A Tutorial on IDSS

for shape retrieval of flexible proteins

Yu-Shen Liu
Department of Mechanical Engineering
Purdue University, West Lafayette, IN, 47907, USA
E-mail: liuyushen00@gmail.com, liu28@purdue.edu

IDSS implements the IDSS (Inner Distance Shape Signature) algorithm for shape retrieval of flexible proteins, rather than the conventional rigid methods.

Availability

The binary executable program for Windows platform and tested database are available from https://engineering.purdue.edu/PRECISE/IDSS .

System Requirements

Source code was written to perform the IDSS for shape retrieval of flexible proteins, as well as visualize the results.

The code was developed and tested using Microsoft Visual C++ 2005 on a Windows XP system.

REQUIRED INSTALLATIONS:
   -Microsoft Visual C++ 2005
   -MICROSOFT FOUNDATION CLASS (MFC) LIBRARY
   -OpenGL

Protein data set

The Database of Macromolecular Movements (MolMovDB) is a set of known proteins with conformational changes in different crystallographic structures.

Web site: http://www.molmovdb.org/cgi-bin/browse.cgi

The database is a collection of data and software pertaining to flexibility in protein and RNA structures, which can be found at: (http://www.molmovdb.org/). The corresponding experimental structures are downloaded from the Protein Data Bank (PDB), and the first chain of each structure is used as the reference structure for superposition.

Alternative, you might find the Hinge Atlas and Hinge Atlas Gold datasets the best curated and easiest to download. You can find the download links and instructions at: http://molmovdb.org/cgi-bin/sets.cgi.
Please refer to Hinge Atlas and/or FlexOracle papers:
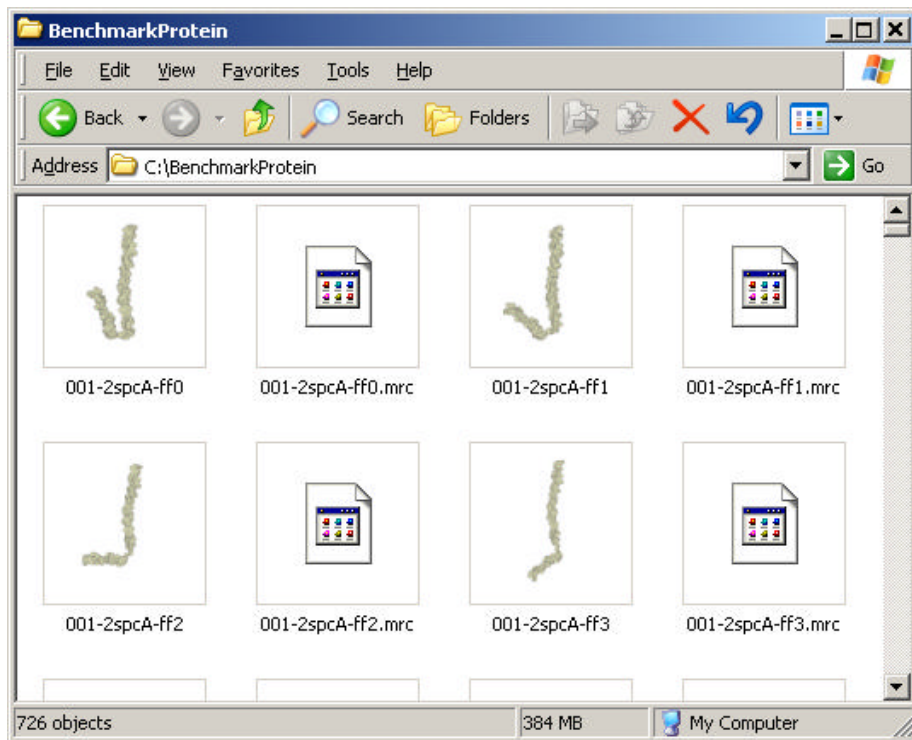
For showing how our algorithm works, we provided a small database link including 360 proteins with 27 groups, which are extracted from MolMovDB with large deformations. The full database and source code can be requested.
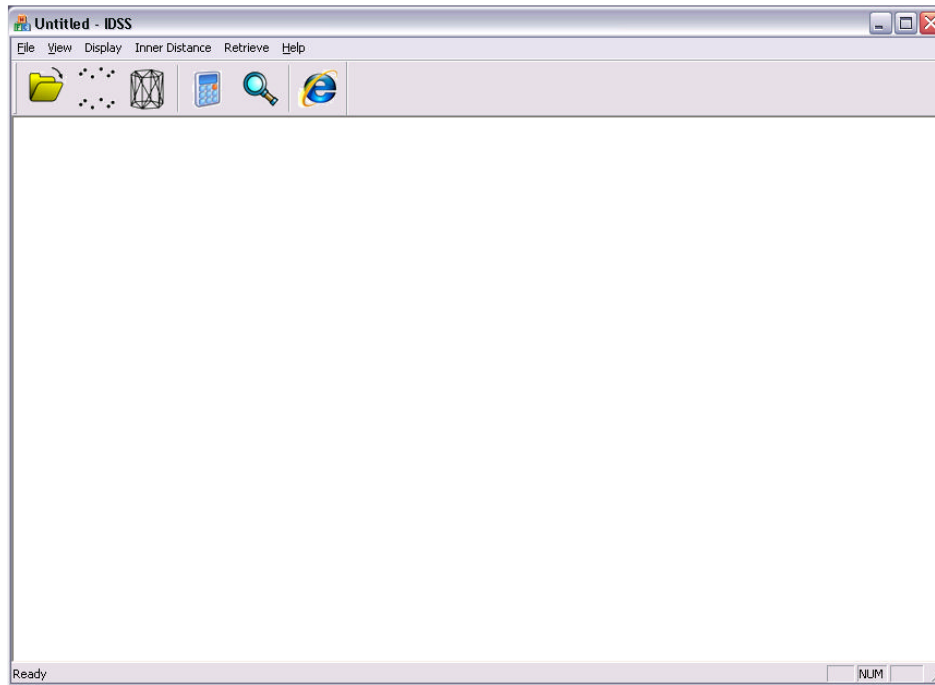
Operation

