Structural bioinformatics

PConPy—a Python module for generating 2D protein maps

Hui Kian Ho¹,²,⁎, Michael J. Kuiper³ and Ramamohanarao Kotagiri¹

¹Department of Computer Science & Software Engineering, The University of Melbourne, Parkville, Victoria 3010,
²NICTA Victoria Research Laboratory, The University of Melbourne and ³Victorian Partnership for Advanced Computing, Carlton South, Victoria 3053, Australia

Received on July 21, 2008; revised on October 17, 2008; accepted on October 27, 2008

Advance Access publication October 31, 2008

Associate Editor: Dmitrij Frishman

ABSTRACT

Summary: PConPy is an open-source Python module for generating protein contact maps, distance maps and hydrogen bond plots. These maps can be generated in a number of publication-quality vector and raster image formats. Contact maps can be annotated with secondary structure and hydrogen bond assignments. PConPy offers a more flexible choice of contact definition parameters than existing toolkits, most notably a greater choice of inter-residue distance metrics. PConPy can be used as a stand-alone application or imported into existing source code. A web-interface to PConPy is also available for use.

Availability: The PConPy web-interface and source code can be accessed from its website at http://www.csse.unimelb.edu.au/~hohkhkh1/pconpy/

Contact: hohkhkh1@csse.unimelb.edu.au

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

Protein contact maps are a reduced representation of protein structure, providing a quick way of visually inspecting structural features. A contact map is a square matrix M where \( M_{ij} = 1 \) if the distance \( \delta_{ij} \) between residues \( i \) and \( j \) is below a predetermined distance threshold \( t \) or \( M_{ij} = 0 \) otherwise [Equation 1].

\[
M_{ij} = \begin{cases} 
0 & \text{otherwise} \\
1 & \text{if } \delta_{ij} < t.
\end{cases}
\]  

(1)

Secondary structure elements appear as bands of consecutive contacts—contacts constituting parallel and anti-parallel sheets run parallel and anti-parallel to the diagonal, respectively. Helices appear as thick bands of contacts immediately surrounding the diagonal. Unlike 3D coordinate representations, contact maps are invariant to translation and rotation, both favourable properties for the comparison of protein structures. The binary nature of a contact map lends itself well to binary machine learning algorithms for structural analysis and classification tasks. The distance constraints implied by contact maps have been shown to be sufficient for reconstructing a 3D structure close to that of the real structure (Bartoli et al., 2008). Consequently, contact map prediction has become an active area of research in recent times.

A distance map is the real-valued version of a contact map [Equation 2], without a distance threshold, whose structural features can be discerned in the same way as in a contact map.

\[
M_{ij} = \delta_{ij}.
\]  

(2)

A hydrogen bond plot (HB-plot) is a visualization of a square matrix \( H \) where \( H_{ij} = 1 \) if there exists a hydrogen bond between residues \( i \) and \( j \), or \( H_{ij} = 0 \) otherwise [Equation 3]. HB-plots can be used to identify residue pairs in a protein that provide structural stability. Bikadi et al. (2007) used HB-plots to study the change in the hydrogen bonding network of a protein during conformational transitions.

\[
H_{ij} = \begin{cases} 
0 & \text{otherwise} \\
1 & \text{if } hbond(i,j)=True.
\end{cases}
\]  

(3)

A number of existing applications can generate contact maps but are primarily aimed at other tasks, such as the statistical analysis of contacts or on 3D visualization, therefore offering limited map customization. Visual molecular dynamics (Humphrey et al., 1996) is used for visualizing protein structures in 3D and possesses an in-built contact map plotting tool. iMolTalk (Diemand and Scheib, 2004) is a web-based structural analysis toolkit that can plot distance maps. SeqX (Biro and Fördös, 2005) and the web-based Construct Map (Chung et al., 2007) are Java applications that focus on statistical analyses of contacts. PROTMAP2D (Pietal et al., 2007) is Python-based and can plot contact and distance maps from molecular dynamics trajectory files. The contact maps produced by these tools are not generated as vectorized images and are limited in their choice of inter-residue distance metrics, metrics which are not necessarily ideal representations of inter-residue contacts.

2 PConPy

We developed PConPy, an open-source Python module for generating contact maps, distance maps and HB-plots in a number of vector and raster image formats. Our aim was to provide an extensible tool for protein researchers to generate 2D protein maps for use in publication or structural analysis. PConPy can be used either as a stand-alone application from the command-line or can be incorporated into existing application source code. As a stand-alone application, the user can choose to generate the plot directly as a vector image (pdf, ps, svg), raster image (eps, jpg, png),...
is the minimum distance between the atom centres of residues (Bartoli et al, 2008) because residues interact with each other via VDW surfaces rather than atomic centres.

PConPy is capable of producing contact and distance maps using all of the aforementioned distance metrics. To illustrate the importance of distance-metric choice, we used PConPy to compare the contact maps of ubiquitin (1ubq) when generated using the minimum VDW and Cα–Cα distance metrics in Figure 1. The same distance threshold of 8 Å was used in both maps, yet there is a clear loss of contact information in the Cα–Cα contact map. The VDW contact map not only retains secondary structure information but also provides contact information between secondary structure elements.

3 IMPLEMENTATION DETAILS

PConPy uses STRIDE (Frishman and Argos, 1995) and DSSP. Hydrogen bond information is obtained from DSSP (Kabsch and Sander, 1983). PConPy is open source and runs on Linux, Mac OS X, and Windows. Installation of Python 2.5, ScientificPython, NumPy, Matplotlib, DSSP and STRIDE are required to run PConPy on all compatible operating systems. Alternatively, the PConPy web interface can be accessed via http://www.csse.unimelb.edu.au/~hohkkh11/pconpy/.

ACKNOWLEDGEMENTS

We thank Dr Anthony Wirth for hosting the PConPy web interface and Alex Stivala for maintaining the local copy of the PDB.

Funding: NICTA.

Conflict of Interest: none declared.

REFERENCES


(Fariselli and Casadio, 1999). The more biologically relevant variation of this metric is the minimum distance between the VDW surfaces of atoms between residues (Wu and Zhang, 2008) because residues interact with each other via VDW surfaces rather than atomic centres.

or ASCII file. PConPy also possesses an interactive mode where the user has the ability to zoom, pan, resize and save the plot manually.

Contact maps can be annotated with secondary structure assignments and hydrogen bond information. For all maps, users can choose specific chains of a protein to be displayed and regions of the protein sequence to highlight. Minimum contact thresholds and sequence separation parameters can also be specified by the user. However, the main advantage PConPy possesses over existing tools is the ability to generate contact and distance maps using a greater range of inter-residue distance metrics, these include the distances between centres-of-mass and side-chain centres-of-mass, minimum distance between atom centres and minimum van der Waals (VDW) distance between atoms, all of which are not implemented by existing tools. The traditional Cα–Cα and Cβ–Cβ distance metrics are also available in PConPy.

The choice of inter-residue distance metric is integral in determining a protein’s contact map. The de facto standards are considered to be the Cα–Cα and Cβ–Cβ distances between residues (Bartoli et al., 2008) and are the primary metrics used by existing tools. However, measuring inter-residue distance using backbone atoms can overlook contacts occurring between side chains. Hydrophobic and hydrophilic interactions, salt bridges and disulphide bonds can all occur between side chains, particularly between secondary structure elements that form structural motifs. Subsequently, inter-residue distance metrics that account for side-chain physical characteristics have been proposed—the distances between centres-of-mass and side-chain centres-of-mass between residues (Bartoli et al., 2008). A more precise metric is the minimum distance between the atom centres of residues