Supplementary material for PBmapclust: Mapping and Clustering the Protein Conformational Space Using a Structural Alphabet

I. Vetrivel[1], L. Hoffmann[2], S. Guegan[1], B. Offmann[2] & A. D. Laurent[1]

[1]Université de Nantes, CEISAM UMR CNRS 6230, 2 chemin de la Houssiniere, 44322 Nantes, France

[2]Université de Nantes, UFIP UMR CNRS 6286, 2 chemin de la Houssiniere, 44322 Nantes, France

1. Description of the steps involved in generating a PBmapclust plot

Generating a PBmapclust plot involves the following three steps, a detailed description of each of the steps follows.

1. Extracting all frames from the trajectory file.
2. Calculating torsion angles and PB assignment.
3. Generating a X PixelMap (XPM) file and subsequently, an Encapsulated PostScript (EPS) file.

In **Step 1**, based on the information available in the *.tpr* file about the duration of the MD simulation and the frequency at which the coordinates are recorded, all the frames from te MD trajectory are written. Later subsets of this can be obtained to focus on specific residues and/or at specific time intervals of the MD simulation. The user has the option to downsample the number of frames to be considered for generating the PBmapclust plot by controlling the parameters in the script. If necessary, all frames from the trajectory can be sampled exhaustively to not loose any structural information.

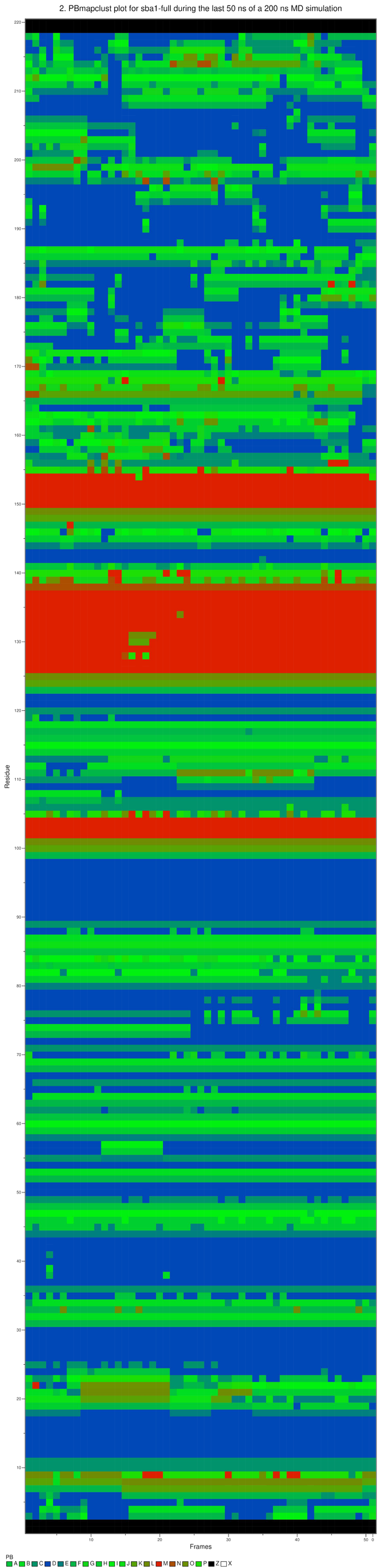
In **Step 2**, the DSSP program is run on each of the above frames and the torsion angles calculated by DSSP is used for PB assignment. This is done for individual chains, in case the MD simulation was run for a multimeric protein or a protein complex. The torsion angles are parsed over a sliding window of eight consecutive values and compared to the standard torsion angles of the 16 PBs as defined in de Brevern et al. [BEH00]. The PB is assigned based on the minimal RMSD (Root Mean Squared Deviation) between the torsion angle being parsed and the standard torsion angle. This process is repeated for all the frames. The process of converting a structure to a sequence of PBs is called PB transform and is available as a web server for converting individual PDB files at http://www.bo-protscience.fr/pbe/?page_id=10 [TSS∗06]. For the purpose of PBmapclust this has been automated in a Perl script.

In **step 3**, the cumulated PB sequences from all the MD trajectory frames are converted into a XPM file and then subsequently into a EPS file for convenient visualization. XPM is a bit map image format developed at the BULL research center at INRIA, France. The XPM file serves to map the 16 PBs to 16 colors. The XPM file is converted to an EPS file using the *xpm2ps* program that comes included by default with a GROMACS distribution.

