

# 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.

The screenshot shows the Dali server web interface. At the top, there is a yellow header with the text "Dali server" and a logo for the "Institute of Biotechnology". Below the header is a navigation bar with five tabs: "SERVICES & TOOLS", "GROUP MEMBERS", "NEWS & VACANCES", "RESEARCH", and "PUBLICATIONS". The main content area is titled "Protein Structure Database Searching by DaliLite v. 3". It contains a paragraph explaining the Dali server's function: "The Dali server is a network service for comparing protein structures in 3D. You submit the coordinates of a query protein structure and Dali compares them against those in the Protein Data Bank (PDB). You receive an email notification when the search has finished. In favourable cases, comparing 3D structures may reveal biologically interesting similarities that are not detectable by comparing sequences." Below this, it states: "Requests can also be submitted by e-mail to dali-server at helsinki dot fi. The body of the e-mail message must contain atomic coordinates in PDB format." It then provides two options: "If you want to know the structural neighbours of a protein already in the Protein Data Bank (PDB), you can find them in the Dali Database." and "If you want to superimpose two particular structures, you can do it in the pairwise DaliLite server." The search form includes a text input field for "Upload a structure:" with an "Avail (1/1000)" button next to it. Below this is a section for "Or enter PDB identifier: 1ctn" with a "chain:" dropdown menu set to "optional". A note below says "(Keyword search for PDB identifiers)". There is a "Job name:" text input field with "optional" next to it. Below that is an "Enter email address for notification:" text input field with "recommended" next to it. At the bottom left of the form is a checkbox labeled "lower priority queue". At the bottom right are two buttons: "submit" and "clear". A footer note at the very bottom says "Most jobs finish within an hour, but if a queue builds up, then it takes longer."

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)

## Summary

	No:	Chain	Z	rmsd	Iali	res	%id	PDB	Description
<input checked="" type="checkbox"/>	1:	<a href="#">1ctn-A</a>	71.7	0.0	538	538	100	<a href="#">PDB</a>	MOLECULE: CHITINASE A;
<input checked="" type="checkbox"/>	2:	<a href="#">2wk2-A</a>	68.6	0.5	538	539	99	<a href="#">PDB</a>	MOLECULE: CHITINASE A;
<input checked="" type="checkbox"/>	3:	<a href="#">2wm0-A</a>	68.6	0.6	538	538	99	<a href="#">PDB</a>	MOLECULE: CHITINASE A;
<input checked="" type="checkbox"/>	4:	<a href="#">2vix-A</a>	68.6	0.5	538	538	99	<a href="#">PDB</a>	MOLECULE: CHITINASE A;
<input checked="" type="checkbox"/>	5:	<a href="#">1eib-A</a>	68.5	0.4	538	540	99	<a href="#">PDB</a>	MOLECULE: CHITINASE A;
<input checked="" type="checkbox"/>	6:	<a href="#">2wly-A</a>	68.4	0.4	538	539	99	<a href="#">PDB</a>	MOLECULE: CHITINASE A;
<input checked="" type="checkbox"/>	7:	<a href="#">1ffz-A</a>	68.2	0.4	538	540	99	<a href="#">PDB</a>	MOLECULE: CHITINASE A;
<input checked="" type="checkbox"/>	8:	<a href="#">1ffq-A</a>	68.1	0.5	538	540	99	<a href="#">PDB</a>	MOLECULE: CHITINASE A;
<input checked="" type="checkbox"/>	9:	<a href="#">1adn-A</a>	68.0	0.4	538	540	99	<a href="#">PDB</a>	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  
CRYSTAL STRUCTURE OF A BACTERIAL CHITINASE AT 2.3 ANGSTROMS RESOLUTION

Examined 86833 entries, (217743 chains) Displaying Matches 1-20 of 36.  
[Back to query](#) [Close](#) [Add pages](#) Sort by: [Score](#) [average by SCOP family](#)

#	Score	Query	Target (PDB entry)						
	Q	P	Z	Rms	Match	Rms	Id	%	Title
1	0.00	100.0	0.00	100	0	100	100	100	CRYSTAL STRUCTURE OF A BACTERIAL CHITINASE AT 2.3 ANGSTROMS RESOLUTION
2	0.99	100.0	0.10	100	0	99	100	99.0	CHITINASE A FROM SERRAIA MARCESCENS 47CCB9 IN COMPLEX WITH CHITOSIN-THIAZOLINE
3	0.99	100.0	0.10	100	0	99	100	100.0	CHITINASE A COMPLEXED WITH TETRA-N-ACETYLCHITOSINE
4	0.99	100.0	0.10	100	0	99	100	100.0	CRYSTAL STRUCTURE OF CHITINASE A MUTANT Y186F COMPLEXED WITH HEXA-N-ACETYLCHITOSAMINE (HAGB)
5	0.99	100.0	0.10	100	0	99	100	100.0	CRYSTAL STRUCTURE OF CHITINASE A MUTANT D193A COMPLEXED WITH OCTA-N-ACETYLCHITOSAMINE (HAGB)
6	0.99	100.0	0.10	100	0	99	100	100.0	CRYSTAL STRUCTURE OF CHITINASE A MUTANT E210G COMPLEXED WITH OCTA-N-ACETYLCHITOSAMINE (HAGB)
7	0.99	100.0	0.10	100	0	99	100	100.0	CRYSTAL STRUCTURE OF CHITINASE A FROM S. MARCESCENS AT 1.05 ANGSTROMS
8	0.97	100.0	0.10	100	0	99	100	100.0	CHITINASE A FROM SERRAIA MARCESCENS 47CCB9 IN COMPLEX WITH CHITOSIN-THIAZOLINE DITHIAMINE
9	0.97	99.0	0.10	100	0	99	100	100.0	CRYSTAL STRUCTURE OF CHITINASE A COMPLEXED WITH ALLOBARBON

