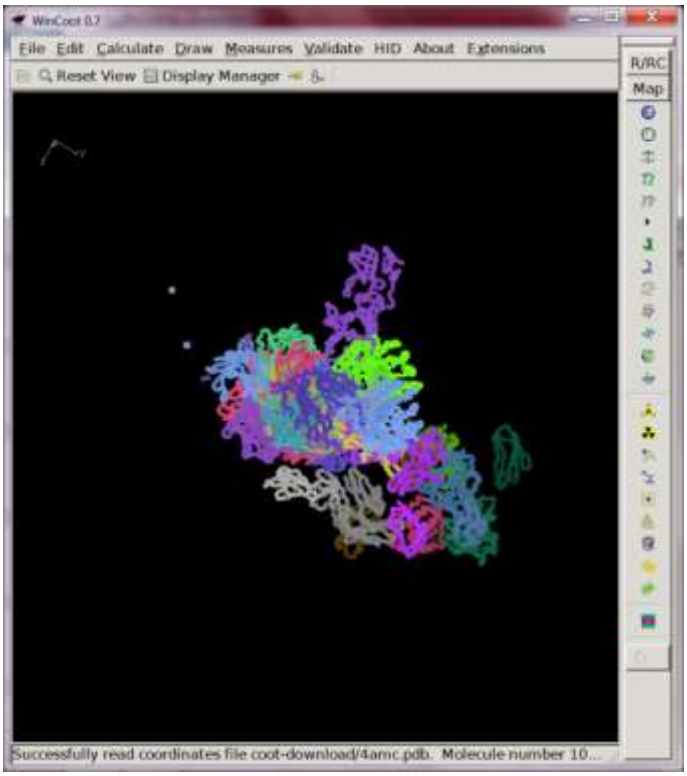


Figure 24

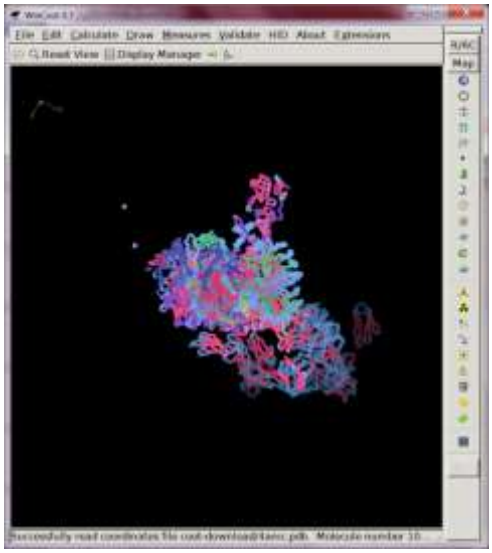


Figure 25

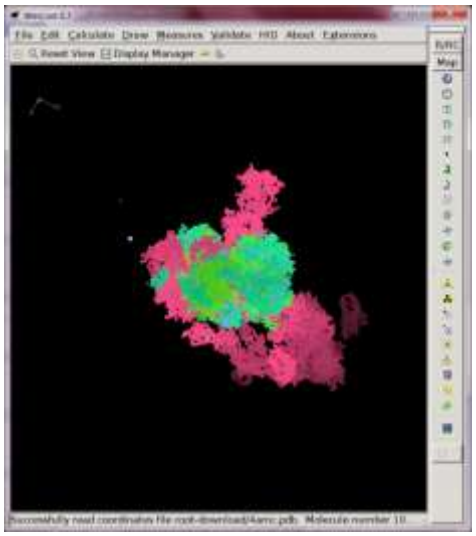


Figure 26

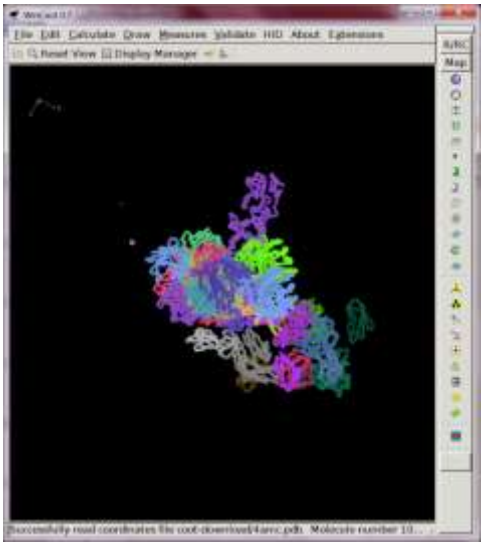


Figure 27