CMPyMOL: An Interactive PyMOL extension for Protein Contact-Map Analysis

Venkatramanan Krishnamani

Department of Molecular Physiology and Biophysics, University of Iowa, Iowa 52242, US

Abstract

Contact-maps are reduced 2D representation of the 3D spatial configuration of a protein. Many valuable structural features like secondary structures, inter- and intra-protein interactions, interacting domains, etc., can be readily identified from these maps. However, it is not straightforward and intuitive to reckon the spatial organization of the contact regions from reduced representation. The CMPyMOL extension for molecular visualization software PyMOL attempts to bridge this gap as an interactive graphical tool for protein contact-maps that interfaces with PyMOL for 3D visualization. Specifically, CMPyMOL helps understand the functional importance of contacts by providing visual overlays of various structural and biochemical properties of a protein on top of its contact-map.

Keywords: contact maps, pymol, biochemical, structure

1. Introduction

A contact-map of a protein is a 2D matrix of pairwise inter-residue distances, typically calculated over distances between C α atoms subject to an arbitrary maximum threshold. By construction this distance matrix is square and symmetrical. Contact-maps have been traditionally used to compare two protein structures/conformations [1], protein-protein-interactions [2, 3], protein folding [4, 5], structure prediction [6] and even reconstruction of the protein's 3D structure [7].

Email address: venky.krishna@icloud.com (Venkatramanan Krishnamani)

9 2. Problems and Background

10 Contact-maps capture high resolution, residue-level information regard-
 11 ing local conformations such as α -helices and β -sheets, and non-local interac-
 12 tions like inter-domain interactions. In this respect, contact-maps are a loss-
 13 less representation of structural information (except for chirality). However,
 14 essential biochemical information such as the residue type and the properties
 15 associated with it are lost during a contact-maps' construction. The interac-
 16 tion type of a particular contact point, such as hydrophobic interactions, salt
 17 bridges and hydrogen bonds, etc., can be crucial in understanding protein
 18 structure and function. A manual assignment to keep track of the residue
 19 number, the residue-type from the protein sequence and its spatial location
 20 can quickly become cumbersome and unmanageable.

21 3. Software Architecture

22 CMPyMOL supplements a contact-map analysis by interfacing with the
 23 powerful 3D visualization capabilities of PyMOL [8]. Launching CMPyMOL
 24 automatically invokes the PyMOL executable and generates a contact-map
 25 (for a specified cut-off distance) for a given PDB file. Visualizing multi-frame
 26 PDB trajectories are also supported. The user is provided with an option
 27 to calculate the distance map between $C\alpha$ or $C\beta$ atoms of residue pairs and
 28 the desired cutoff distance. The main program window (**Figure 1**) displays
 29 the contact-map in gray scale, shaded according to the distance between the
 30 pair of $C\alpha$ atoms.

31 CMPyMOL allows for manual selection of interacting residues on the
 32 2D contact map while the program highlights the corresponding residues
 33 in the PyMOL 3D visualization. This provides a intuitive bridge between
 34 the 2D and 3D representations of the protein. It also provides visual over-
 35 lays of structural and biochemical properties of the amino acid residues.
 36 Secondary structure (using STRIDE [9, 10] embedded within CMPyMOL),
 37 charge-charge interactions, hydrophobic interactions, B-factor and custom
 38 selected residue pair interaction sites (see **Figure 1**) are the currently avail-
 39 able overlays in CMPyMOL. The program also calculates pairwise residue
 40 interaction heat-map and residue-wise contact density plots as an alternate
 41 representations of the contact-map data.

42 4. Highlighted Software Functionalities

43 4.1. Substructure Selection

44 Contact points can be selected individually (left-click) or as contact re-
 45 gions (rectangular selection by left-click and drag) interactively over the
 46 contact-map. When a substructure of interest is selected, the correspond-
 47 ing residues are synchronously highlighted in the PyMOL window (**Figure**
 48 **2A and 2B**). The secondary structure overlay shows that our selection is
 49 a contact formed between two helical segments (*red* rectangles in **Figure**
 50 **2C**). Since the example protein is a homodimeric protein (PDBID: 2MG4)
 51 [11], the gray square represents the inter-protein interaction region and the
 52 selection corresponds to the monomer-monomer contact.

