

configuration_analysis

August 14, 2018

Analyse the distribution of side chain dihedral angles: Intro The resultant cystine residue from the formation of the covalent disulfide bond between the thiol groups of two cysteine side chains is comprised of five dihedral angles. Based on X-ray structures we look at the distribution of these angles (X1, X2 and X3) to define their different conformations.

```
/home/david/coding/lib/python2.7/site-packages/ipykernel_launcher.py:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

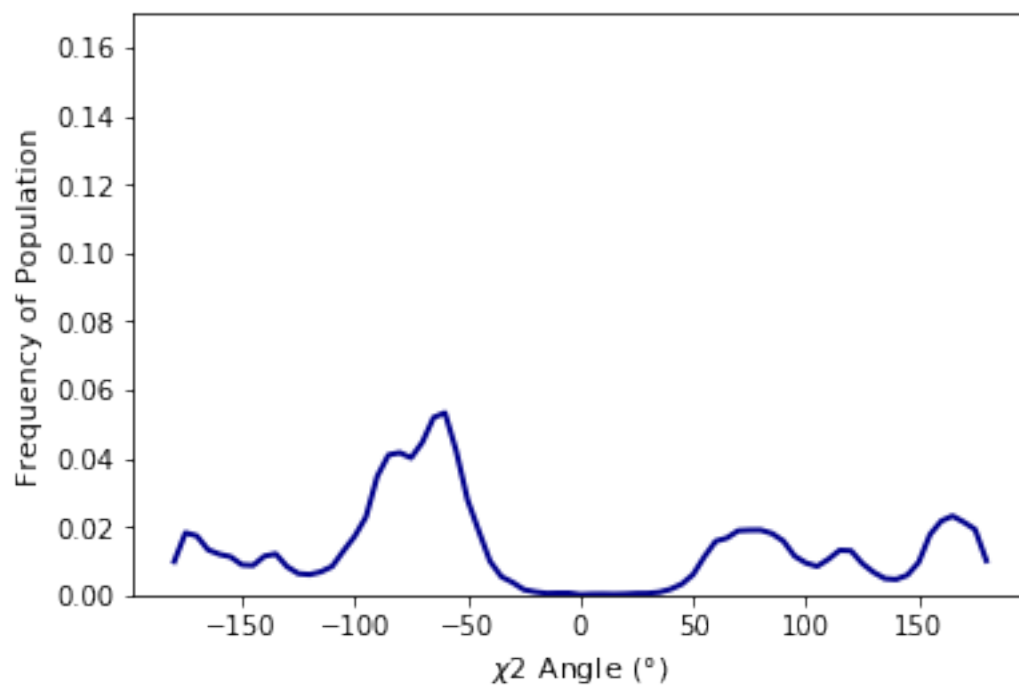
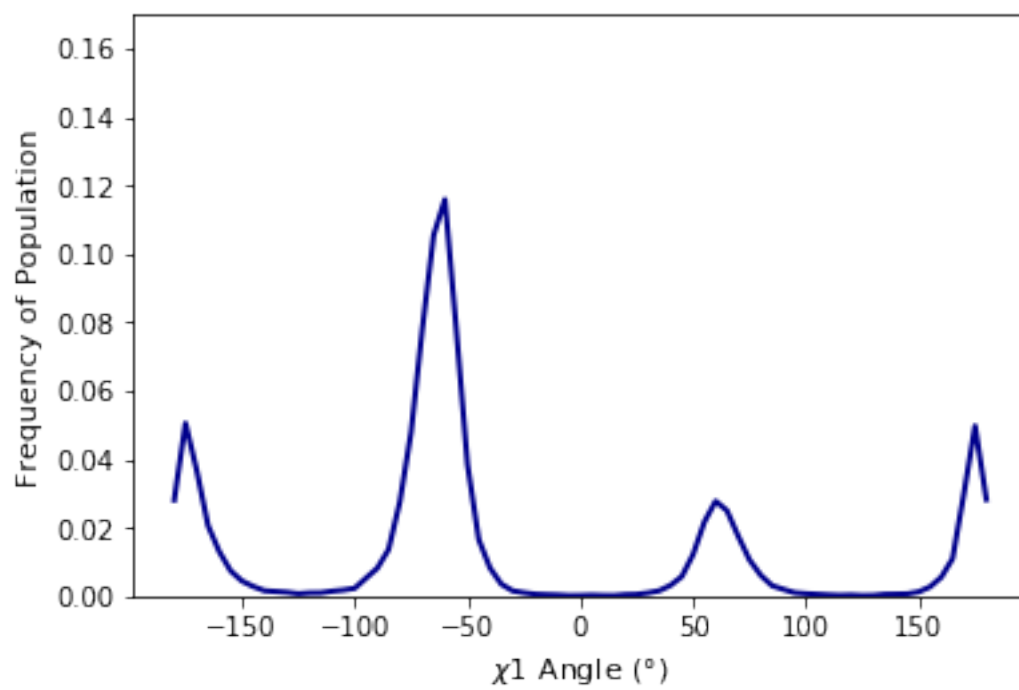
```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
    if __name__ == '__main__':
/home/david/coding/lib/python2.7/site-packages/ipykernel_launcher.py:10: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

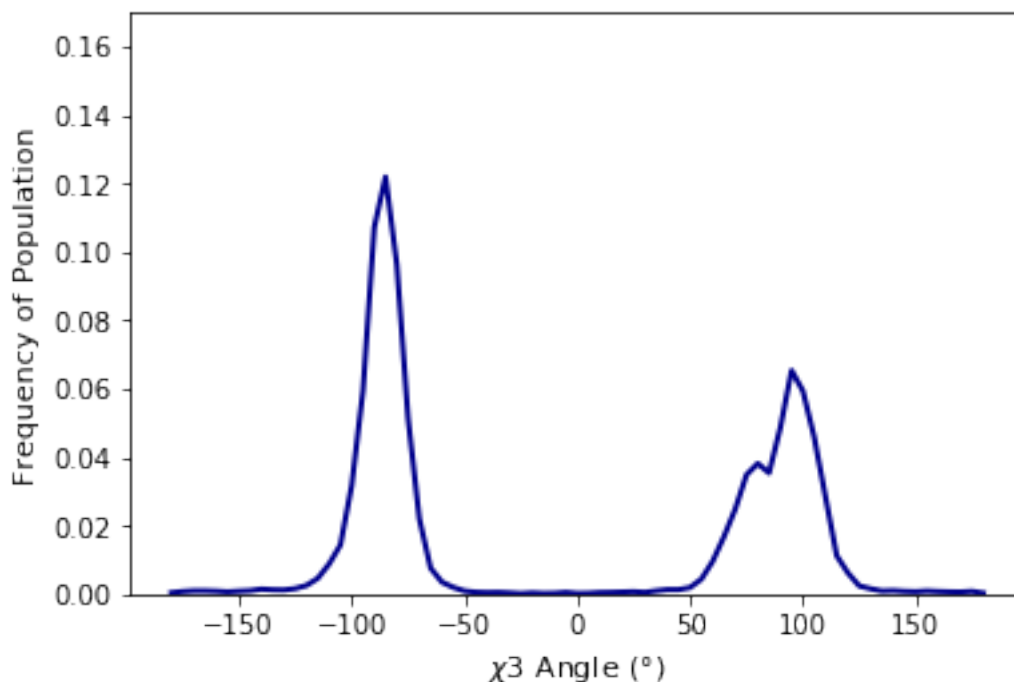
```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
    # Remove the CWD from sys.path while we load stuff.
/home/david/coding/lib/python2.7/site-packages/ipykernel_launcher.py:11: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
    # This is added back by InteractiveShellApp.init_path()
/home/david/coding/lib/python2.7/site-packages/ipykernel_launcher.py:12: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
    if sys.path[0] == '':
/home/david/coding/lib/python2.7/site-packages/ipykernel_launcher.py:13: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
    del sys.path[0]
```





Cystine dihedral distribution: Results In agreement with previous analysis, can see that X1 adopts the 'Gauche-, Gauche+ and Trans' conformations with a very narrow distribution. Therefore defined as: Gauche+ (+60) = +30 to +90 Gauche- (-60) = -90 to -30 Trans (180) = 150 to 210

X2 also adopts these conformations, however a greater distribution is observed. Therefore the following ranges were used for classification Gauche+ (+60) = +30 to +120 Gauche- (-60) = -120 to -30 Trans (180) = 150 to 210

X3 angle distribution follows the previously identified 'Right Handed (+90)' or 'Left Handed (-90)' Right Handed (+90) = +60 to +120 Left Handed (-90) = -120 to -60

Frequency of Configurations: Intro Based on the definitions above, there is a possible 90 configurations a cystine can adopt when considering all five X angles (X1, X2, X3, X2, X1). Below calculates the frequency of those configurations If a disulfide had a dihedral angle out of the defined ranges above it was excluded from analysis

Total Number of Disulfides: 19605

Total Number of Disulfides in Defined Configurations for Structural Analysis: 15538

