NAME

seq2itp.pl

Generate a MARTINI coarse-grained topology for proteins.

USAGE

seq2itp.pl [options] protein.seq protein.ssd

DESCRIPTION

A perl script to build coarse-grained MARTINI topology is generated from the aminoacid sequence and secondary structure information by selecting matching bead types, correct force constants and appropriate constraints. As input it requires the amino acid sequence (.seq file) and the secondary structure information (.ssd) file. Use the -h option to see the help.

ARGUMENTS

Protein sequence (protein.seq)

The aminoacid sequence of the protein, in FASTA format. The FASTA format used for the aminoacid sequence file has two main elements: chain identifiers and aminoacid sequences. Each chain is defined as a separate entity containing one line with the chain identifier and a second line with the aminoacid sequence of the residues in the chain. E.g.:

```
>1PGA:A

MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE
```

Secondary structure (protein.ssd)

The secondary structure of the residues in SSDUMP format (see for example the program do_dssp in GROMACS). SSDUMP format is used to define the secondary structure of the residues. The first line has the number of residues in the whole protein and the second line has either H, E or C for each residue marking, accordingly, either helix, extended or coil structure. E.g.:

```
56

    EEEEEEEEECCEEEEEEEEECCHHHHHHHHHHHHHHCCCCCEEEEECCCEEEEEEE
```

MODIFICATIONS

The script has been modified in order to include the parameters for hydroxyproline residues and for the collagen triple helical secondary structure. In the protein sequence input file (protein.seq) hydroxyproline residues must be indicated with the code "O". In the secondary structure input file (protein.ssd) triple helical conformation must be indicated by the code "D".