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: *23/04/2013*

Assignment 3

Tasks 1 & 2

In this third assignment we were asked to continue the previous assignment 2 where we translated in the same Cartesian coordinates both the proteins of Class A and B (see Ass. 2) and represented them in the 3-dimensional system. The tasks for this third assignment are to find and collocate the resemblances and differences of the geometric characteristics that are conserved or non-conserved (Task 1). Taking that in account, the second task is to record and evaluate the geometric descriptors of these characteristics (Task 2).

At first, we searched how to do this job using WinCoot and we found 3 useful options in the menu. In specific the options suitable for this job were: “Ramachandran plot”, “Kleywegt plot” and “Geometry Analysis” under the “Validate” top menu bar option. But, because of the intractable of them being separated and for having to combine all these option results we decided for one more time to learn how to use something new, clearly for the acedemic purpose of this assignment.

The new tool we chose to use is the MolProbity (also discussed in class). So, to begin with, the first thing needed to do for the first task of the assignment is to find and collocate the resemblances and differences of the geometric characteristics. To do that, we open the MolProbity webtool and from the section of the suggested tools we selected the tool with name “Analyze geometry with all-atom contacts” which provides us with information about:

- All-atom steric contacts (clashlist, clash score, contact dots)
- Protein geometry evaluation (Ramachandran plot, rotamers, Cβ deviations)
- Nucleic acid geometry (base-phosphate perpendiculars, suite names)
- Evaulate protein and nucleic acid bond lengths and angles
- Multi-criterion chart and kinemage displays

Figures 1–4 show the steps we follow to take the results that we will afterwards compare among each other.