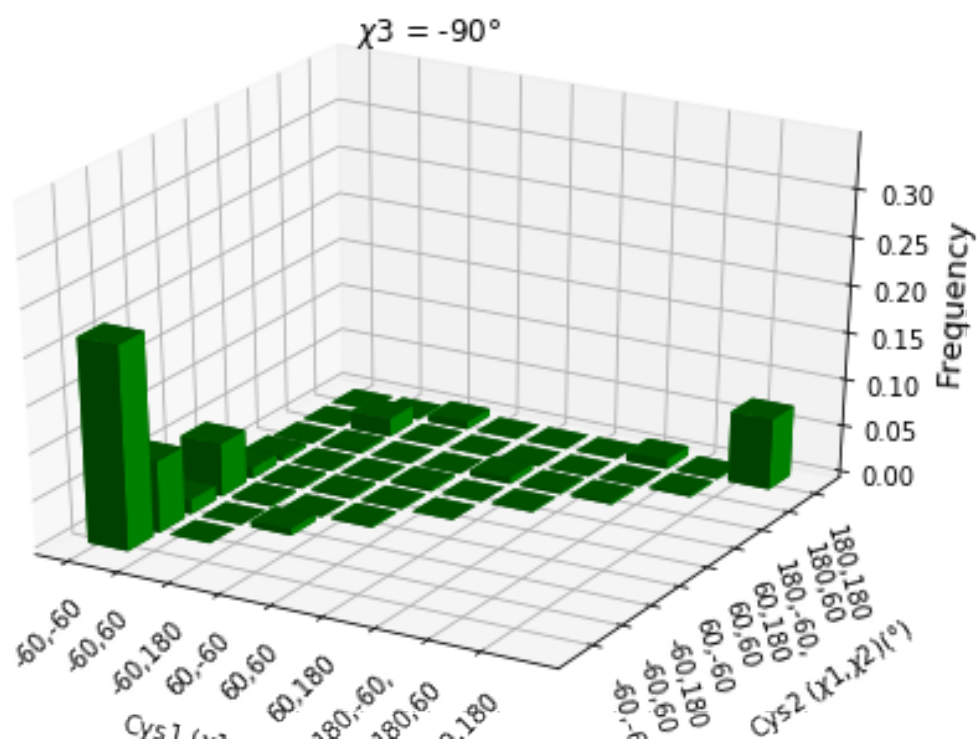
```
/home/david/coding/lib/python2.7/site-packages/ipykernel_launcher.py:192: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

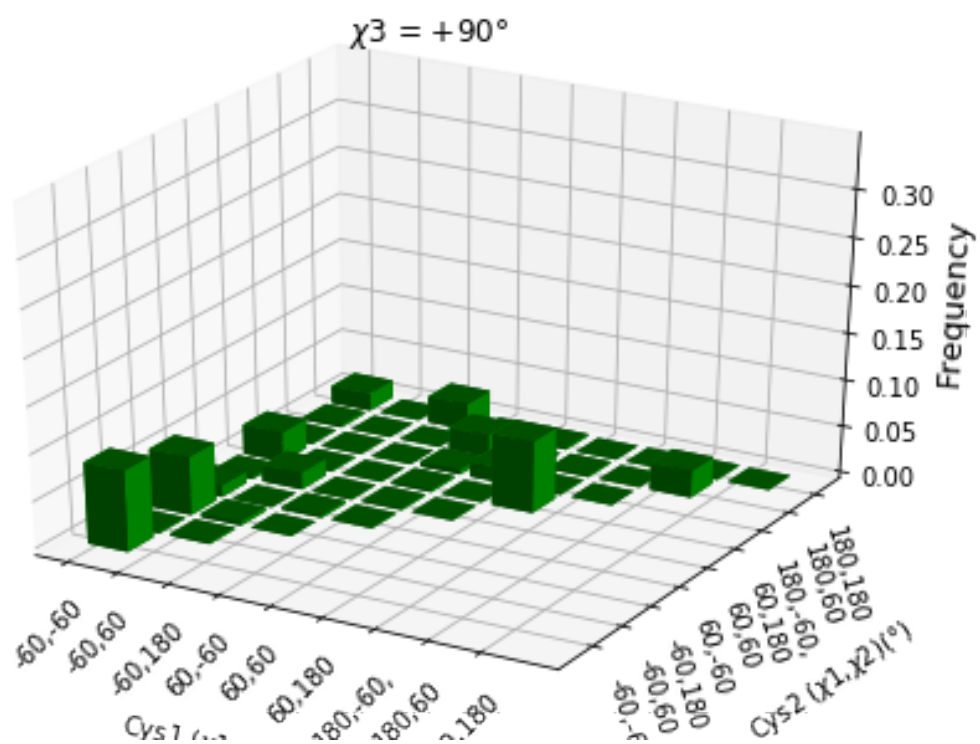
See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#>

Generate Configuration Frequency Graphs Generate a 3D plot tha shows the frequency of pop-
ulation for each configuration

1 Generate configuration graphs for all cystines

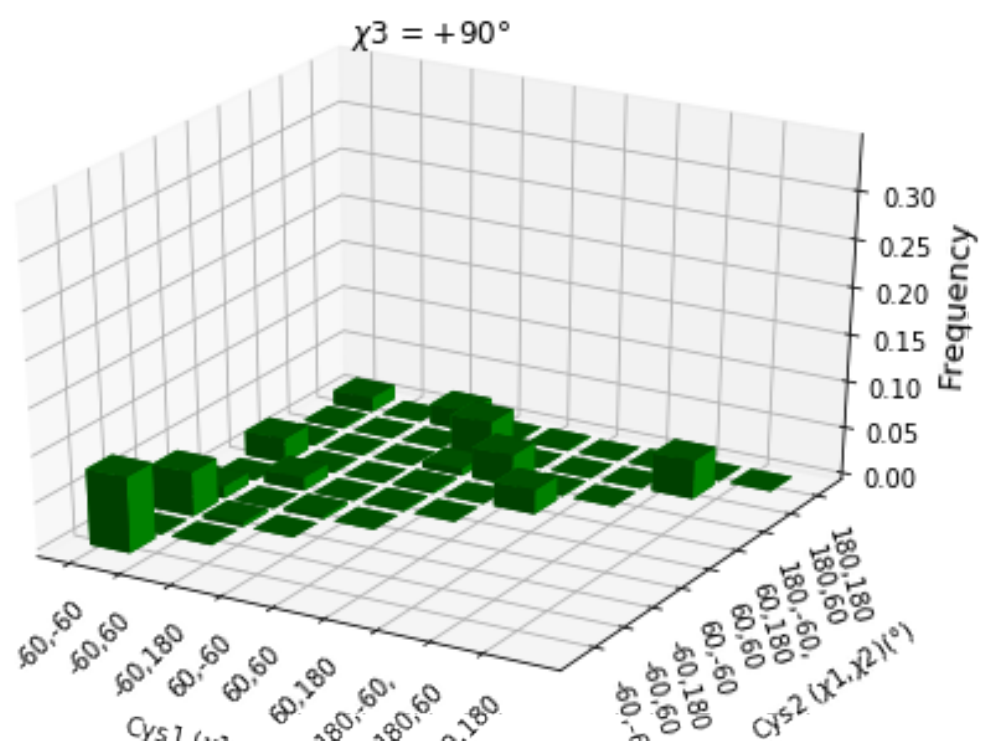
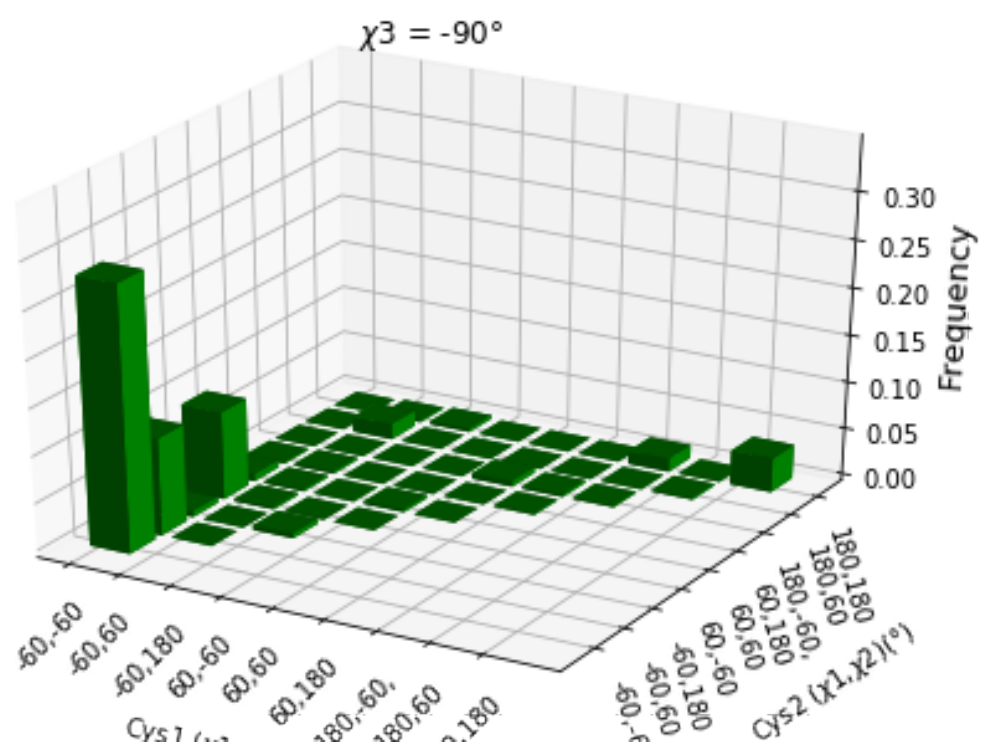
Out [5] : ()





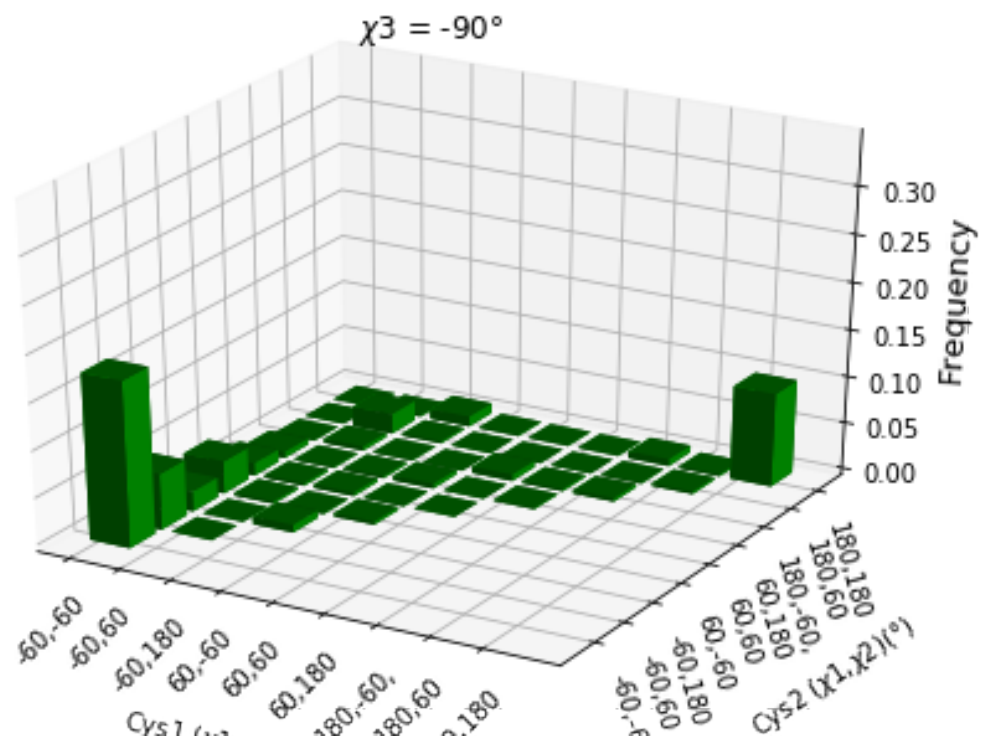
1.1 Generate configuration graphs for peptide cystines

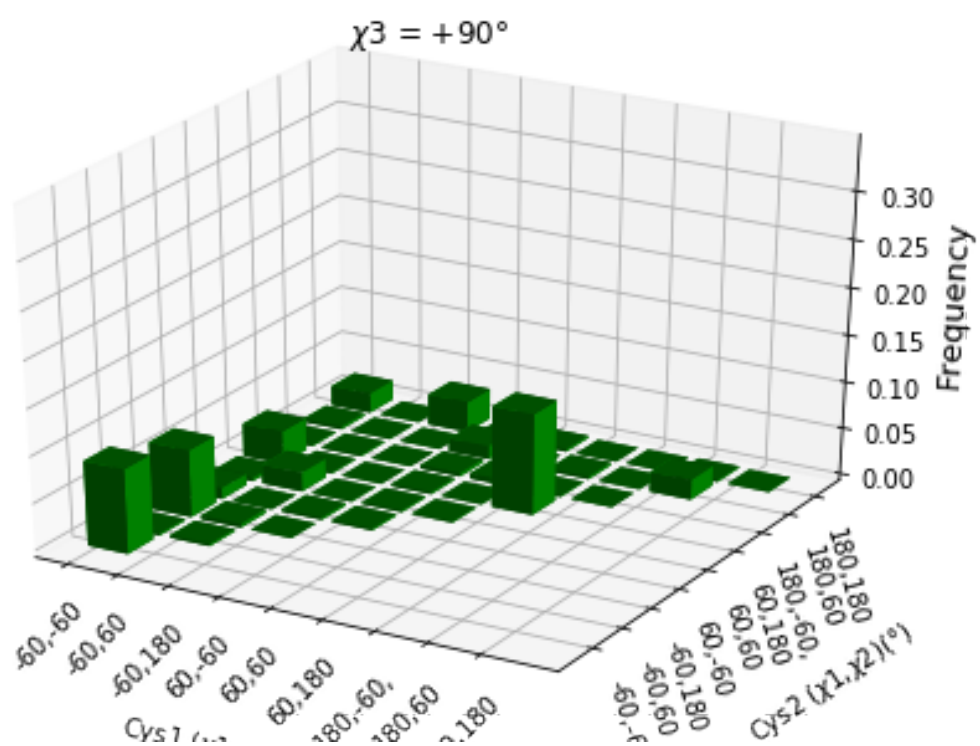
Out[6]: ()



