

# CMPyMOL 2.0

Contact Map plugin for Python Molecular Viewer

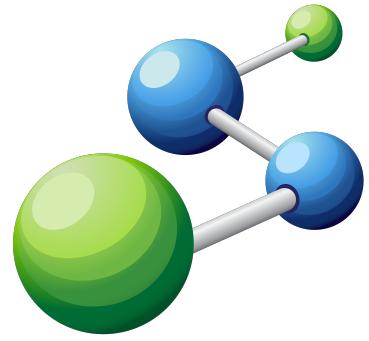
**Dr. Venkatramanan Krishnamani**

Copyright © 2016 Venkatramanan Krishnamani

[HTTPS://GITHUB.COM/EMPTYEWER/CMPyMOL/RELEASES](https://github.com/emptyewer/CMPyMOL/releases)

THIS SOFTWARE IS PROVIDED UNDER “THE MIT LICENSE” (MIT). *See Section 1.3*

*First printing, January 2016*



# Contents

I

## Part One

<b>1</b>	<b>Installation Instructions</b>	<b>7</b>
<b>1.1</b>	<b>Pre-compiled Binaries</b>	<b>7</b>
1.1.1	Download Link .....	7
1.1.2	Prerequisites for pre-compiled binaries .....	7
1.1.3	Mac OS X Compatibility .....	7
1.1.4	Windows Compatibility .....	7
1.1.5	Linux Compatibility .....	7
<b>1.2</b>	<b>Running from Source</b>	<b>8</b>
1.2.1	Pre-requisite Softwares .....	8
1.2.2	Pre-requisite Python Libraries .....	8
<b>1.3</b>	<b>Open Source License</b>	<b>8</b>

II

## Part Two

<b>2</b>	<b>Basic Usage</b>	<b>11</b>
<b>2.1</b>	<b>Launch CMPyMOL</b>	<b>11</b>
<b>2.2</b>	<b>CMPyMOL Interface</b>	<b>11</b>
<b>2.3</b>	<b>Parameters to Calculate Contact Map</b>	<b>12</b>
2.3.1	Load PDB file .....	13
2.3.2	PDB Format .....	14

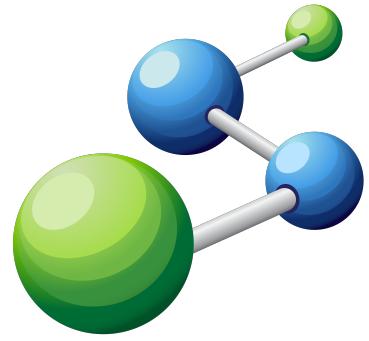
<b>3</b>	<b>Functionality</b>	<b>17</b>
<b>3.1</b>	<b>Contact Map</b>	<b>17</b>
3.1.1	Variance Contact Map	18
<b>3.2</b>	<b>Selections</b>	<b>18</b>
<b>3.3</b>	<b>Overlays</b>	<b>18</b>
3.3.1	Secondary Structure	19
3.3.2	Charged Interactions	19
3.3.3	Hydrophobic Interactions	20
3.3.4	B-factor Overlay	20
3.3.5	User-defined Pairwise Aminoacid Interactions	20
<b>3.4</b>	<b>Plots</b>	<b>20</b>
3.4.1	Pairwise Heatmap	20
3.4.2	Contact Density Histogram	22
<b>Bibliography</b>		<b>25</b>
<b>Books</b>		<b>25</b>
<b>Articles</b>		<b>25</b>
<b>Index</b>		<b>27</b>



# Part One

1	Installation Instructions .....	7
1.1	Pre-compiled Binaries	
1.2	Running from Source	
1.3	Open Source License	





## 1. Installation Instructions

### 1.1 Pre-compiled Binaries

#### 1.1.1 Download Link

Platform-specific compiled binaries (*Mac OS X and Windows*) of **CMPyMOL** can be downloaded from the below URL.

<https://github.com/emptyewer/CMPyMOL/releases>

#### 1.1.2 Prerequisites for pre-compiled binaries

---

# 1. Download and Install PyMOL (according to your platform)  
<https://www.pymol.org/>

---

#### 1.1.3 Mac OS X Compatibility

Mac OS X (10.10+) Yosemite and above

#### 1.1.4 Windows Compatibility

64-bit or 32 bit Windows 7 and above.

Note that CMPyMOL itself is a 32-bit software.

#### 1.1.5 Linux Compatibility

64-bit Linux Binaries not available for this release. The next release will include the linux binaries.



Other machine specific binaries will be provided upon request.

## 1.2 Running from Source

CMPyMOL source can be downloaded from the following URL.

---

<https://github.com/emptyewer/CMPyMOL>

---

The following libraries are necessary to run CMPyMOL from source.

-  All necessary libraries and software dependencies (except PyMOL) are included in the compiled binaries, this requirement is only for running CMPyMOL from source

### 1.2.1 Pre-requisite Softwares

1. Python 2.7
2. PyMOL in system \$PATH (for all platforms)
3. STRIDE secondary structure prediction software properly configured to be in \$PATH

---

<http://webclu.bio.wzw.tum.de/stride/>

---

-  Stride for Mac OSX and Windows is included in the precompiled binary.

