KU LEUVEN

Bio-Molecular Model Building

Exam Exercise

Spring 2015

Authors:
Cedric Lood
Yi Ming Gan

Supervisors:
Marc De Maeyer
Joren De Raeymaecker
Xiaoyu Qing

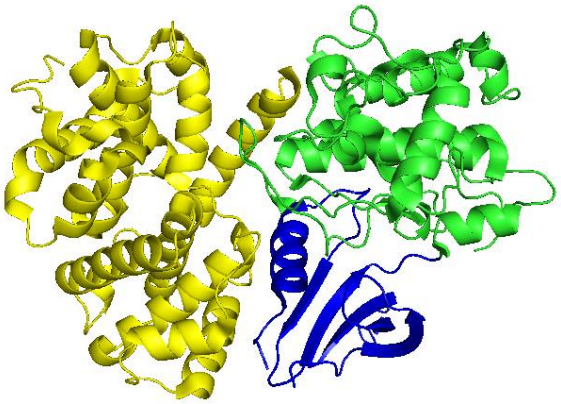
May 22, 2015

Contents

Question 1

Part a

Chain A is the kinase domain (CDK2), N-terminal is colored in blue and C-terminal is colored in green. The regulatory domain (cyclin-A2) is chain B, colored in yellow.



Part b

This is the list of residue sequences associated with their corresponding secondary structures::