1. Uncompress 'BenchmarkProtein.zip' into the system root directory C:\

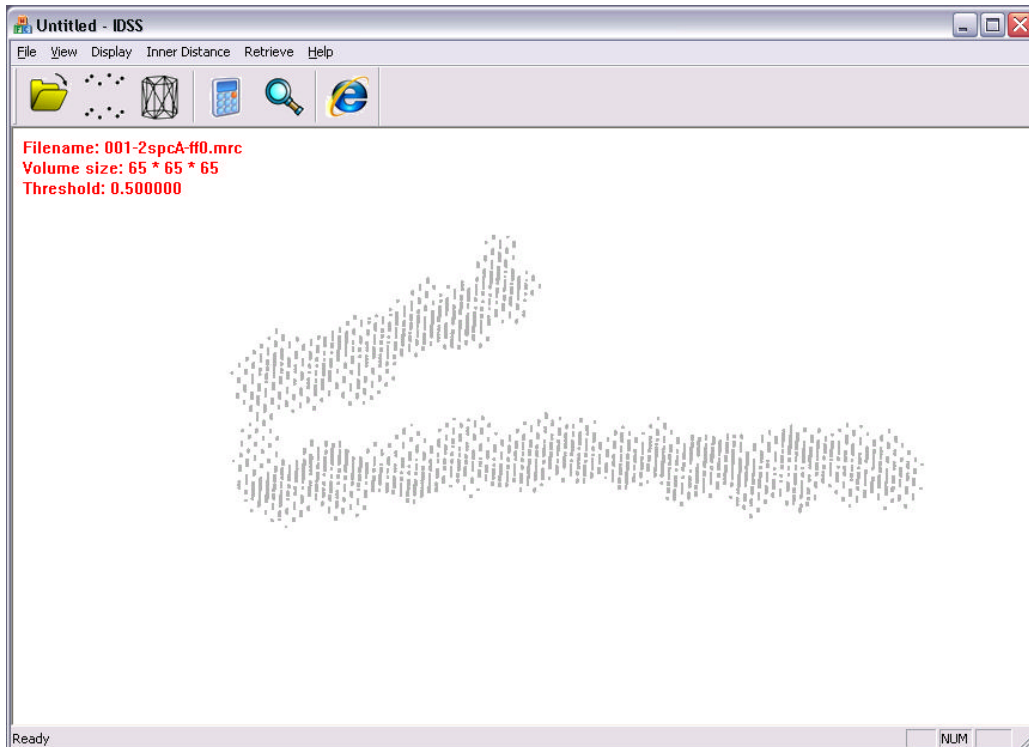For example, the directory is like as follows:

