User Guide for *CSPro*

(Dec. 2007)

This work has been published in Bioinformatics

(http://bioinformatics.oxfordjournals.org/content/vol23/issue24/index.dtl)

Identification and visualization of cage-shaped proteins

Min Hu; Junhui Wang; Qunsheng Peng

Bioinformatics 2007, 23: 3400 - 3402.

Contents

Part 1 Run *CSPro*

STEP1: Get *CSPro* online

1) Download the ***CSPro*** package at http://www.cad.zju.edu.cn/home/humin

2) Unpack ***CSPro***

3) Run **CSPro.exe**

The main menu appears. The upper left portion of the interface is as follows:

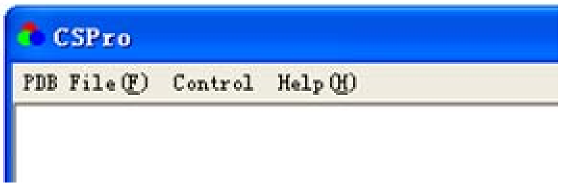


Fig.1 Main menu of ***CSPro***

Now it is ready for use ***CSPro***.

STEP2: Input structural data

There are two ways to detect if proteins are cage-shaped or not. One is for single protein structure，the other is for a group of protein structures simultaneously. They can be switched by controlling the checkbox of “**Batch Operation**”, which is the first item of the main menu “**Control**”. The default status of the checkbox of “**Batch Operation**” is off. (See Fig.2)

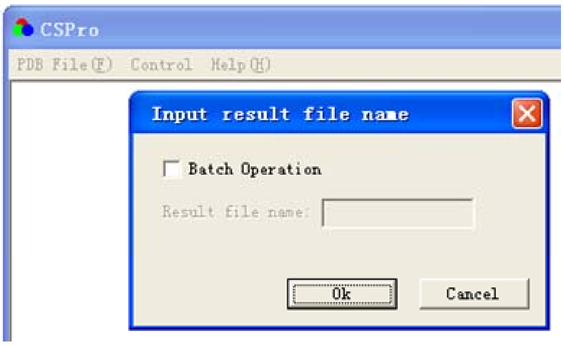


Fig.2 The status of the checkbox of “**Batch Operation**” is off.

◆ For single protein structure

Assume the status of the checkbox of "**Batch Operation**" is off , drop down the first column of the main menu, and open a PDB format protein structure file, with suffix .ENT or .PDB.

◆ For batch operation

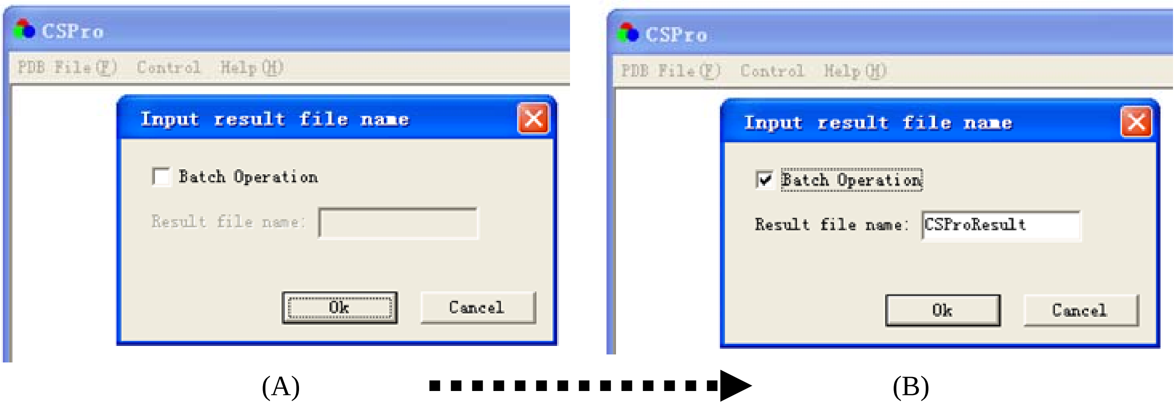
1) Prepare an index file by inputting a sequence of names for protein structures with the Microsoft tool of Notepad. For example, we would like to detect if a group of structures are cage-shaped or not, including PDB1dps, PDB1mfr, PDB1g65. Assume that all these PDB structural files are pre-stored under a directory called "**testdata**" in ***CSPro***, we key in the three names one line by one into a document called the index file as follows:

   PDB1dps.ENT
   PDB1mfr.ENT
   PDB1g65.ENT

   The last char of the index file is a return char, which means the end of the index file. The index file should also be stored in the same directory as those structural files reside in. The demo index file named as **BatchTemplate.txt** is available under the directory "**testdata**" in the ***CSPro*** package.

Note that the maximum number of structures to be detected is up to 20, considering the interaction of the software.

2) Drop down the second column of the main menu. A dialog of "Input result file name" appears. Pick on the check box of **Batch Operation** (see Fig.3(A)), you may input a file name for storing the result of the coming batch operation (see Fig.3(B)). The default name of ***CSPro*** is "**CSProResult**" under the same directory of "**testdata**".