53 It should be noted that CMPyMOL only highlights the residues that are
 54 in contact and within the current selection. By construction, the lower and
 55 upper half of the contact-map separated by the main diagonal are symmetric,
 56 hence the lower half of the contact map is not shown by default.

57 4.2. Overlays and Plots

58 A novel feature in CMPyMOL are the overlays of structural and bio-
 59 chemical properties that highlights the nature of interaction of each contact
 60 point. This allows clear distinction, discovery and classification of contacts
 61 of a specific biochemical/structural type. For example, turning on “Charged
 62 Interactions” displays a *blue* overlay on all the contacts points where two
 63 charged residues are in proximity, i.e. in contact (while “Hydrophobic Inter-
 64 actions” as displayed as an *yellow* overlay) (**Figure 2D**). “B-factor” draws
 65 a *red* highlight of residues that have a Debye-Waller factor [12, 13] above
 66 a certain cutoff (specified by the slider, **Figure 1**) and that are in contact
 67 (*not shown*). Additionally, users can select any pair of amino-acids, listed on
 68 the right, to highlight those residues selections in PyMOL window that are
 69 within the defined cutoff (*see manual*). This is a powerful tool for quickly
 70 and efficiently identifying specific interactions types and simultaneously vi-
 71 sualizing their spatial orientation.

72 The “Pairwise Heat-map” and “Contacts Histogram” (**Figure 3**), calcu-
 73 lates and plots the number of pairwise residue contacts and contact density
 74 of each residue, respectively. Of these maps, the contact density map is
 75 interactive, mouse selections of a particular contact density highlights the lo-
 76 cation in the PyMOL window. When a multi-frame PDB trajectory is loaded
 77 into the software, the user can choose to view either the contact-map or the

78 cumulative variance contact-map calculated for all the frames starting from
79 frame 1 to the current frame selection.

80 5. Implementation

81 CMPyMOL is developed using the Python programming language and
82 open source libraries—*PyQT4*, Numeric Python (*numpy*) and *matplotlib*. It
83 is provided under an open source license (The MIT License) as source code
84 and as pre-compiled binaries for Windows and Mac OS X operating systems
85 (Linux users will be able to run CMPyMOL directly from source). The
86 detailed installation instructions are listed in the user guide (**Table 1**).

87 CMPyMOL provides a much needed add-on to the PyMOL software pack-
88 age, a tool which is typically a built-in part of other molecular visualization
89 programs, such as VMD [14]. There exists at least one other interactive
90 tool for visualizing contact-maps, but it is limited in terms of displaying
91 biochemical and structural information on the contact-maps and does not
92 support multi-frame PDB trajectories [15]. CMPyMOL is intended to re-
93 place an existing PyMOL plugin, Contact Map Visualizer (available from
94 <http://www.pymolwiki.org>), that was co-developed by the author (VK) in
95 collaboration with Thomas Holder (Schrödinger, LLC).

96 6. Illustrative Example

97 Using an NMR structure (10 models) of a designed homodimeric pro-
98 tein (PDBID: 2MG4), one workflow highlighting three core functionalities of
99 CMPyMOL will be demonstrated in this section [11]. CMPyMOL user guide
100 is downloadable from the software’s `github` page. It provides detailed de-
101 scriptions of other functionality of CMPyMOL. <https://github.com/emptyewer/CMPyMOL>.
102 The PDB file (PDBID: 2MG4) used in this example is distributed along with
103 the source code.

104 6.1. Identifying Chemical Nature of Secondary Structure Contacts

105 After the CMPyMOL window is initialized and the PDB file is loaded,
106 selecting the toggle button (named Secondary Structure) on the right-hand
107 side of the contact map main-display overlays the secondary structure infor-
108 mation. The secondary structure overlay superposes α -helical and β -sheet as
109 red and green translucent rectangles, respectively (**Figure 2C**). The loop re-
110 gions are uncolored. Selecting the contact points within the α -helix- α -helix

111 interaction region brings to focus in the corresponding 3D structure in the Py-
 112 MOL window (**Figure 2B**). Further examining the selection in the PyMOL
 113 window reveals the biochemistry of the interaction is electorstatic/charged in
 114 nature (see *blue* squares in **Figure 2D** corresponding to the selection where
 115 GLU, LYS and ARG within 8Å).