1.2 Generate configuration graphs for protein cystines

Out[7]: ()





1.3 For Vicinal Disulfides

Out [8] : ()

1.4 CONFIGURATION FUNCTION

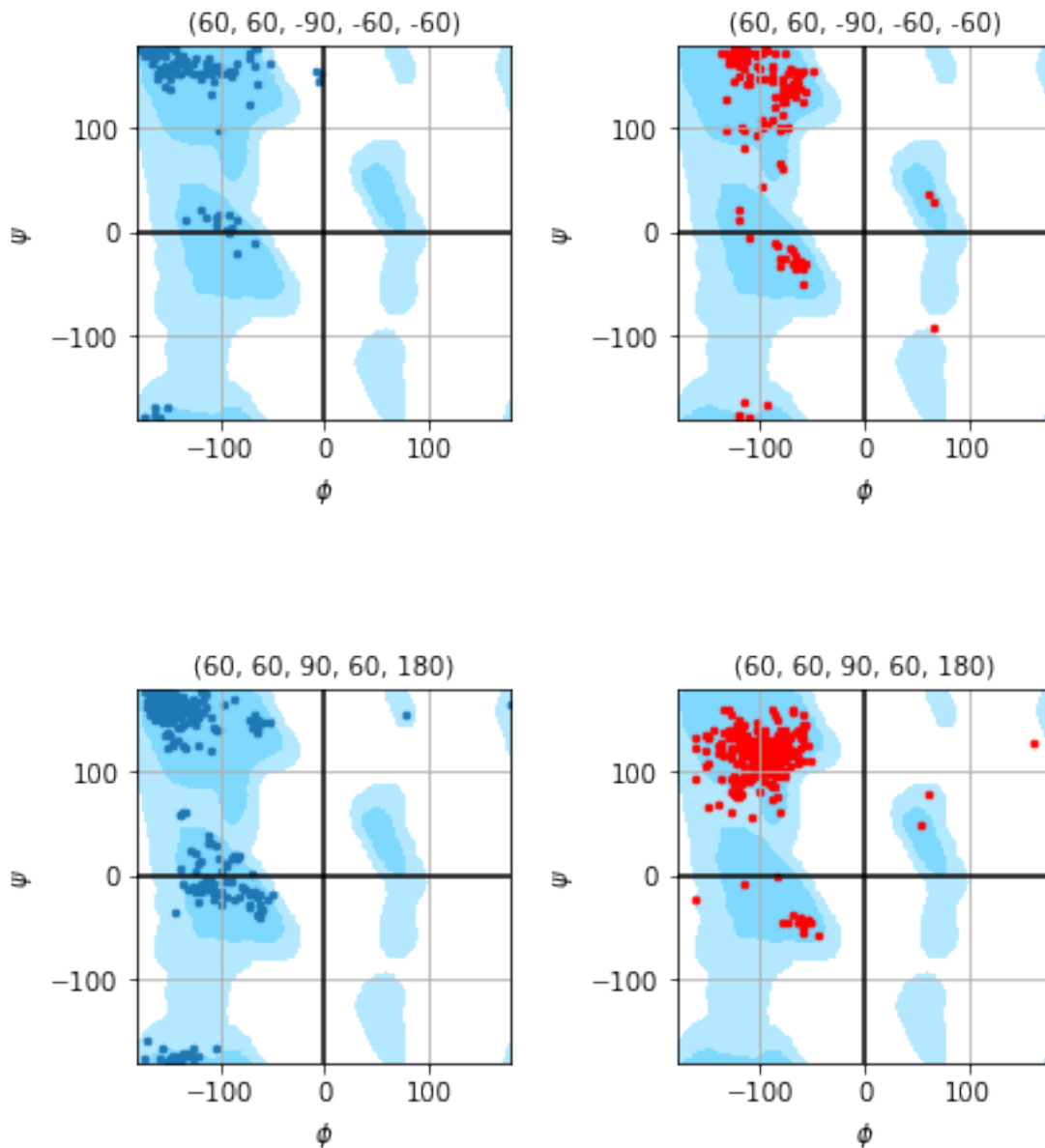
1.4.1 The below function returns individual dataframes for each configuration

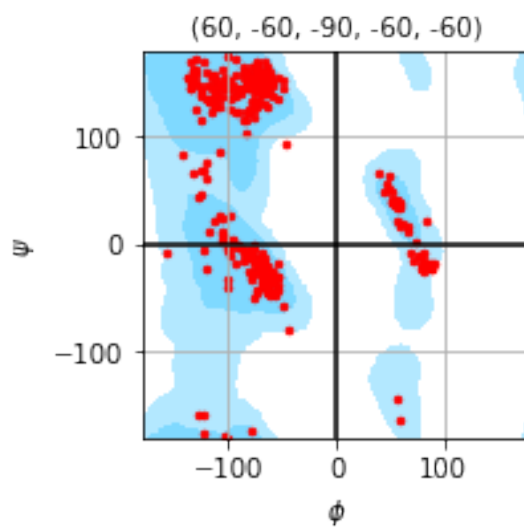
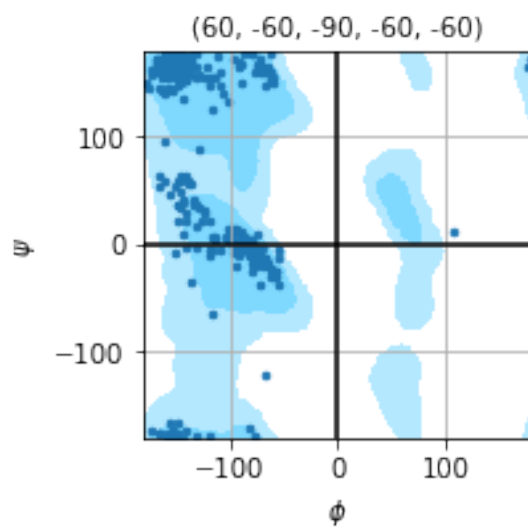
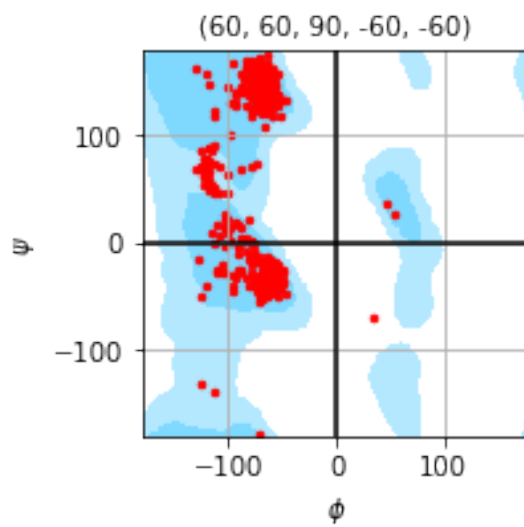
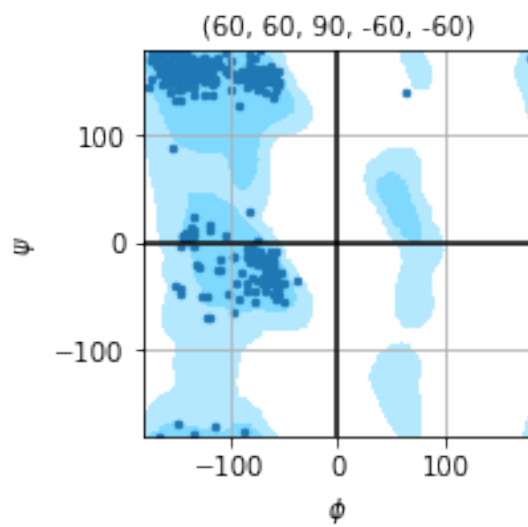
1.4.2 It will be called to generate figures that investigate structural features based on configuration

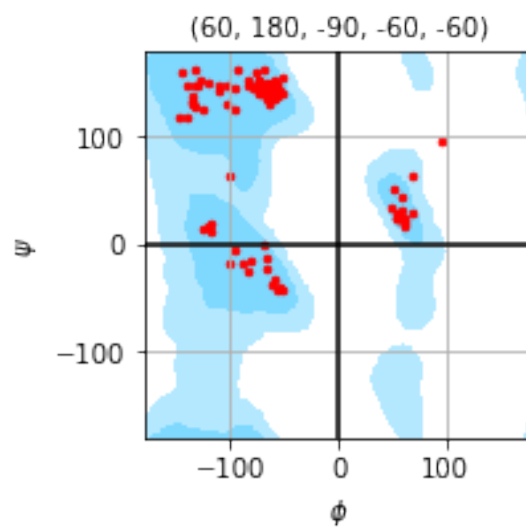
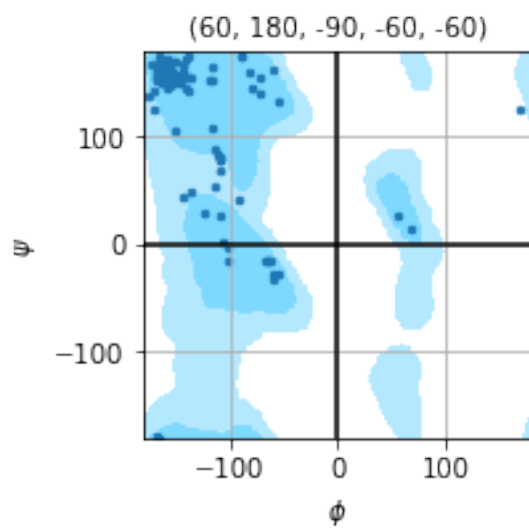
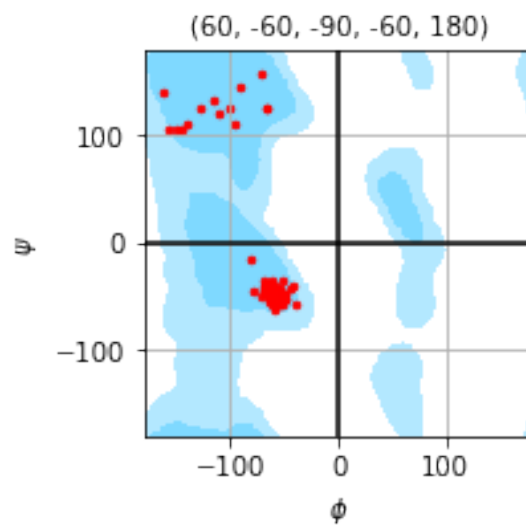
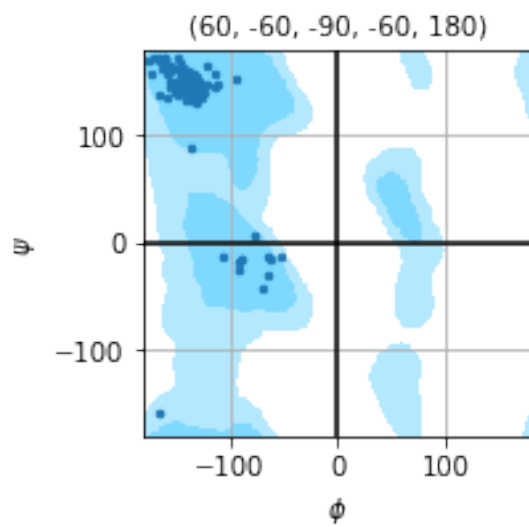