Figure 1

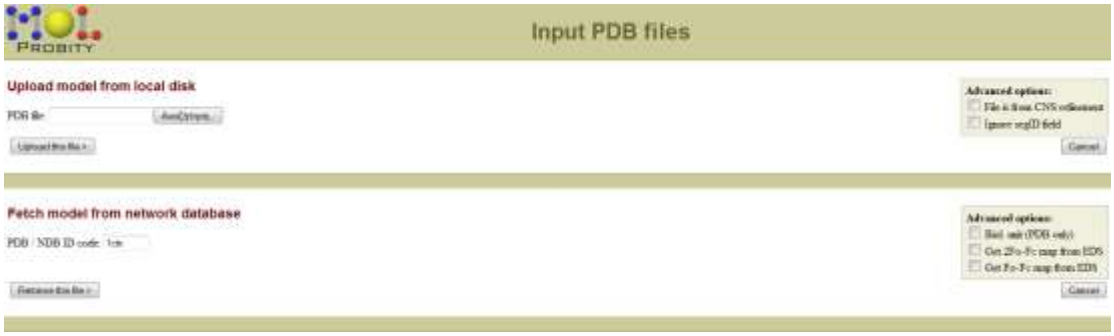

Figure 2

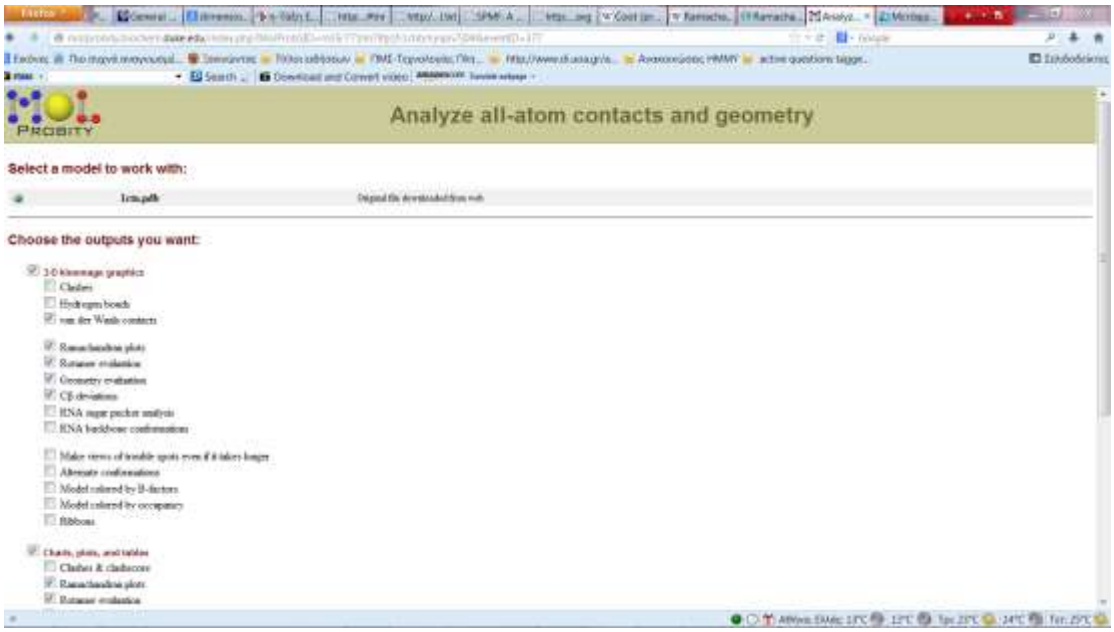

Figure 3

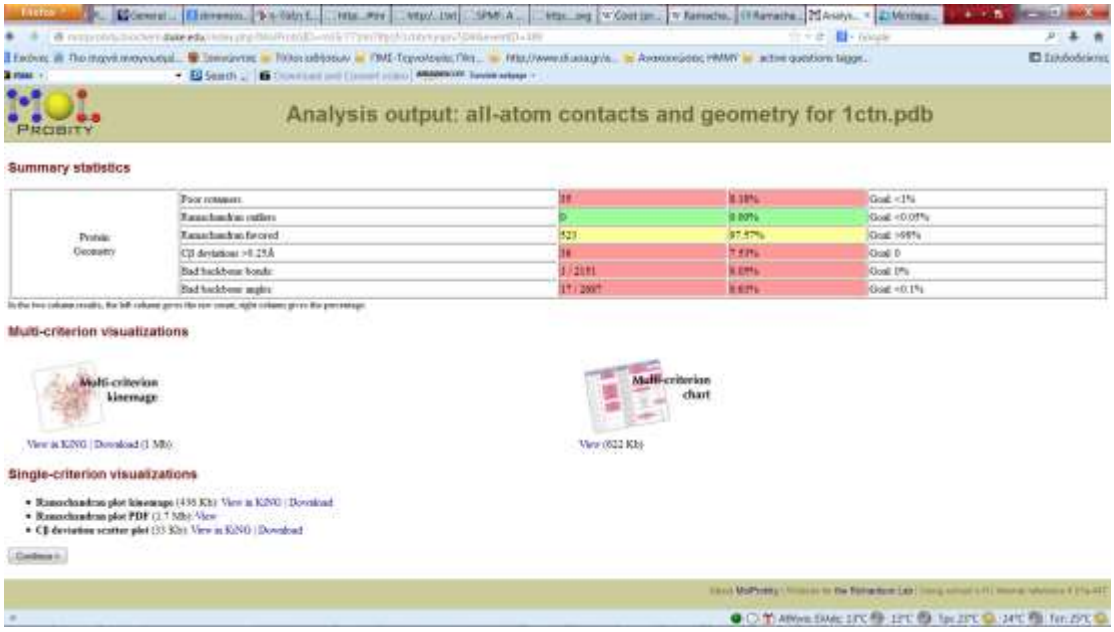

Figure 4

The above four steps were followed for every structure homologue of the A and B classes (see assignment 2). In Figure 1 we select the tool we want to use, then in Figure 2 we select the structure homologue that we want to fetch from the PDB data bank or upload from our computer. Figure 3 shows in checkbox selections what do we want to be output, and finally Figure 4 shows the statistics and some more features we will explicate later. Some of these features are the kinemage applet, showing the 3D structure with all the geometric utilities letting the user experiment with them (Figure 5), the Ramachandran plot in a pdf file format and in a kinemage java applet format (Figure 6).