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
<input type="checkbox"/> 1:	<a href="#">1ctn-A</a>	71.7	0.0	538	538	100	<a href="#">PDB</a>	MOLECULE: CHITINASE A;
<input type="checkbox"/> 37:	<a href="#">1itx-A</a>	46.9	1.7	391	419	32	<a href="#">PDB</a>	MOLECULE: GLYCOSYL HYDROLASE;
<input type="checkbox"/> 38:	<a href="#">1lg2-A</a>	44.3	1.6	337	359	26	<a href="#">PDB</a>	MOLECULE: CHITOTRIOSIDASE;
<input type="checkbox"/> 39:	<a href="#">1hkk-A</a>	44.2	1.6	340	364	26	<a href="#">PDB</a>	MOLECULE: CHITOTRIOSIDASE-1;
<input type="checkbox"/> 41:	<a href="#">1l14-A</a>	44.0	1.5	355	392	32	<a href="#">PDB</a>	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
<input type="checkbox"/>	962:	<a href="#">3faw-A</a>	4.1	3.6	89	775	13	<a href="#">PDB</a>	MOLECULE: RETICULOCYTE BINDING PROTEIN;
<input type="checkbox"/>	963:	<a href="#">3qs9-E</a>	4.0	3.1	61	405	10	<a href="#">PDB</a>	MOLECULE: SL CYTOKINE;
<input type="checkbox"/>	967:	<a href="#">3o4o-C</a>	3.4	3.5	65	316	11	<a href="#">PDB</a>	MOLECULE: INTERLEUKIN-1 BETA;
<input type="checkbox"/>	968:	<a href="#">2vzv-A</a>	3.3	4.7	211	857	8	<a href="#">PDB</a>	MOLECULE: EXO-BETA-D-GLUCOSAMINIDASE;
<input type="checkbox"/>	970:	<a href="#">4amc-A</a>	2.4	4.1	87	1019	5	<a href="#">PDB</a>	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.