1.4.3 It re-writes the dataframe to ensure $\text{Cys1} - \text{Cys2} == \text{Cys2} - \text{Cys1}$ based on configuration order

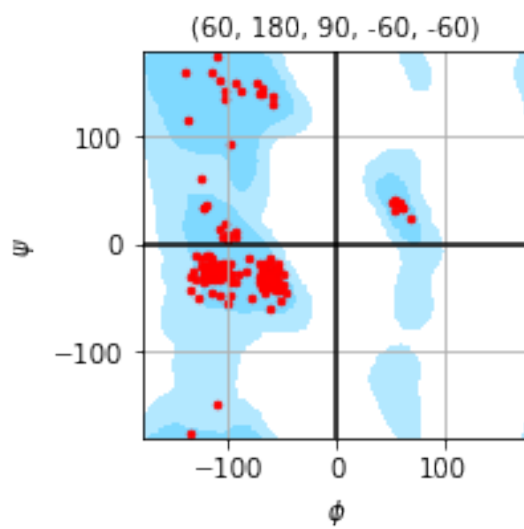
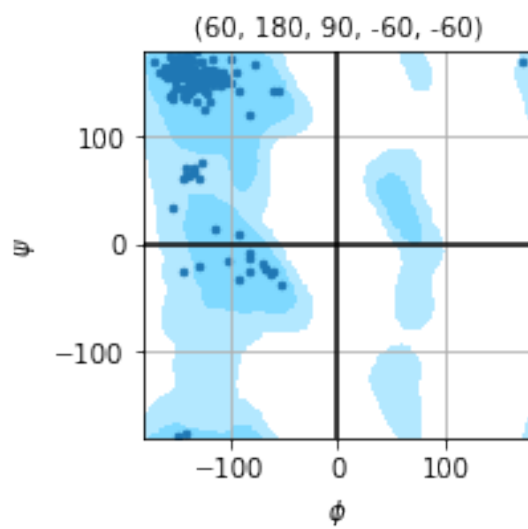
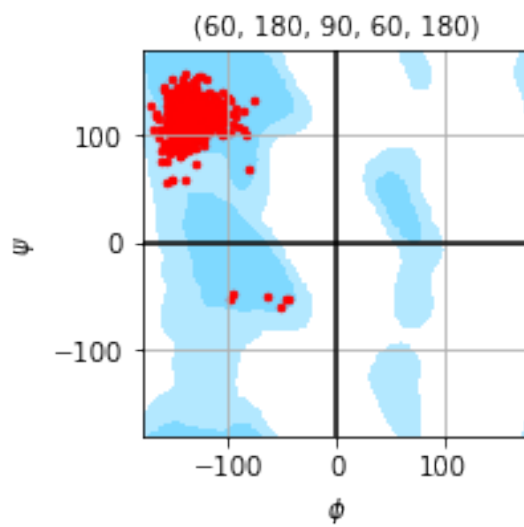
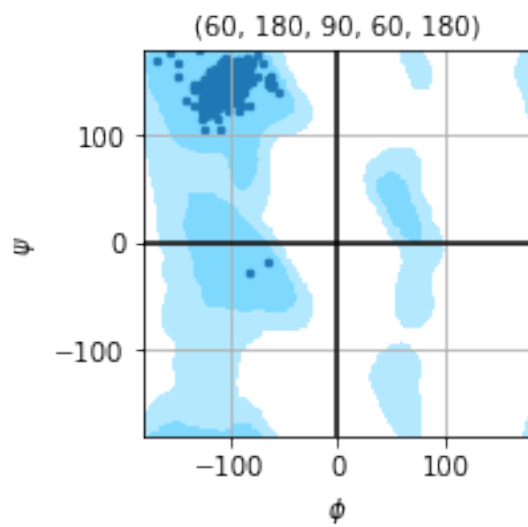
1.5 Function to generate Ramachandran Plots

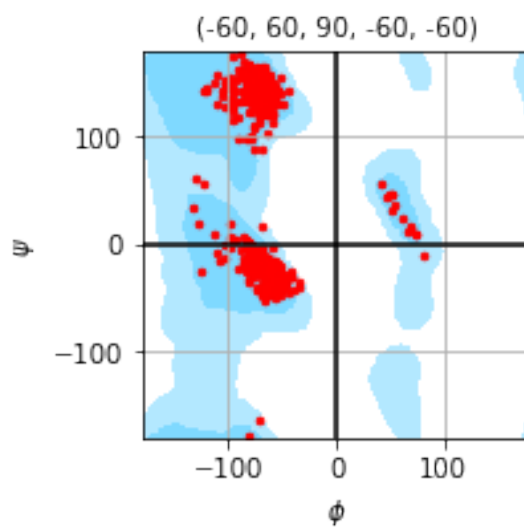
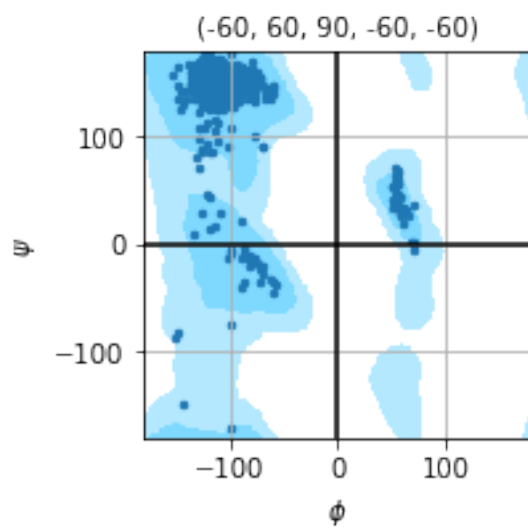
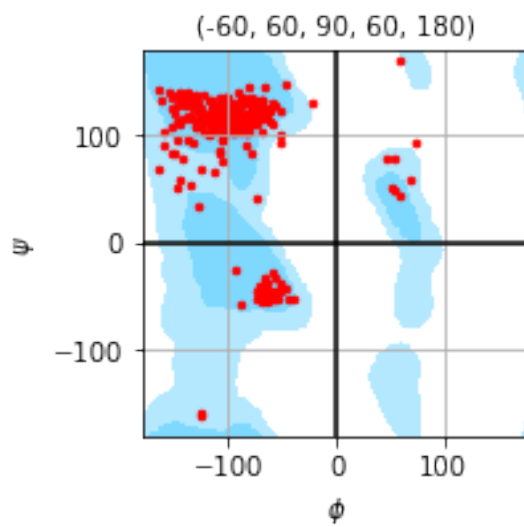
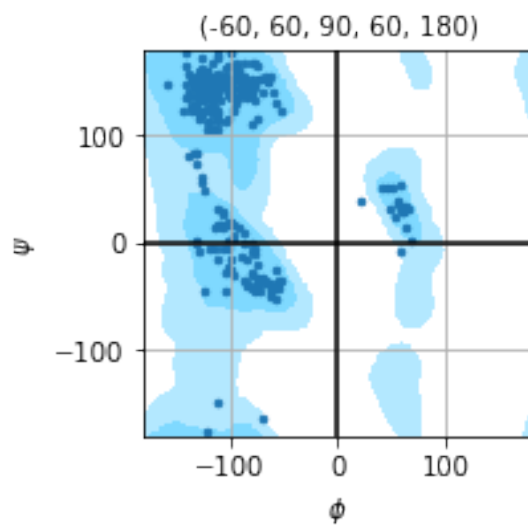
1.6 Call each configuration and

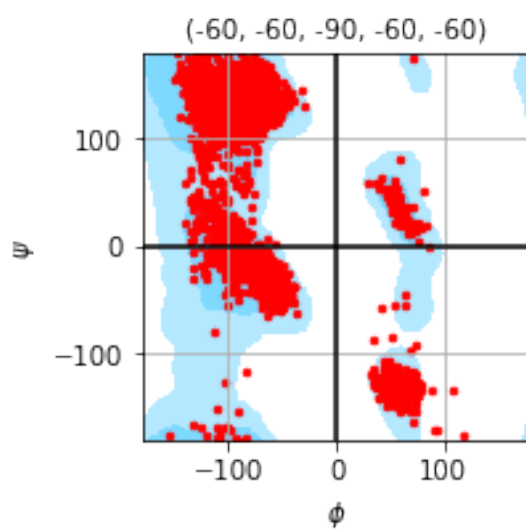
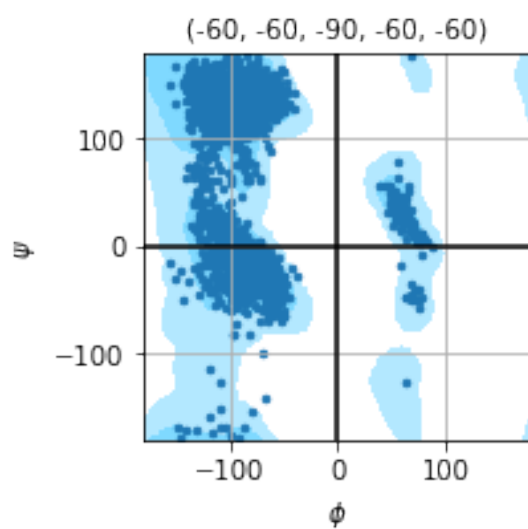
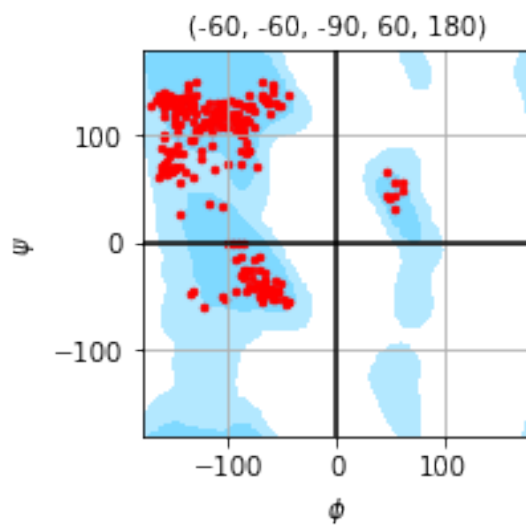
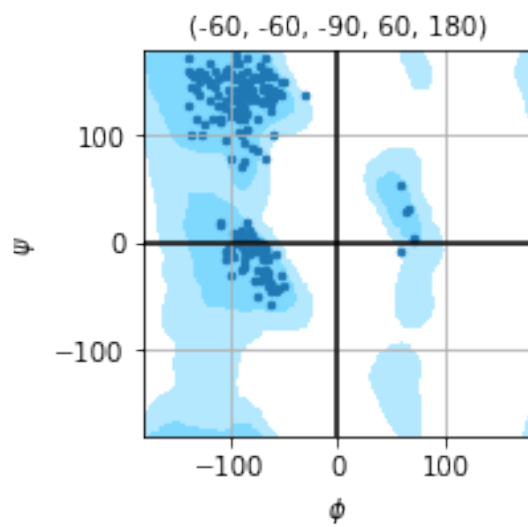


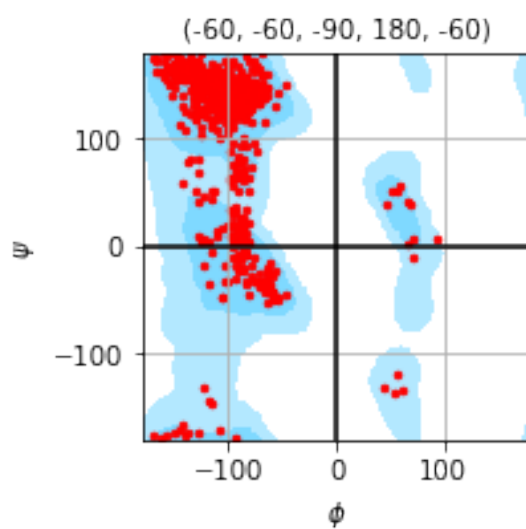