### 1.2.2 Pre-requisite Python Libraries

1. PyQt4
2. matplotlib
3. xmlrpclib
4. numpy

## 1.3 Open Source License

---

The MIT License (MIT)

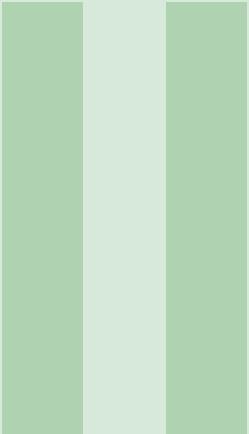
Copyright (c) 2016 Venkatramanan Krishnamani

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

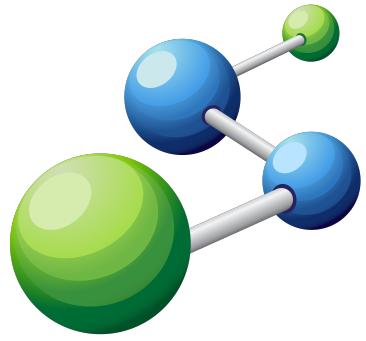
---



# Part Two

<b>2</b>	<b>Basic Usage .....</b>	<b>11</b>
2.1	Launch CMPyMOL	
2.2	CMPyMOL Interface	
2.3	Parameters to Calculate Contact Map	





## 2. Basic Usage

Currently both single frame and multi-frame PDB files are supported for analysis by CMPyMOL.

### 2.1 Launch CMPyMOL

CMPyMOL can be launched after installation by clicking on the executable with an icon (appropriately for Windows and Mac OS X platform). To launch from source simply type `python CMPyMOL_2.0.py` in machine appropriate command line shell.

- R** If PyMOL executable is not automatically located by CMPyMOL, a pop-up will request the location of PyMOL executable. Select either the location of `MacPyMOL.app` (for Mac OS X) or `pymol.exe` (for Windows). The user has to select the location of PyMOL executable only once, after which the path is remembered. In subsequent launching of CMPyMOL if PyMOL is not automatically launched, check the troubleshooting section **??**.

### 2.2 CMPyMOL Interface

After launching CMPyMOL, the user will be presented with a window similar to Figure 2.1. This interface provides all the functionality of the software with no additional pop-ups that distract the user.

The overall organization of the interactive elements in this interface is such that all overlay's that are displayed on top of contact map are on the right side. The parameters for calculating the contact map and other information are provided on the left.

All the maps (contact map, heat map of pairwise amino acid contacts and contact density histogram) are plotted as tabs in the central widget. These tabs will be added to the CMPyMOL interface when a PDB file is loaded. The functionatily of each interactive element will be discussed further in the following sections.

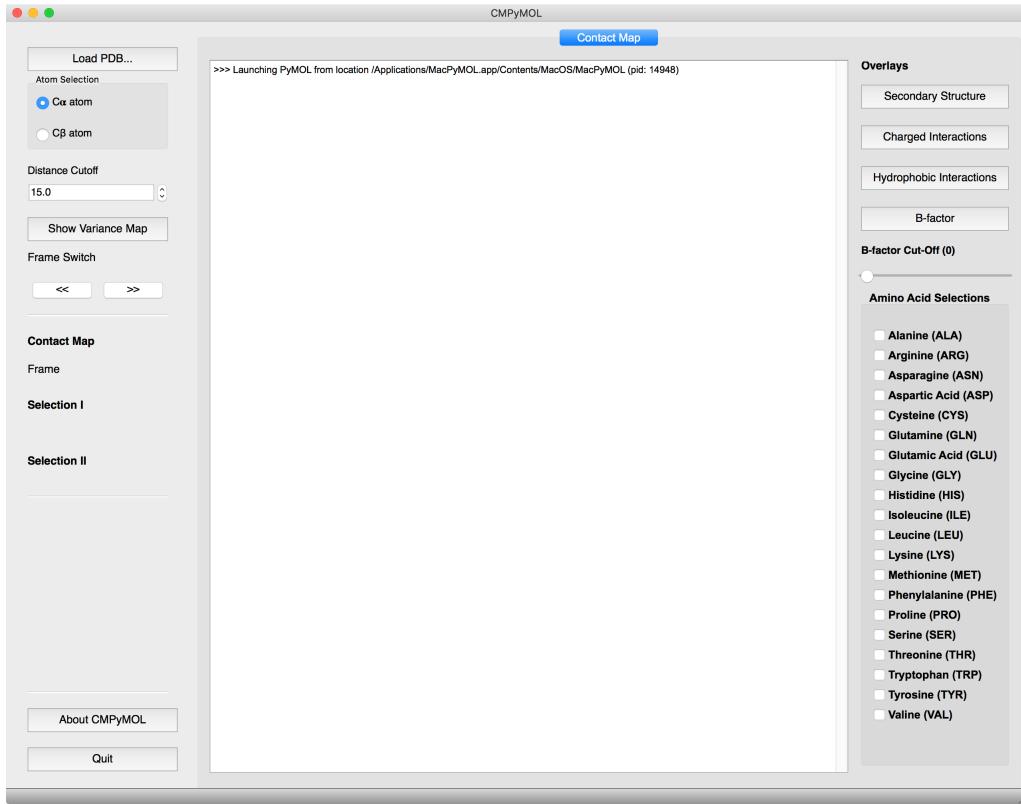


Figure 2.1: Main Interface of CMPyMOL.