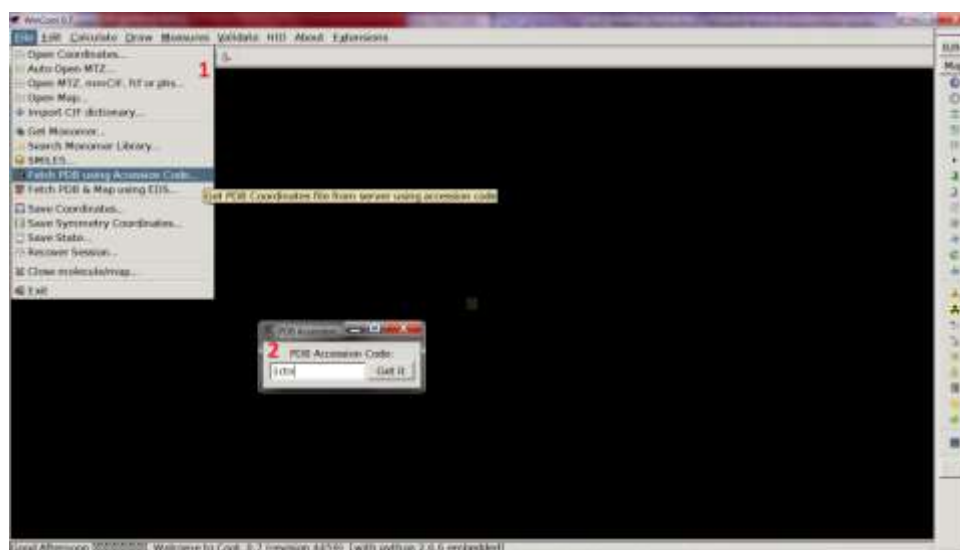


Figure 4

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

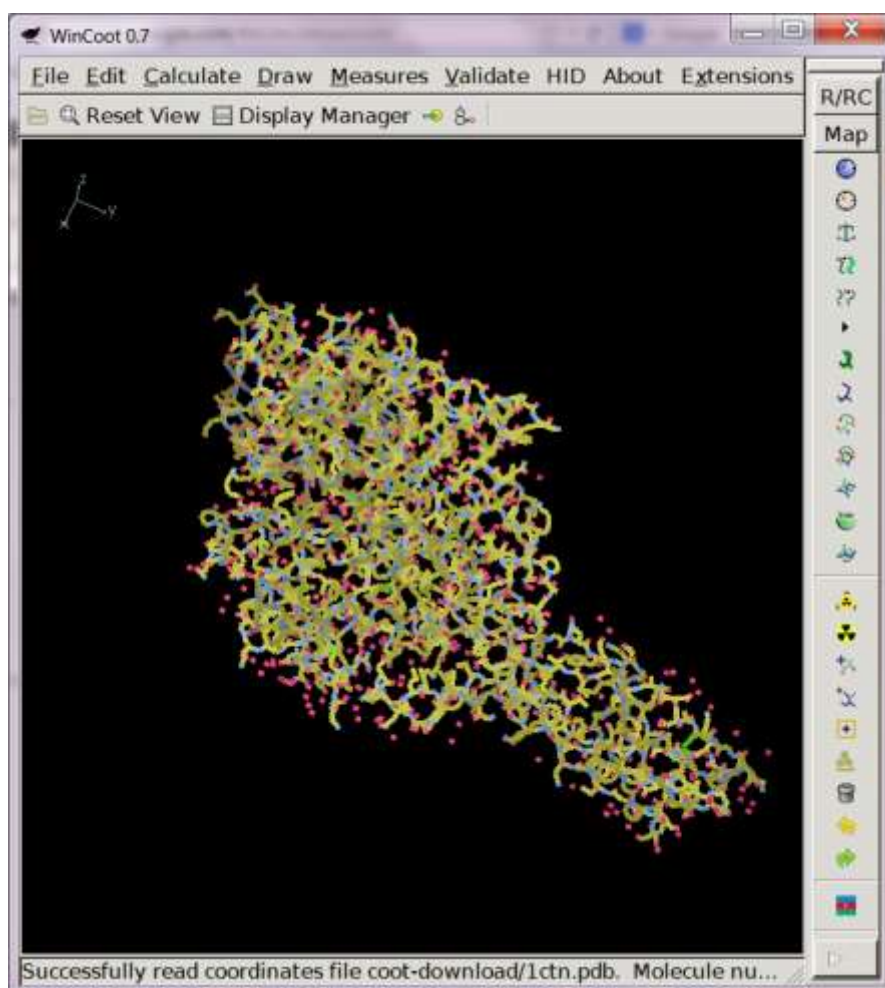


Figure 5

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  $180^\circ$  degrees keeping the y axon stable, in order to provide a good view of the rest four enzymes. Figure 7 has been rotated  $-90^\circ$  degrees keeping the y axon stable, in order to provide a good view of the rest five enzymes.