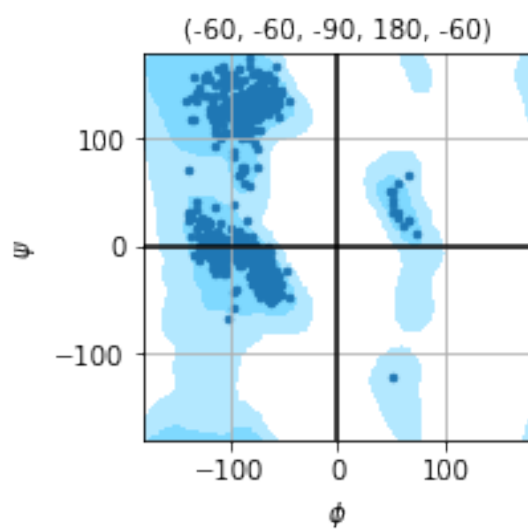
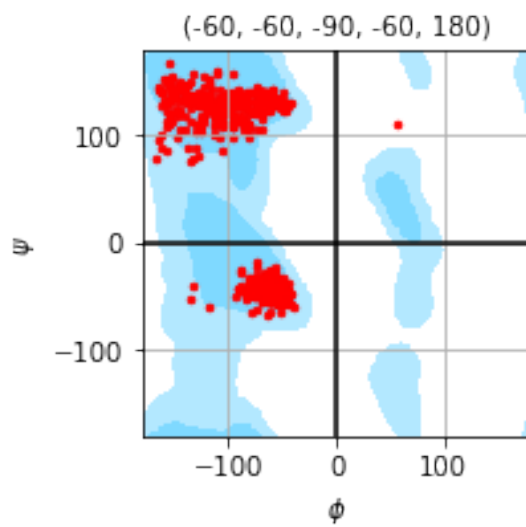
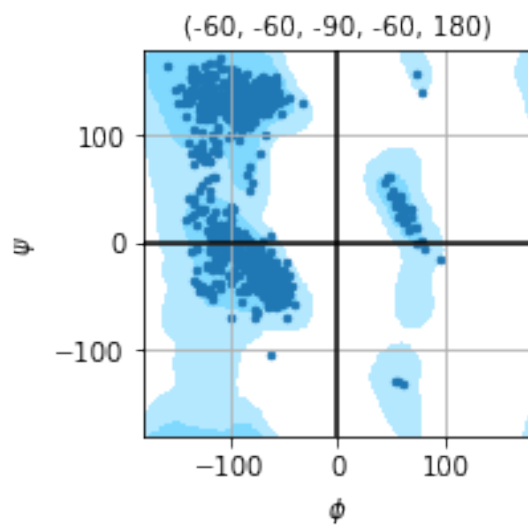


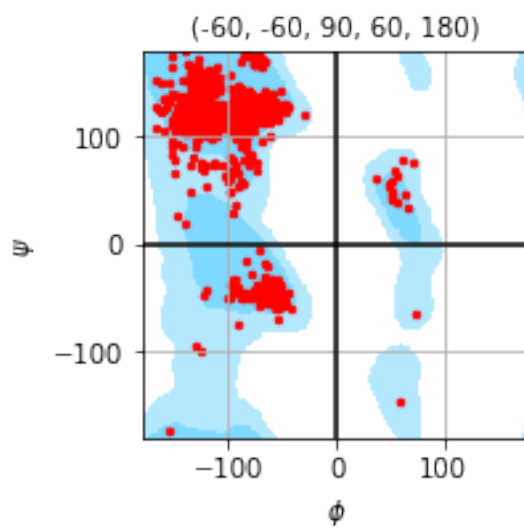
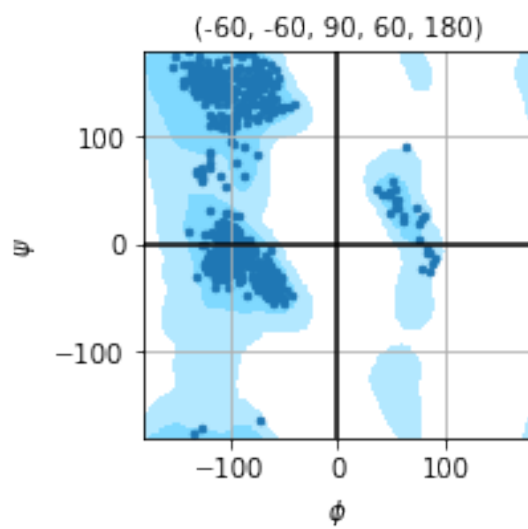
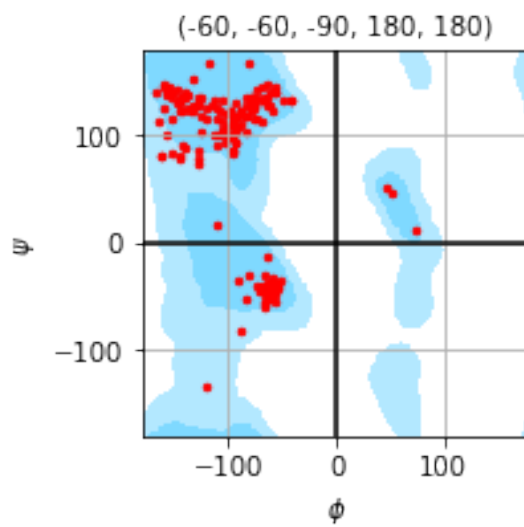
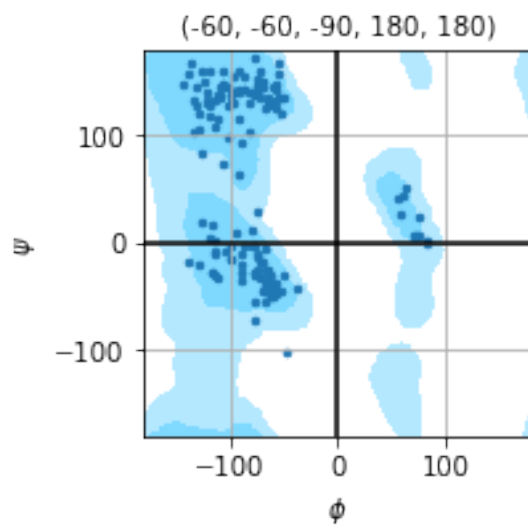


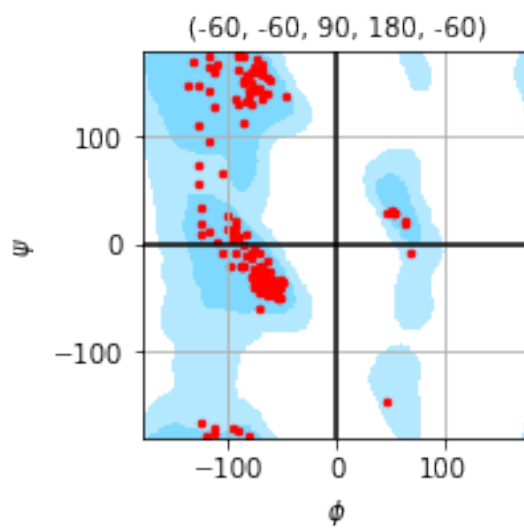
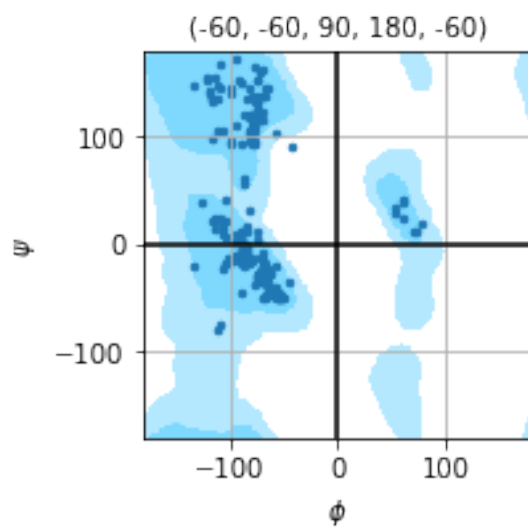
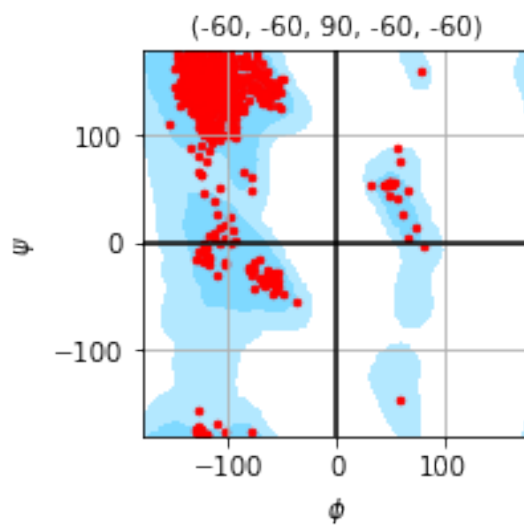
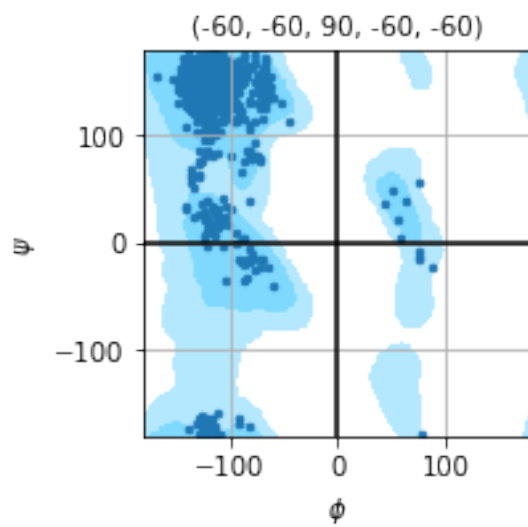


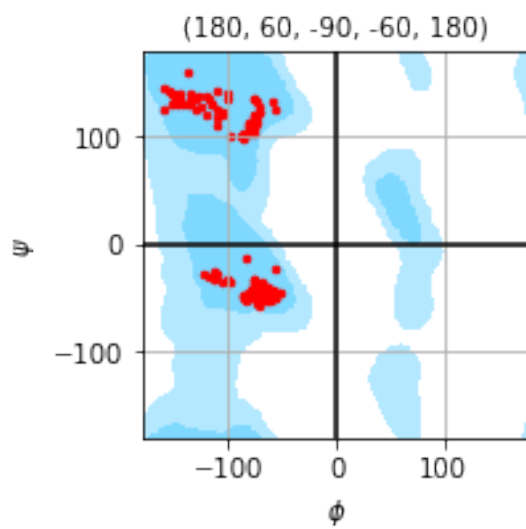
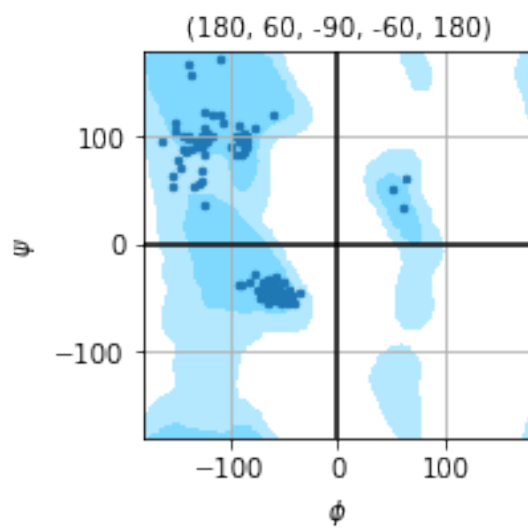
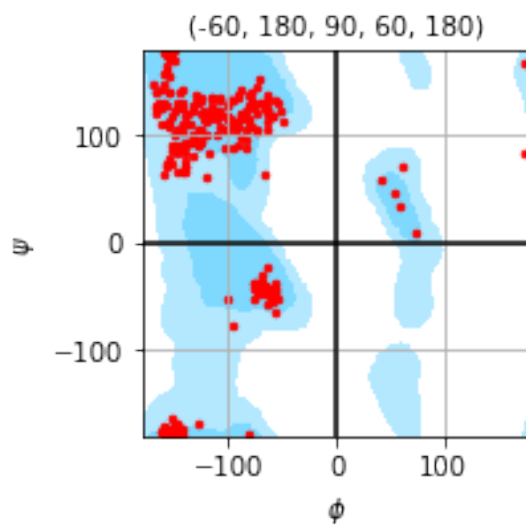
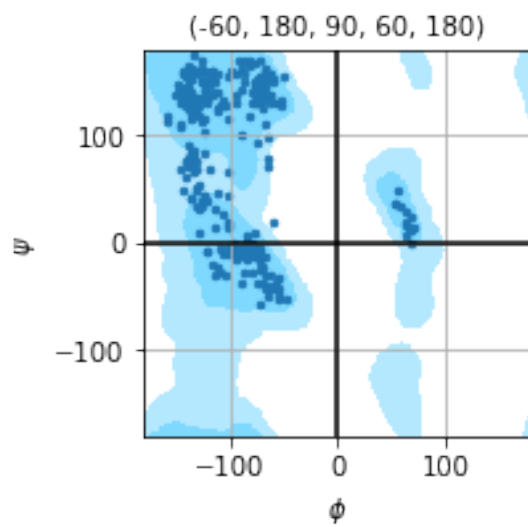


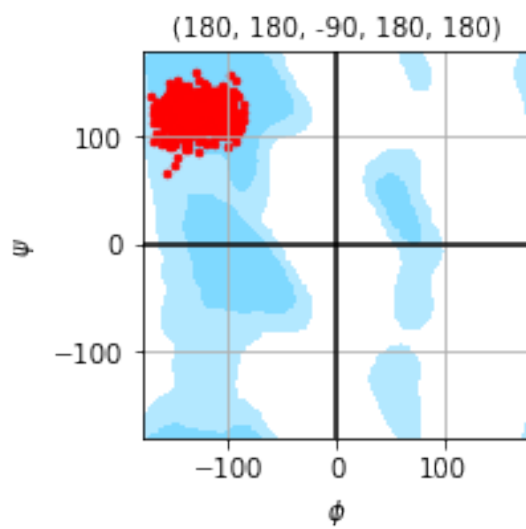
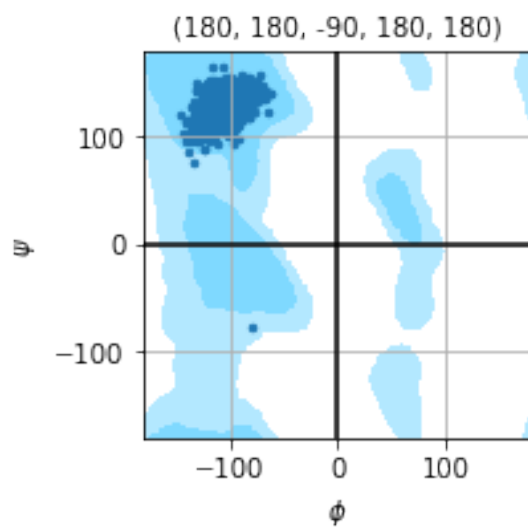