2. Execute IDSS.exe

3. Open a MRC file.

Menu: File → Open.

Or click the button [folder icon] for opening MRC file.

For example, open a MRC example file '**001-2spcA-ff0.mrc**' attached.

The left-top text shows the names of PDB files and the number of total residues:

3. Compute IDSS of a flexible protein

Menu: Inner Distance → **Compute Inner Distance**,

Or click the button for computing Inner Distance.

After choosing '**Compute Inner Distance',** the following dialog box appears with two extra parameters: (a) '**Sample number**' is the number of sample points and (b) '**Sample methods**' contain '**Random**' and '**Uniform**'.
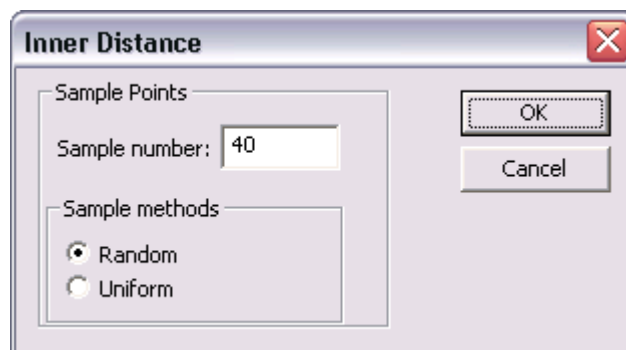
Note: IDSS computes the inner distance between all point pairs, so the algorithm complexity is $O(m^3)$, m is the number of sample points. We suggest you choose a small number, such as *20~100* for showing the fast demo. We have pre-computed the signatures of all proteins in the database with *m=500* using a console program.