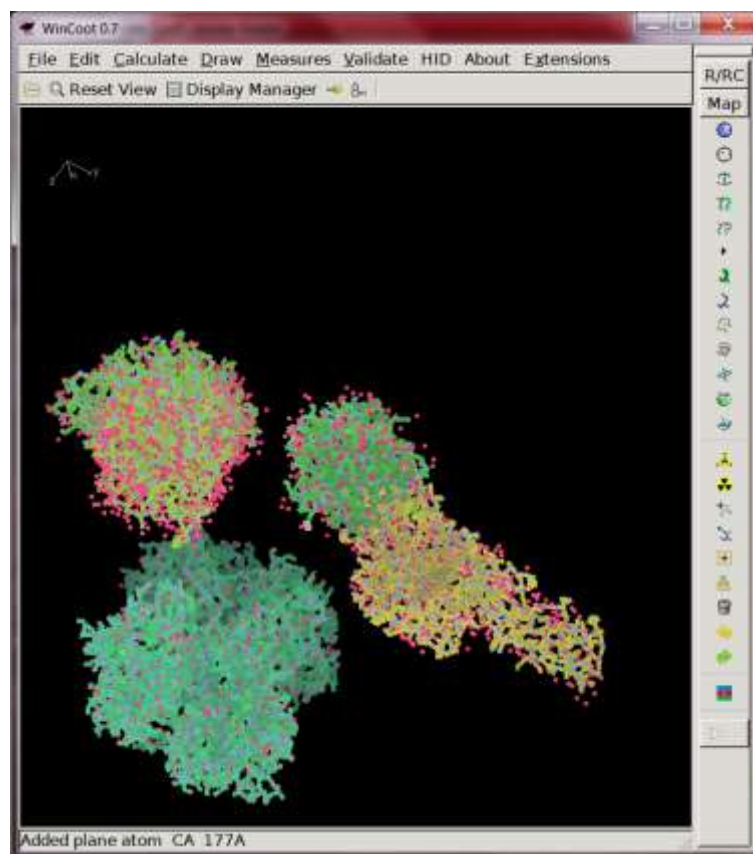


Figure 6

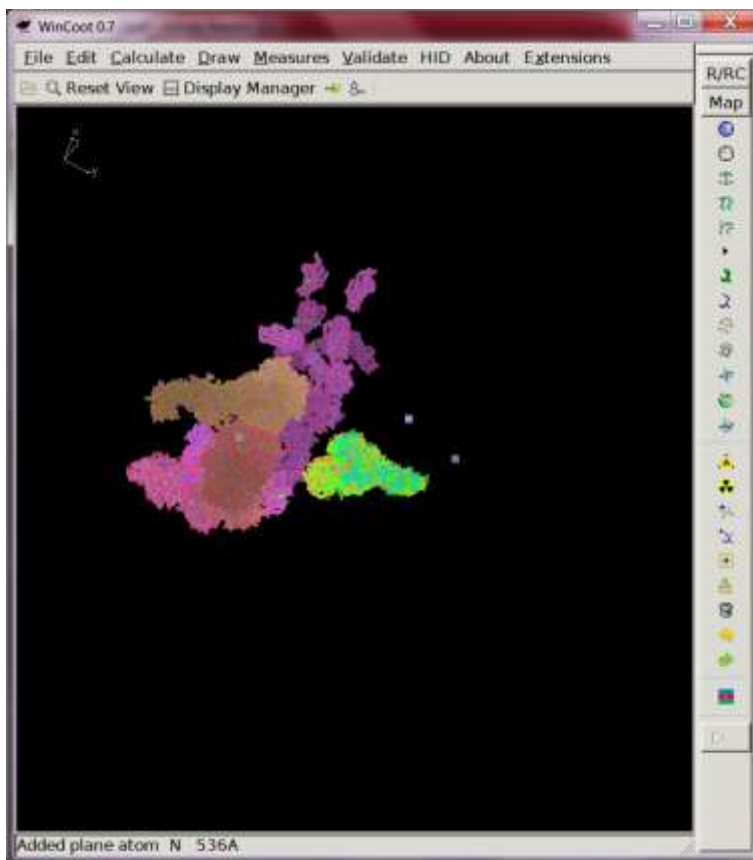


Figure 7



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.

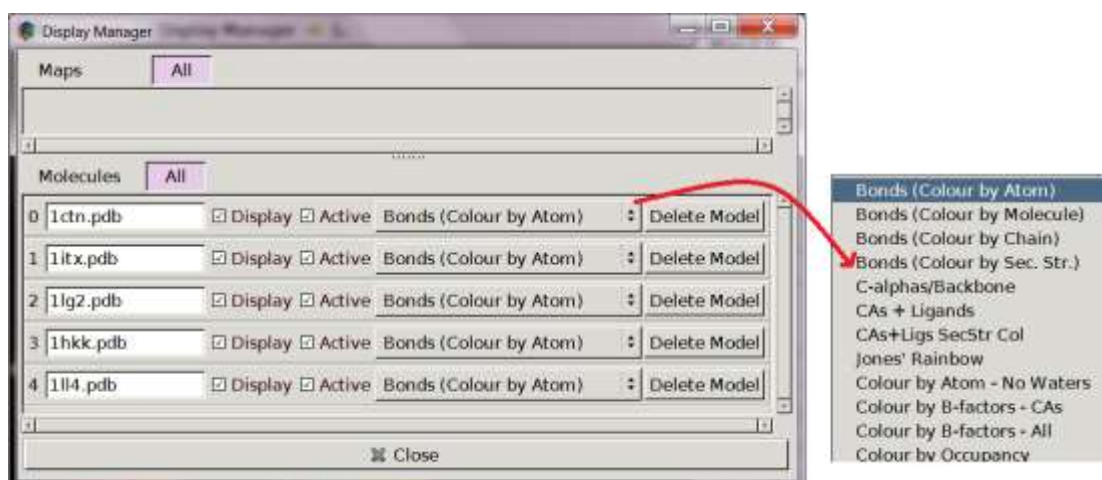


Figure 8

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.