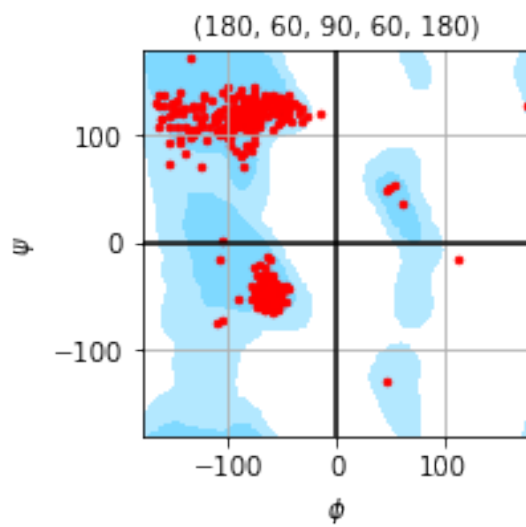
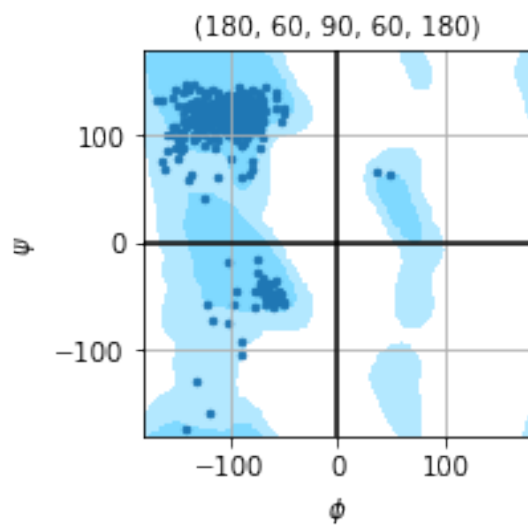


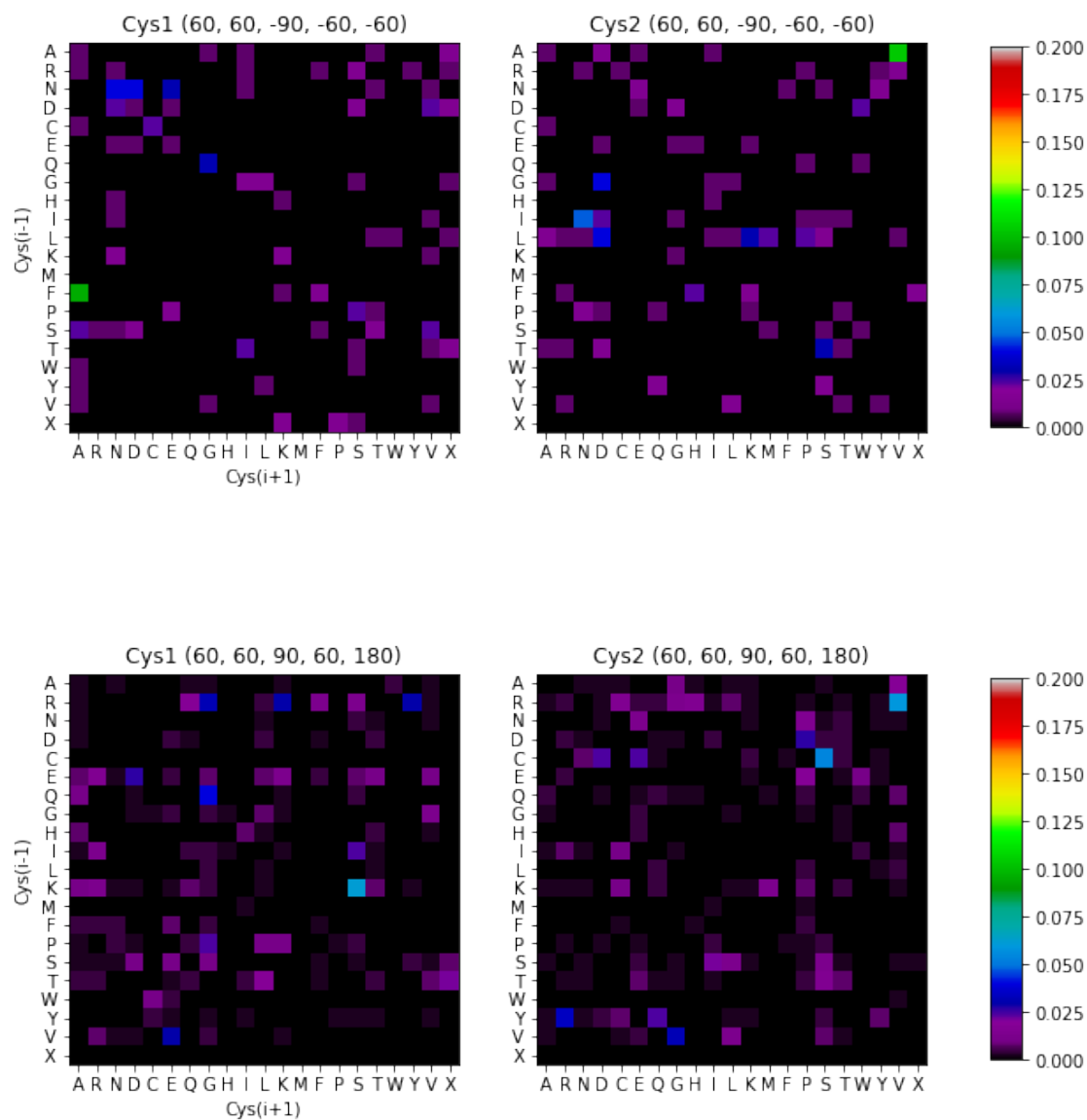


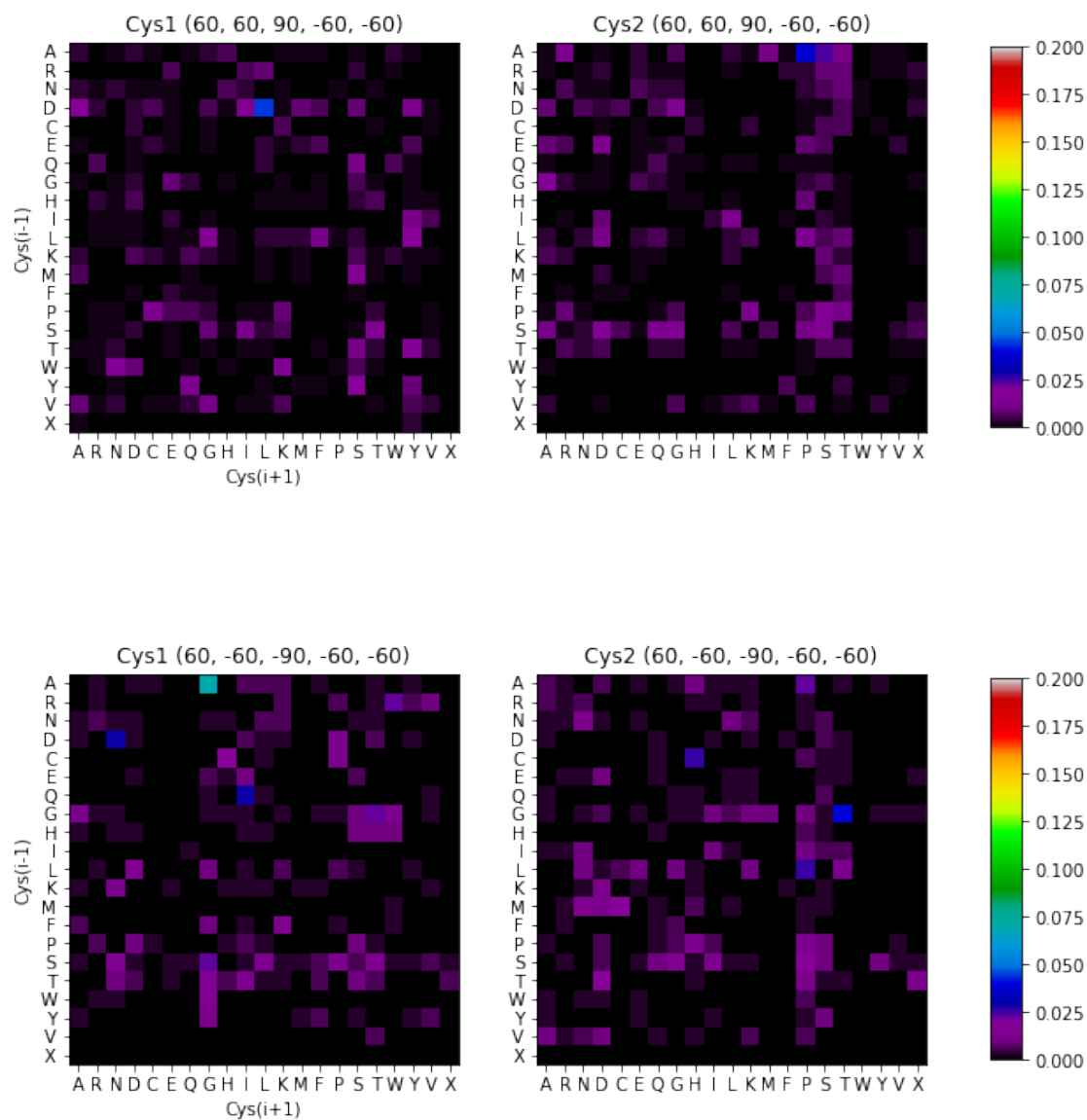


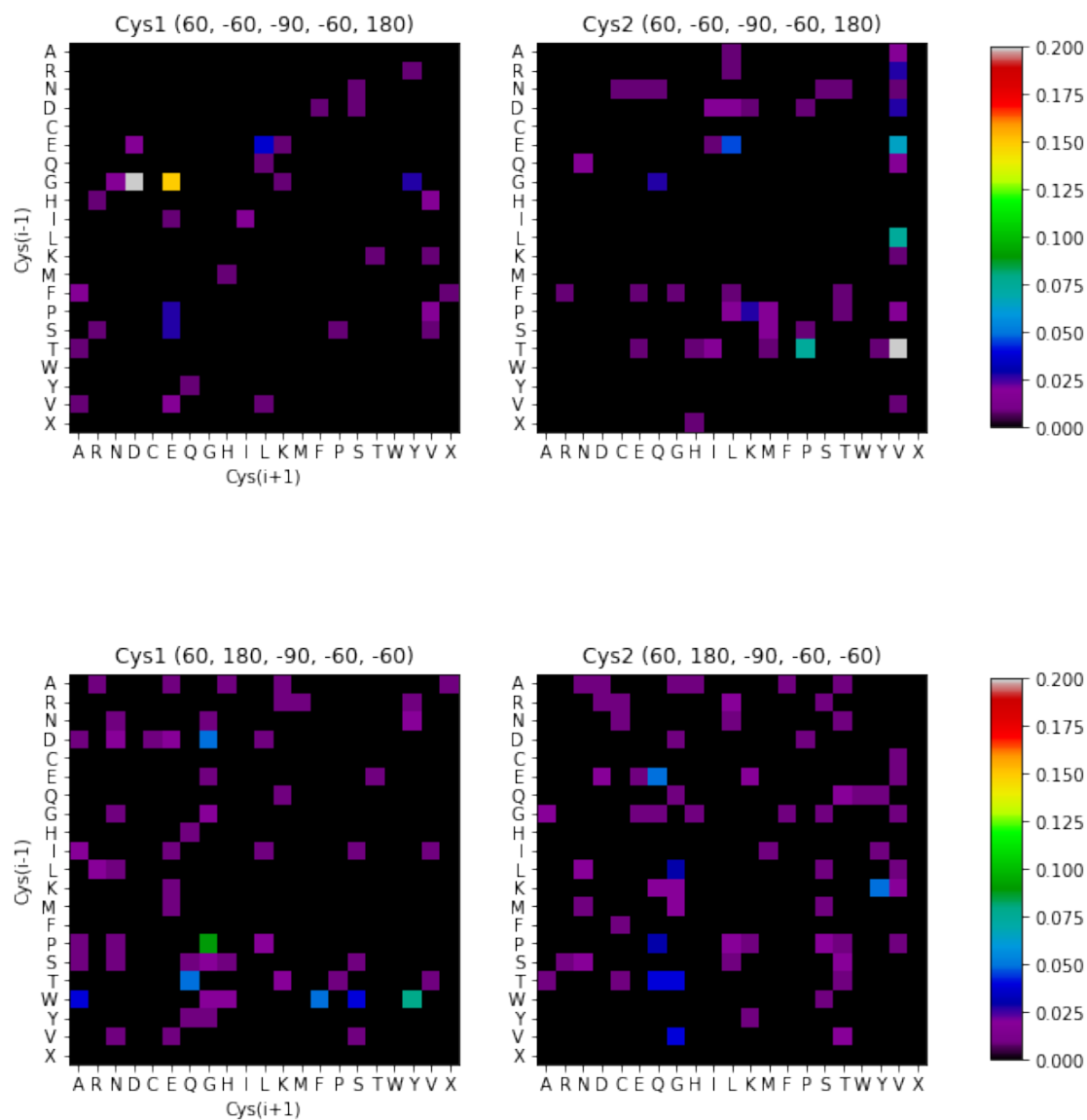


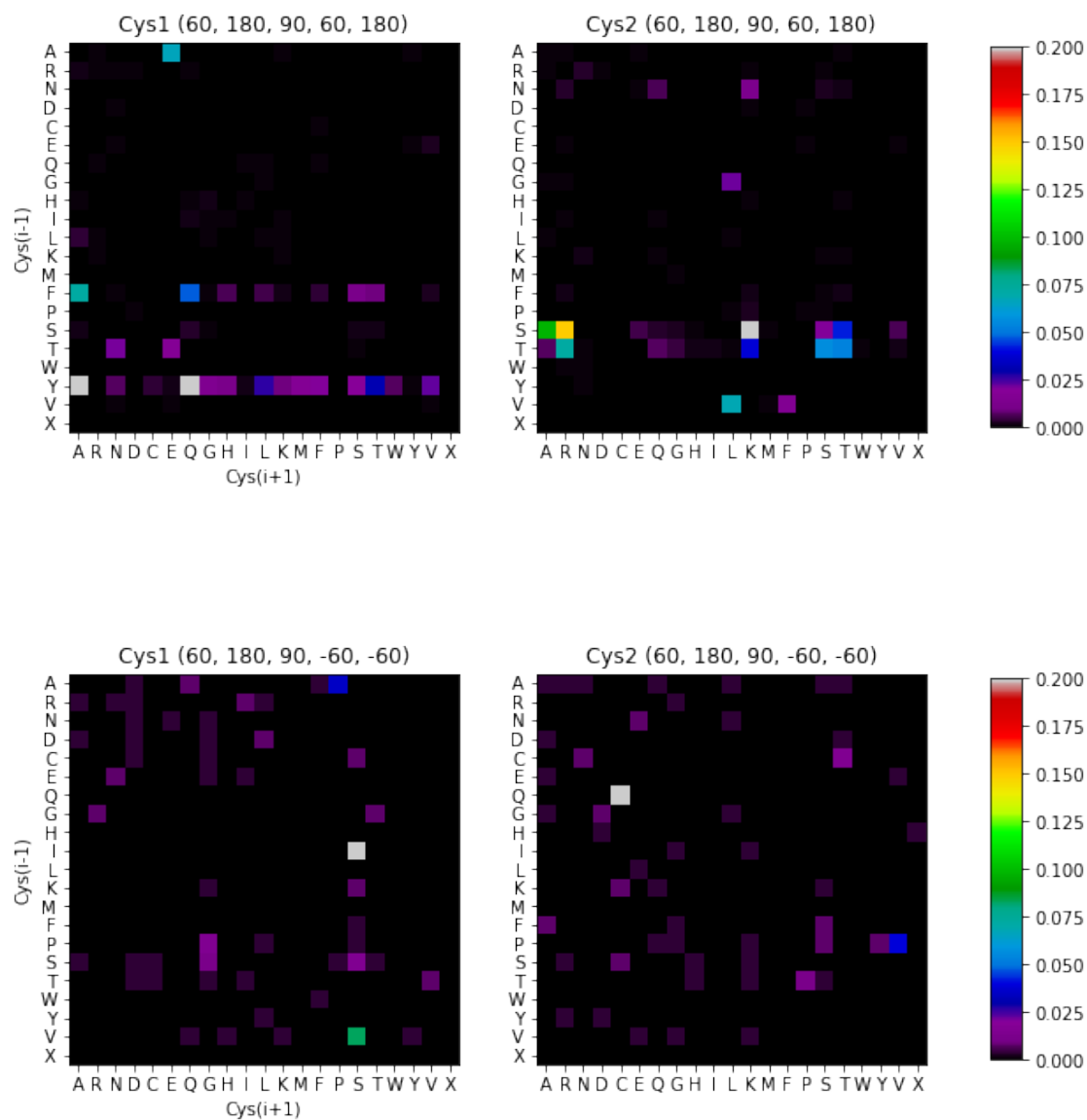


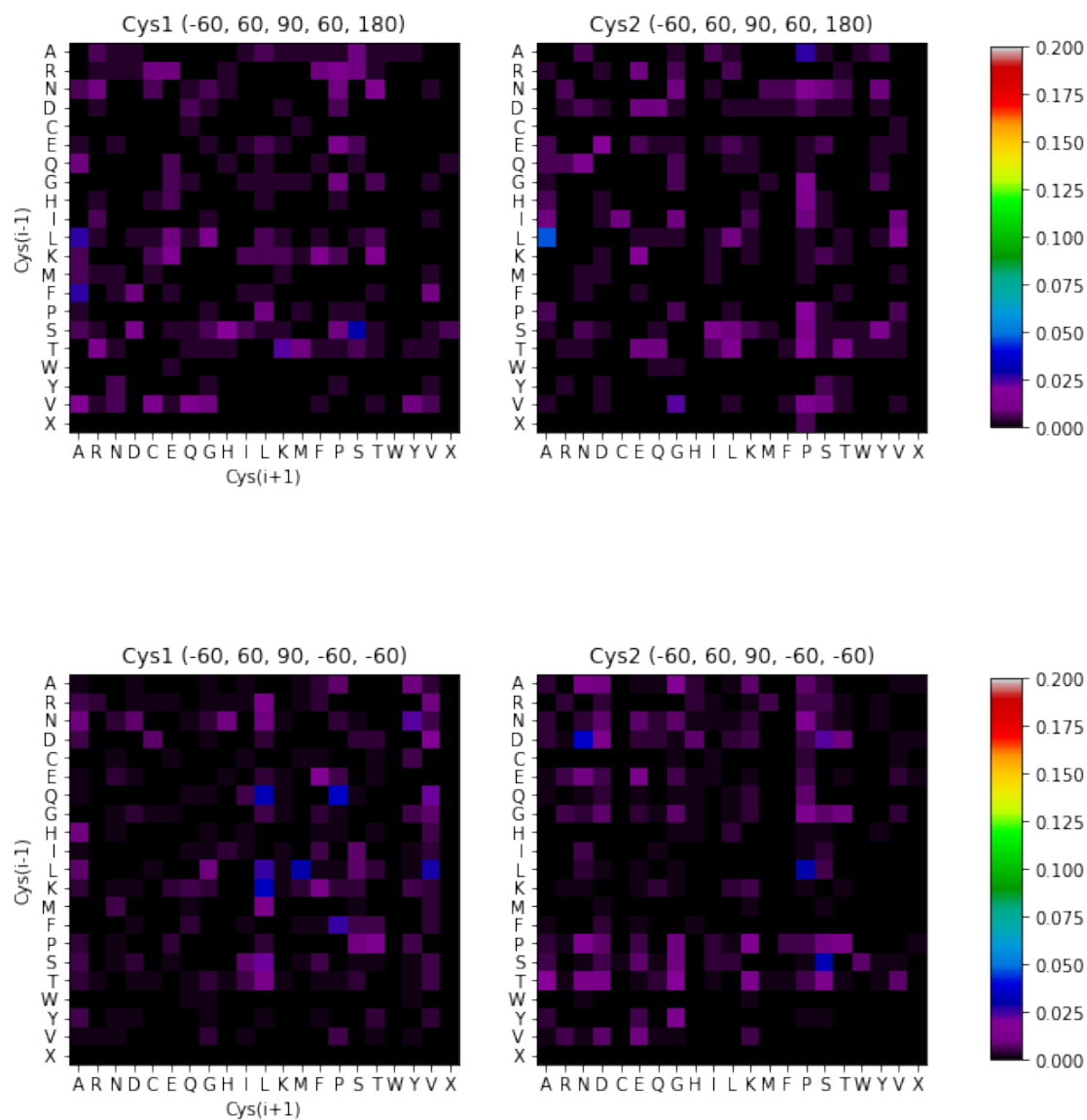


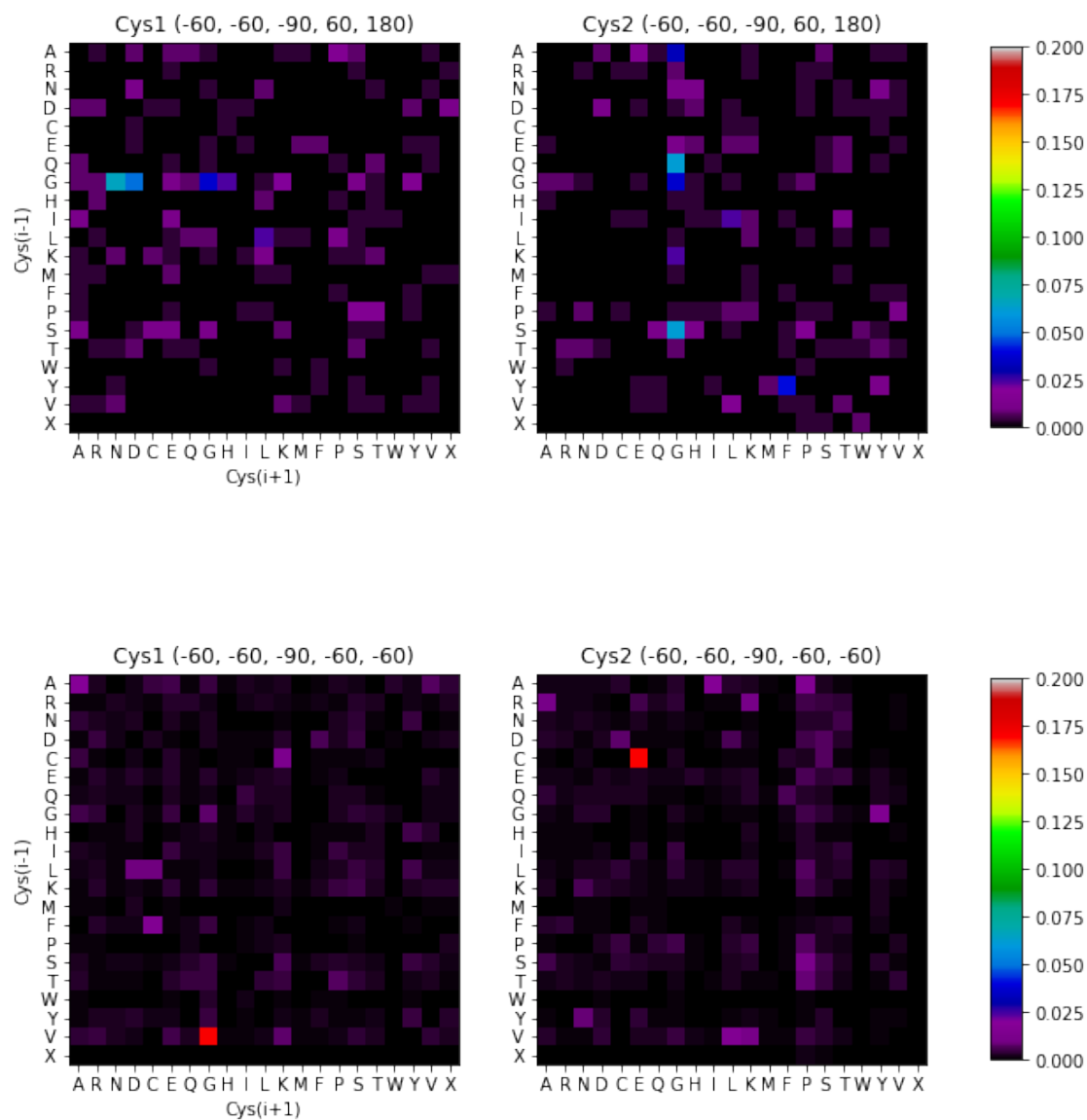


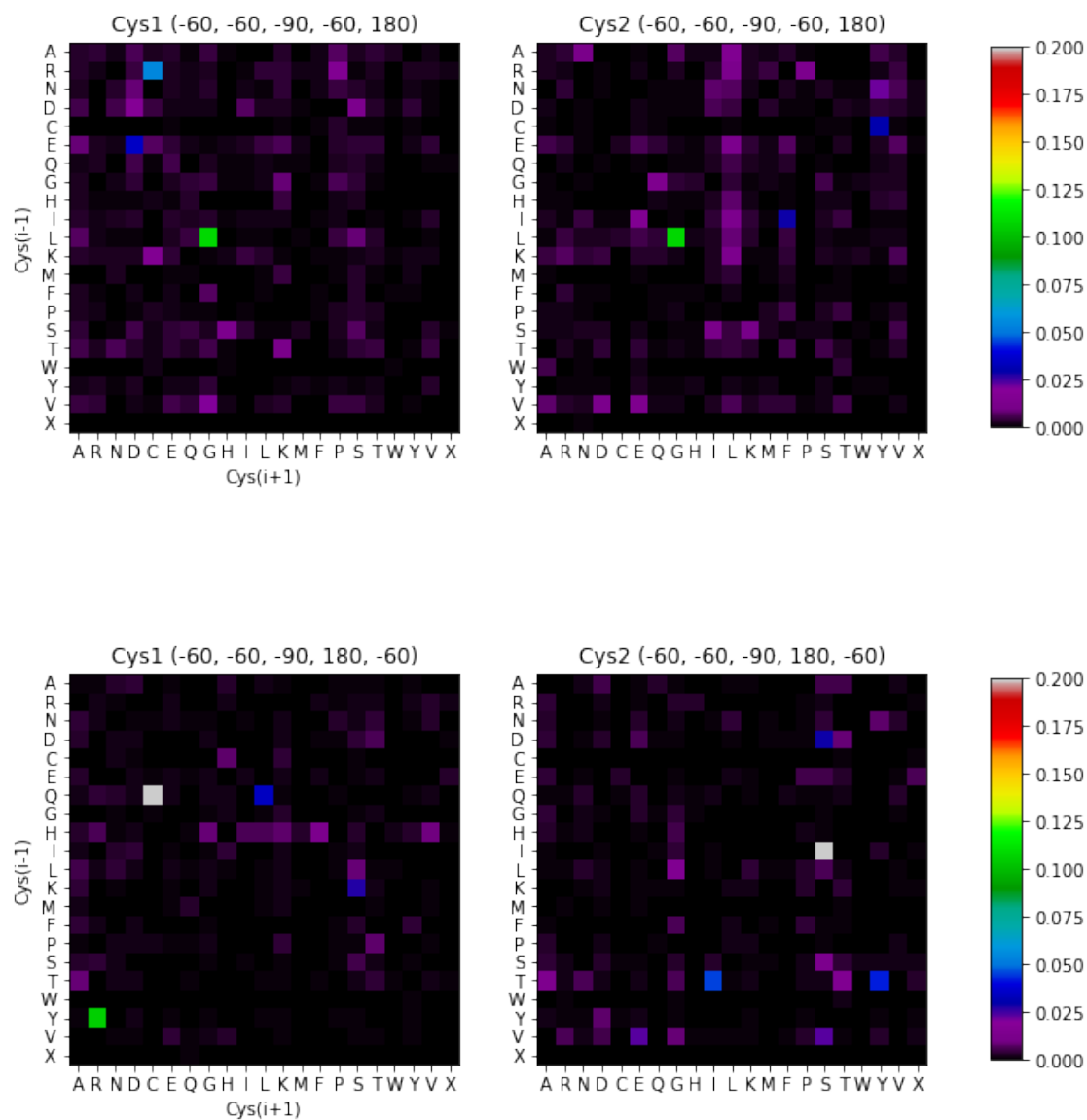


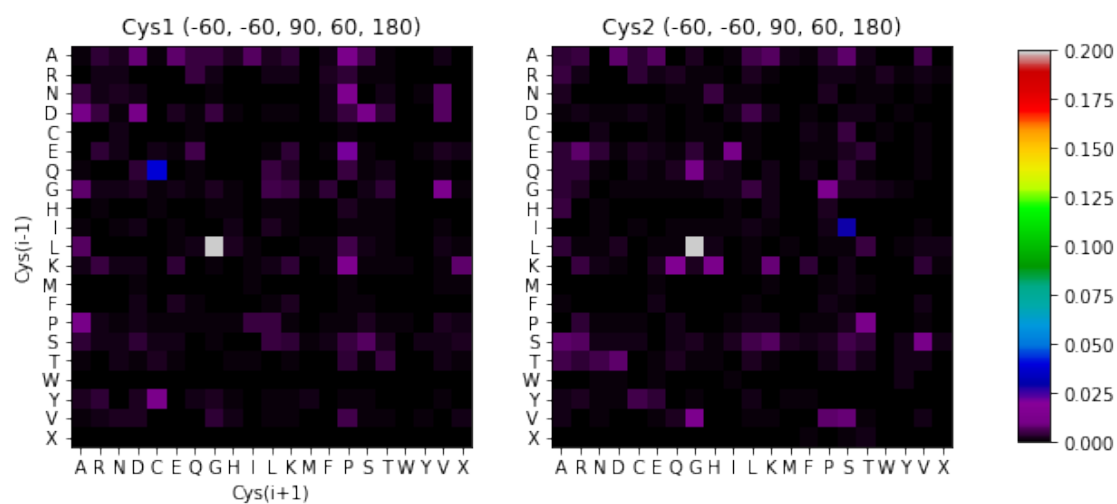
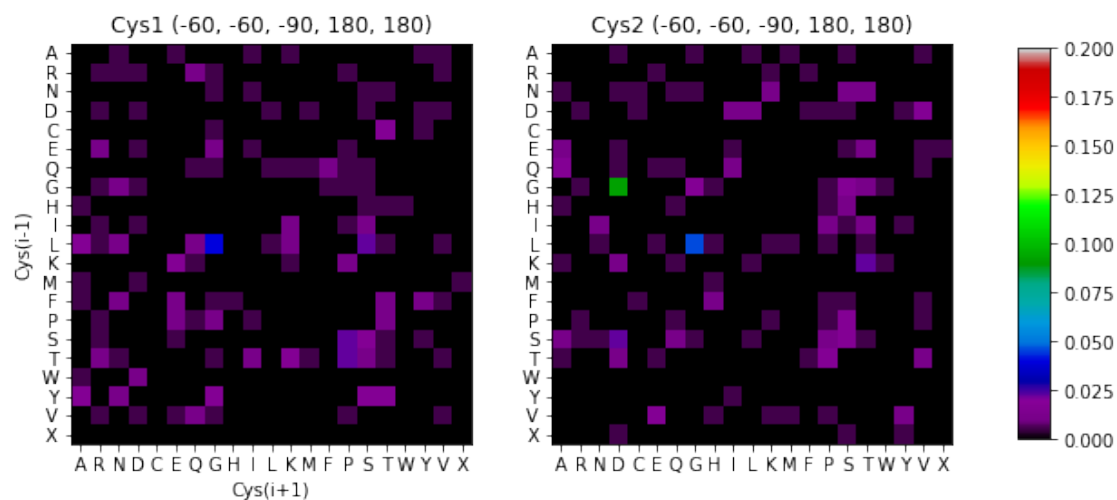


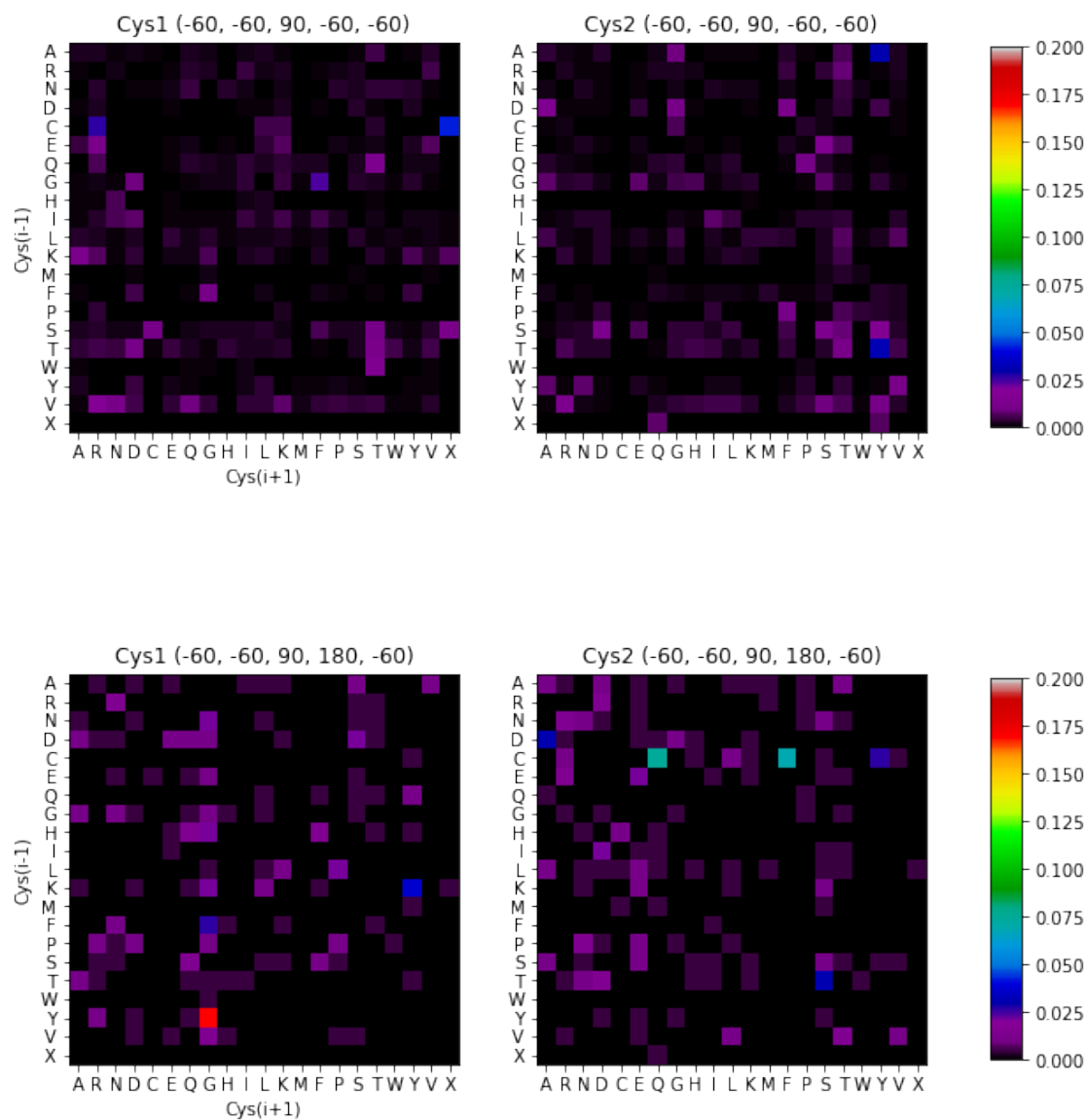


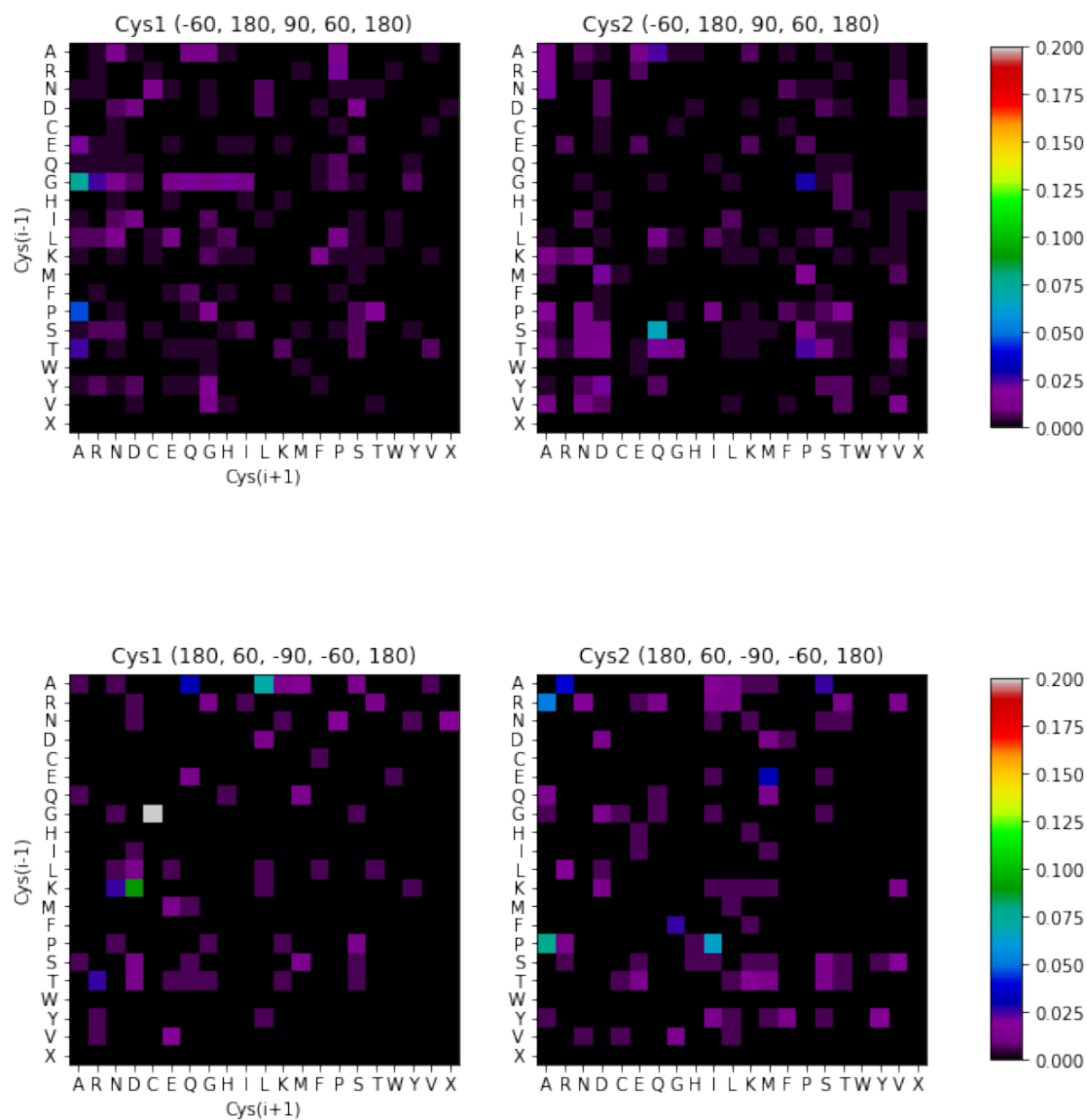


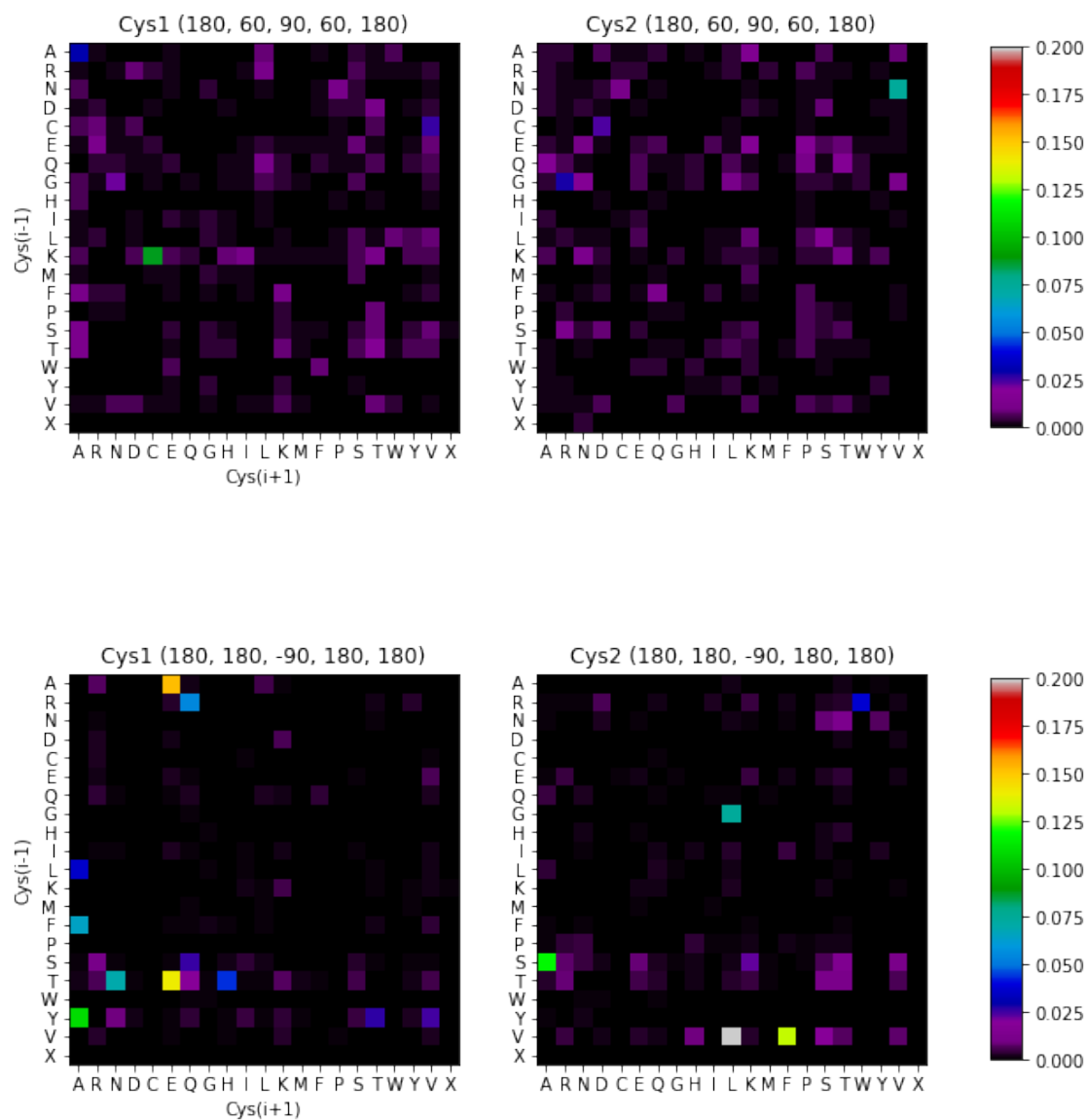




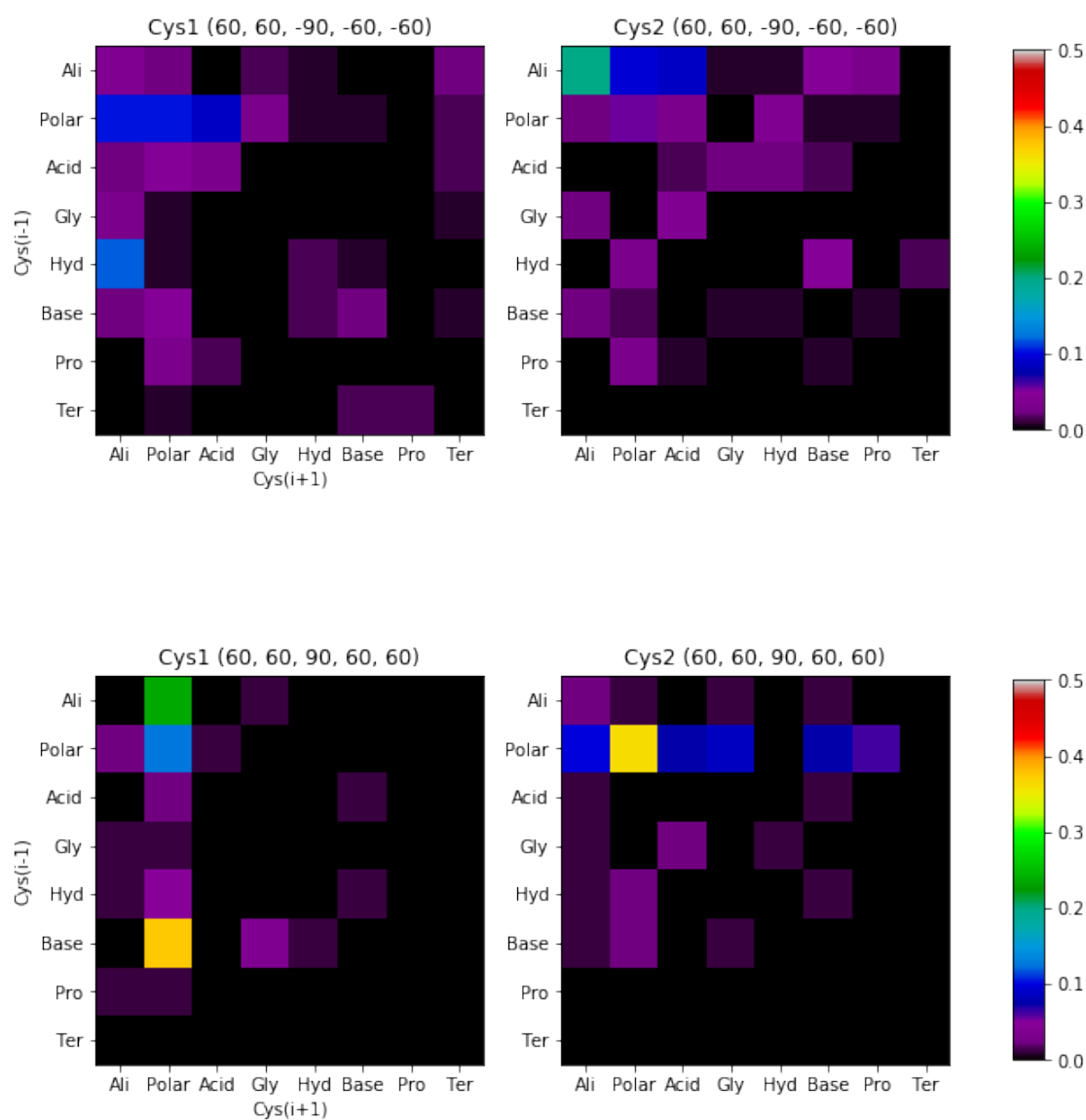


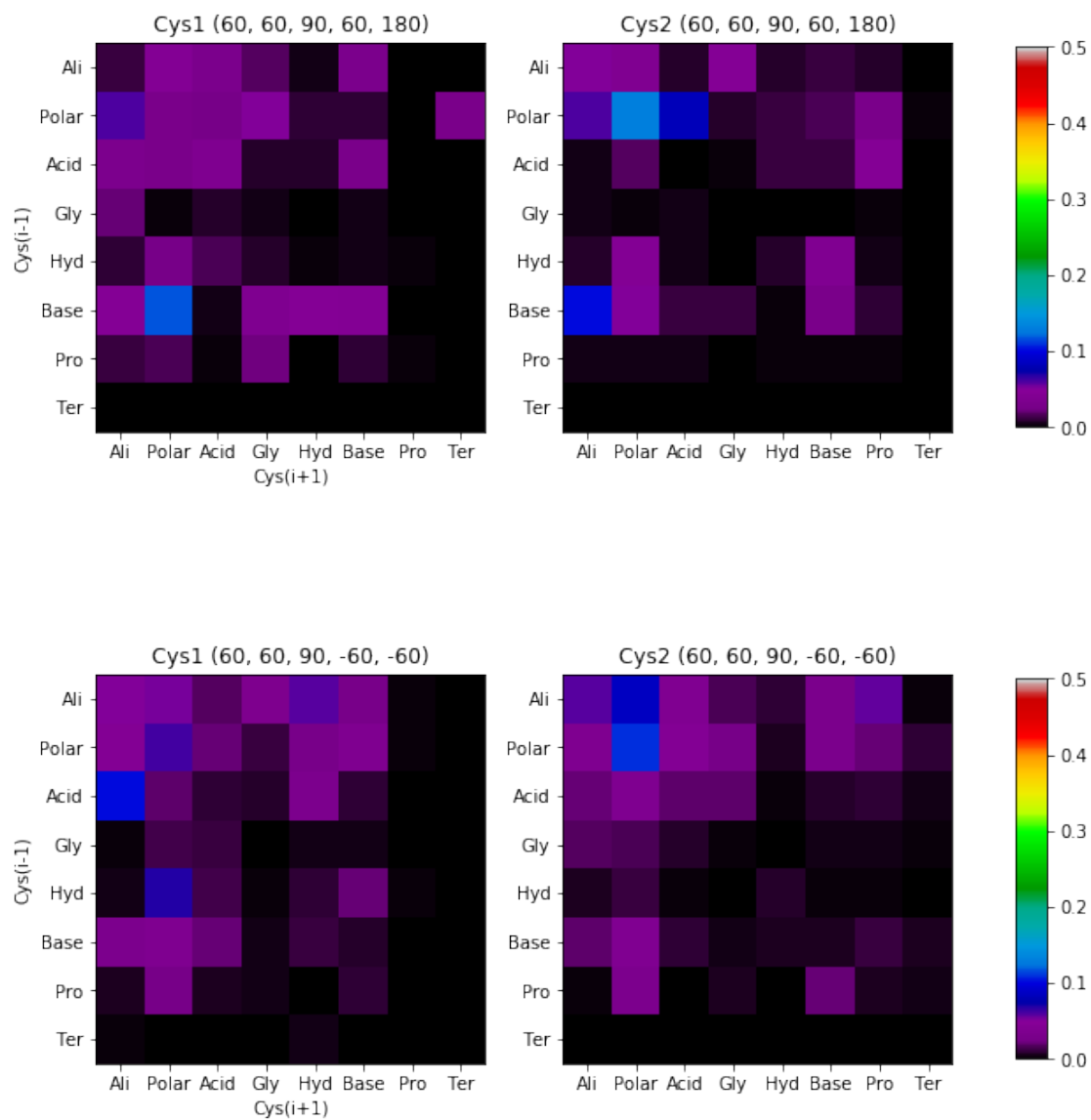