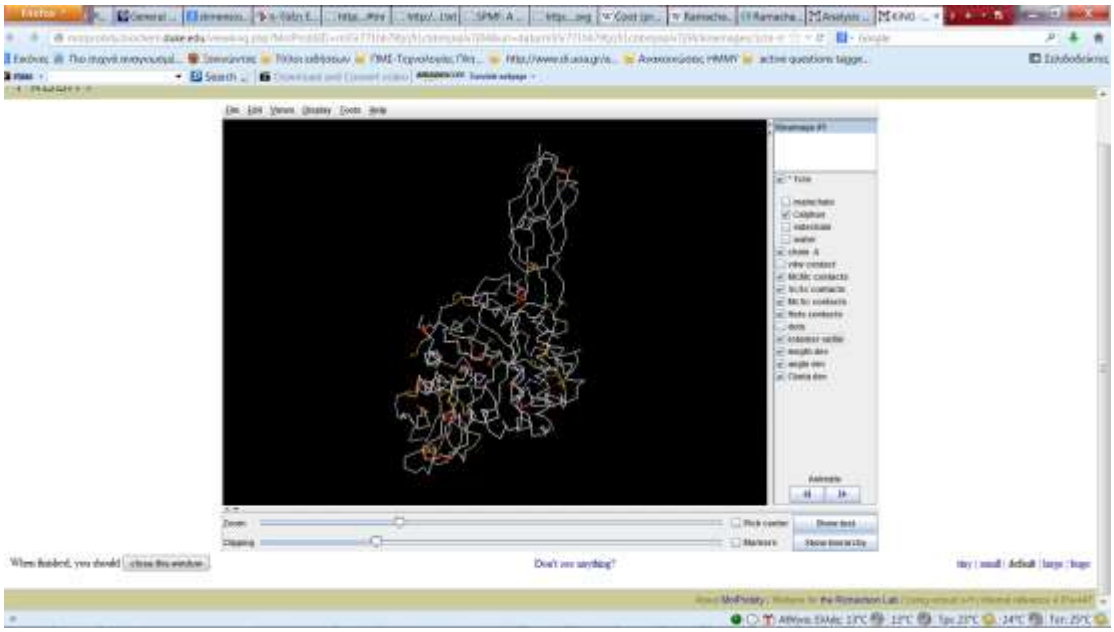

Figure 5

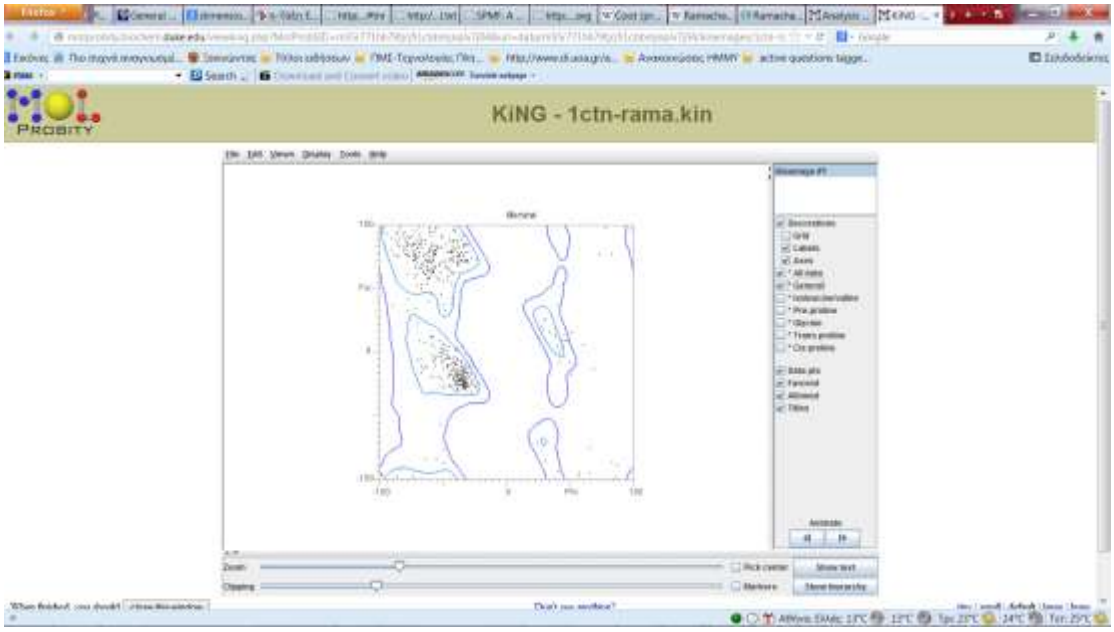

Figure 6

All the files used and shown above and later in the next figures are stored and provided in the deliverable zip file. The files are separated according to the class A or class B where they belong. Class A's files can be found in folder "chrysina_ass3_nbegetis\files with exported data\molprobity - high five" whereas Class B's files can be found in folder "chrysina_ass3_nbegetis\files with exported data\molprobity - low five". As now we are referring to the deliverable, we kept in purpose the figures from the previous assignment in the folder "chrysina_ass3_nbegetis\figures" so that we can compare the statistics we found with the real 3D superimposed structures over 1ctn chitinase. All the images of this assignments have a name analogous to Fig3_x.png, where x is a number. Finally all the ramachandran plots for all cases of every structure of every class are plovided in a .pdf file format in the folder "chrysina_ass3_nbegetis\files with exported data\rama plots".