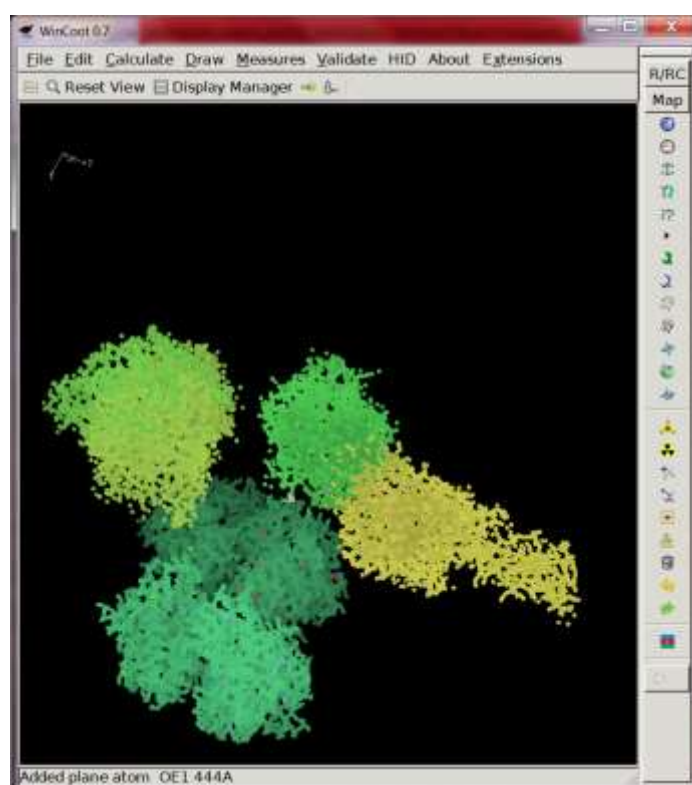


Figure 9

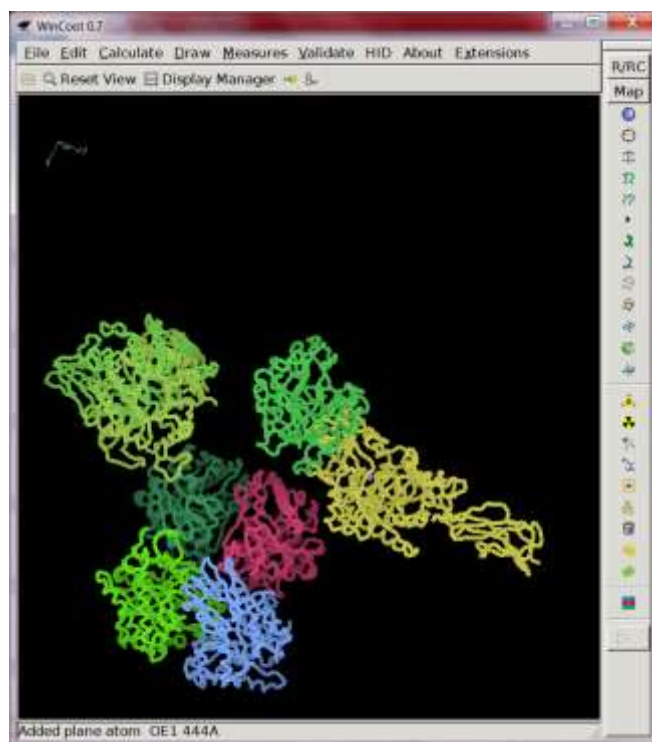


Figure 10

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.