### 2.3 Parameters to Calculate Contact Map

A protein contact map represents the pairwise distance between all possible amino acid residue pairs of a three-dimensional protein structure. These distances are typically calculated between either C $\alpha$  or C $\beta$  atom. The cut-off distance eliminates all the pairwise distances that are larger than that value.

The user has the choice to calculate the contact map based on specific parameters (Figure 2.2). The choice of C $\alpha$  or C $\beta$  atom as the basis atom and the cut-off distance (defaults to 15) for calculating the contact map will be can be customized by the user.

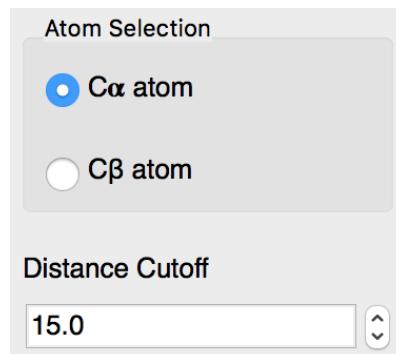


Figure 2.2: Parameters for Contact Map Calculation

- R** The parameters to calculate contact maps can be set by the user before or after loading the PDB for on-the-fly updates of maps and overlays.

### 2.3.1 Load PDB file

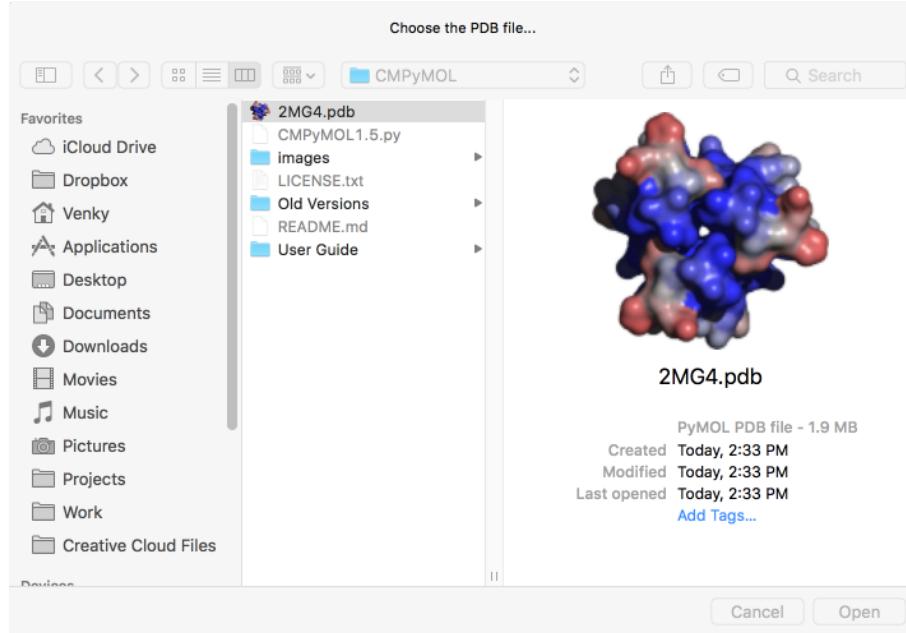


Figure 2.3: Pop-up dialog for file selection.

Either a single frame PDB file or a multi-frame PDB from MD simulation trajectory or NMR trajectory can be loaded at this step. The correct format of the PDB file is provided in Section 2.3.2 for reference. Other structure formats are currently not supported.

Click on the to invoke a pop-up to locate the PDB file. A file dialog (Figure 2.3) will facilitate this choice. Throughout this guide PDB ID: 2MG4 will be used for illustration. 2MG4 is an NMR structure of an artificially designed protein that is a symmetric protein homodimer [1].

After loading the PDB file, there will be two windows presented to the user as shown in Figure 2.4 and 2.5.

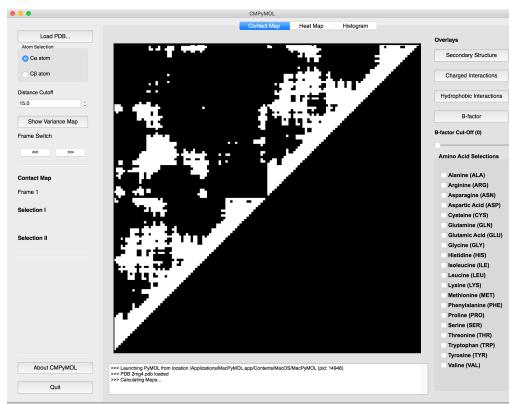


Figure 2.4: CMPyMOL window.