Now that we finished with the procedural matters we continue to the main task 1 of this assignment. Following the above steps shown in Figures 1-4 we get at first all the primitive information about the homologues' position in space. Figures 7-11 show this information for each protein of class A. And figures 12-16 show the same information for structures of class B.

It is worthy saying that we cannot compare the uploaded structure homologues as they are because they all are in a different resolution loaded (angstrom), so the comparison should be made according to the final summary statistical analysis.

Class A

Figure 7

Figure 8

Figure 9

Figure 10

Figure 11

Class B

Figure 12

Figure 13

Figure 14

Figure 15

Figure 16

Now that we presented all the primitive information for the homologue structure we can provide the statistical analysis for all these structures. Figures 17-21 show the statistical analysis of Class A, while figures 22-26 show the statistical analysis for Class B's structures.

CLASS A

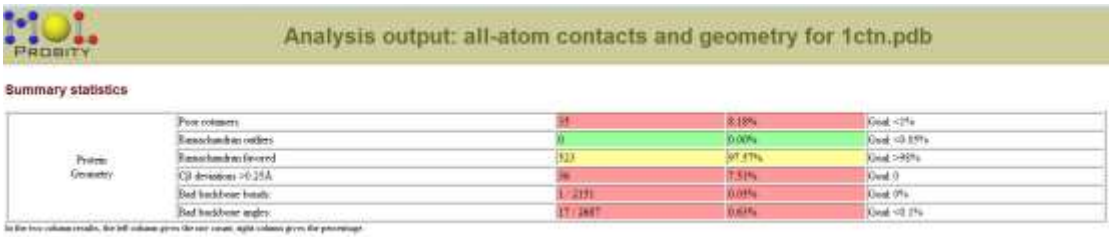

Figure 17

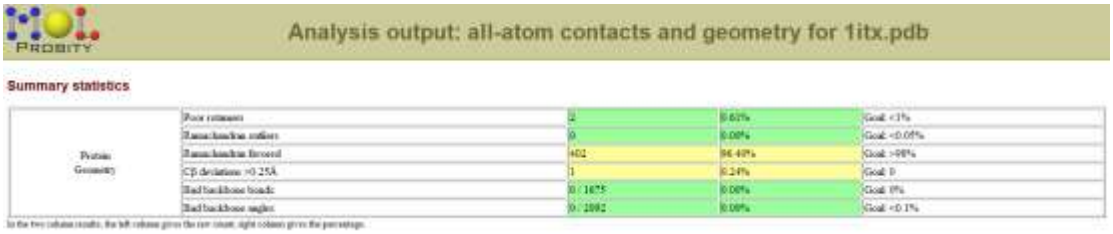

Figure 18

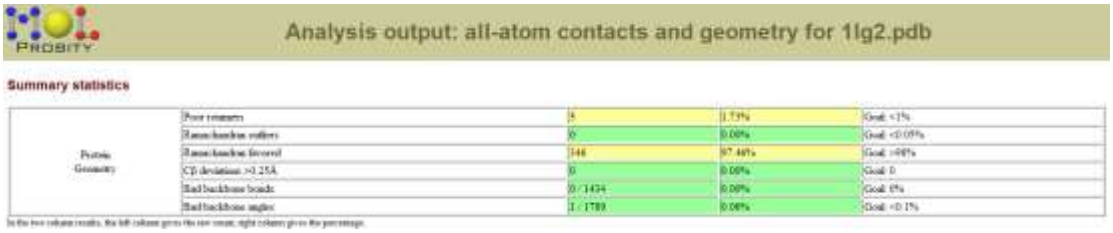

Figure 19

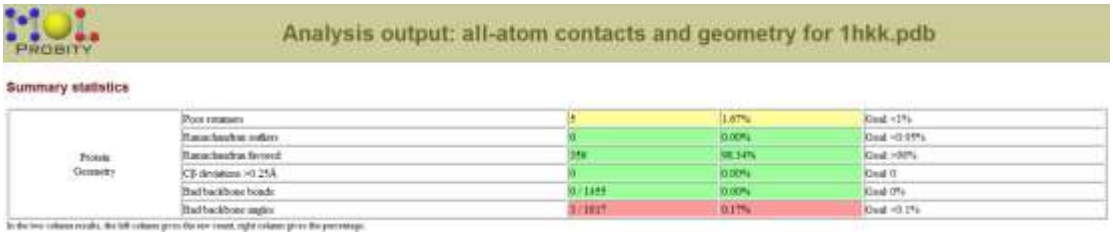

Figure 20

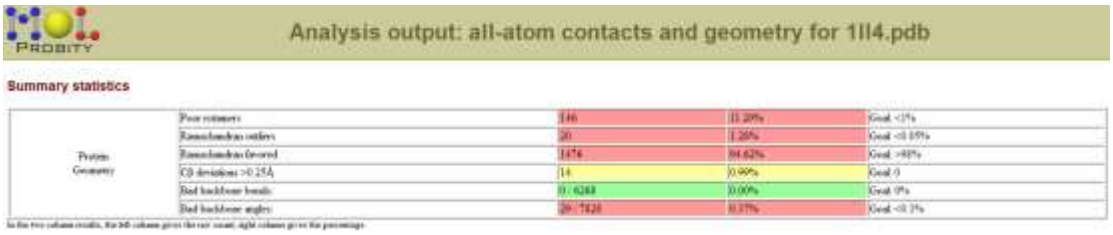

Figure 21

Class B

Figure 22

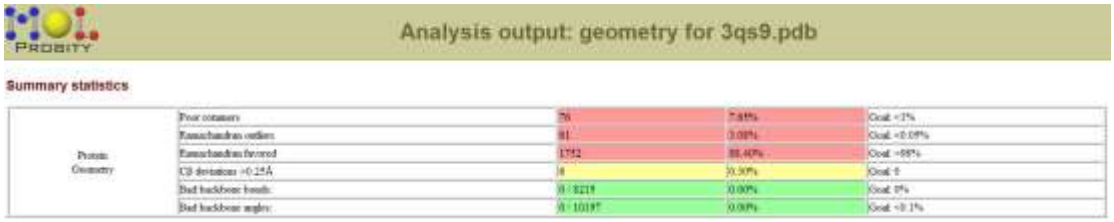

Figure 23

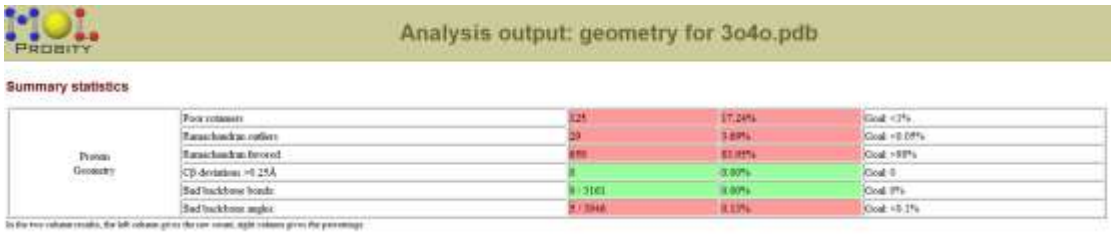

Figure 24

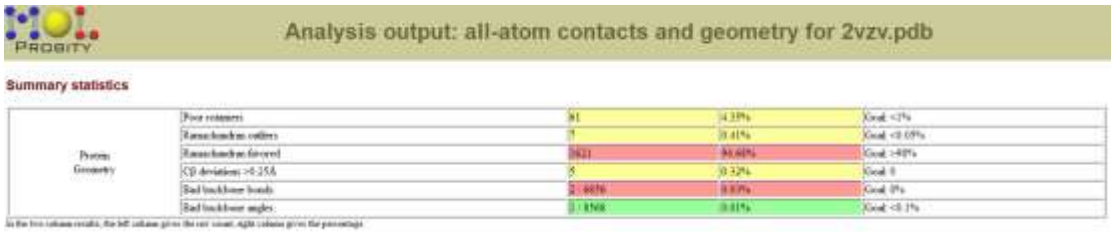

Figure 25

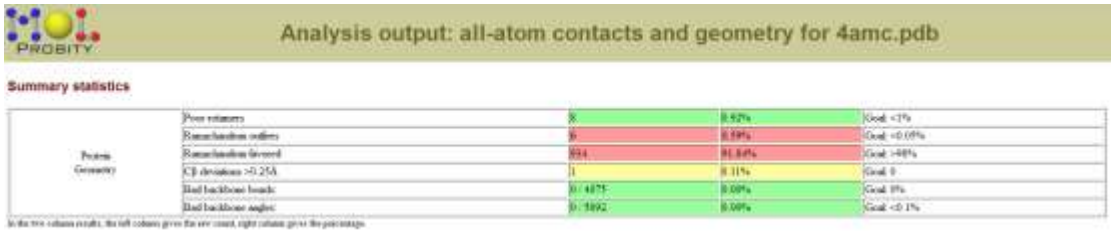

Figure 26