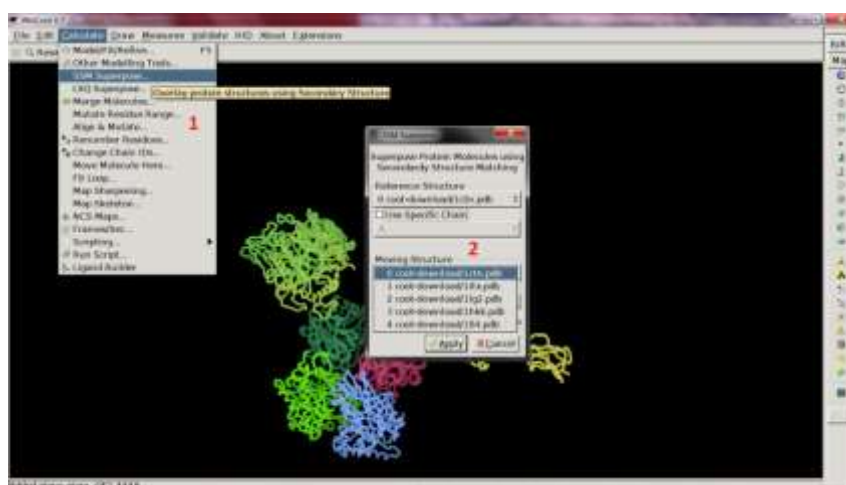


Figure 11



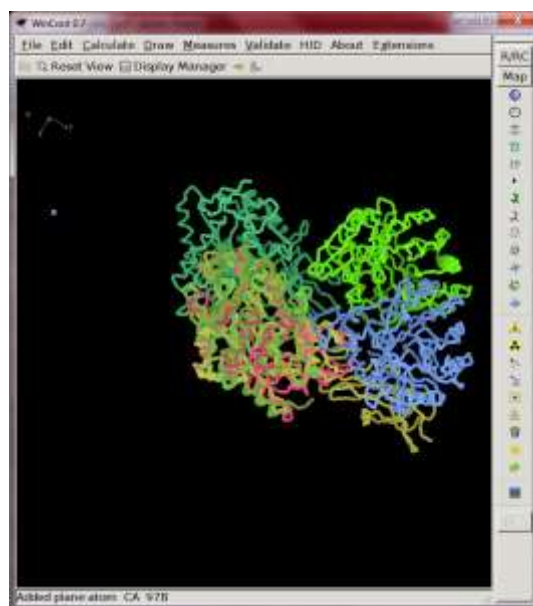


Figure 12



Figure 13

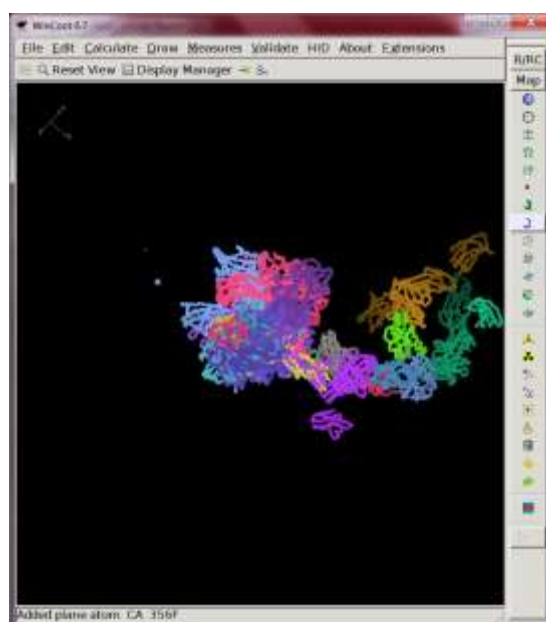


Figure 14

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).