After click '**OK'**, finish **Inner Distance** computation.
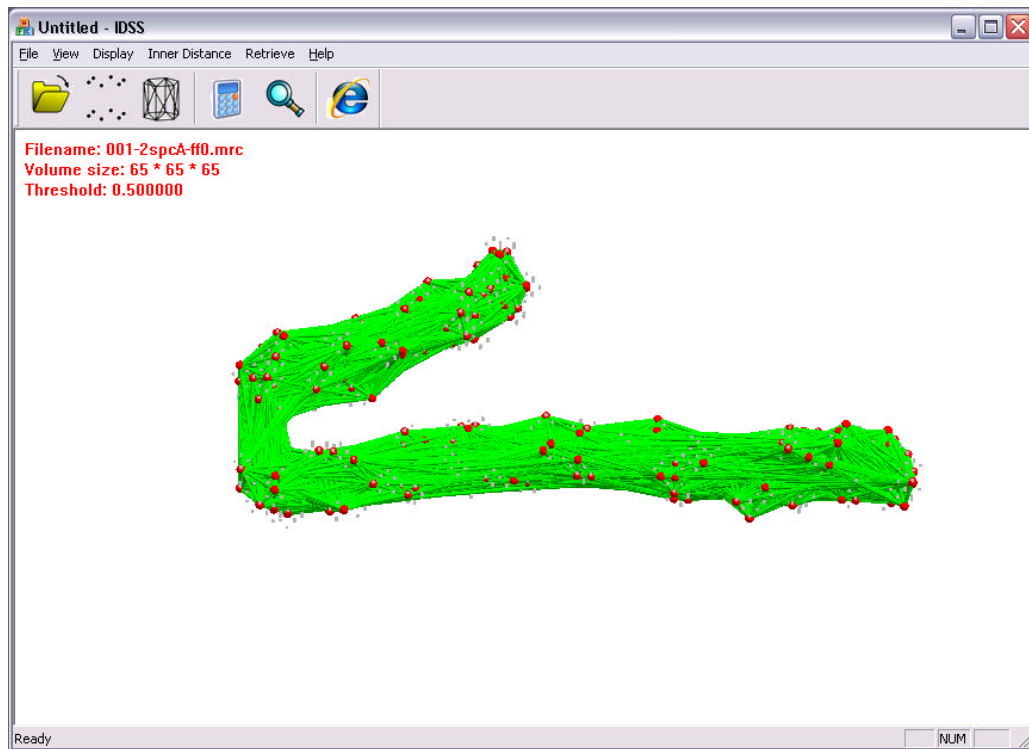
Note: the red points are the sample points and the green lines are the inner distance between all sample point pairs.

The left-top text shows the input model information as follows

where '**Filename'** is the input MRC file name, **'Volume size'** is the size of input MRC volume model, and '**Threshold'** is the threshold of density map.

4. Optionally, retrieve the similar protein in the database.

Menu: Retrieve → ID Retrieve

Or click the button to retrieve the similar proteins in the database ('**C:\BenchmarkProtein**').

For example, the query protein is retrieved as follows:
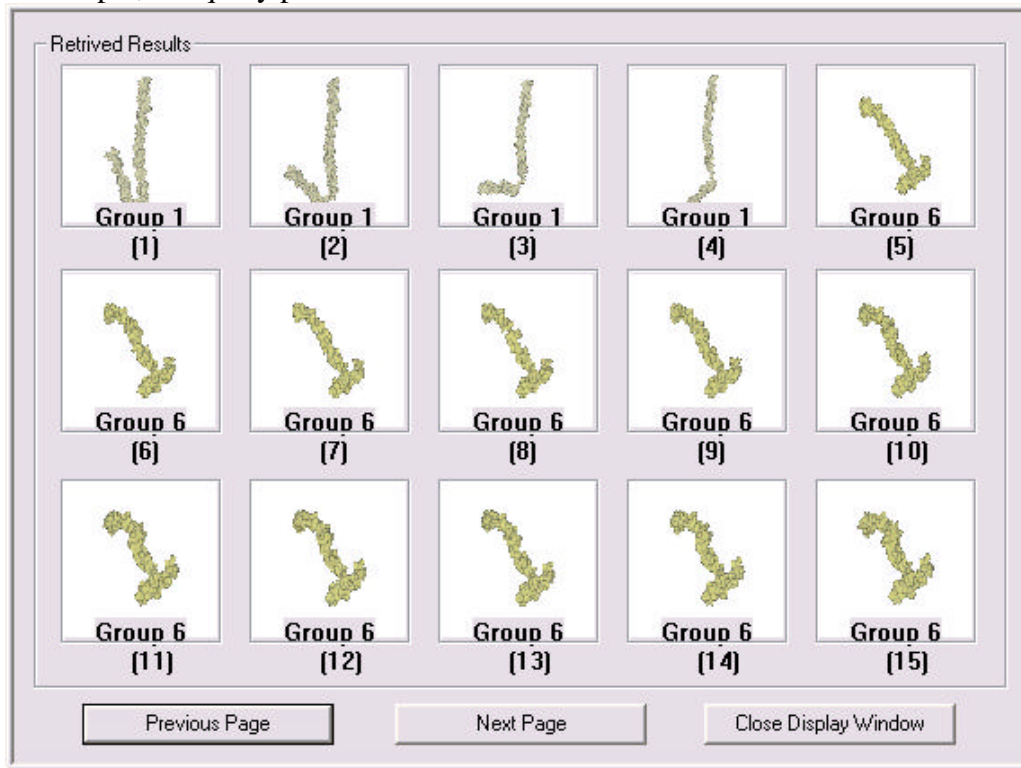
Here, the query example '**001-2spcA-ff0.mrc**' from **Group 1** returns first the full four proteins from the same group. Note that the retrieved results contain different deformation shapes.