Beginning from task 2, from the above results we can infer that almost all the structure homologues have got the C_β deviations > *0.25A* (4th line in summary statistics), meaning that they have these deviations in **good** regions. We can verify the above inference if we ask to view the C_β deviations in scatter plot in KiNG applet view. There we can see that almost all structures have very reasonable distributions, which tells us that we have found the global minimums and we have not been stuck in some local minimums.

Moreover, except for 1ctn chitinase almost all of the structure homologues have got a good percentage of poor rotamens (1st line in summary statistics) which means that **only a few** restricted rotations around single bonds occur, which is have we like to happen.

Ramachandran outliers (2nd line in summary statistics) indicate values that almost reach the edges of the correct geometrical areas for a residue to be found. So the less outlier we have, the best. As it is shown from above, Class A's structure homologues have a **good** percentage of outliers (except for 1ll4.pdb, and this is explained because 1ll4 has residues in more than one chains). On the contrary, the low score structure homologues of Class B, have got a big percentage of ramachandran outliers (**bad**), which is also explained because the z-score with which we sorted the structures in the 2nd class assignment is related to the ramachandran outliers.

Furthermore, ramachandran favored (3rd line in summary statistics) makes the differences between the two class (Class A - high scores structure homologues, Class B - low score structure homologues) now very clear. The favored areas are those **inside the azure** color (Figures 27 and 28) and they indicate the allowed, and even we can say "generously allowed" areas in where a molecule can be angled. The conclusion about the two classes we can infer is that the class A has a lot in common in its primary and secondary structure with 1ctn chitinase and this fact makes the class A residues more dense inside the azure lines of the ranachandran plot. On the contrary, because class B's structure homologues differ a lot from the 1ctn chitinase in the primary structure, but has a lot in common in the secondary structure (refer for that in the previous 2 assignments), that is why it does not have a good percentage in the ramachandran favored.

Finally from the 5th and 6th line we can observe that no conclusion can be educed, as both in class A and B are found **good and bad** percentages of backbone angles and bonds. This is because backbone chain may differ a lot in the edges of the chain regardless of keeping the main chain stable.

Having now evaluated the above six geometry descriptors and before moving on to task 1, it is important to notice that as indicated from what we wrote the important information in the structure analysis is not found in the overall scores, but in the specific good or bad local regions that produce them.

To take a deeper view and to answer in task's 1 question, we gathered all the ramachandran plots of each of the two classes and we deposited them in one view, in specific that of the general case. Figures 27 and 28 show the results, respectively for Class A and B. The answer is provided in the last page of this document.