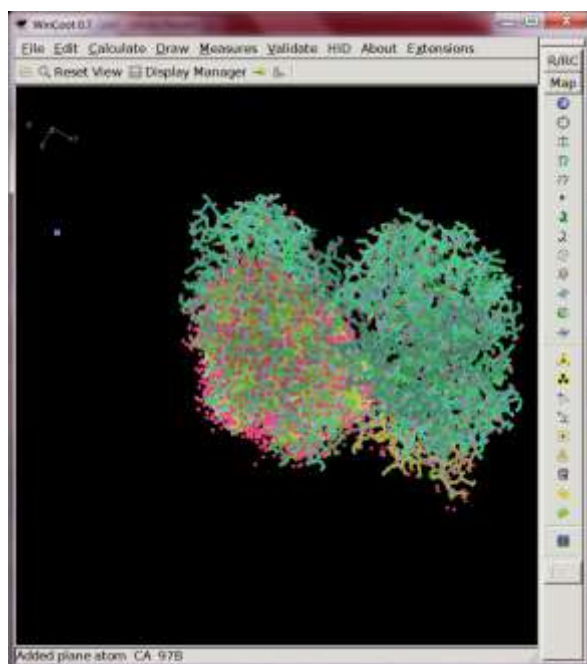


Figure 15

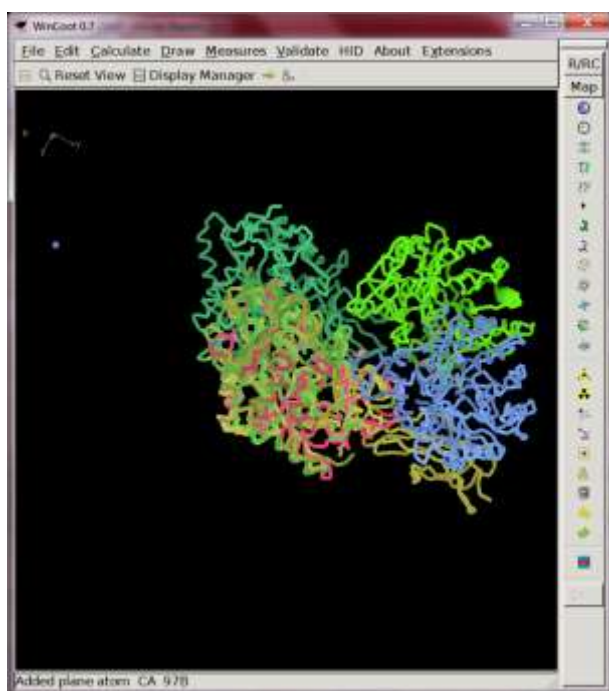


Figure 16

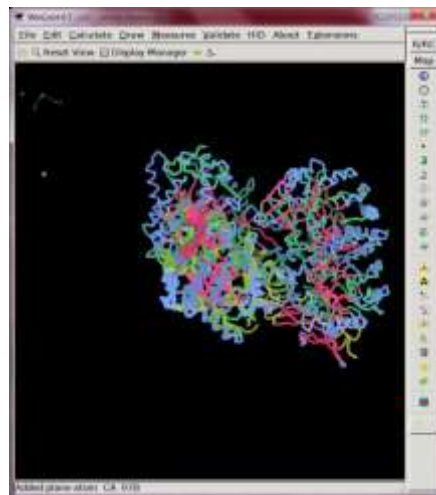


Figure 17

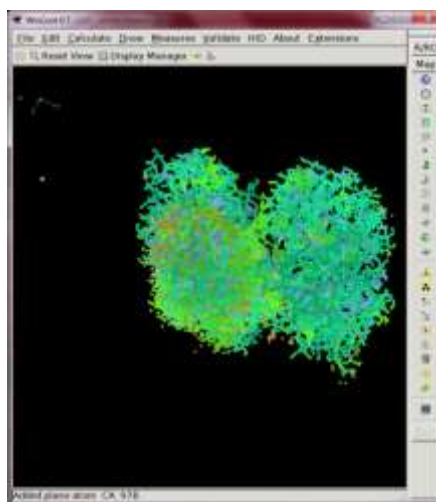


Figure 18

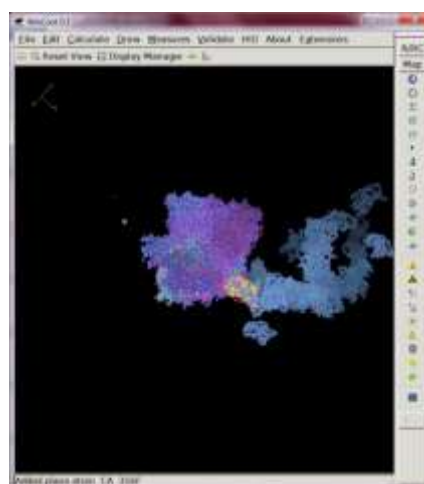


Figure 19

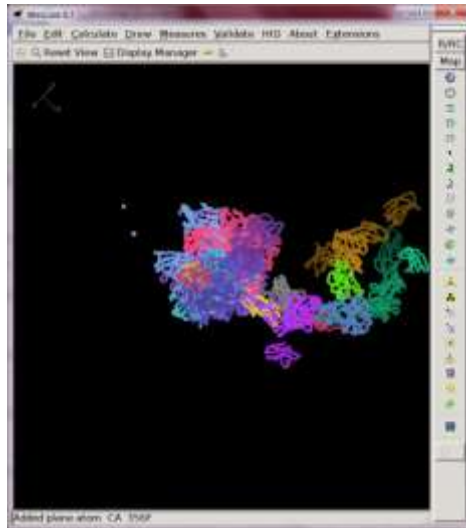


Figure 20

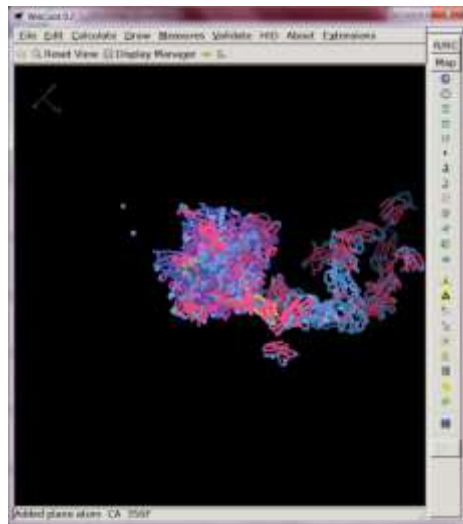


Figure 21

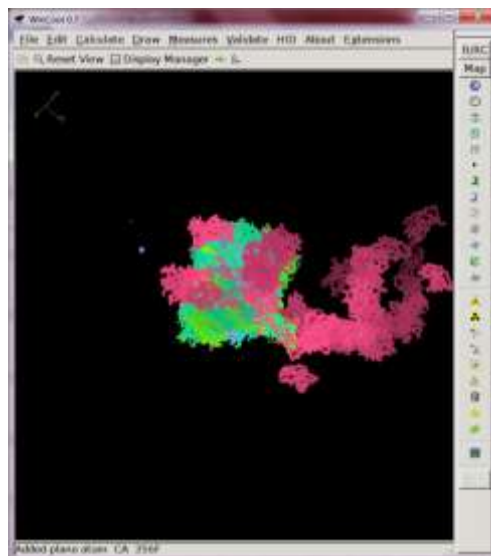


Figure 22

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”.