1-4 Loop

5-11 Beta sheet

12-16 Loop

17-23 Beta sheet

24-28 Loop

29-36 Beta sheet

37-45 Loop

46-57 Alpha helix

58-65 Loop

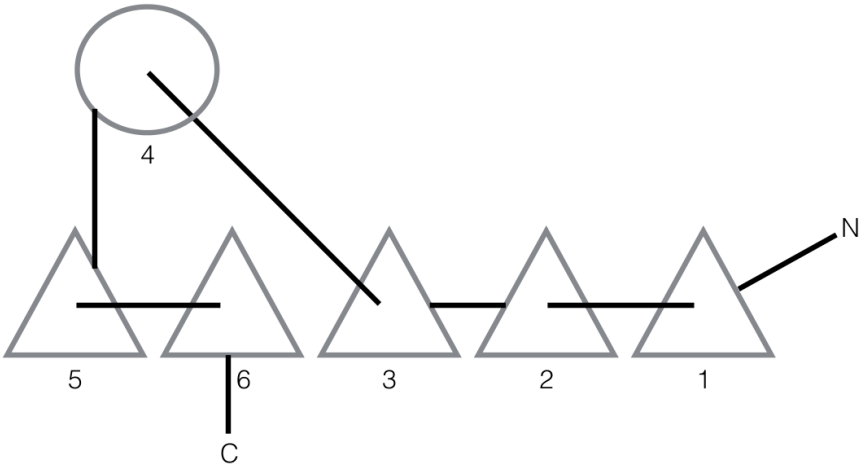
66-71 Beta sheet

72-74 Loop

75-81 Beta sheet

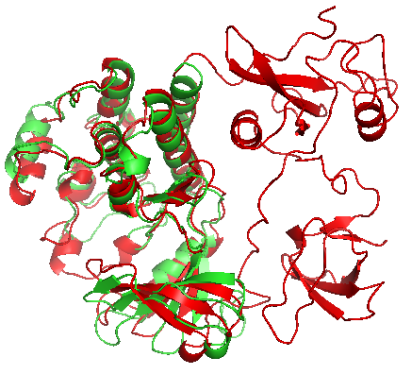
82-84 Loop

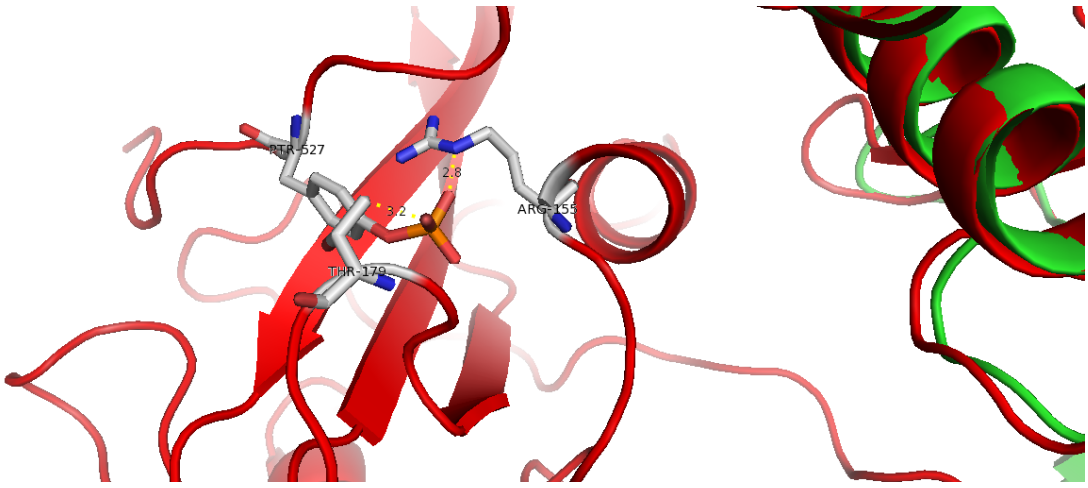
And the topology diagram of the N-terminal domain:



Question 2

The role of regulatory domains in the kinases are to induce conformational changes that switch the kinase from one form (inactive or active) to the other [1].In the cartoon below, the red structure is the 2SRC, the green one is 3LCK.





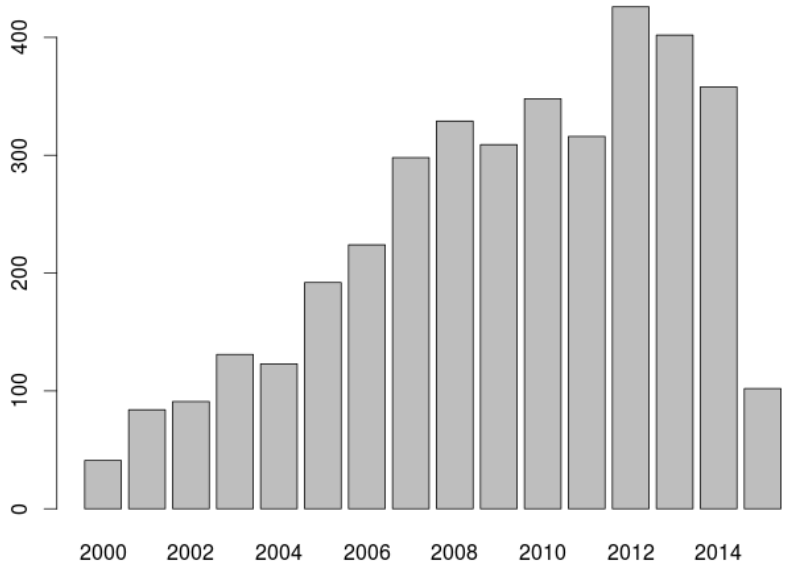
Question 3

We extracted the data from the protein database using the following query:

```
Holdings : Molecule Type=ignore Experimental Method=X-RAY

and

DepositDateQuery: database_PDB_rev.date_original.comparator=between
database_PDB_rev.date_original.min=2000-01-01 database_PDB_rev.date_original.max=2015-05-13
database_PDB_rev.mod_type.comparator=< database_PDB_rev.mod_type.value=1

and

StructTitleQuery: struct.title.comparator=contains struct.title.value=Kinase
```

We did export the results to a CSV file to further our analysis using R and obtained the following summary table and associated graph:

```
> table(years) years 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
 2010 2011 2012 2013 2014 2015 41 84 91 131 123 192 224 298 329 309
 348 316 426 402 358 102 > barplot(table(years))
```



We think that the steady increase in the number of kinases being researched is linked to the fact that Kinases are involved in various pathways whose defects lead to diseases.

