AbIni3D

AbIni3D - *Ab inition* folding.

Problem: The program is intended for calculating 3D structure of proteins, provided that 3D structures of individual parts (fragments) of the protein are known, while phi and psi angles between the fragments should be found. This problem may arise when constructing a protein structure from fragments, whose structures were obtained using the search for homology of their primary sequences.

Method: The angles are calculated by genetic algorithm. The target optimization function is comprised by two additive contributions: (a) energy of the short-range interaction between the fragments and (b) the energy of phi/psi angles constructed basing on statistics of the angles between fragments of secondary structures in protein 3D structures from PDB database.

Results: Testing using seven natural proteins (with lengths from 58 to 135 aa; each protein consisted of several fragments) demonstrated that the program restores the native structure with a mean accuracy of 5.3.6.7 A. The prediction accuracy depends on individual protein and program operation mode: for three best proteins, the mean value of RMSD between the restored and native structures over ten runs amounted to 1.9, 2.3, and 2.6 A.

HELP in questions and answers on the AbIni3D program

Q: For what purpose the program is intended?

A: For calculating protein spatial structures basing on the fragments of whole structure that can be obtained by use of search for homology.

Q: How are the fragments selected?

A: Fragments of protein sequence (homologous regions) should be selected so that they would completely span the whole sequence of the target protein and, on the other hand, should not overlap. The program joins the fragments into a single chain and by use of genetic algorithm, optimizes phi and psi angles at the sites where the fragments were joined to find the conformation displaying a minimal energy.

Q: What are the launching parameters, input, and output formats?

A: The program has two mandatory parameters and one optional: these are the input COV file, output PDB file, and optional parameter-the number of computing cycles for genetic algorithm (default value, 500).

Q: How the run-time should be selected?

A: This depends on the number of fragments-more fragments require a longer run-time. For example, 50 cycles are sufficient for optimizing two fragments.

Q: What is the input COV format?

A: This is a specialized format for the program in question containing information on the primary structure of the fragments, alignments for covering of the target sequence, and "pieces" of PDB files corresponding to the covering fragments.

   Example:

```
========================================================================

*****  SET 1 *****
>1NDDB qb=0 pb=25 le=20 Sc=98.9
aaaa            bbbbb
MSANFTDKNGRQSKGVLLLR
IKERVEEKEGIPPQQQRLIY
aaaaaaaaa       bbbbb
ATOM     794  N   ILE B 126      37.162  -0.022  40.293  1.00 12.67           N
ATOM     795  CA  ILE B 126      35.962  -0.674  39.781  1.00 11.72           C
ATOM     796  C   ILE B 126      35.671  -0.073  38.399  1.00 12.39           C
ATOM     797  O   ILE B 126      35.366  -0.799  37.452  1.00 14.47           O
ATOM     798  CB  ILE B 126      34.746  -0.424  40.696  1.00 13.18           C
ATOM     799  CG1 ILE B 126      35.033  -0.951  42.107  1.00 14.02           C
ATOM     800  CG2 ILE B 126      33.499  -1.074  40.094  1.00 15.53           C
ATOM     801  CD1 ILE B 126      33.908  -0.706  43.107  1.00 14.94           C
ATOM     802  N   LYS B 127      35.806   1.249  38.282  1.00 11.60           N
ATOM     803  CA  LYS B 127      35.581   1.929  37.006  1.00 11.37           C


....    ...  ..  ... . ...      ......  .....  ......  .... .....            .


ATOM     964  CZ  TYR B 145      25.681  -2.498  47.587  1.00 17.99           C
ATOM     965  OH  TYR B 145      25.481  -3.704  48.220  1.00 20.22           O
>2PDZA qb=20 pb=31 le=17 Sc=93.1
b
TLAMPSDTNANGDIFGG
```

```
KIFKGLAADQTEALFVG
b    aaaa
ATOM    498  N   LYS A  32      -1.097 -3.476 -1.916  1.00  0.00           N
....     ...  ..  ... . ...      ......   .....   ......   ....  .....                .
TER
```
===
There may be several variants of coverings (SETs); therefore, each new variant starts from the corresponding keyword, for example, "SET 1"; next, "SET 2"; etc.

Q: How is it possible to create a COV file?

A: The file mandatory starts with the keyword "SET" with any number, for example, 1, 2, etc., followed one after another by the "pieces" of spatial structures in PDB format. The fragments are separated from one another by an empty string.

Example: suppose, you want to "disrupt" the native structure of a protein (and you have this structure in PDB format) to test then how it will be restored using this program. For this purpose, copy your PDB file, for example, YourProtein.pdb, into the file with a name, for example, YourProtein.cov, and introduce the corresponding changes:

- Put the text, for example, " SET 1 ", into the first string (it is important that the first string would contain the word SET in capitals) and

- Add empty strings at the points where you want to destroy the protein structure (i.e. break the conformation of the main chain); several breaks (empty strings) are recommended, for example, tree-five.

```
     Example:
******* SET 1 *******
REMARK   MSI WebLab Viewer PDB file
REMARK   Created:  Fri Oct 25 07:58:42 ‡€Њтр'™ЬЊ'  LЋ›  (ħ›˜') 2002
CRYST1   57.810   29.700  106.090  90.00 101.99  90.00 A2
ATOM     1  N   GLY A   1      15.740  11.178 -11.733  1.00  0.00
ATOM     2  CA  GLY A   1      15.234  10.462 -10.556  1.00  0.00
ATOM     3  C   GLY A   1      16.284   9.483  -9.998  1.00  0.00
ATOM     4  O   GLY A   1      17.150   8.979 -10.709  1.00  0.00
....     ...  ..  ... . ...      ......   .....   ......   ....  .....
ATOM   310  N   LEU A  40       6.658  -4.909  19.830  1.00  0.00
ATOM   311  CA  LEU A  40       6.751  -5.839  20.961  1.00  0.00
ATOM   312  C   LEU A  40       5.510  -6.747  21.050  1.00  0.00
ATOM   313  O   LEU A  40       5.642  -7.969  21.132  1.00  0.00
ATOM   314  CB  LEU A  40       6.968  -5.086  22.286  1.00  0.00
ATOM   315  CG  LEU A  40       7.926  -5.898  23.179  1.00  0.00
ATOM   316  CD1 LEU A  40       8.886  -4.973  23.944  1.00  0.00
ATOM   317  CD2 LEU A  40       7.121  -6.784  24.145  1.00  0.00
           // Empty line - a point of a break
ATOM   318  N   GLU A  41       4.357  -6.093  21.040  1.00  0.00
ATOM   319  CA  GLU A  41       3.066  -6.778  21.082  1.00  0.00
ATOM   320  C   GLU A  41       2.967  -7.863  19.997  1.00  0.00
ATOM   321  O   GLU A  41       2.821  -9.046  20.315  1.00  0.00
ATOM   322  CB  GLU A  41       1.903  -5.775  20.992  1.00  0.00
ATOM   323  CG  GLU A  41       1.986  -4.741  22.132  1.00  0.00
ATOM   324  CD  GLU A  41       0.577  -4.464  22.689  1.00  0.00
ATOM   325  OE1 GLU A  41      -0.227  -5.435  22.661  1.00  0.00
ATOM   326  OE2 GLU A  41       0.371  -3.298  23.120  1.00  0.00
TER
```

Parameters:

Input	
Data	*.cov file, containing one or more sets of protein fragments

Options	
Number of Sets	Protein fragments sets number
Number of Steps	Number of cycles of optimisation (usually 100 - 1000).