116 6.2. Identifying Contacts that Stabilizes Protein Dimer

117 In **Figure 2A** the gray box on the top left represents the contacts from
 118 inter-protein monomer-monomer interaction. The secondary structure over-
 119 lay readily identifies most of the contact between the protein partners are α -
 120 helix- α -helix interaction (contact points within the region where two red rect-
 121 angles intersect) with some interactions between two loops (contacts within
 122 uncolored region) and a few interactions between α -helix and loop **Figure**
 123 **2C**. Further the interactions stabilizing the dimer are hydrophobic (*yellow*
 124 squares within *dotted-gray* box **Figure 2D**) and *two* charged interactions
 125 (*blue* squares **Figure 2D**). Note that even though there are four *blue* spots
 126 within the *gray* box, due to diagonal symmetry of the contact map the actual
 127 charged interactions in the protein is two.

128 6.3. Identifying Regions of Maximum Flexibility

129 Since the PDB in this example is an NMR structure, CMPyMOL can
 130 calculate the variance of the distances between pairwise contacts along the
 131 trajectory. With such a representation, selecting a region displaying highest
 132 variance in intra-protein interaction region reveals that the contacts belong
 133 to a particular loop on each monomer (*not shown*). This is described in
 134 more detail in the CMPyMOL user manual that can be downloaded from
 135 the official site. <https://github.com/emptyewer/CMPyMOL/releases>

136 7. Conclusions and Limitations

137 CMPyMOL integrates 2D contact-maps augmented with biochemical in-
 138 formation and powerful 3D Visualization of PyMOL. This provides an intu-
 139 itive platform for simultaneously exploring protein interaction sites and its
 140 3D structure.

141 Currently, CMPyMOL only supports importing locally available PDB
 142 file-formatted files. Since each pixel of the contact-map image represents
 143 an interacting residue pair, the number of residues of a protein that can
 144 be comfortably displayed on a computer screen is limited to approximately

145 1000px by 1000px or 1500px by 1500px (varies by screen resolution). This
146 software is under active development, so users can request new features and
147 report bugs on the CMPyMOL [github](#) repository. The next release (2.1)
148 will support further data formats.

149 Acknowledgements

150 I would like to thank my former postdoctoral advisor Dr. Markus Deserno
151 at Carnegie Mellon University for his support and insightful suggestions.
152 *Funding:* Volkswagen Foundation for supporting this project within the frame-
153 work of their program “New Conceptual Approaches to Modeling and Sim-
154 ulation of Complex Systems”

155 References

- 156 [1] A. Caprara, R. Carr, S. Istrail, G. Lancia, B. Walenz, 1001 Optimal
157 PDB Structure Alignments: Integer Programming Methods for Finding
158 the Maximum Contact Map Overlap, *J Comput Biol* 11 (1) (2004) 27–
159 52. doi:10.1089/106652704773416876.
- 160 [2] R. C. de Melo, C. E. R. Lopes, F. A. Fernandes, Jr, C. H. da Silveira,
161 M. M. Santoro, R. L. Carceroni, W. Meira, Jr, A. d. A. Araújo, A Con-
162 tact Map Matching Approach to Protein Structure Similarity Analysis,
163 *Genet Mol Res* 5 (2) (2006) 284–308.
- 164 [3] R. C. Melo, C. Ribeiro, C. S. Murray, C. J. M. Veloso, C. H. da Silveira,
165 G. Neshich, W. Meira, Jr, R. L. Carceroni, M. M. Santoro, Finding
166 Protein-Protein Interaction Patterns by Contact Map Matching, *Genet*
167 *Mol Res* 6 (4) (2007) 946–63.
- 168 [4] M. Vendruscolo, E. Domany, Protein Folding Using Contact Maps, *Vi-*
169 *tam Horm* 58 (2000) 171–212.
- 170 [5] K. Park, M. Vendruscolo, E. Domany, Toward an Energy Function for
171 the Contact Map Representation of Proteins, *Proteins* 40 (2) (2000)
172 237–48.
- 173 [6] P. Di Lena, K. Nagata, P. Baldi, Deep Architectures for Protein Contact
174 Map Prediction, *Bioinformatics* 28 (19) (2012) 2449–57. doi:10.1093/
175 *bioinformatics/bts475*.