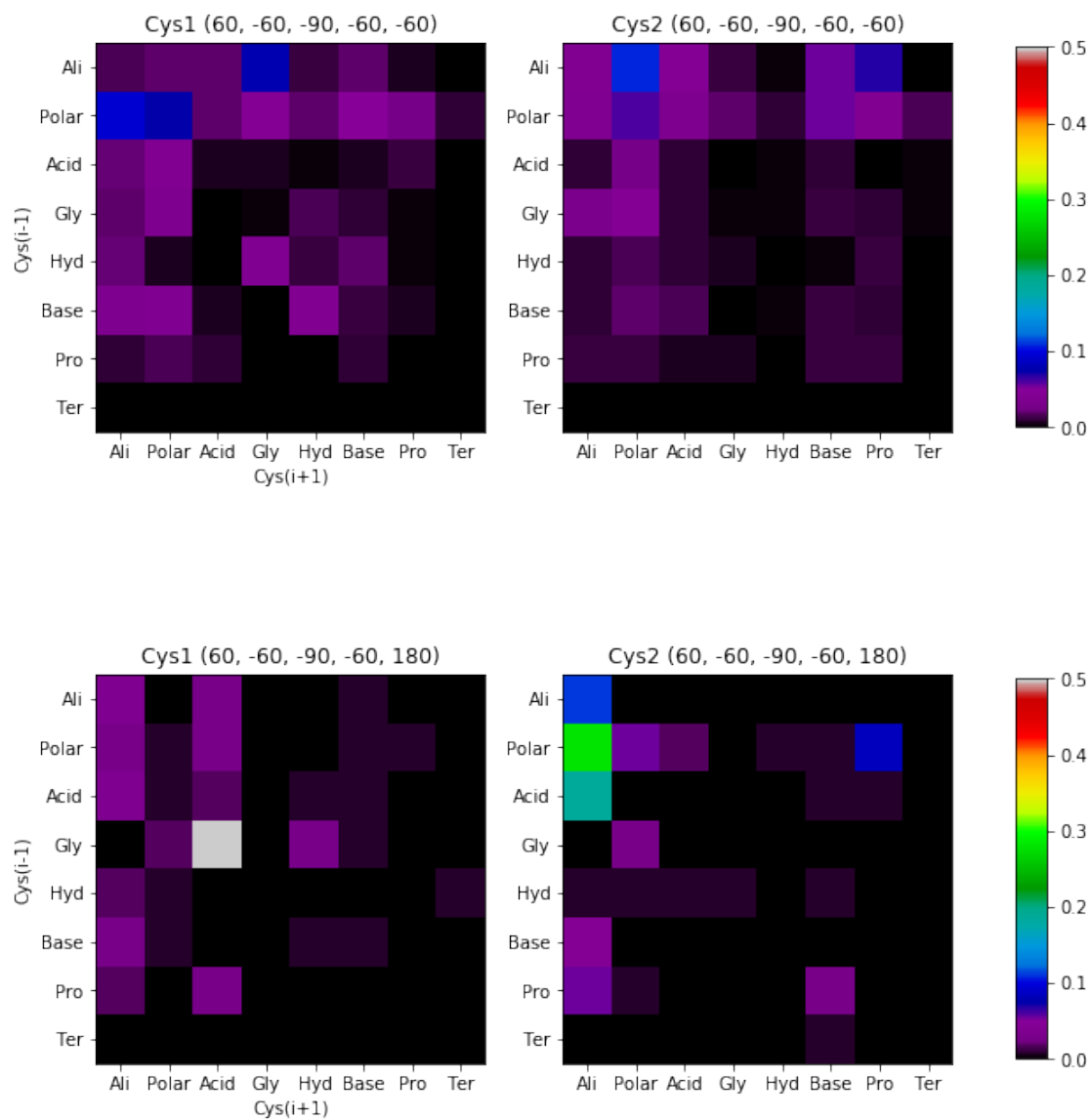


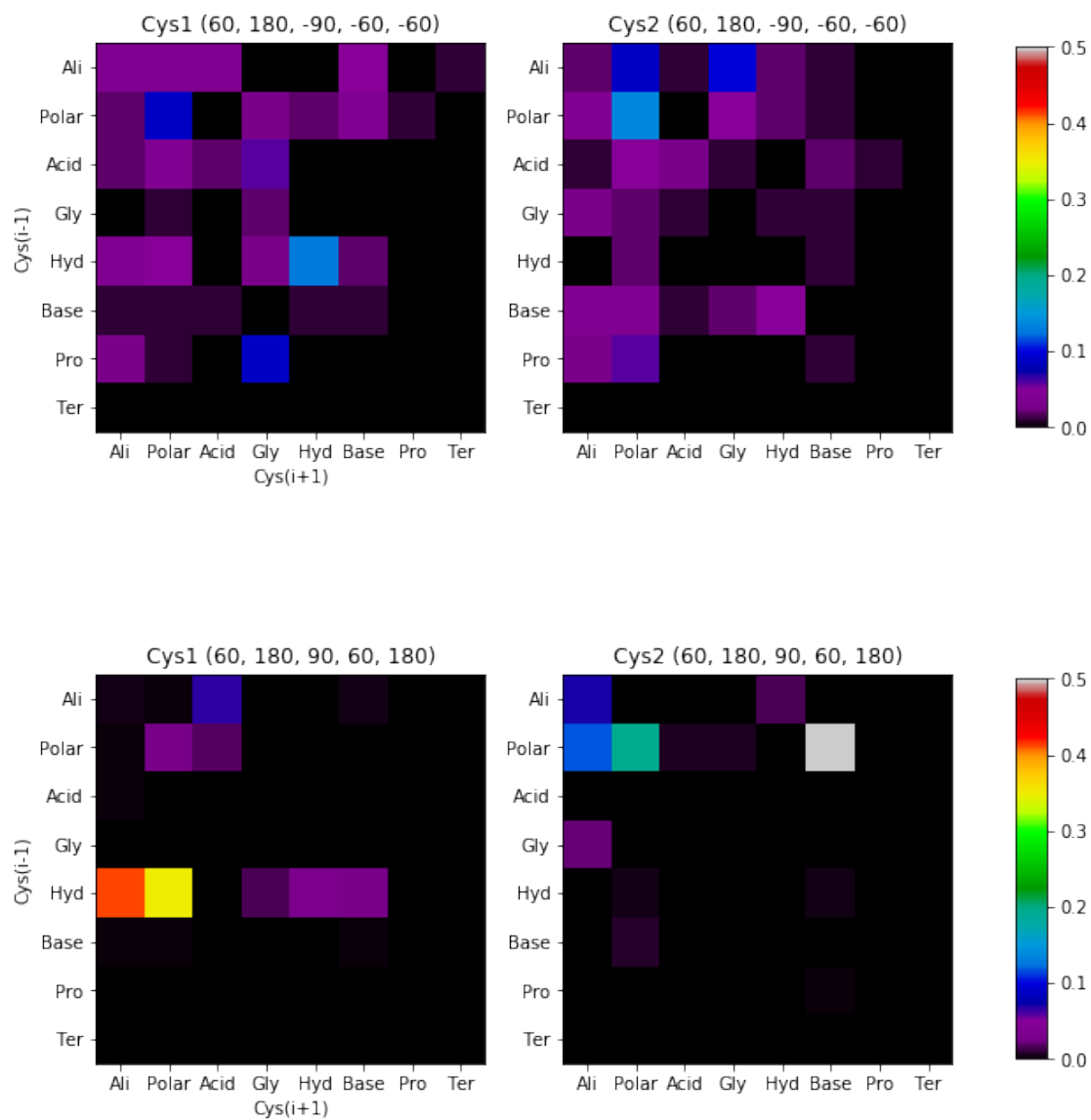


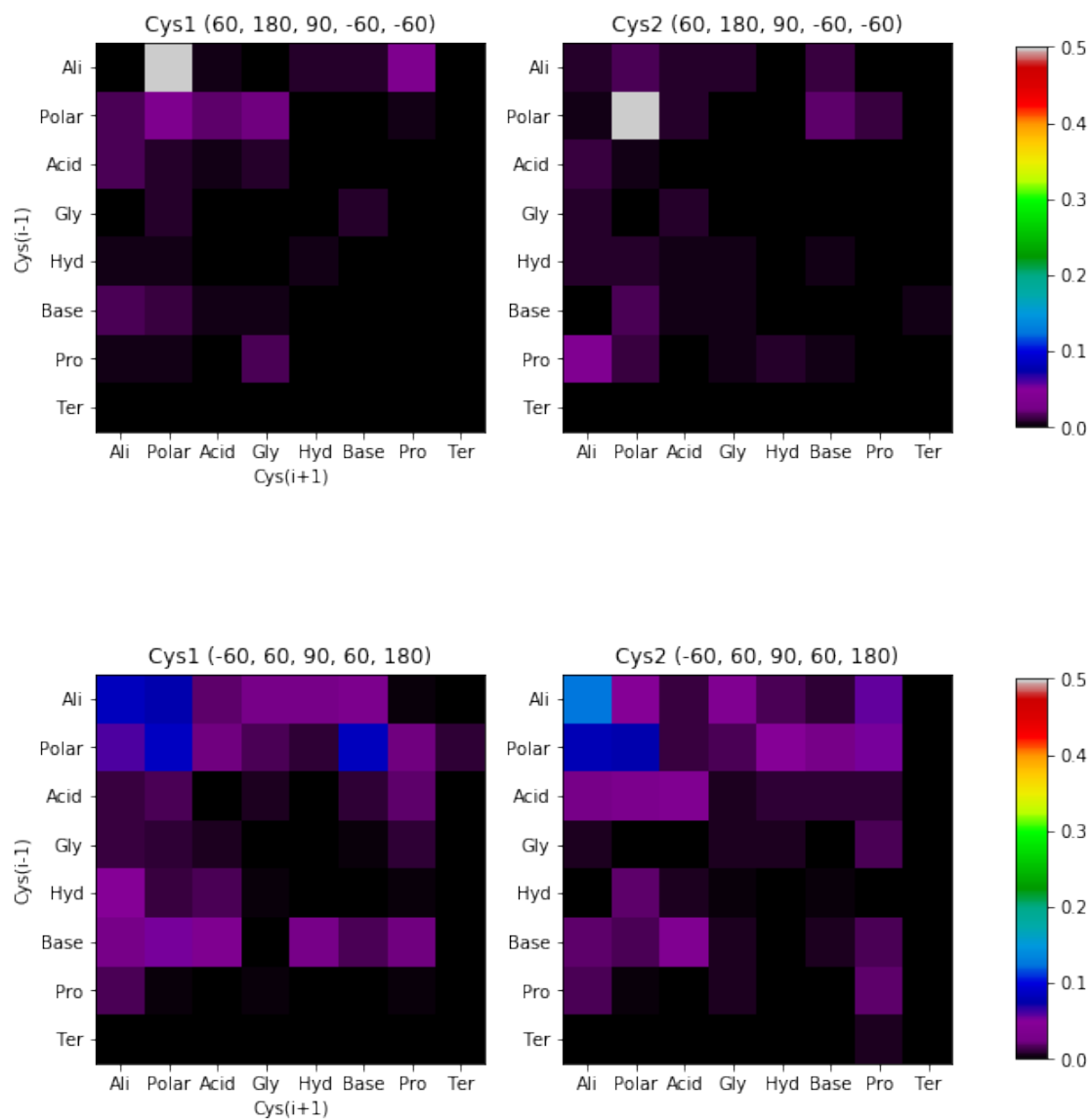
1.6.1 Frequency of Amino Acid by Biochemical Property

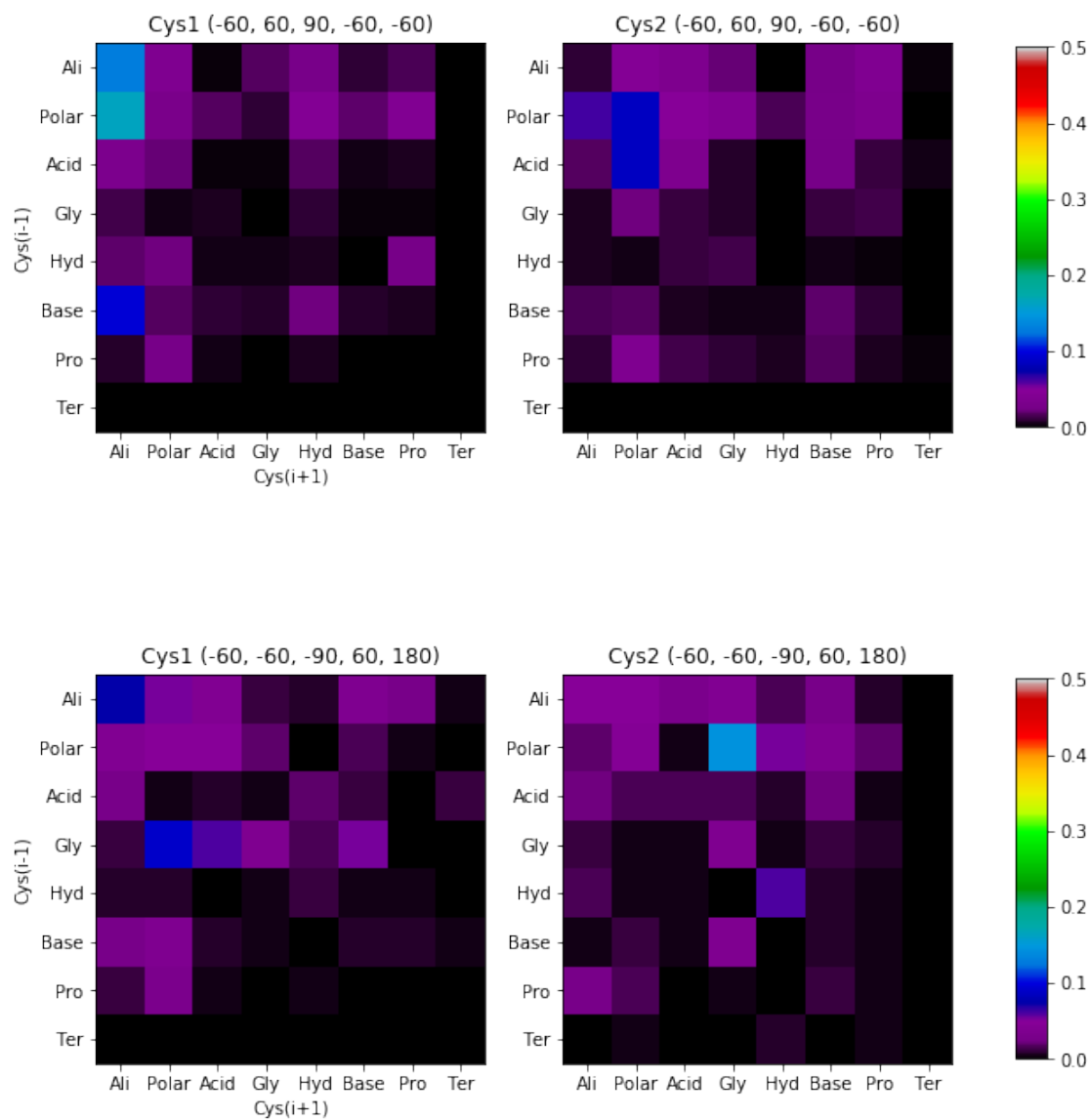


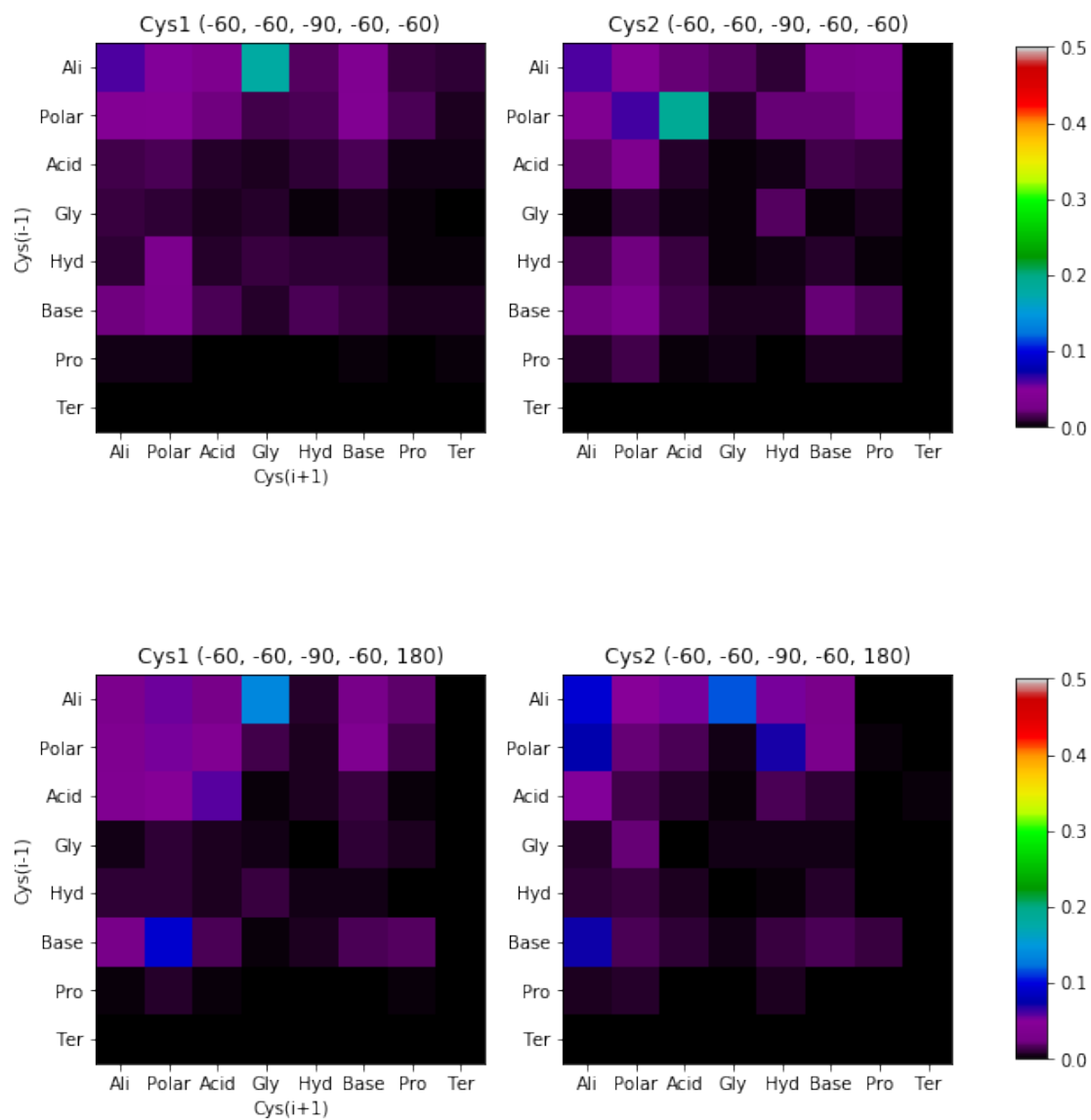


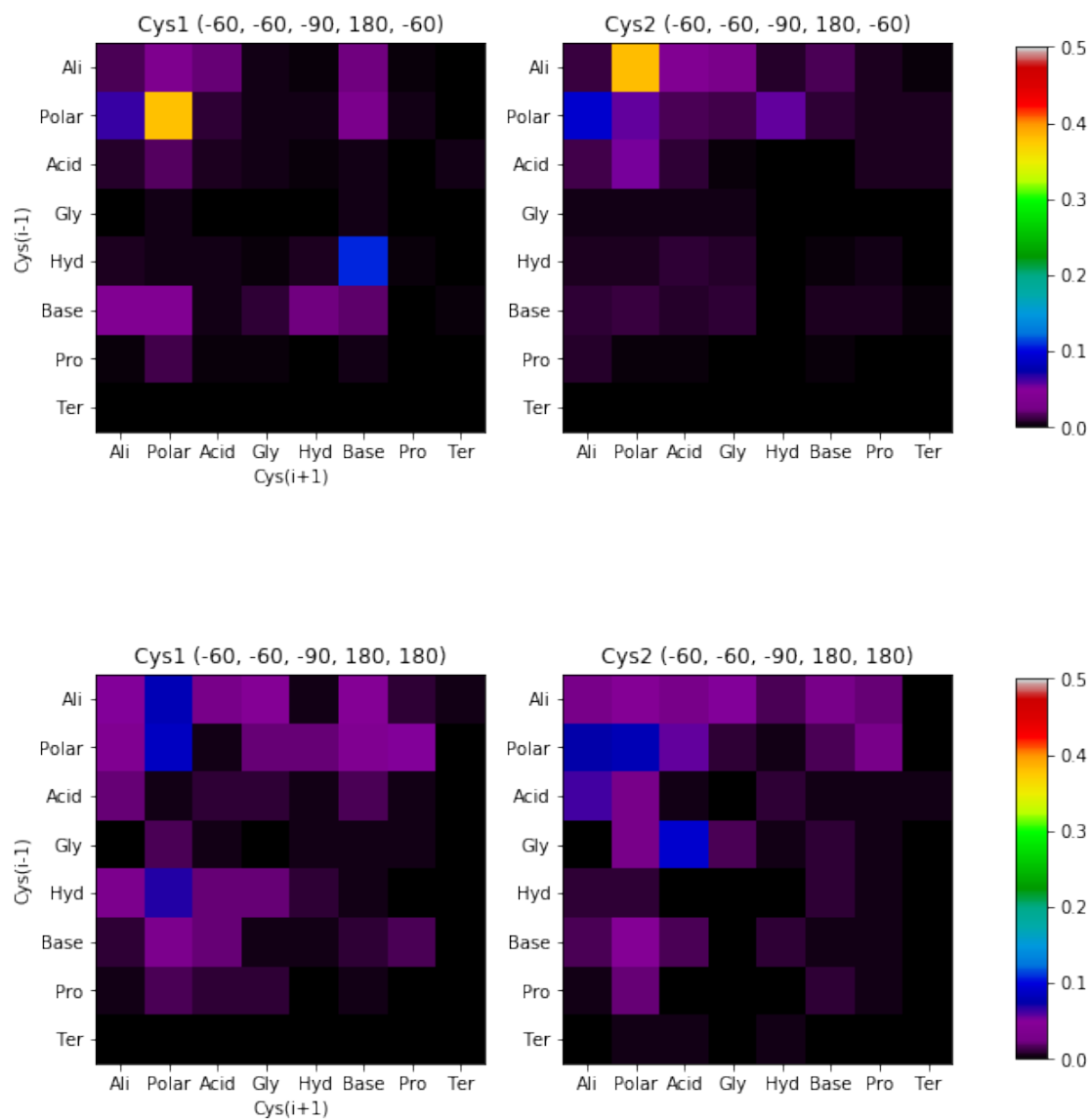


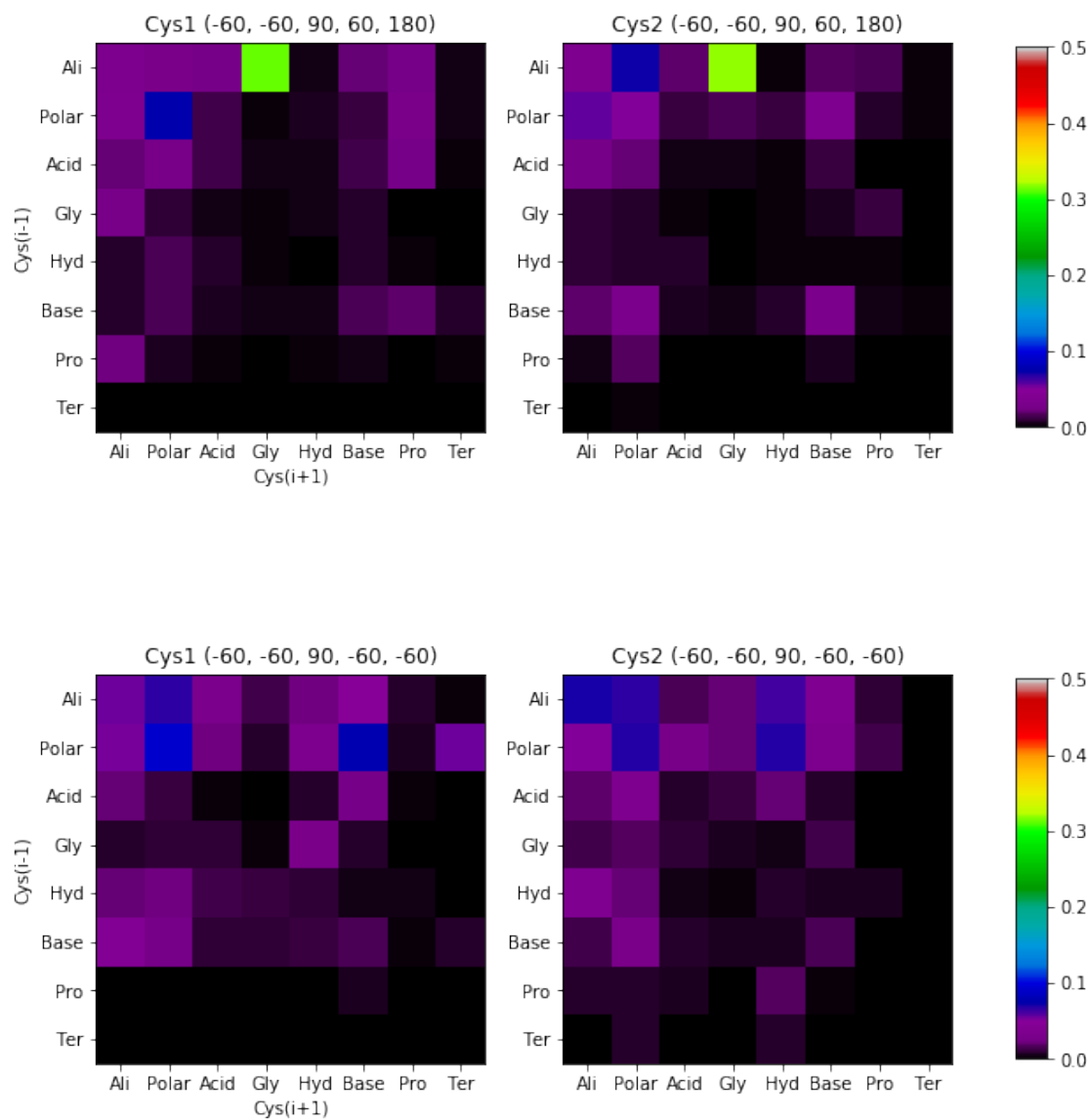


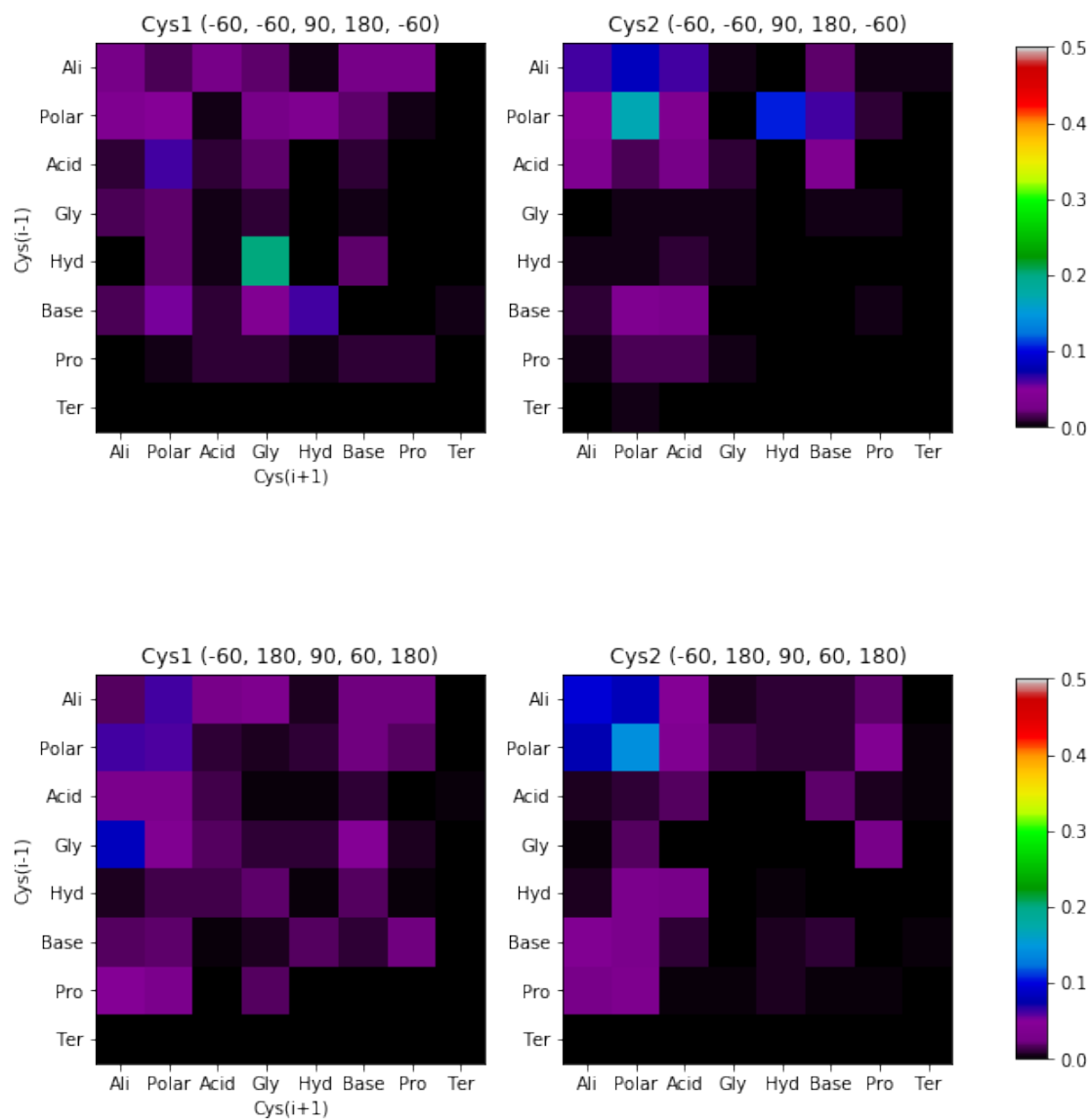


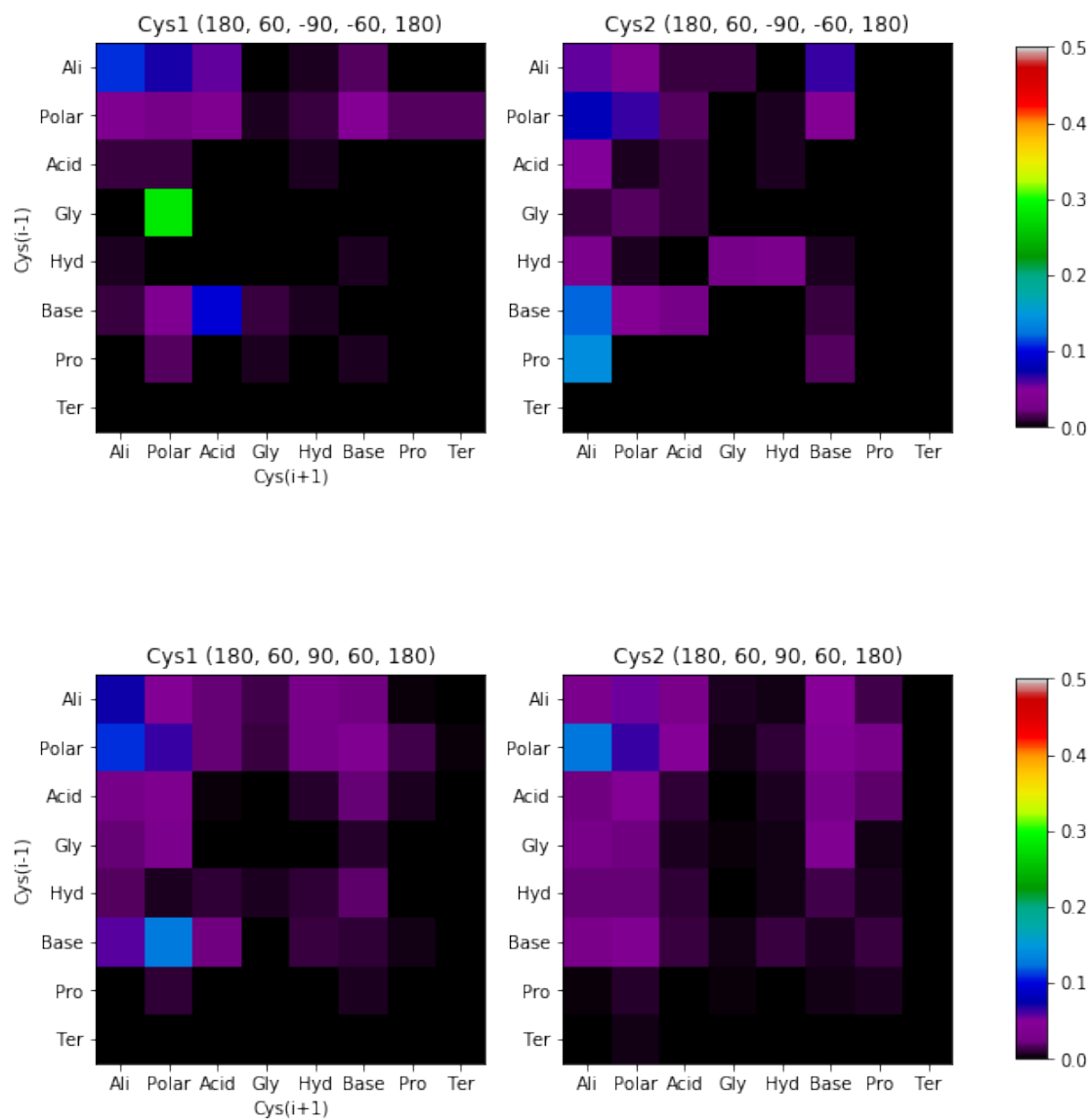


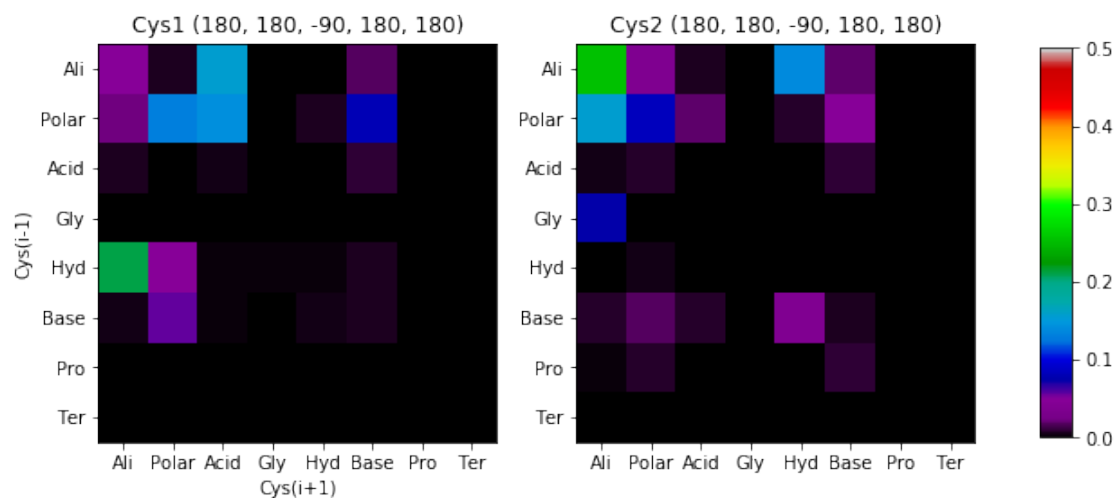




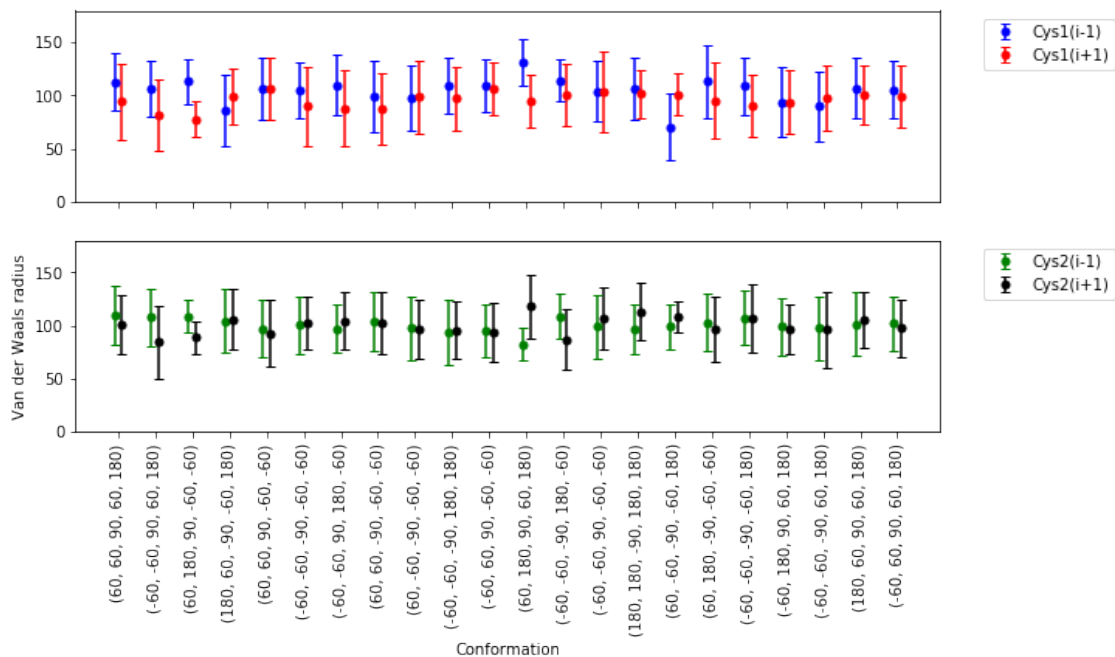








1.6.2 Frequency of Amino Acid by VDW Radi



Out[14]: ()

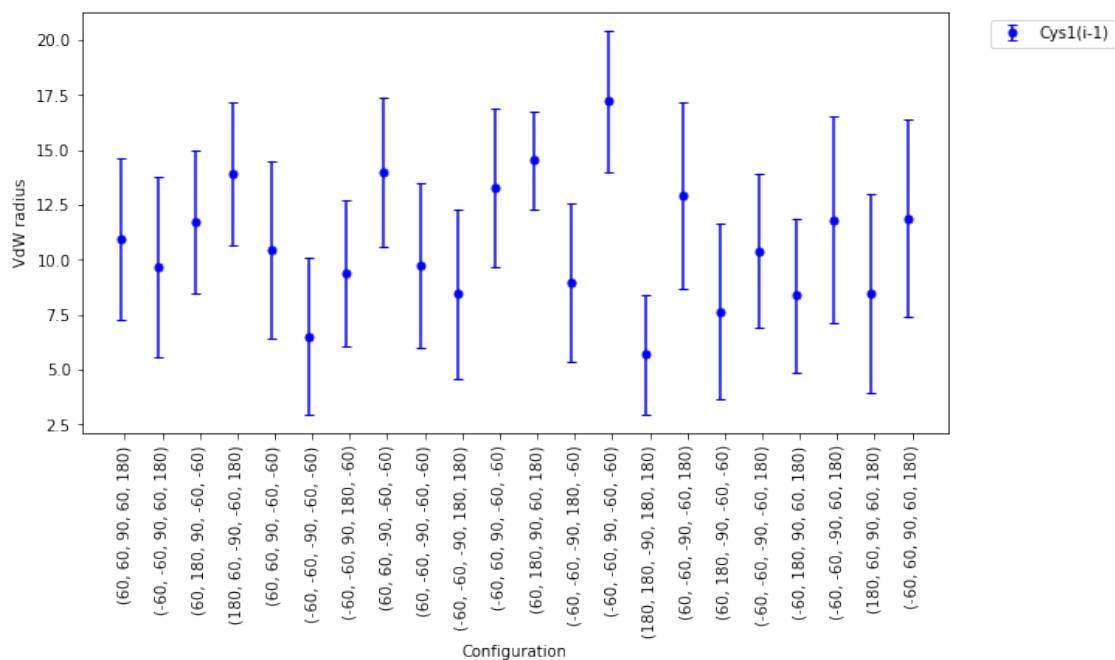
10.94146046003315
 9.673237960921325
 11.707496947589744
 13.951319572960529