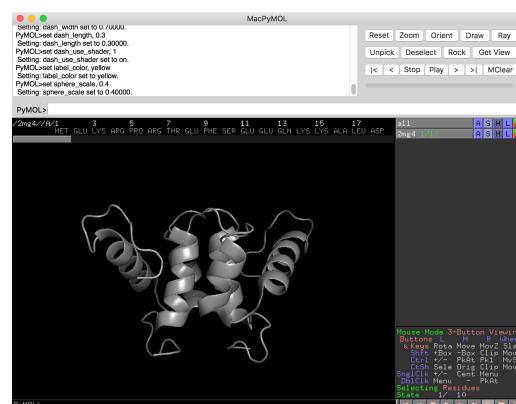


Figure 2.5: PyMOL window

### 2.3.2 PDB Format

---

```
MODEL X
```

```
.
```

```
.
```

```
ATOM ...
```

```
ATOM ...
```

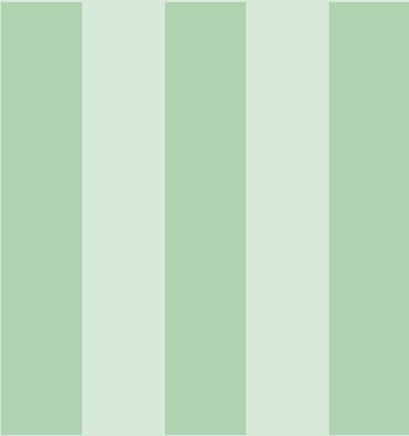
```
ATOM ...
```

```
.
```

```
.
```

```
ENDMDL
```

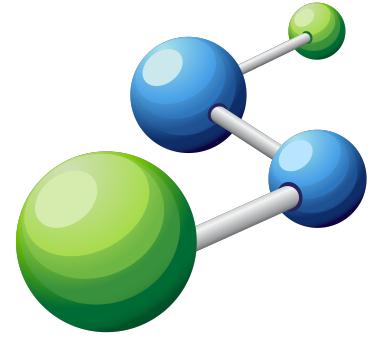
---



# Part Two

<b>3</b>	<b>Functionality</b>	17
3.1	Contact Map	
3.2	Selections	
3.3	Overlays	
3.4	Plots	
	<b>Bibliography</b>	25
	Books	
	Articles	
	<b>Index</b>	27





### 3. Functionality

The overlays in CMPyMOL enriches the interpretation of contact maps.

#### 3.1 Contact Map

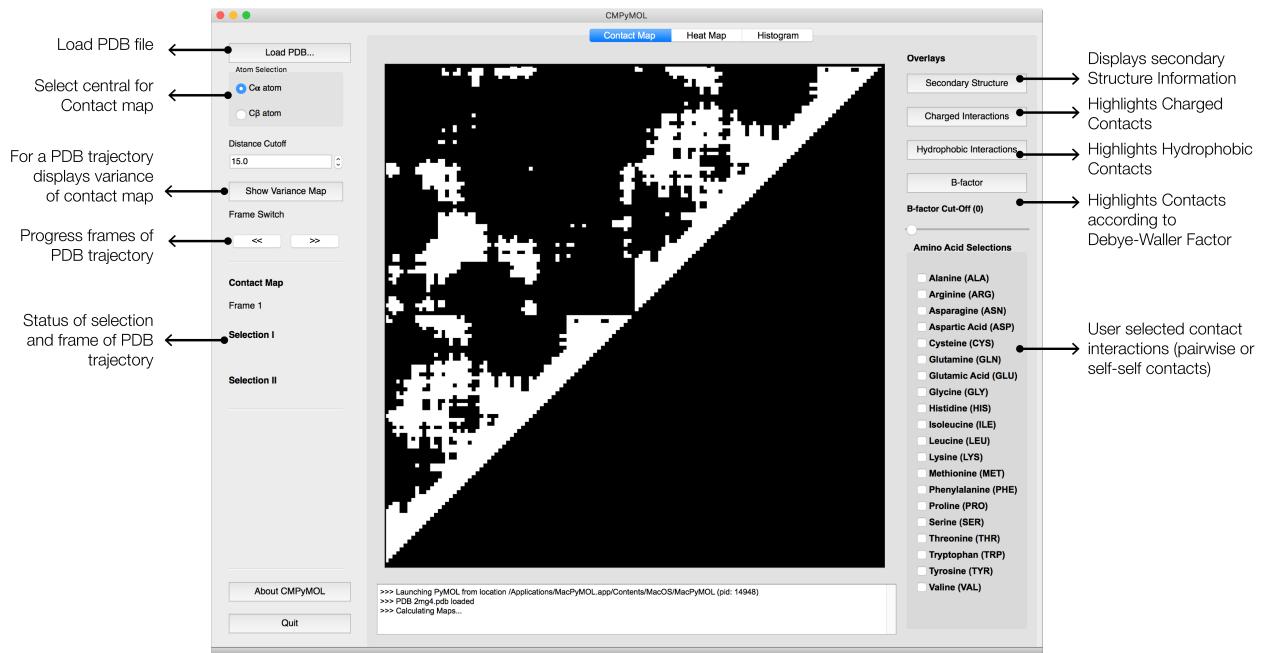


Figure 3.1: Main CMPyMOL Window and Its Functionality

The main window of CMPyMOL (Figure 3.1) provides controls for all the selection, overlay and plots to analyze contact maps. The overlays (the toggle buttons on the right of the contact map) superpose chemical and structural information on top of the contact map when activated. The plots

(buttons on the left side of the contact map) pops open a new window that provides an overview of the nature of contacts.

### 3.1.1 Variance Contact Map

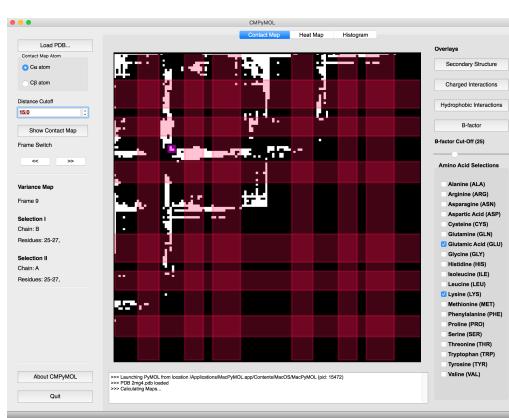


Figure 3.2: Variance Contact Map. Showing the regions of highest flexibility.