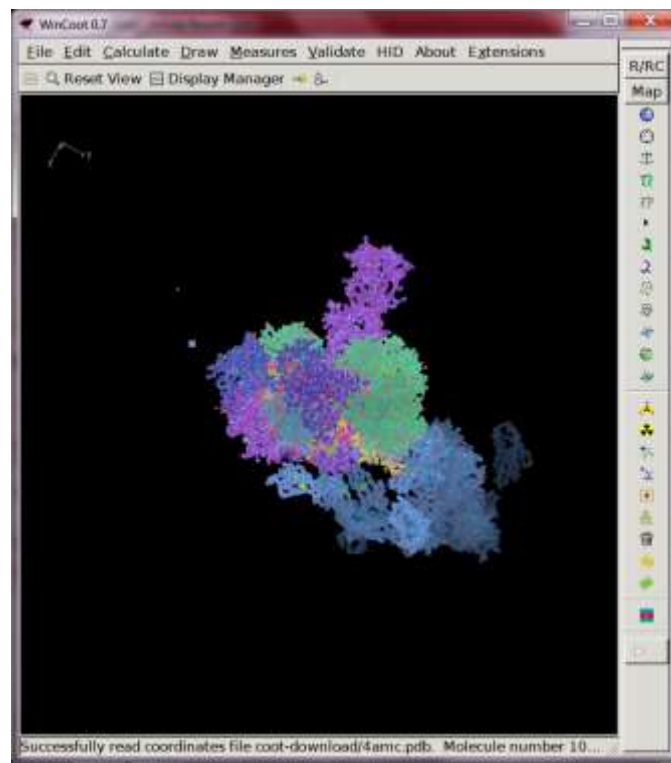


Figure 23

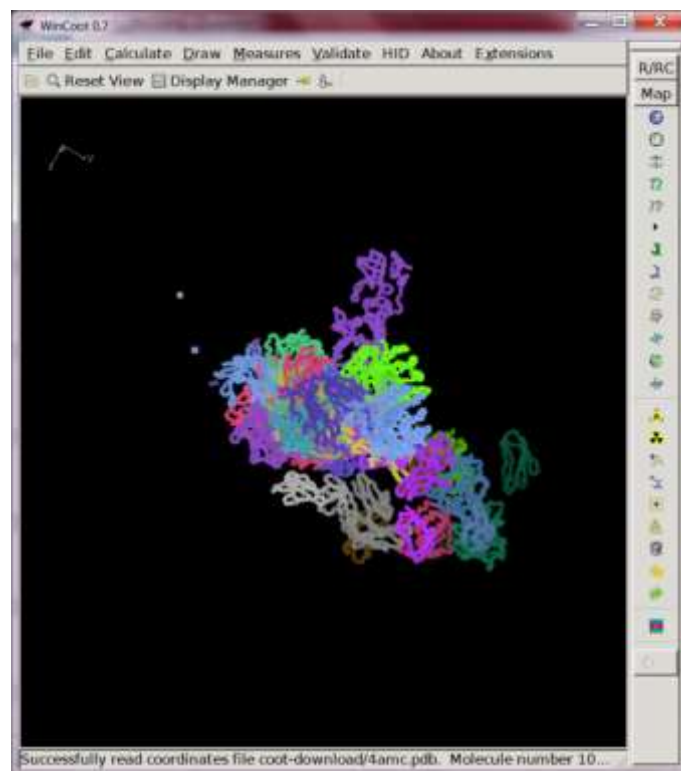


Figure 24

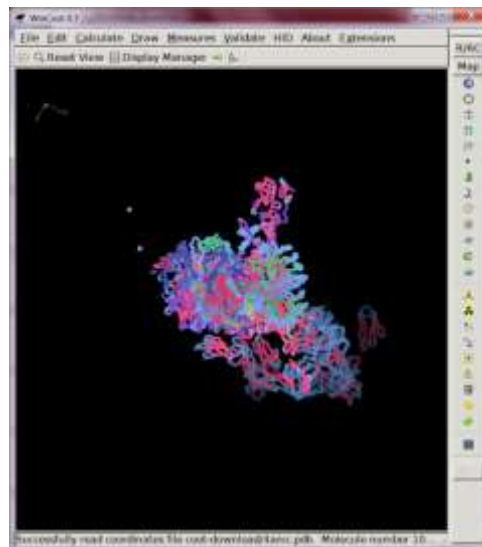


Figure 25

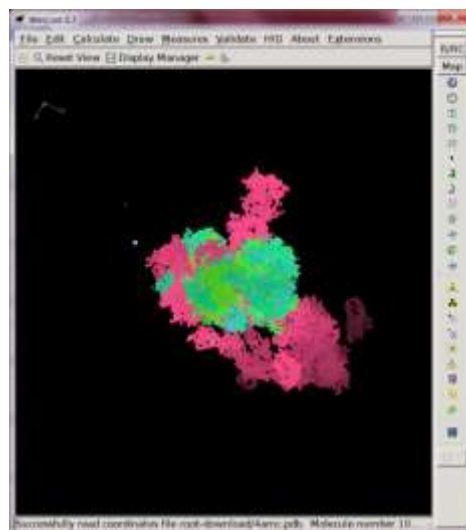


Figure 26

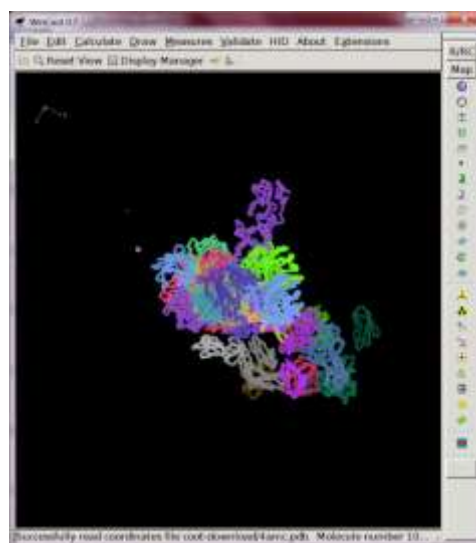


Figure 27