Question 4

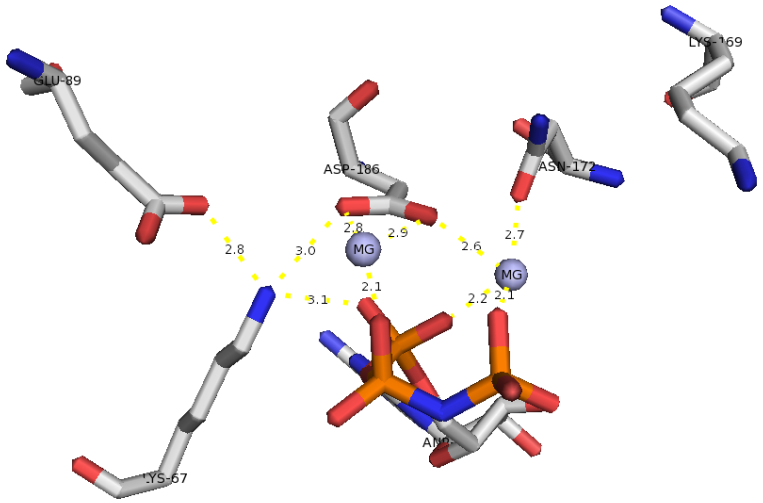
Part a

The molecule bound is Phosphoaminophosphonic Acid-Adenylate Ester, or ANP. Along with 2 Magnesium Ions.

Part b

ANP is an analog of ATP that cannot be hydrolized by the kinase. Therefore it stays bound to the active site of the kinase and allows for the crystal structure of the molecule to be established.

Part c



As shown in the figure, a salt bridge is formed between Glu89 and Lys67. Lys67 forms a salt bridge directly with the α-phosphate oxygen of the ANP molecule. Asp186 forms H-Bond with the Lys67 and also coordinates the Magnesium Ion, that in turns coordinates the β-phosphate oxygen of the ANP molecule. Asn172, in collaboration with Asp186 coordinates the second Magnesium Ion which interacts with the α and γ-phosphate oxygens of the ANP molecule [2].

Part d

The AUTHOR section from the PDB file reveals the same list of names as the list of the article's authors:

```
AUTHOR     K.C.QIAN,L.WANG,E.R.HICKEY,J.STUDTS,K.BARRINGER,C.PENG,
AUTHOR    2 A.KRONKAITIS,J.LI,A.WHITE,S.MISCHE,B.FARMER
```

Question 5

We found 2 proteins of interest: 1JKK and 3F5U. Both have reasonable resolutions and R-Free values are similar. 1JKK boasts a 'up to 1.5 A' resolution of the catalytic domain.

However, after visualizing the B-Factors using pymol, we decided to go with 3F5U.

Part a

The AUTHOR section from the PDB file reveals the same list of names as the list of the article's authors:

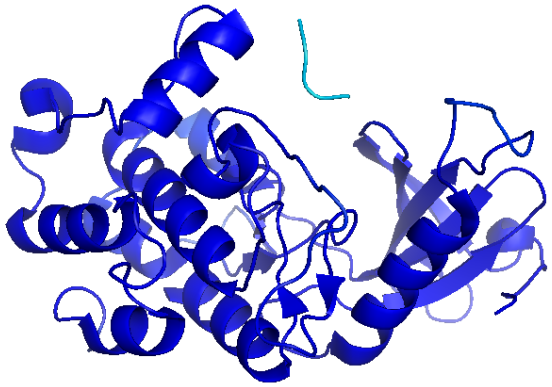
```
AUTHOR    L.K.MCNAMARA,D.M.WATTERSON,J.S.BRUNZELLE
```

Part b

Arg156 and Glu107 have zero occupany. Since the orientation of the residues in space could not be determined, the interaction of zero occupancy residue with neighbouring residues is unknown. This may in turn affect the protein model tertiary structure. Gln223 has multiple alternative rotamer conformations.

Part c

Overall, this structure has low B-factors. There is however a loop region, located around amino acids 291-294, which displays higher B-Factors (around 100). The reason for this seems to be lying in the fact that this loop is part of a flexible regions at the C-Terminus, which is not captured by the X-Ray crystalography.