Variance map can be activated by clicking on the button **Show Variance Map**. The red rectangles in figure 3.2 overlays the secondary structure information of the PDB (more in section 3.3.1).

The PDB loaded as an example here is a NMR structure with multiple-frames (10 models). When a multi-frame PDB is loaded, the “Variance Contact Map” button is enabled and it allows for

calculating the variance in contact distance for a residue pair. Using **<>** buttons the user can progress through the different models. Such frame switch will update all the maps in CMPyMOL along with updating the model itself in the PyMOL window.

This allows for readily identifying flexible regions of the protein. In figure 3.2 a selection highlights the intra-protein flexible region with high variance. In the PyMOL window it can be verified that indeed the region of high variance is on a loop (Figure 3.3).

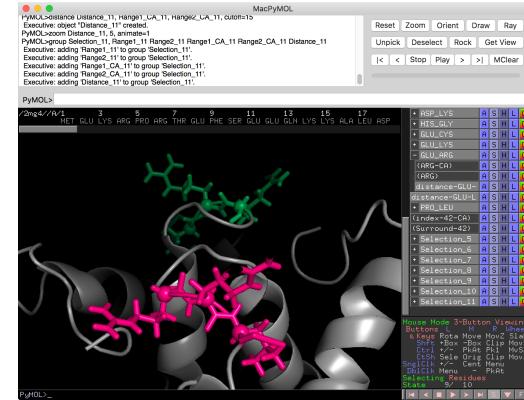


Figure 3.3: Selecting a high variance region in the intra-subunit quadrant reveals in the PyMOL window that the region is a flexible loop.

## 3.2 Selections

Selecting a particular region by clicking-n-dragging on a portion of the CMPyMOL generated contact map (shown as a magenta selection box Figure 3.4) highlights the corresponding region in the PyMOL window (Figure 3.5). Since contact map is a pairwise interaction matrix, the two interacting regions are colored hotpink and limegreen. The atom based on which the distances for the contact map are highlighted as “spheres”.

## 3.3 Overlays

Overlays in CMPyMOL provide a intuitive superposing of chemical and structural information on top of the contact map. This allows of correlating interacting residues with its type (chemical or structural).

### 3.3 Overlays

19

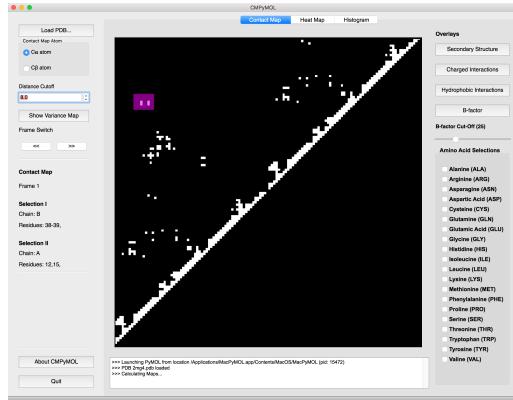


Figure 3.4: Selection in CMPyMOL window.

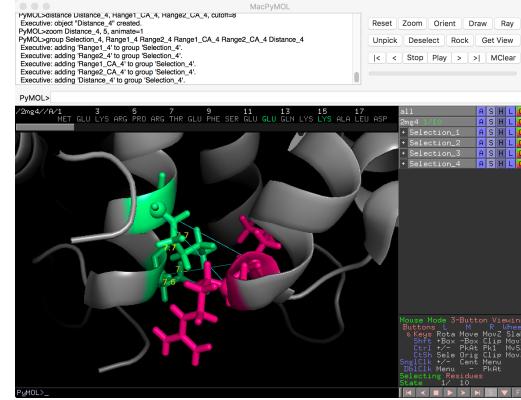


Figure 3.5: Focus on the selected region in PyMOL window.

#### 3.3.1 Secondary Structure

The button **Secondary Structure** toggles the overlaying the secondary structure of protein onto the contact map (Figure 3.6).

The secondary structure overlay superposes  $\alpha$ -helical and  $\beta$ -sheet as red and green translucent rectangles, respectively. Notedly, the selection from the last section (Section 3.2) is in a  $\alpha$ -helix –  $\alpha$ -helix interaction, which can be confirmed by the selection highlighted in PyMOL window (Figure 3.5).