```

10.43890017018778
6.524539662137568
9.405272451410811
14.020357311082645
9.729912143632106
8.447279527899408
13.290180158957144
14.546032212688422
8.957282444078768
17.22122359748012
5.687167499307144
12.91536445628846
7.642602000650002
10.409521748311489
8.379706433324529
11.819749037119658
8.473521595474157
11.888714858685315

```

Out[15]: ()



```

6.38426746905e-05
6.28439113194e-05
6.25292284655e-05
6.31073564367e-05

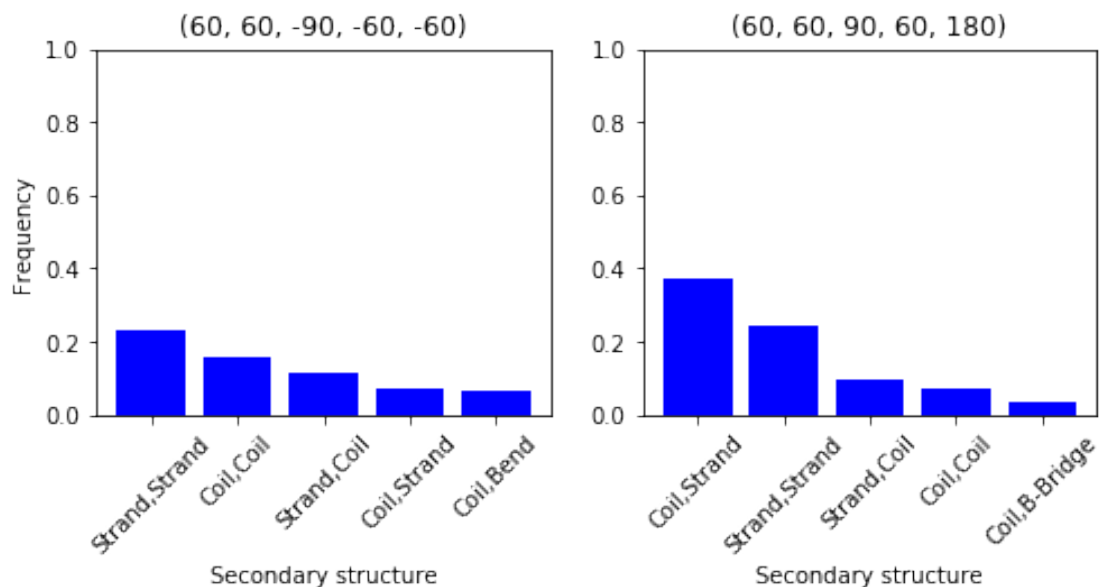
```

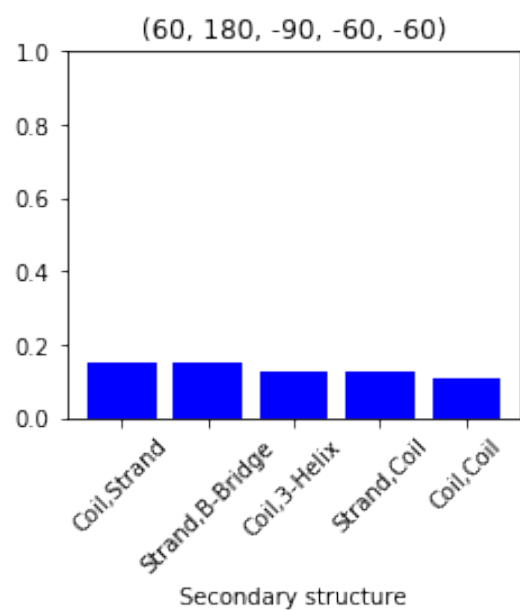
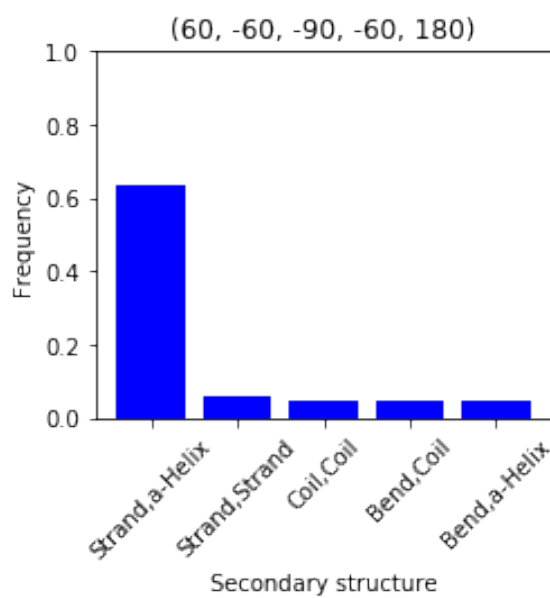
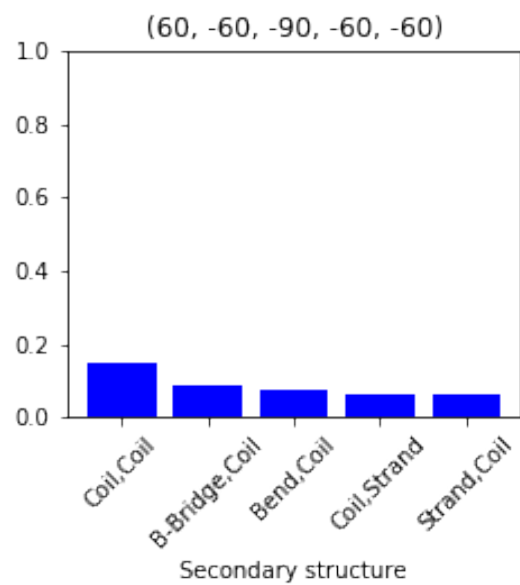
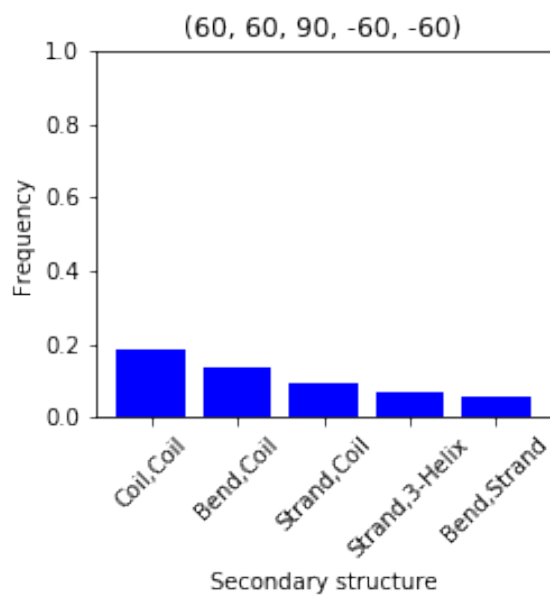
6.39166761082e-05
 6.3941362304e-05
 5.95712018805e-05
 6.33189074595e-05
 6.31439203915e-05
 6.23323494512e-05
 6.3363749433e-05
 5.178978153e-05
 5.96018810105e-05
 6.07755587178e-05
 6.36457561741e-05
 6.043232711e-05
 5.89836968941e-05
 6.35843437867e-05
 6.32089752941e-05
 6.37195293393e-05
 6.250508746e-05
 5.96364139177e-05

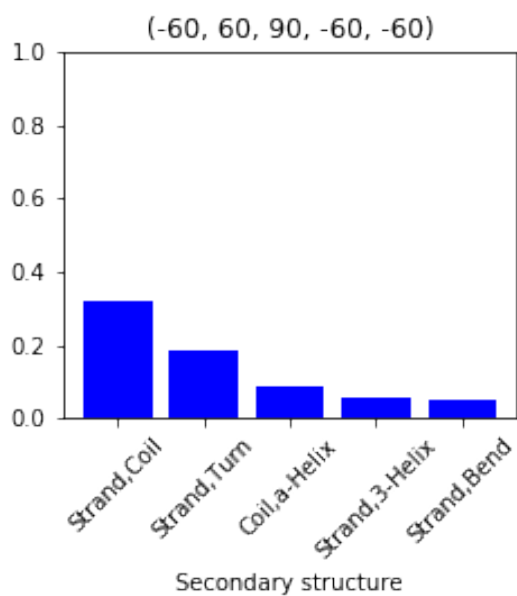
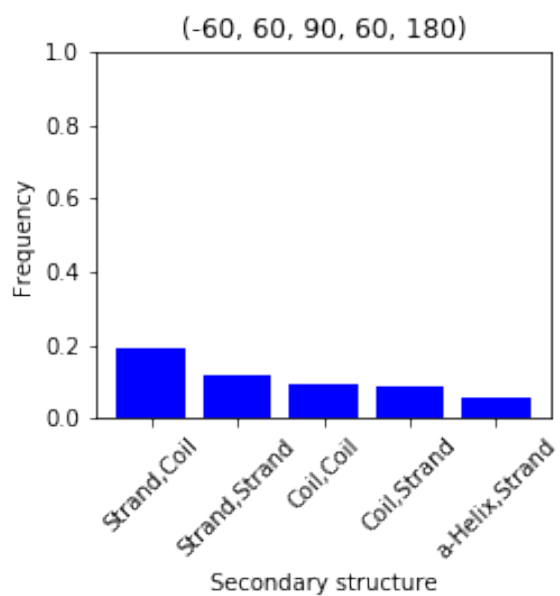
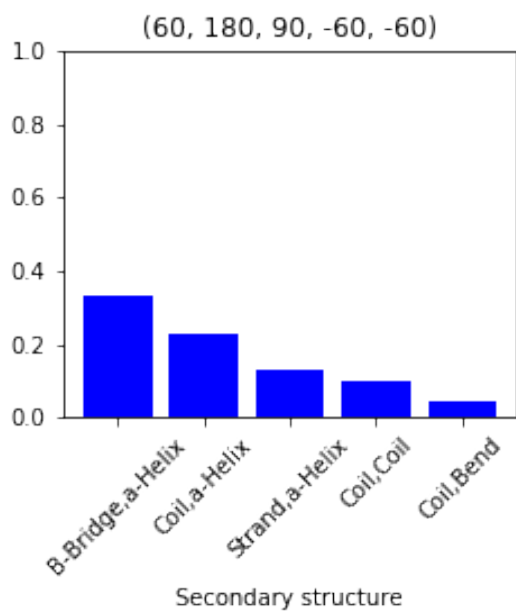
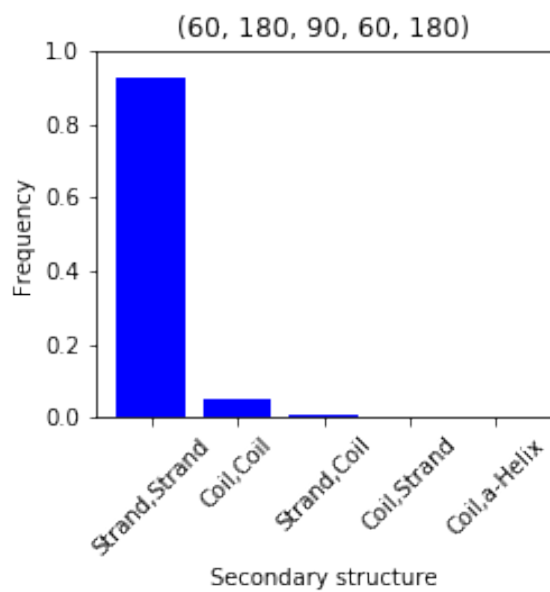
1.7 Secondary Structure

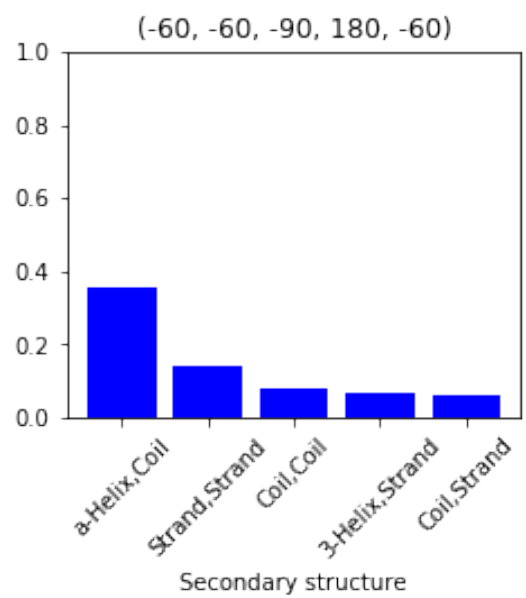
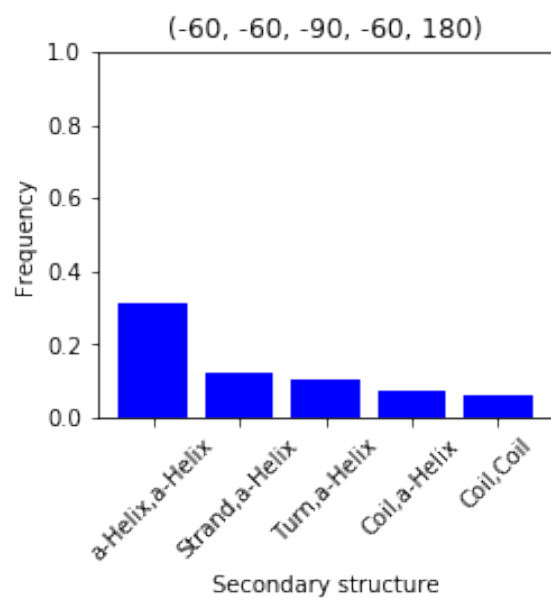
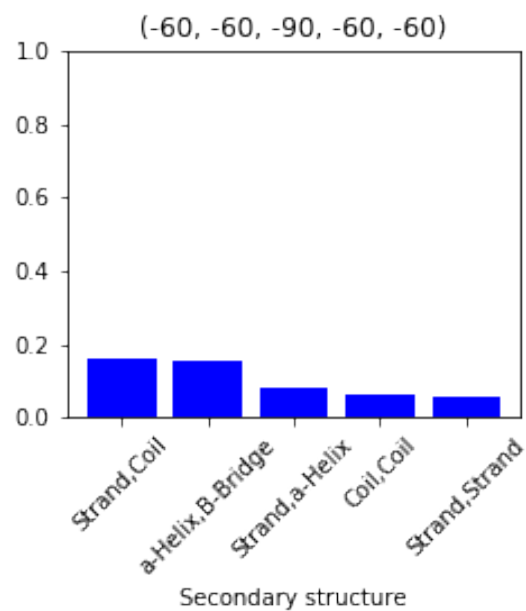
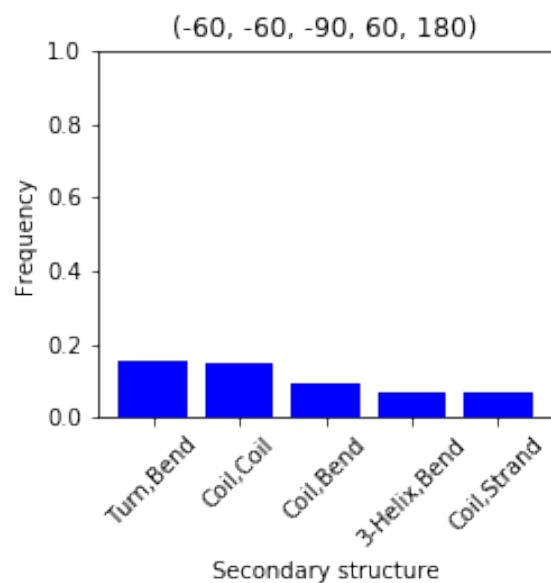
/home/david/coding/lib/python2.7/site-packages/ipykernel_launcher.py:90: SettingWithCopyWarning:
 A value is trying to be set on a copy of a slice from a DataFrame.
 Try using `.loc[row_indexer,col_indexer] = value` instead

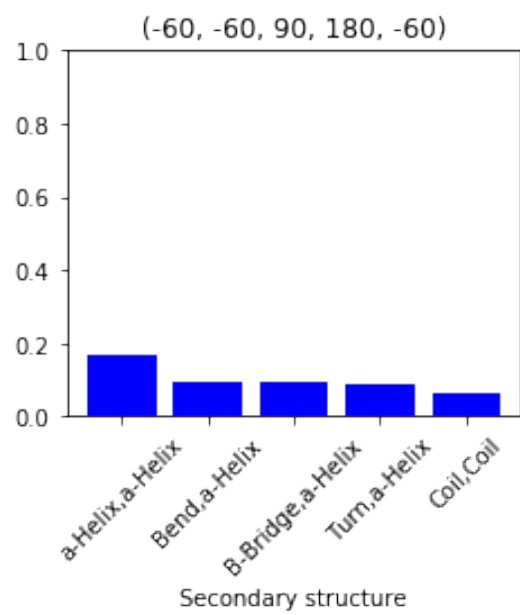
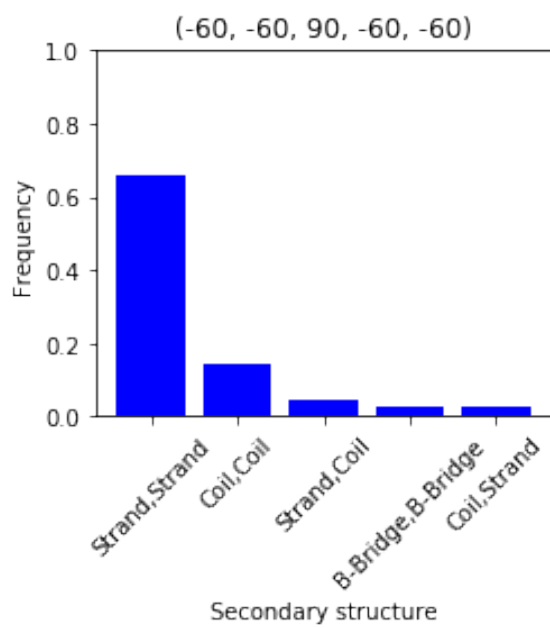
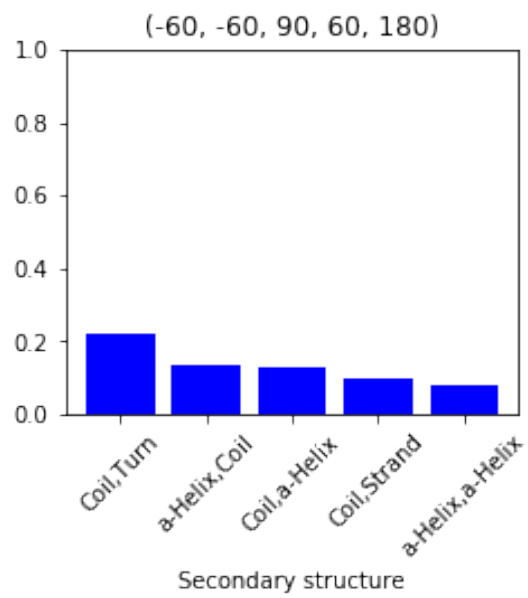
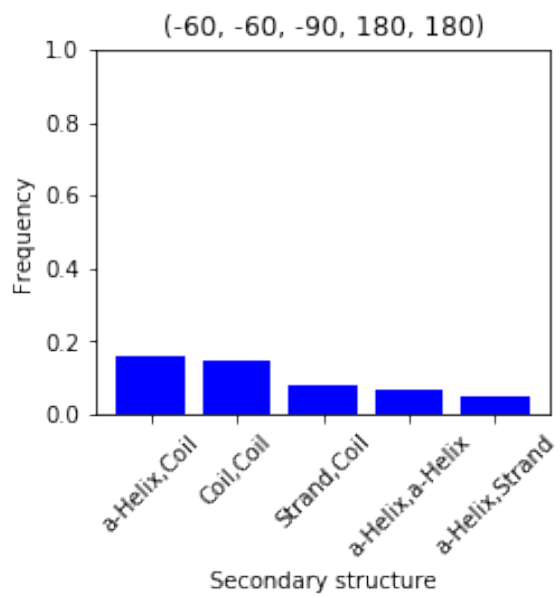
See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#>

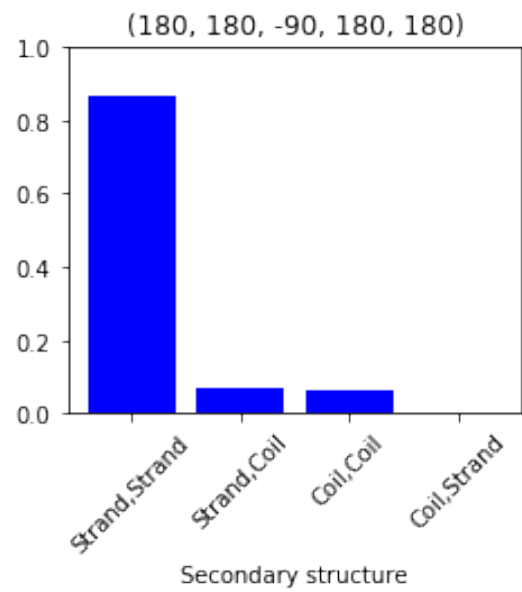
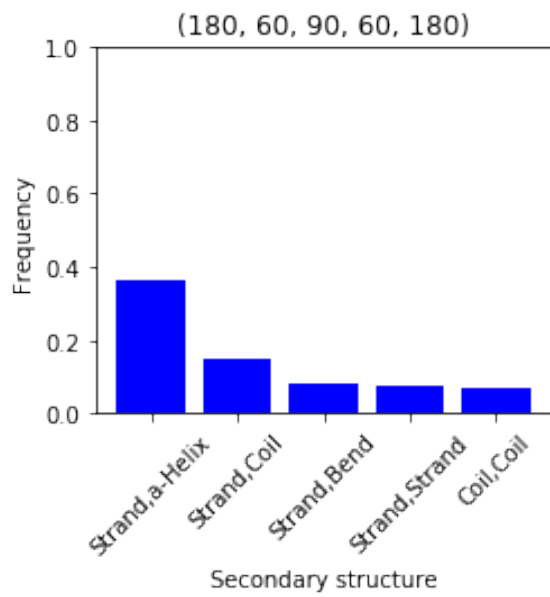
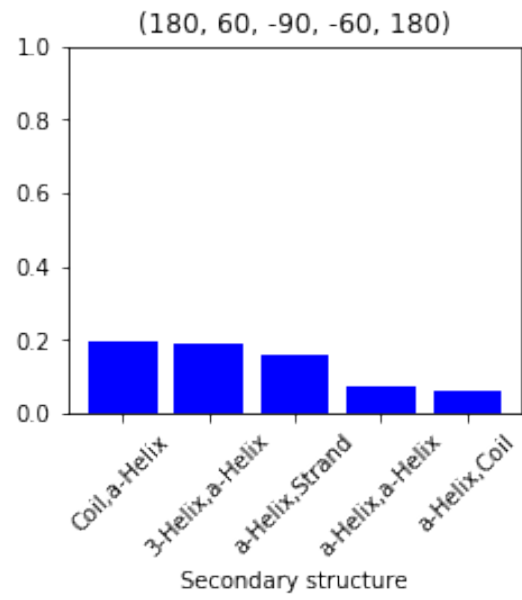
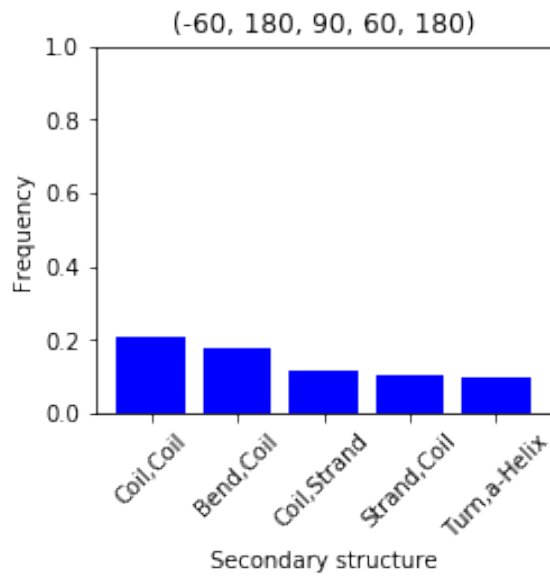


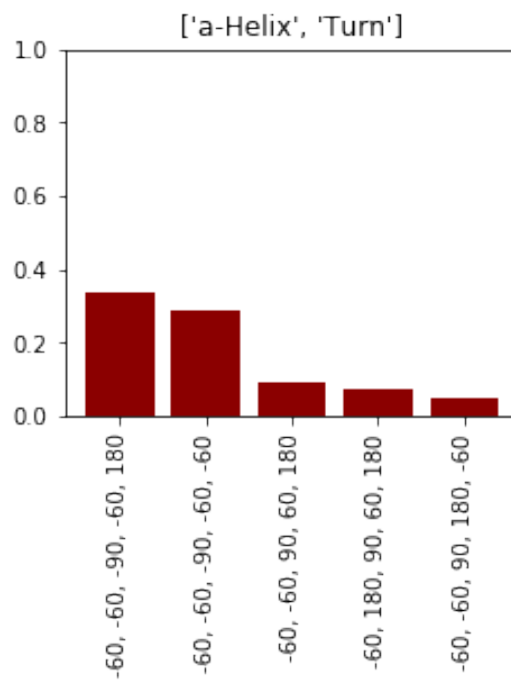
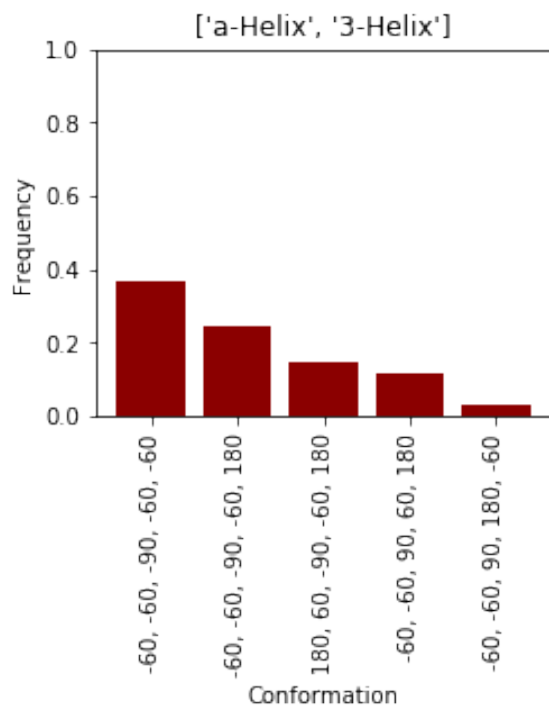
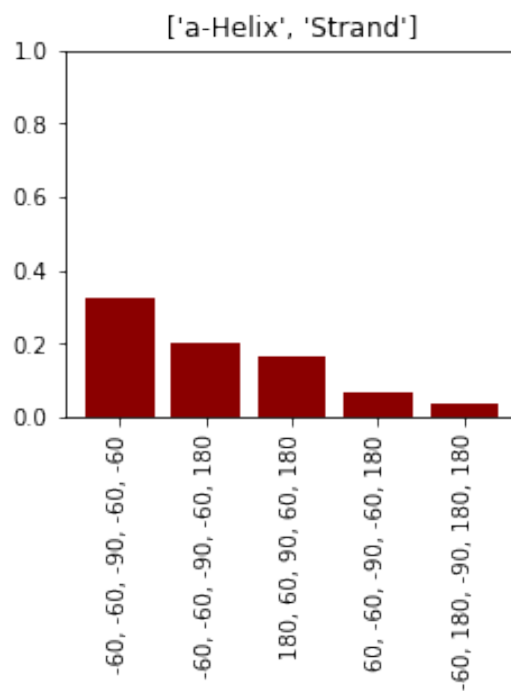
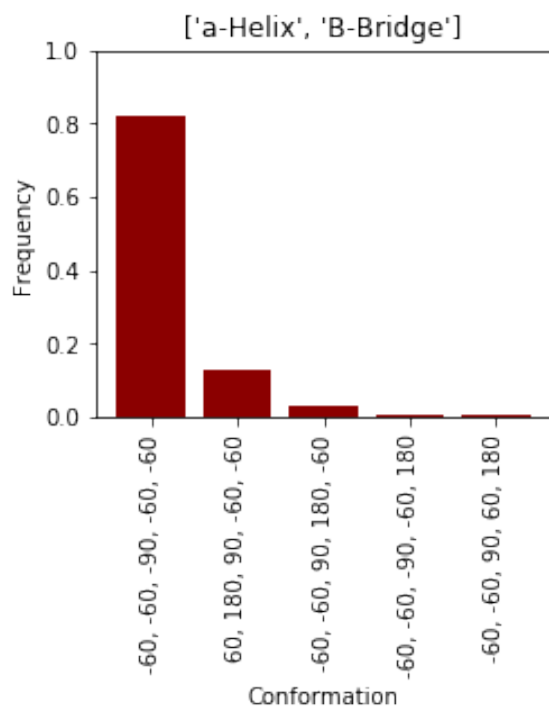












1.8 Dihedral configuraiton by Secondary Structure

1.9 Energy analysis