Part d

The domains present in DAPK1-Human [1] are:

- Kinase whose function is to catalize the transfer of phosphate groups to specific substrates.
- Ankyrin domains (multiple found) mediate protein-protein interactions. [2]
- Roc domain, which is a GTPase domain. GTP binding to the ROC domain activates kinase activity. [3]
- The death domain (DD) is a protein interaction module composed of a bundle of six alpha-helices. [4]

Part e

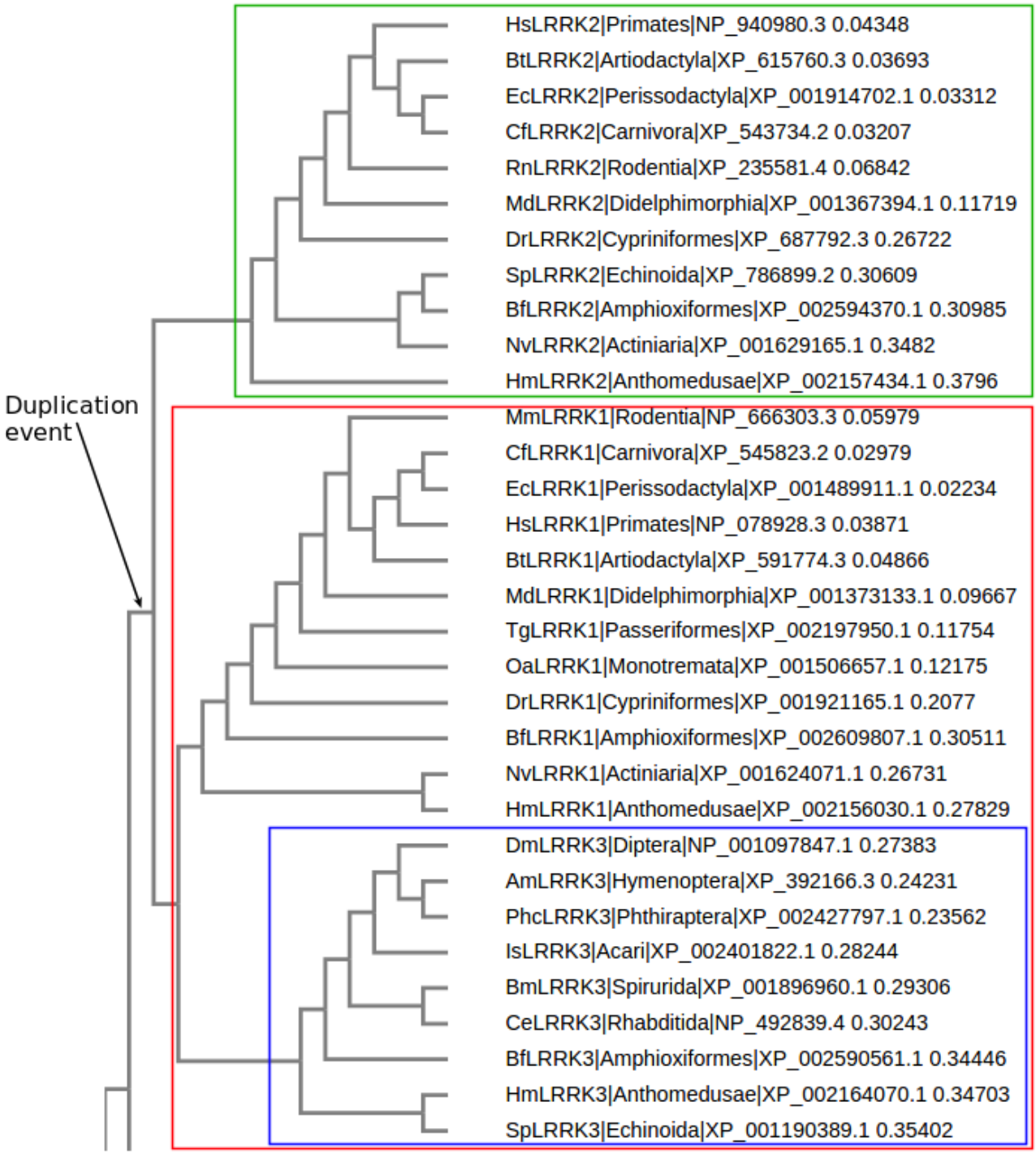
The sequences provided are all homologous and display sequence similarity.Some of the organism in which the protein is found belong to varied kingdom such as Animal, Bacteria, Plants, which indicates that it must originates from an ancient gene that existed even before the split between Eukaryotes and Prokaryotes on the tree of life.