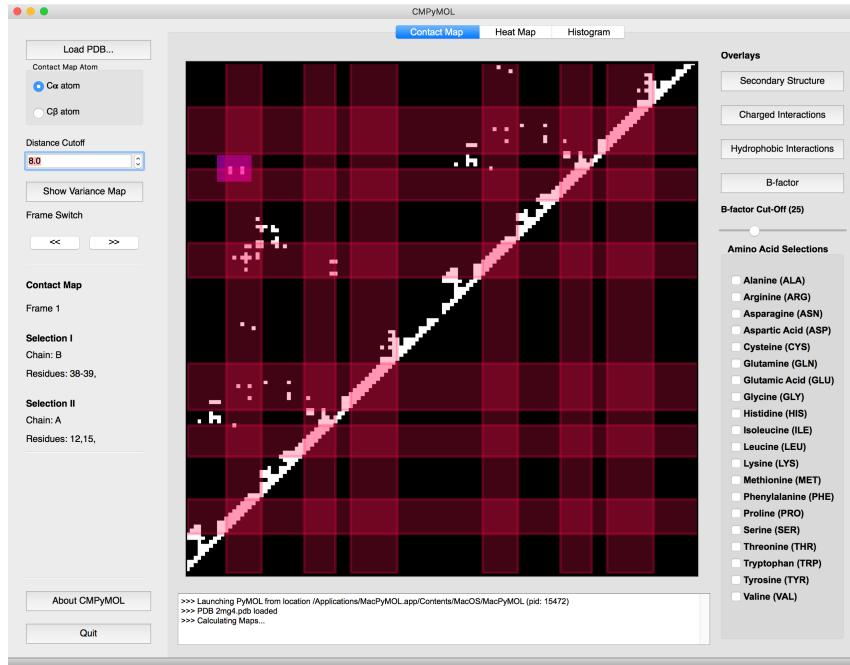


Figure 3.6: Secondary structure overlay on top of contact map.

#### 3.3.2 Charged Interactions

Similar to secondary structure overlay, by clicking on the charged interaction toggle button **Charged Interactions** draws a overlay that highlights contact points where two charged residues

are located within the cut-off distance specified in Section 2.3 (Figure 3.7). These charged interaction points are highlighted as **marine** pixels on top of the contact map.

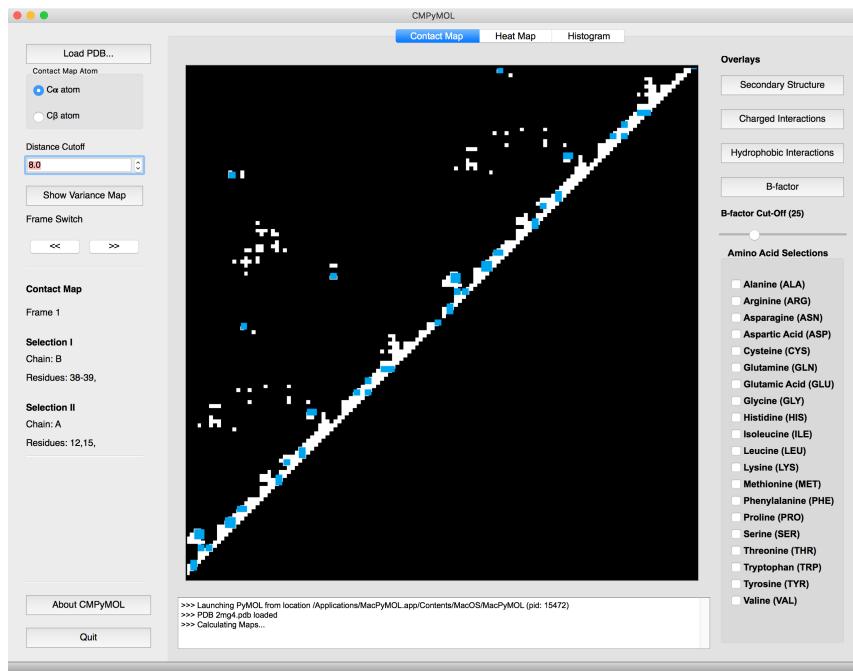


Figure 3.7: Charge-charge interactions highlighted on top of contact map.

### 3.3.3 Hydrophobic Interactions

Hydrophobic interactions overlay can be toggled by clicking the button **Hydrophobic Interactions**. This highlights the points of contact map where two hydrophobic residues are coming in contact. These interaction points are highlighted as **yellow** over the contact map (Figure 3.8).

### 3.3.4 B-factor Overlay

In a similar fashion B-factor (if those values are listed in the PDB) highlights the points with a larger value than set with the slider below the B-factor button (in the case shown in Figure 3.9, the highlighted contact points have a b-factor larger than 25). These interaction points are highlighted as **cyan** over the contact map.

### 3.3.5 User-defined Pairwise Aminoacid Interactions

In addition to the above overlays, the user can choose to highlight interaction points of two specific aminoacids. In Figure 3.10, the interactions of Lysine and Glutamic acid residues are highlighted in the PyMOL window (Figure 3.11).

## 3.4 Plots

Plots in CMPyMOL provide statistical analysis of the contact points.

### 3.4.1 Pairwise Heatmap

This heatmap counts the number of pairwise contacts of a given aminoacid to the rest of the other aminoacids (Figure 3.12). The order of aminoacids in this plot are arranged according to

