USER MANUAL for the PentUnFOLD algorithm

Prediction from the **PDB file**

1. To start the prediction, you have to upload your PDB file into the area in the middle of the screen, or click the icon on the "file.pdb" input field and select your file from a folder, and click on the button "open".
2. In the drop-down menu you have to choose a chain from the pdb file. Automatically the first chain (usually Chain A) from the pdb file will be selected for predictions.
3. Select checkbox with the target version of the PentUnFOLD algorithm ("1D", "2D", or "3D") for predictions. Actually, you may select three of them simultaneously.
4. Turn on the switch if you want to get the result from DSSP server according to the name of the pdb file (the name must be the same as in Protein Data Bank).
5. Click on the button "Get result".
6. Download the results of prediction in xlsx format.
7. Open the downloaded files with MS Excel or Libre Office Calc.
8. Press "Ctrl+Alt+F9" (for MS Excel) or "Ctrl+Shift+F9" (for Libre Office Calc) in case if those spreadsheets do not recalculate all the cells automatically.
9. See the result of prediction.
10. For "2D" spreadsheet you may replace amino acids and change secondary structure (in the columns "User input"). Do not forget that recalculation may take some time.

Prediction from the **amino acid sequence**

1. To start the process, you need to switch to the "AMINO ACID SEQUENCE".
2. Paste amino acid sequence in the window. Please, DO NOT paste any other symbols, such as FASTA heading.
3. Click on the button "Get result".
4. Download the results of prediction in xlsx format.
5. Open the downloaded files with MS Excel or Libre Office Calc.
6. Press "Ctrl+Alt+F9" (for MS Excel) or "Ctrl+Shift+F9" (for Libre Office Calc) in case if those spreadsheets do not recalculate all the cells automatically.
7. See the result of prediction.

Secondary structure determination

1. Our server sends the pdb file to the DSSP server (https://www3.cmbi.umcn.nl/xssp/) and receives its results.
2. DSSP server may fail with pdb file processing. So, you have an option to try to receive it via the name of the pdb file (PDB ID). In this case, you need to turn on the switch.
3. In case if DSSP server doesn't work, our own script will determine secondary structure of the protein.
4. The principle of our own script is exactly the same as the one for DSSP in terms of helices and beta strands finding. The only difference is in the way to estimate positions of hydrogen atoms. That is why minimal differences in lengths of alpha helices and beta strands are possible.
5. In "1D" version of the PentUnFOLD algorithm secondary structure is predicted from the amino acid sequence.

3D contacts counting

1. Our script follows the principles of the PIC server (http://pic.mbu.iisc.ernet.in/) and finds the same types of intraprotein contacts.
2. Minor differences with PIC server may appear in cation-pi and ionic contacts, since our script calculates distances from Nitrogen atoms to centroids of aromatic residues and Oxygen atoms of carboxylic groups, respectively.

Click on the button "Clean out" to start the next prediction.

Interpretation of the predictions

For "1D" spreadsheets one can analyze the following columns:

№	amino acid residue number
AA	the name of amino acid residue (in one-letter code)
HELIX	stability of predicted alpha helices: "HS" is a stable alpha helix; "HN" is a nonstable alpha helix (prone to turn into random coil)
SHEET	stability of predicted beta strands: "ES" is a stable beta strand; "EN" is a nonstable beta strand (prone to turn into random coil)
COIL	stability of predicted random coils: "C" is a stable random coil; "CE" is a nonstable random coil (prone to turn into beta strand); "CH" is a nonstable random coil (prone to turn into alpha helix)
Disorder	shows by "D" residues that are prone to turn into the disordered state

For "2D" spreadsheets one can analyze the following columns:

№	amino acid residue number
AA	the name of amino acid residue (in one-letter code)
User input	type a name of amino acid residue (in one-letter code) in a cell in this column, and it will be taken into calculations
SS	secondary structure according to the DSSP: "H" is alpha helix; "E" is beta strand; "C" is random coil (neither "H", nor "E")
User input	type a name of secondary structure element (in one-letter code) in a cell in this column, and it will be taken into calculations
HELIX	stability of alpha helices: "HS" is a stable alpha helix; "HN" is a nonstable alpha helix (prone to turn into random coil)
SHEET	stability of beta strands: "ES" is a stable beta strand; "EN" is a nonstable beta strand (prone to turn into random coil)
Disordered coil	stability of random coils: "C" is a stable random coil; "CE" is a nonstable random coil (prone to turn into beta strand); "CH" is a nonstable random

coil (prone to turn into alpha helix); "D" is a disordered coil (prone to turn into the disordered state)

Disordered helix	shows by "D" residues in alpha helices that are prone to turn into the disordered state
Disordered sheet	shows by "D" residues in beta strands that are prone to turn into the disordered state
Completely disordered coil	shows by "D" residues in random coils that are especially prone to turn into the disordered state
Completely disordered helix	shows by "D" residues in alpha helices that are especially prone to turn into the disordered state
Completely disordered sheet	shows by "D" residues in beta strands that are especially prone to turn into the disordered state

For "3D" spreadsheets one can analyze the following columns:

№	amino acid residue number
AA	the name of amino acid residue (in one-letter code)
SS	secondary structure according to the DSSP: "H" is alpha helix; "E" is beta strand; "G" is 3/10 helix; "I" is π helix; "B" is beta bridge; "T" is turn; "S" is bend. Our own script finds just: "H"; "G"; "I"; "B"; and "E".
Disorder	shows by "D" residues that are prone to turn into the disordered state, and by "O" residues that are not prone to turn into the disordered state