To further our analysis, we used the guide tree created when we performed the multiple sequence alignment, along with the results of the alignment. We found that the gene had had multiple duplication event during evolution. For example, we analyzed the duplication event (indicated on the figure) that gave rise to the proteins indicated by LRRK1, LRRK2, and LRRK3. All of which can be fount in organism belonging to the Eukaryotic domain. On the picture below, you can see that we emphasized 3 sub-clusters. The green and the red

[1] http://www.uniprot.org/uniprot/P53355
[2] https://en.wikipedia.org/wiki/Ankyrin_repeat
[3] http://www.copewithcytokines.de/cope.cgi?key=ROC%20domain
[4] https://en.wikipedia.org/wiki/Death_domain

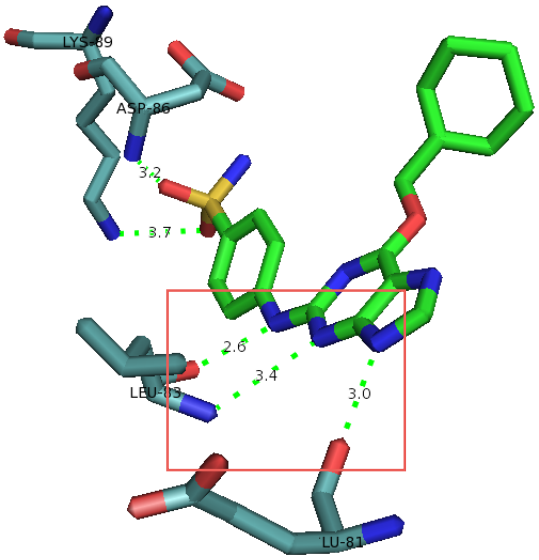
cluster show the LRRK1 and LRRK2+LRRK3 groups (the blue cluster can be hypothesized to have come from a duplication event early in the Animals/Invertebrates branch).

Using the results from the sequence alignment, you can observe that the alignment scores between the protein within the same cluster (for example LRRK1) are better than the ones from another cluster (LRRK2 in our example). This makes sense when you think about the evolutionary distance between the 2 genes coding for these paralog genes.

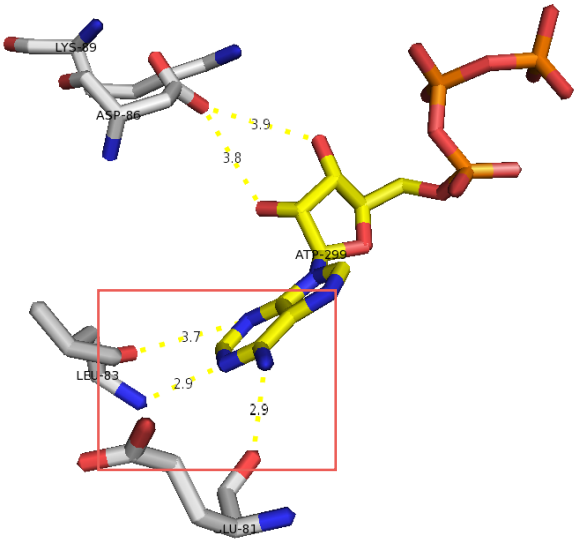


Question 6

The essential features are indicated in the red frame on the cartoon below. They are all hydrogen bridges.