Note: We have pre-calculated all inner distance of queries on the database and display them using images in the dialog box of retrieval results. All ID signatures are stored in 'ID.idx'. The group names are kept in 'types.txt', and 'PRcurves.txt' is the average precision-recall curves.

In addition, the user can open the retrieved result by double clicking the image in the dialog box.

5. Mouse:

 Left: rotation
 Right: Zoom in/out

6. Menu → Display for assistant displaying

For example: Menu →Display→Border Points: show border points of volume (default).

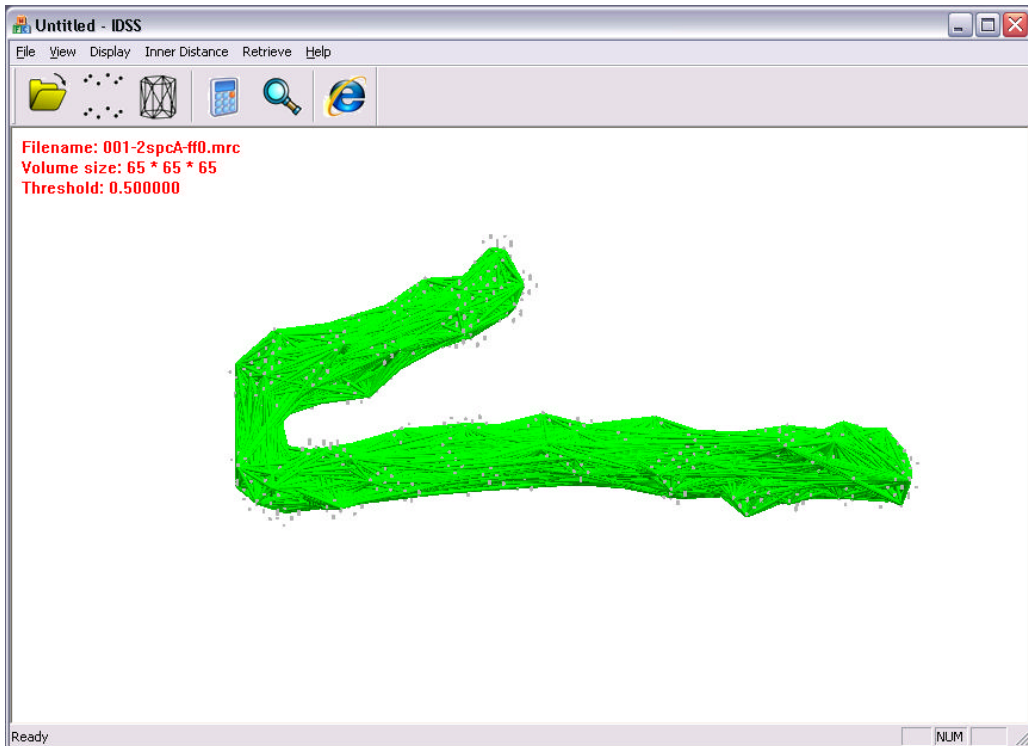Menu →Display→Full Points: show full points including border and inside points.