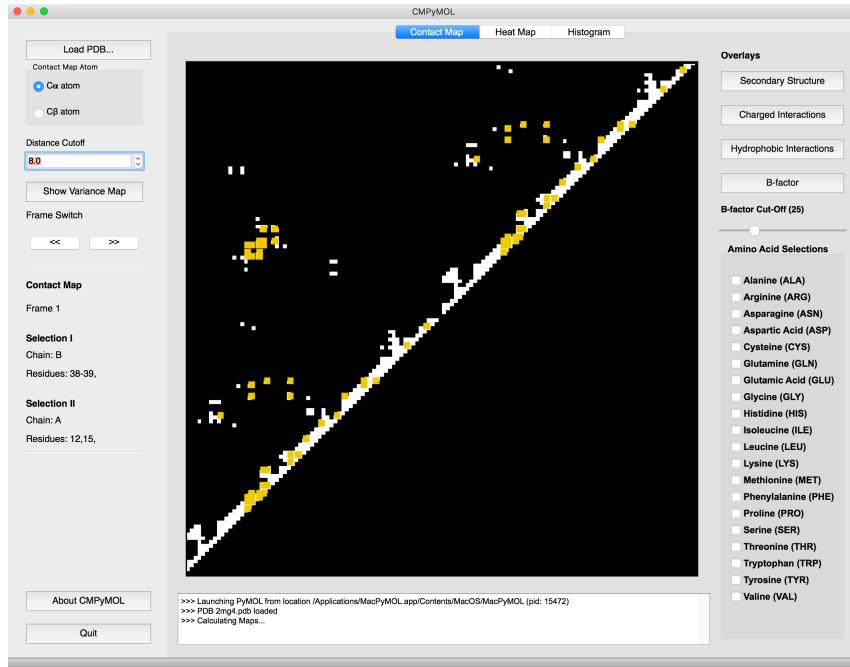


Figure 3.8: Hydrophobic interactions highlighted on top of contact map.

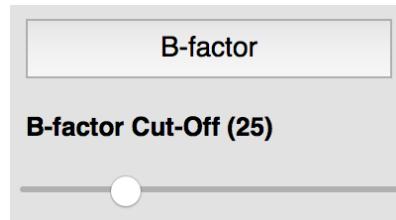


Figure 3.9: Parameters to draw an overlay to display B-factor. The contact points with b-factor larger than cutoff (in brackets) is colored in cyan.

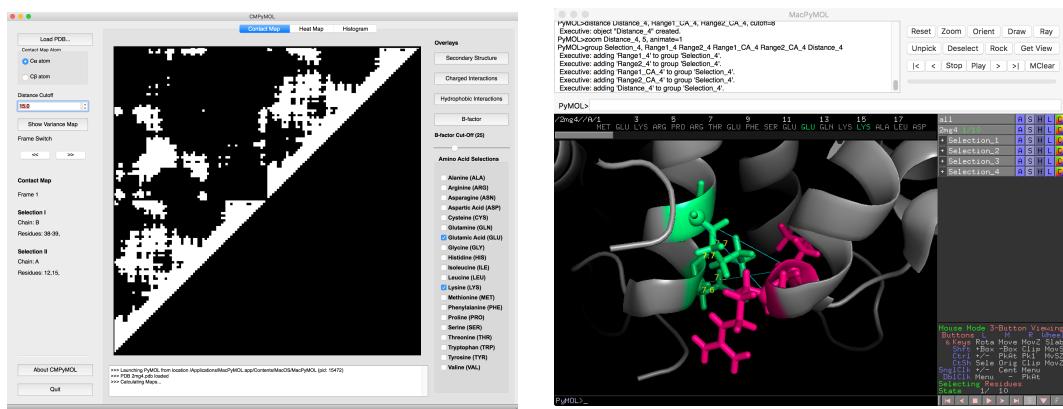


Figure 3.10: User selected interaction between Glutamic Acid and Lysine residues with a a cutoff distance of 15.

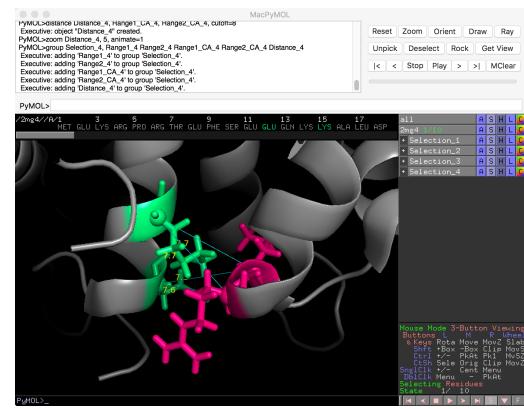


Figure 3.11: The residues corresponding with the selection (Figure 3.10) are highlighted in the PyMOL window.

their hydrophobicity. The color scale shows the number of each pairwise contacts in the protein. The plot is interactive and clicking on any “box” will correspondingly update the pymol window