☐ Batch Operation
Result file name:

☑ Batch Operation
Result file name: CSProResult

(A) (B)

Fig.3 The batch operation

3) Open the index file we have prepared, for example, open **BatchTemplate.txt** under the directory of **"testdata"**, then the system starts to compute. When the computation is over, ***CSPro*** will prompt users.

STEP3: Obtain the results

◆ For single protein structure

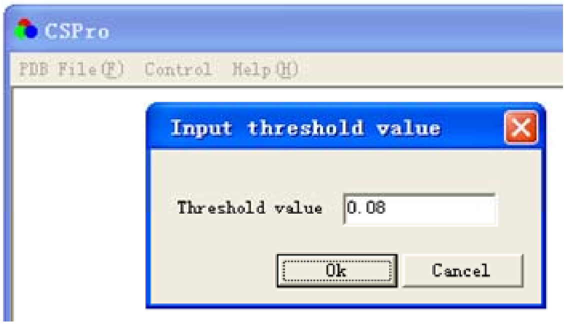
The detected result is given interactively. For cage-shaped protein, the volume of the cage and the ratio of the cage volume to the protein volume are given directly on the screen.

◆ For batch operation

The run results can be achieved by opening the result file, which is stored under the same directory of the structural data of proteins. The default name “**CSProResult”** is under the directory of “**testdata**”. Users can obtain the detailed results by opening the result file with the Microsoft tool of the tablet.

Part2 Threshold for filtering cage-shaped proteins

Cage-shaped proteins are filtered by the threshold ***T***. Users can change it manually. By choosing the second item of the menu “**Control**”, users can modify the threshold value (see Fig.4). The default threshold value set by ***CSPro*** is 0.08, as the ratio of the volume of the cavity to that of the protein for most of the cage-shaped proteins is greater than 0.08.



Threshold value 0.08

Fig.4 Input the threshold value for filtering cages

The threshold value keeps working until it is changed again.

Part3 Some snapshots for cage-shaped proteins

Fig.5 and Fig.6 depict some detected cage-shaped proteins with two possible visualization modes.

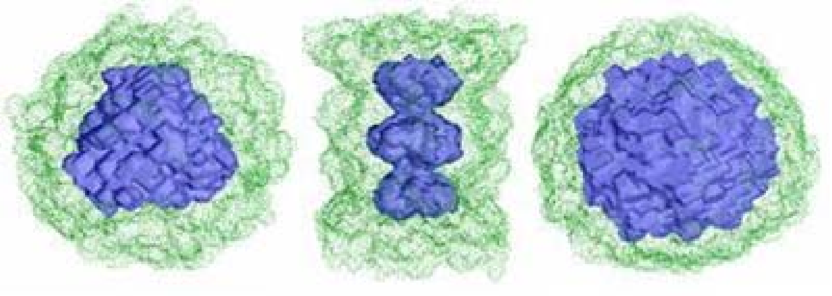


Fig.5 From left to right are PDB1dps, 1g65, 1mfr

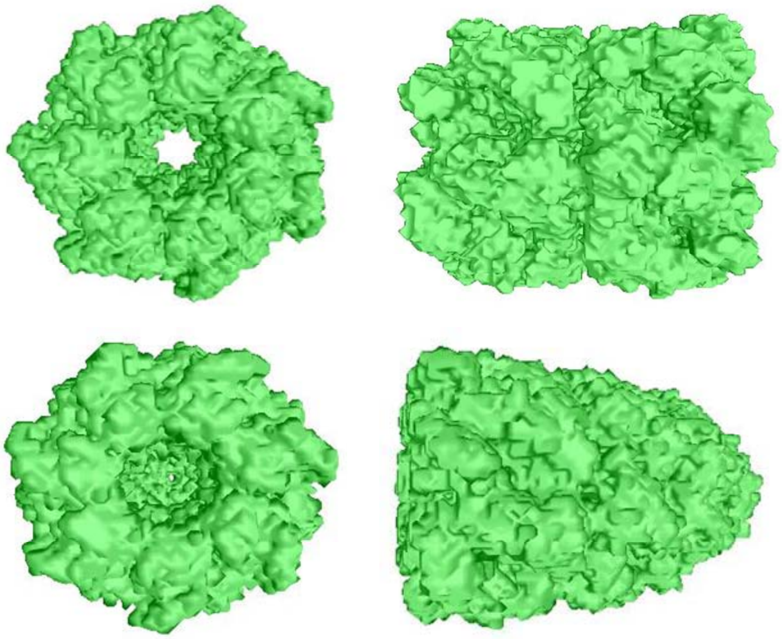


Fig.6 Two detected proteins with big openings are drawn in their outer surface, with two images from different views. The upper row is for PDB1kpo and the lower row is for PDB1svt.

Contact information

Min HU: humin@cad.zju.edu.cn