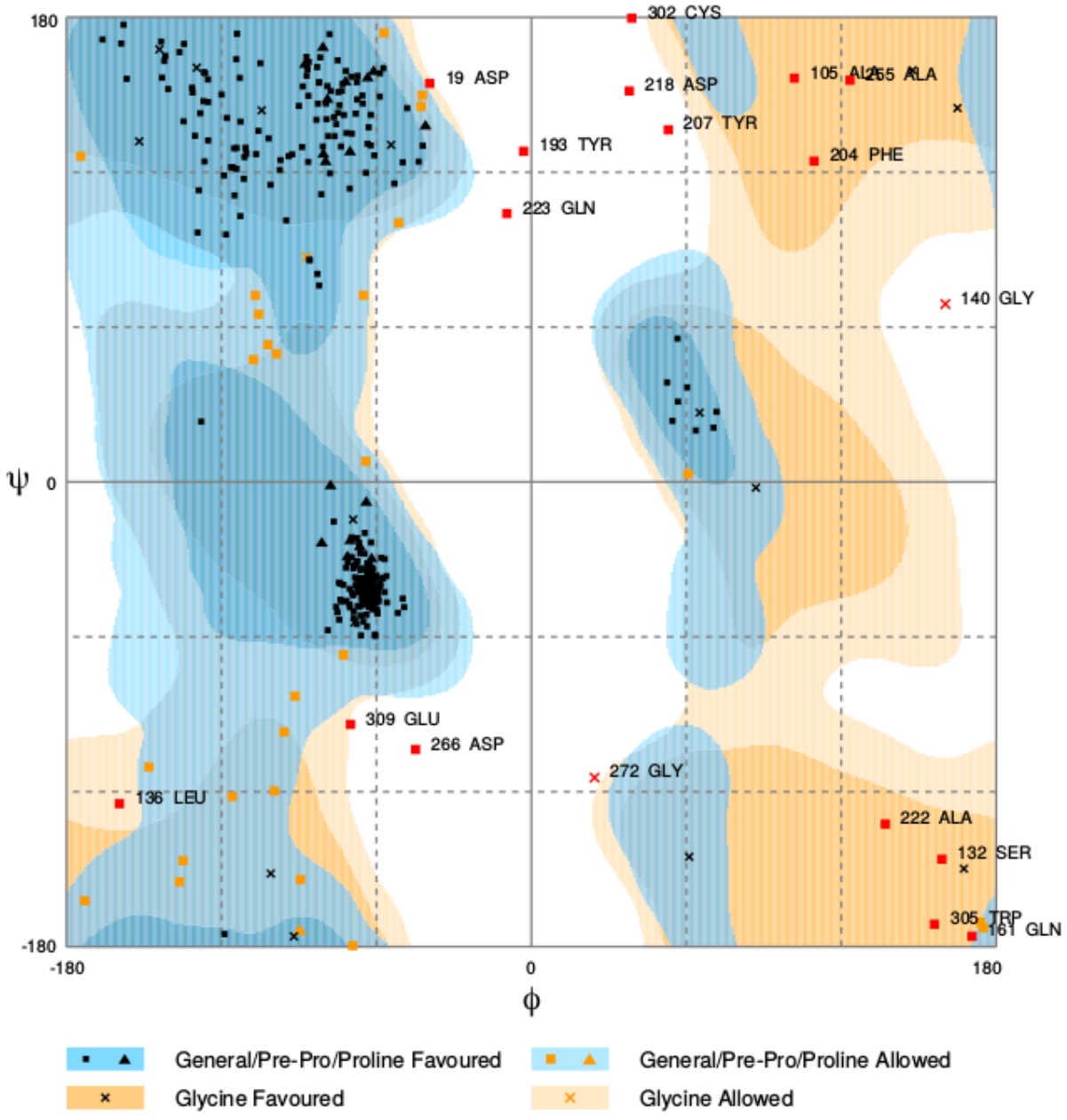
2C6O



1FIN

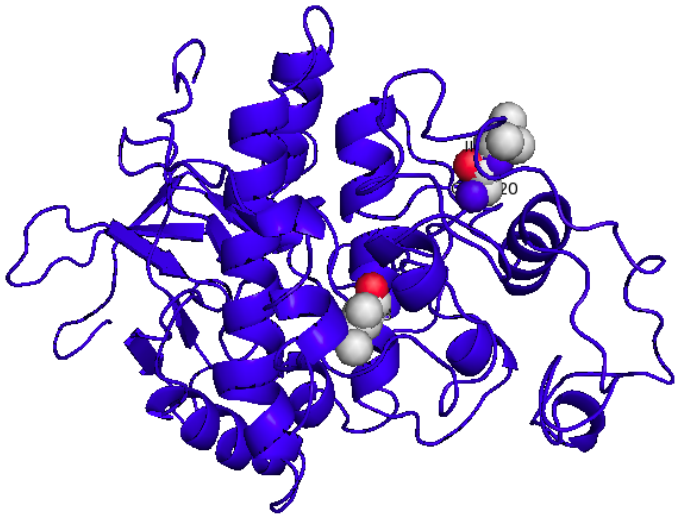
Question 7

Part a



Number of residues in favoured region (~98.0% expected)	:	303 (86.8%)
Number of residues in allowed region (~2.0% expected)	:	28 (8.0%)
Number of residues in outlier region	:	18 (5.2%)

Part b



```
CLUSTAL 2.1 multiple sequence alignment

lrrk2_mut2      ICGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDL 60
lrrk2_mut3      ICGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDL 60
lrrk2_wt        ICGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDL 60
lrrk2_mut1      ICGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDL 60
                ************************************************************

lrrk2_mut2      ILADLPRNIMLNNDELEFEQAPEFLLGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQ 120
lrrk2_mut3      ILADLPRNIMLNNDELEFEQAPEFLLGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQ 120
lrrk2_wt        ILADLPRNIMLNNDELEFEQAPEFLLGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQ 120
lrrk2_mut1      ILADLPRNIMLNNDELEFEQAPEFLLGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQ 120
                ************************************************************

lrrk2_mut2      ELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIALHVA 180
lrrk2_mut3      ELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIALHVA 180
lrrk2_wt        ELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIALHVA 180
lrrk2_mut1      ELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIALHVA 180
                ************************************************************

lrrk2_mut2      DGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYSIAQYCCRMGIKTSEGTPGFR 240
lrrk2_mut3      DGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGTAQYCCRMGIKTSEGTPGFR 240