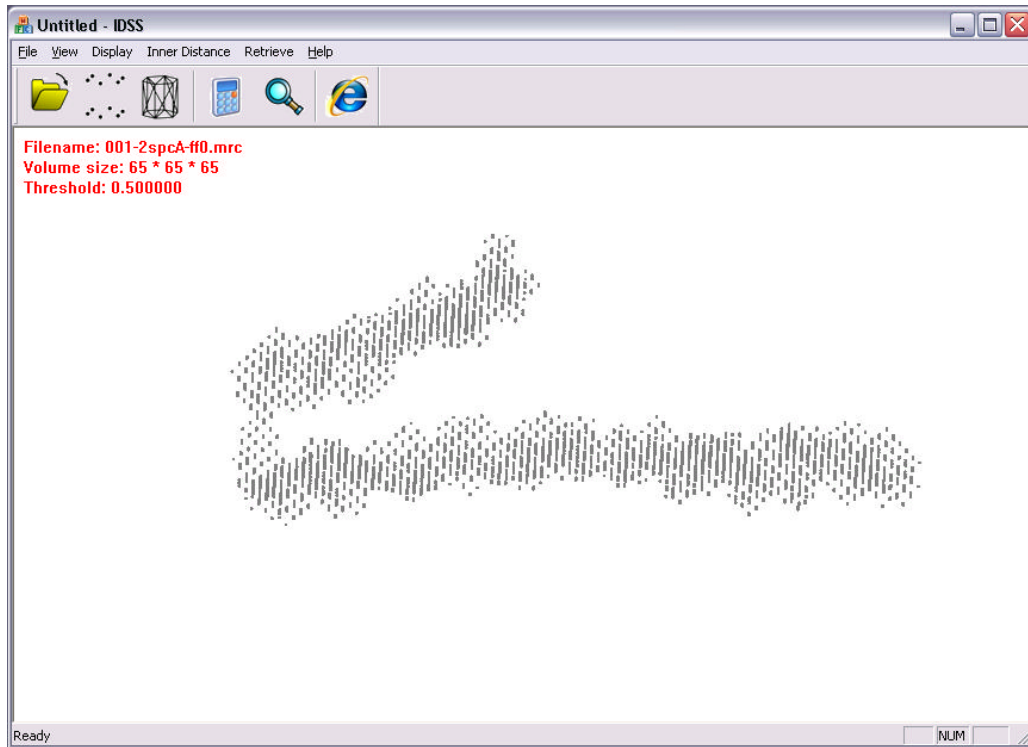
Menu: Display➔Surface: show the iso-surface of volume based on Marching Cube.
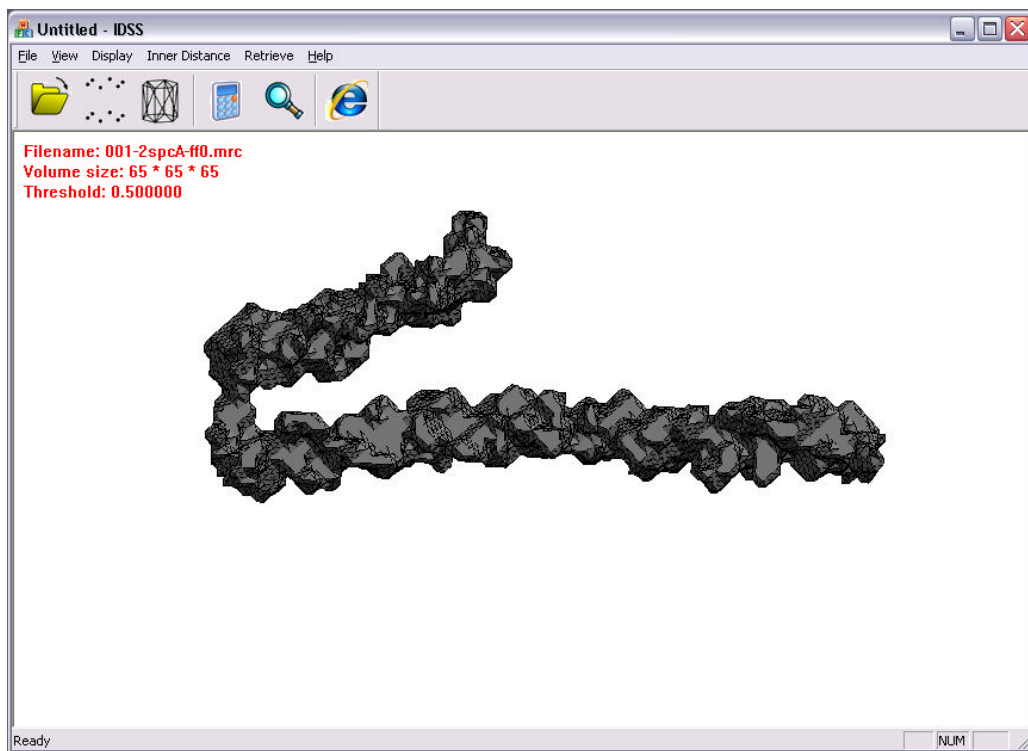
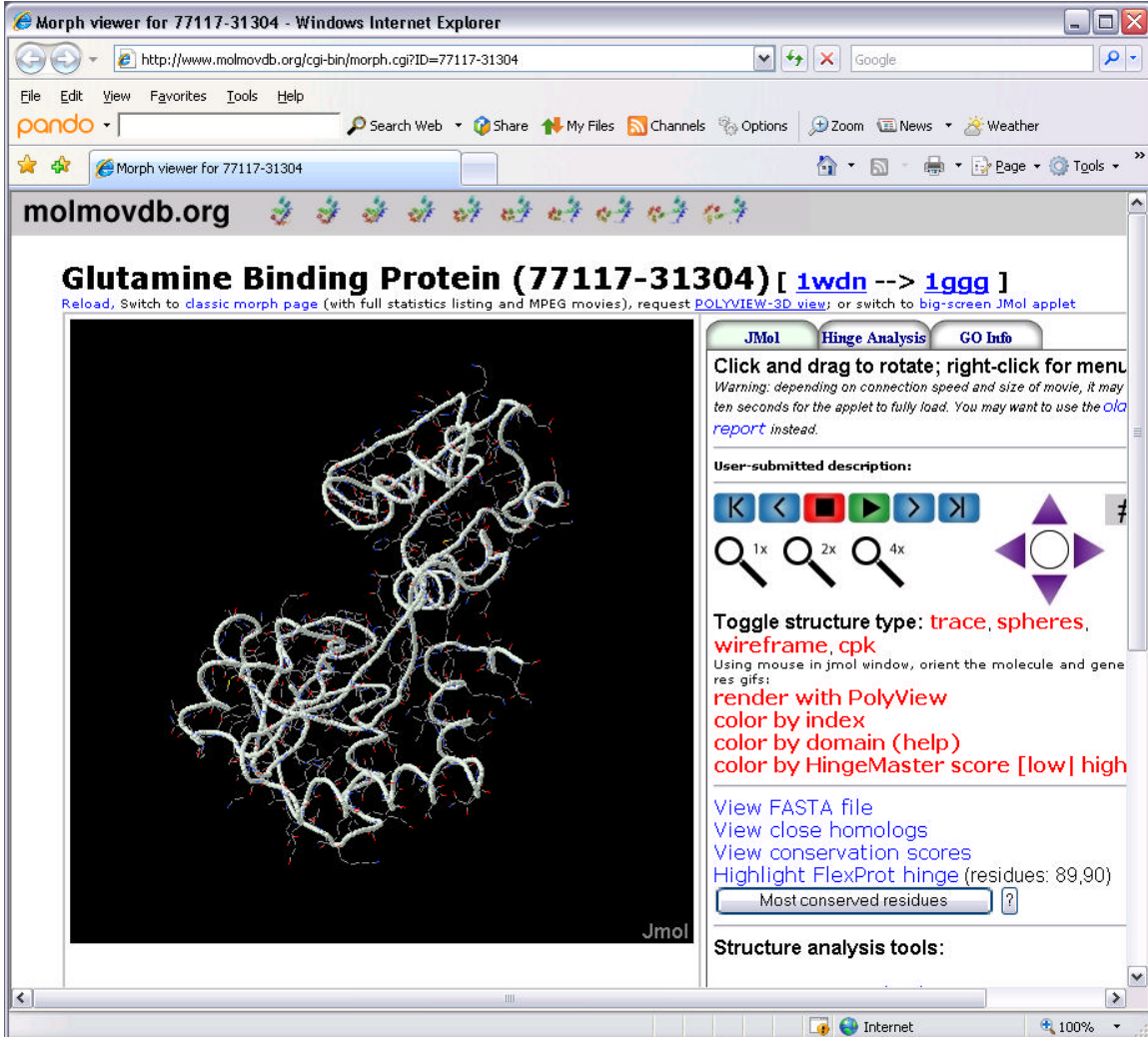Menu: Display➔Sample Points or click the button ⠿ : show/hide the Sample Points.
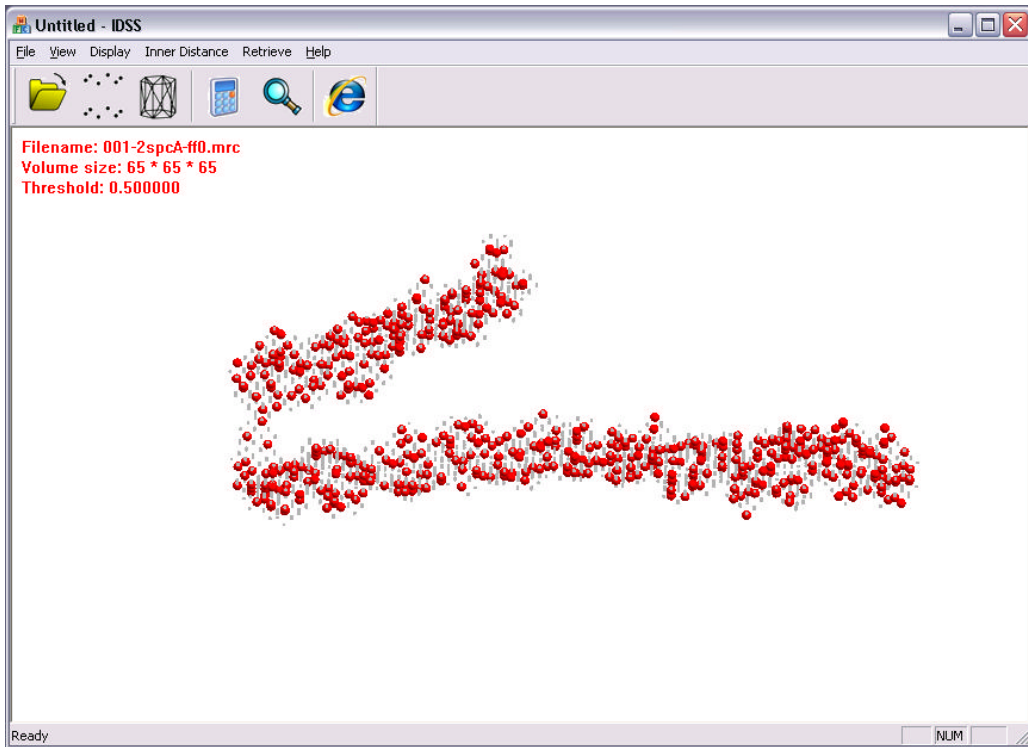
Menu: Display→Line Visibility or click the button ⬡ : show/hide the Inner Distance.

7. Menu ➔ Retrieve ➔ Link Website

Or click the button : link the website in the MolMovDB database to see how the protein deformation works.

For example, we open the file "77117-31304-ff0.mrc" and click this button and link the following website, where the PDB ID of the deformation protein is given.