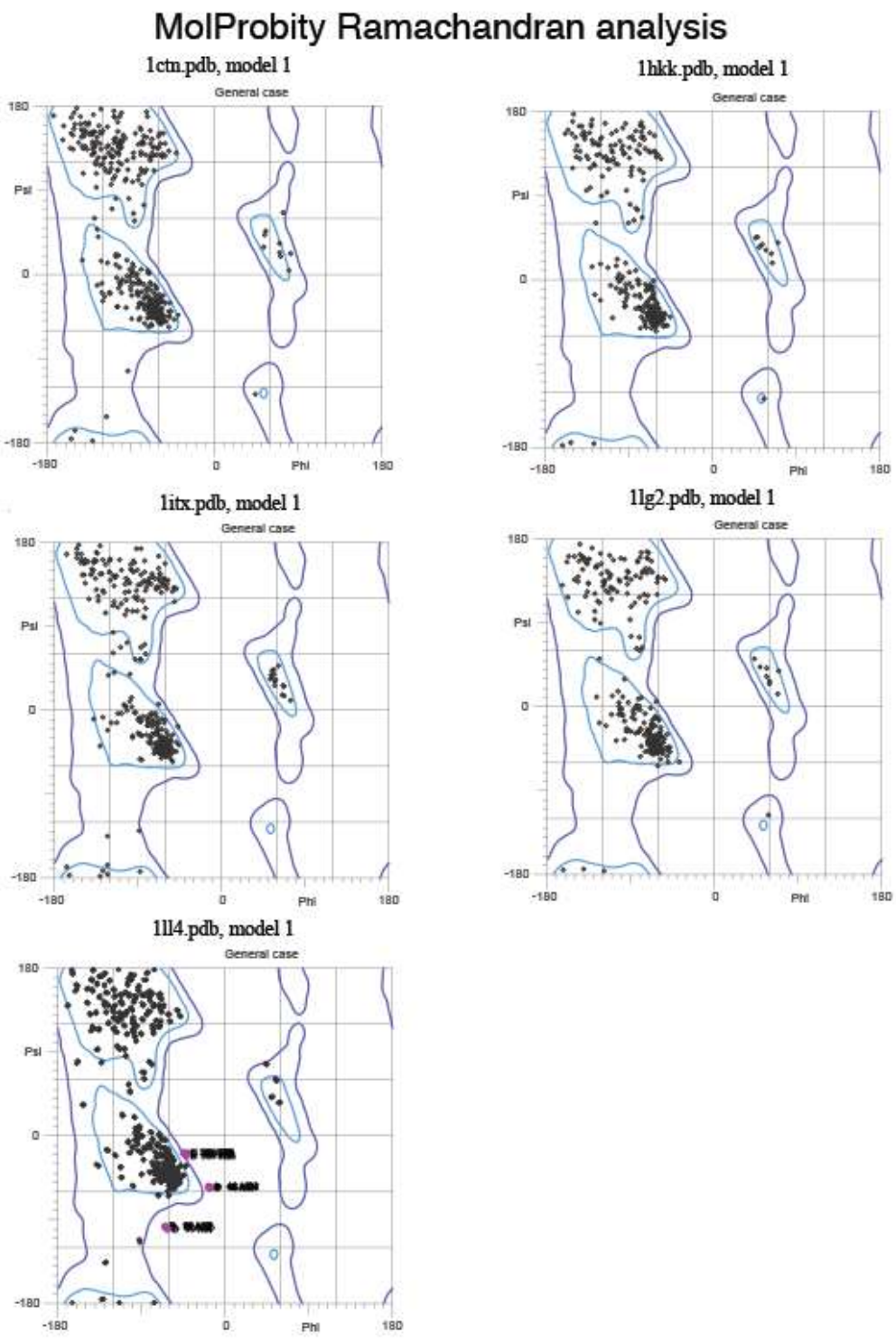
MolProbity Ramachandran analysis
1ctn.pdb, model 1
1hkk.pdb, model 1
1itx.pdb, model 1
1lg2.pdb, model 1
1ll4.pdb, model 1
General case
Psi
Phi
180
0
-180