lrrk2_wt        DGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGIAQYCCRMGIKTSEGTPGFR 240
lrrk2_mut1      DGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAITAKIADYGIAQYCCRMGIKTSEGTPGFR 240
                ******************************** ******. ******************

lrrk2_mut2      APEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPVKEY 300
lrrk2_mut3      APEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPVKEY 300
lrrk2_wt        APEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPVKEY 300
lrrk2_mut1      APEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPVKEY 300
```

```
                ************************************************************

lrrk2_mut2      GCAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILLPKNV 351
lrrk2_mut3      GCAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILLPKNV 351
lrrk2_wt        GCAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILLPKNV 351
lrrk2_mut1      GCAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILLPKNV 351
                ************************************************
```

Part c

The multiple sequence alignment below reveals mutations in the Amino Acids at position 212, 219, and 220, also indicated in the cartoon as spheres. These mutations convert hydrophobic amino acids (Iso, Gly) to polar amino acids (Ser, Thr) who are prone to establish H-bonds, making the transition to the active or inactive form of the molecule more difficult.by reducing the flexibility of the region.

References

[1] Huse M. and Kuriyan J. The conformational plasticity of protein kinases. *Cell*, 2002.

[2] Kevin C. Qian, Lian Wang, Eugene R. Hickey, Joey Studts, Kevin Barringer, Charline Peng, Anthony Kronkaitis, Jun Li, Andre White, Sheenah Mische and Bennett Farmer. Structural basis of constitutive activity and a unique nucleotide binding mode of human pim-1 kinase. *The Journal of Biological Chemistry*, 2005.