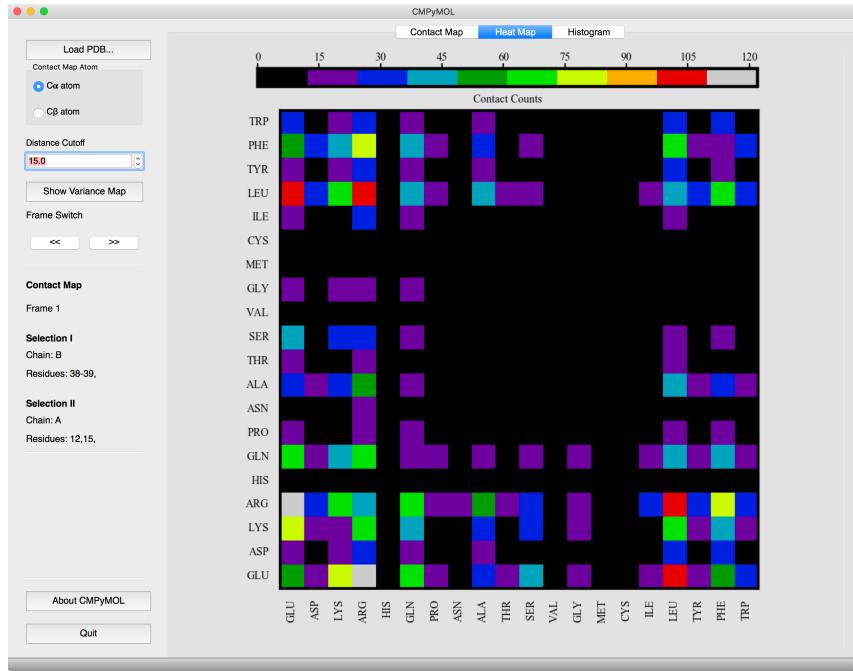


Figure 3.12: Heatmap of pairwise residue-residue interaction map. The residues are listed according to increasing hydrophobicity

corresponding to the selection.

### 3.4.2 Contact Density Histogram

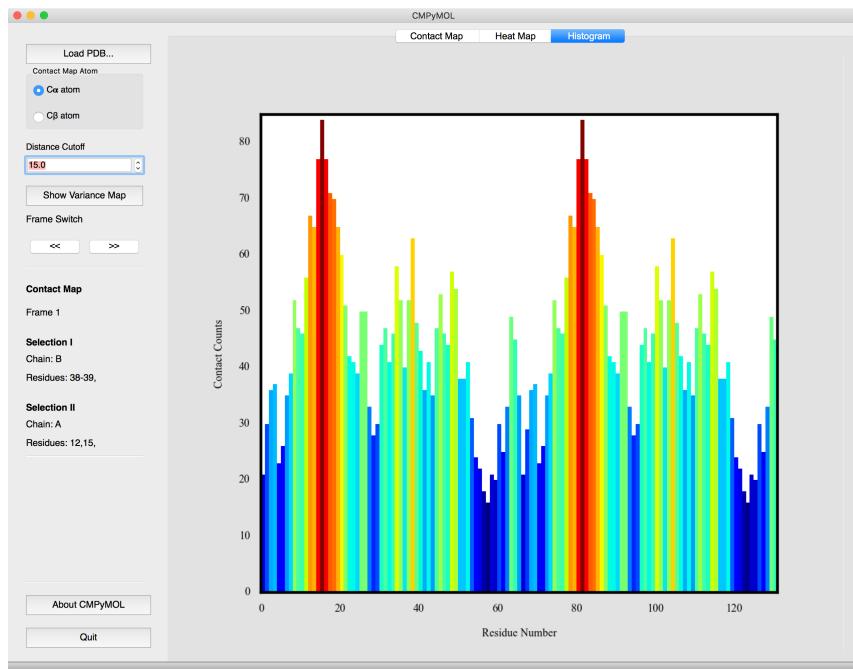
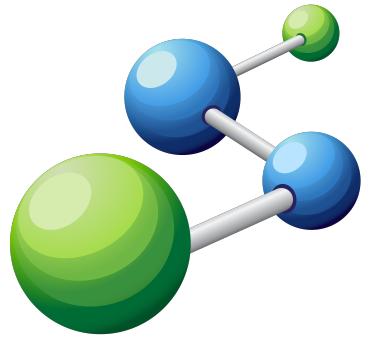


Figure 3.13: The residue-wise density of contacts in the PDB.

The contact histogram plot, graphs the density of contacts with respect to residue position (Figure 3.13). This plot is interactive and clicking on a particular bar selects the corresponding

residues and the surrounding contacts in the PyMOL window.





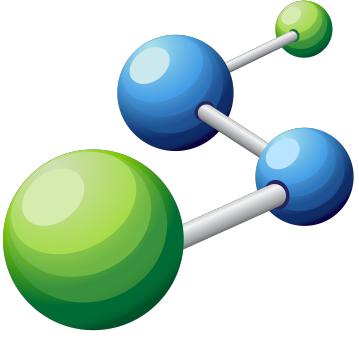
## Bibliography

### Books

### Articles

- [Mou+15] Yun Mou et al. “Computational design and experimental verification of a symmetric protein homodimer”. In: *Proceedings of the National Academy of Sciences* 112.34 (2015), pages 10714–10719 (cited on page 13).





# Index

## C

Calculate Contact Map .....	12
Charged Interactions .....	19
CMPyMOL Functionality .....	17
CMPyMOL Interface .....	11
CMPyMOL Usage .....	11
Contact Map .....	17
Contacts Density Histogram.....	22

## D

Debye-Waller Factor .....	20
Download Link .....	7

## H

Hydrophobic Interactions .....	20
--------------------------------	----

## I

Installation .....	7
--------------------	---

## L

Launch CMPyMOL .....	11
Linux Compatibility .....	7
Load PDB File.....	13

## M

Mac OS X Compatibility .....	7
------------------------------	---

## O

Open Source License .....	8
---------------------------	---

## P

Pairwise Amino Acid Selections .....	20
Pairwise Heatmap .....	20
PDB Format .....	14
Plots .....	20
Pre-requisite Python Libraries .....	8
Pre-requisite Softwares .....	8

## R

Running from Source .....	8
---------------------------	---

## S

Secondary Structure .....	19
Selecting Residue Ranges .....	18
source .....	7

## W

Windows Compatibility .....	7
-----------------------------	---