References

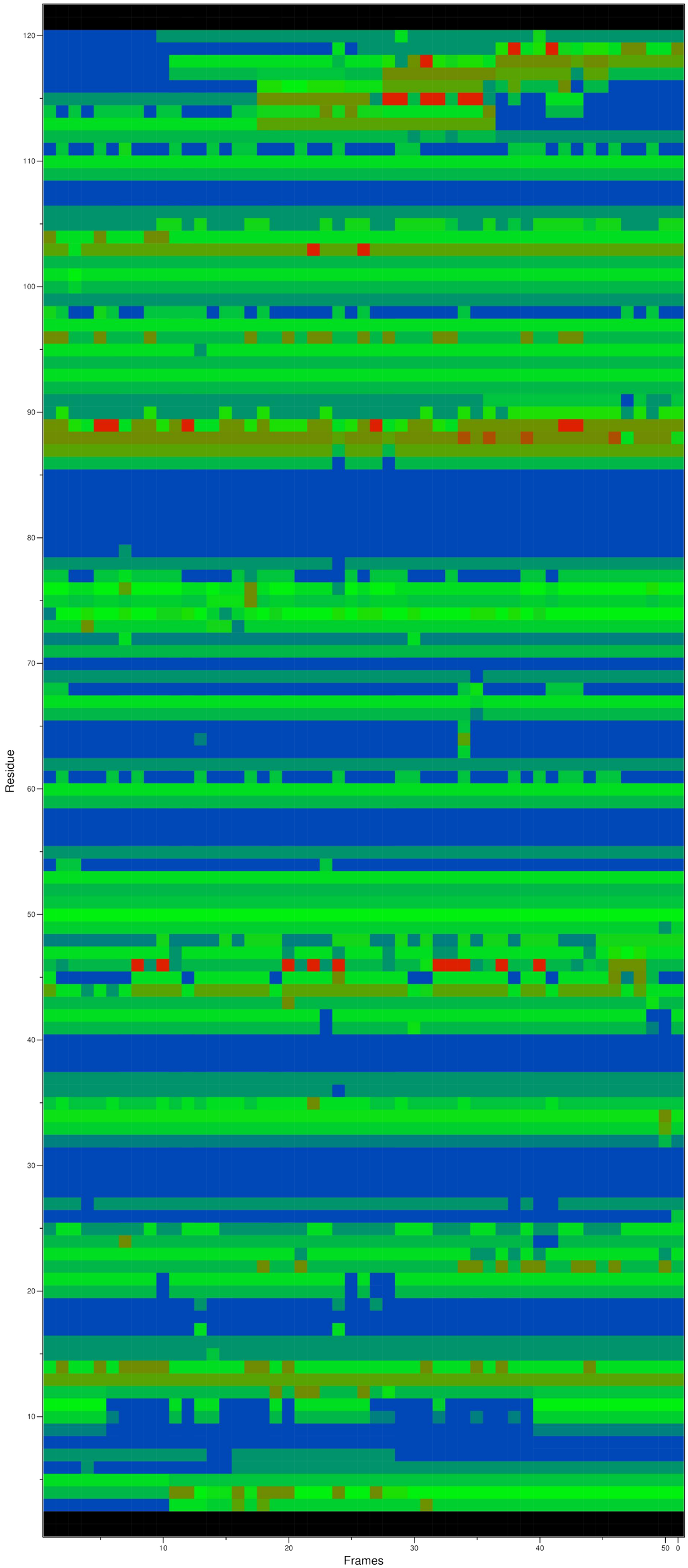
[BEH00]
Brevern A. G. D., Etchebest C., Hazout S.: Bayesian
Probabilistic Approach for Predicting Backbone. Proteins 287, January (2000), 271–287.

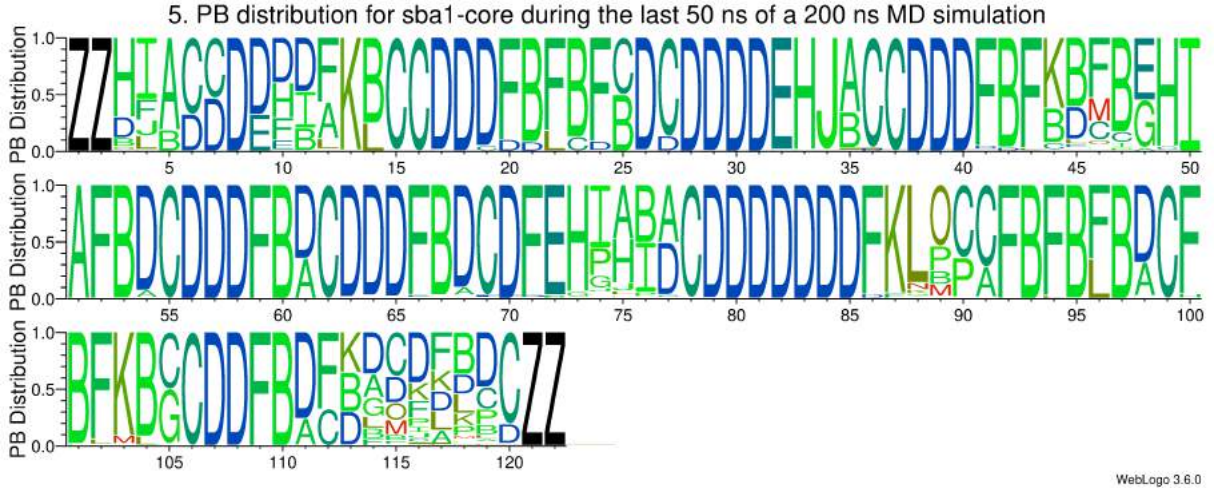
[TSS∗06] Tyagi M., Sharma P., Swamy C. S., Cadet F., Srinivasan N., de Brevern A. G., Offmann B.: Protein Block Expert (PBE): A web-based protein structure analysis server using a structural alphabet.. Nucleic Acids Research (2006).


2. PBmapclust plot for sba1-full during the last 50 ns of a 200 ns MD simulation
Residue
Frames
PB
A B C D E F G H I J K L M N O P Z X


3. PB distribution for sba1-full during the last 50 ns of a 200 ns MD simulation
PB Distribution
1.0
0.5
0.0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
WebLogo 3.6.0

4. PBmapclust plot for sba1-core during the last 50 ns of a 200 ns MD simulation




5. PB distribution for sba1-core during the last 50 ns of a 200 ns MD simulation
PB Distribution
1.0
0.5
0.0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
WebLogo 3.6.0

6. RMSD plot for 200 ns MD simulation of sba1-core and sba1-full