- 176 [7] J. M. Duarte, R. Sathyapriya, H. Stehr, I. Filippis, M. Lappe, Optimal
177 Contact Definition for Reconstruction of Contact Maps, BMC Bioinfor-
178 matics 11 (2010) 283. doi:10.1186/1471-2105-11-283.
- 179 [8] W. DeLano, The PyMOL Molecular Graphics System.
- 180 [9] D. Frishman, P. Argos, Knowledge-based Protein Secondary Struc-
181 ture Assignment, Proteins 23 (4) (1995) 566–79. doi:10.1002/prot.
182 340230412.
- 183 [10] M. Heinig, D. Frishman, STRIDE: A Web Server for Secondary Struc-
184 ture Assignment from Known Atomic Coordinates of Proteins, Nucleic
185 Acids Res 32 (Web Server issue) (2004) W500–2. doi:10.1093/nar/
186 gkh429.
- 187 [11] Y. Mou, P.-S. Huang, F.-C. Hsu, S.-J. Huang, S. L. Mayo, Computa-
188 tional design and experimental verification of a symmetric protein ho-
189 modimer, Proceedings of the National Academy of Sciences 112 (34)
190 (2015) 10714–10719.
- 191 [12] I. Waller, Zur Frage der Einwirkung der Wärmebewegung auf die In-
192 terferenz von Röntgenstrahlen, Zeitschrift für Physik A Hadrons and
193 Nuclei 17 (1) (1923) 398–408.
- 194 [13] P. Debye, Interferenz von Röntgenstrahlen und Wärmebewegung, An-
195 nalen der Physik 348 (1) (1913) 49–92.
- 196 [14] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual Molecular Dynam-
197 ics, J Mol Graph 14 (1) (1996) 33–8, 27–8.
- 198 [15] C. Vehlouw, H. Stehr, M. Winkelmann, J. M. Duarte, L. Petzold,
199 J. Dinse, M. Lappe, CMView: Interactive Contact Map Visualization
200 and Analysis, Bioinformatics 27 (11) (2011) 1573–4. doi:10.1093/
201 bioinformatics/btr163.

202 Current executable software version

S1	Current software version	2.0
S2	Permanent link to executables of this version	https://github.com/emptyewer/CMPyMOL/releases
S3	Legal Software License	MIT License
S4	Computing platform/Operating System	Windows and Mac OS X
S5	Installation requirements & dependencies	<i>pymol</i>
S6	User manual	https://github.com/emptyewer/CMPyMOL/releases
S7	Support email for questions	<i>venky.krishna@icloud.com</i>

Table 1: Software metadata

Figure 1: The main window of CMPyMOL 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.

Figure 2: **A)** The purple rectangular selection highlights a contact in the inter-protein interaction region (gray square) of the contact map. The contact map was generated with a cutoff of 8Å between Cα atoms of PDB:2MG4. **B)** This selection also brings to focus the selected region in the PyMOL window. The interacting residues within the selection are highlighted in red and blue with the sidechains represented as sticks. **C)** Activating the secondary structure overlay identifies the previous selection is a contact between two α-helices. **D)** To overlay the chemical nature, namely, charged (blue) and hydrophobic (yellow), highlights the corresponding contacts.

Figure 3: **A)** Heatmap of pairwise residue-residue interaction map. This map counts the number of pairwise contacts of a given aminoacid to the rest of the other aminoacids. The order of aminoacids in this plot are arranged according to their hydrophobicity. The color scale shows the number of each pairwise contacts in the protein. **B)** The residue-wise density of contacts in the protein (PDBID:2MG4). The contact histogram plot, graphs the density of contacts with respect to residue position. Both plots are interactive and clicking on a particular residue or residue-pair highlights the corresponding selection in the PyMOL window.