Figure 27

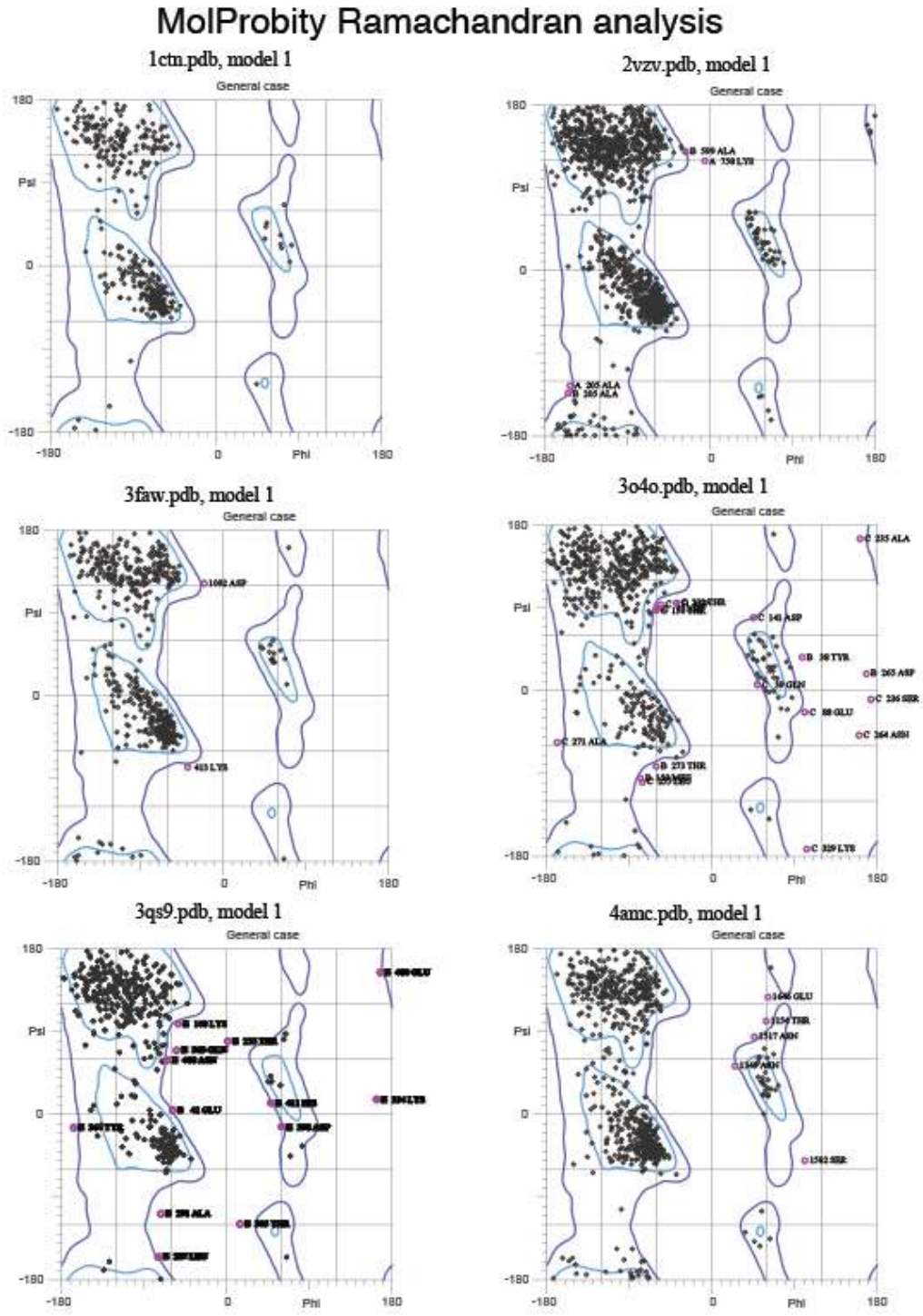

Figure 28

Now that we saw the ramachandar plots in the general case for all the structure homologues of both classes (more plots per occasion for each structure are found in the deliverable) it is easy and clear to notice that the main structure of both classes consists of α-helices and β-sheets. As we learned in class both of these secondary structures are depending of the ϕ and ψ angle that amino-acid residues form among them. As we know for the β-sheets, they are found between (ϕ_1, ψ_1) = (–135°, 135°) and (ϕ_2, ψ_2) = (–180°, 180°) angles, whereas, α-

helices are found ranging from (ϕ_3, ψ_3) = (-90°, -15°) to (ϕ_4, ψ_4) = (-35°, -70°). In the above diagrams we can easily observe that all of the structure homologues have preserved the main secondary structure in the conserved areas. The top-left areas show the β-sheets, while the center-left areas show the α-helices.

The only two things that it is important to touch on are that Class A structure homologues present almost no-Lα helices (lefthanded helices, shown the middle-right of the ramachandran plot), while the Class B structure homologues have a big part of their residues expressed as lefthanded helices. The reason can be once more imputed to the fact that the primary structure between 1ctn and Class A's structures differs a lot from the primary structure of the Class B's structures. The second important notice is to refer once more that as it is shown from the ramachandran plots the Class A's structure homologues residues have a more dense representation in the middle of the azure areas indicating the ramachandran favored areas we analyzed above. On the other hand Class B's structure homologues residues have a more sparse representation inside the blue wider areas, which are the permissible areas for the angles ϕ and ψ.

At this time I forestalled to miss once more the deadline and there are a couple of hours left for the deadline to expire B-)