Description of the DSSP program

DSSP was written by Wolfgang Kabsch and Chris Sander

Function

Definition of secondary structure of proteins given a set of 3D coordinates.

Description

The DSSP program defines secondary structure, geometrical features and solvent exposure of proteins, given atomic coordinates in Protein Data Bank format. The program does NOT PREDICT protein structure. According to the Science Citation Index (July 1995), the program has been cited in the scientific literature more than 1000 times.

Authors of the DSSP method

Wolfgang Kabsch and Chris Sander, MPI MF, Heidelberg, 1983.
Reference: Kabsch,W. and Sander,C. (1983) Biopolymers 22, 2577-2637

Usage and command line options

```
dssp [-na] [-v] pdb_file [dssp_file]
dssp [-na] [-v] -- [dssp_file]
dssp [-h] [-?] [-V]
```

Command line options:

-na
: Disables the calculation of accessible surface.

-c
: Classic (pre-July 1995) format.

-v
: Verbose.

--
: Read from standard input.

-h -?
: Prints a help message.

-l
: Prints the license information.

-V
: Prints version, as in first line of the output.

Examples

In this example verbose mode was turned on to see the progress of execution for the large *photoreaction center* (1prc) input file.

```
unix% dssp -v 1prc.pdb 1prc.dssp
 !!! Backbone incomplete for residue ALA  333 C
    residue will be ignored !!!

 !!! Residue SER  273 L has  3 instead of expected   2 sidechain atoms.
    last sidechain atom name is  OXT
    calculated solvent accessibility includes extra atoms !!!

 !!! Residue LYS  323 M has  6 instead of expected   5 sidechain atoms.
    last sidechain atom name is  OXT
    calculated solvent accessibility includes extra atoms !!!

 !!! Residue LEU  258 H has  5 instead of expected   4 sidechain atoms.
    last sidechain atom name is  OXT
    calculated solvent accessibility includes extra atoms !!!

 !!! Polypeptide chain interrupted !!!
Inputcoordinates done        1189
```

```
    Flagssbonds done
    Flagchirality done
    Flaghydrogenbonds done
    Flagbridge done
    Flagturn done
    Flagaccess done
    Printout done
```

Output file is 1ppt.dssp

In this example the coordinates of *avian pancreatic polypeptide* (1ppt) were first converted from star format to pdb format and then piped into dssp.

```
    unix% star2pdb 1ppt.star | dssp -- > 1ppt.dssp
     !!! Residue TYR   36   has  9 instead of expected   8 sidechain atoms.
        last sidechain atom name is  OXT
        calculated solvent accessibility includes extra atoms !!!
```

Output file is 1ppt.dssp

Output

The output from DSSP on file myprotein.dssp contains secondary structure assignments and other information, one line per residue. Extract from 1est.dssp (simplified):

```
HEADER    HYDROLASE   (SERINE PROTEINASE)         17-MAY-76   1EST
...
  240  1  4  4  0 TOTAL NUMBER OF RESIDUES, NUMBER OF CHAINS,
                   NUMBER OF SS-BRIDGES(TOTAL,INTRACHAIN,INTERCHAIN)               .
 10891.0   ACCESSIBLE SURFACE OF PROTEIN (ANGSTROM**2)
  162 67.5   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(J)  ; PER 100 RESIDUES
    0  0.0   TOTAL NUMBER OF HYDROGEN BONDS IN     PARALLEL BRIDGES; PER 100 RESIDUES
   84 35.0   TOTAL NUMBER OF HYDROGEN BONDS IN ANTIPARALLEL BRIDGES; PER 100 RESIDUES
...
   26 10.8   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+2)
   30 12.5   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+3)
   10  4.2   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+4)
...
  #  RESIDUE AA STRUCTURE BP1 BP2  ACC   N-H-->O  O-->H-N  N-H-->O  O-->H-N
    2   17   V  B 3   +A  182   0A   8  180,-2.5 180,-1.9   1,-0.2 134,-0.1

                                    TCO  KAPPA ALPHA  PHI   PSI    X-CA   Y-CA   Z-CA
                                  -0.776 360.0   8.1 -84.5 125.5  -14.7   34.4   34.8

....;....1....;....2....;....3....;....4....;....5....;....6....;....7..
    .-- sequential resnumber, including chain breaks as extra residues
    |    .-- original PDB resname, not nec. sequential, may contain letters
    |    |   .-- amino acid sequence in one letter code
    |    |   |  .-- secondary structure summary based on columns 19-38
    |    |   |  | xxxxxxxxxxxxxxxxxxx recommend columns for secstruc details
    |    |   |  | .-- 3-turns/helix
    |    |   |  | |.-- 4-turns/helix
    |    |   |  | ||.-- 5-turns/helix
    |    |   |  | |||.-- geometrical bend
    |    |   |  | ||||.-- chirality
    |    |   |  | |||||.-- beta bridge label
    |    |   |  | ||||||.-- beta bridge label
    |    |   |  | |||||||   .-- beta bridge partner resnum
    |    |   |  | |||||||   |   .-- beta bridge partner resnum
    |    |   |  | |||||||   |   |.-- beta sheet label
    |    |   |  | |||||||   |   ||   .-- solvent accessibility
    |    |   |  | |||||||   |   ||   |
  #  RESIDUE AA STRUCTURE BP1 BP2  ACC
    |    |   |  | |||||||   |   ||   |
   35   47   I  E     +     0   0    2
```

```
  36   48   R  E >  S- K   0  39C  97
  37   49   Q  T 3  S+     0   0   86    (example from 1EST)
  38   50   N  T 3  S+     0   0   34
  39   51   W  E <   -KL  36  98C   6
```

Line length of output is 13x characters. Lines end in a number or a period.

Histograms:
the number 2 under column ’8’ in line ’residues per alpha helix’ means: there are 2 alpha helices of length 8 residues in this data set.

For definitons, see above BIOPOLYMERS article.

In addition note:

Each line contains the following residue information

RESIDUE

two columns of residue numbers. First column is DSSP’s sequential residue number, starting at the first residue actually in the data set and including chain breaks; this number is used to refer to residues throughout. Second column gives crystallographers’ ’residue sequence number’,’insertion code’ and ’chain identifier’ (see protein data bank file record format manual), given for reference only.

AA

one letter amino acid code, lower case for SS-bridge CYS.

S (first column in STRUCTURE block)

compromise summary of secondary structure, intended to approximate crystallographers’ intuition, based on columns 19-38, which are the principal result of DSSP analysis of the atomic coordinates.

BP1 BP2

residue number of first and second bridge partner followed by one letter sheet label

ACC

number of water molecules in contact with this residue *10. or residue water exposed surface in Angstrom**2.

N-H-->O etc.

hydrogen bonds; e.g. -3,-1.4 means: if this residue is residue i then N-H of I is h-bonded to C=O of I-3 with an electrostatic H-bond energy of -1.4 kcal/mol. There are two columns for each type of H-bond, to allow for bifurcated H-bonds.

TCO

cosine of angle between C=O of residue I and C=O of residue I-1. For alpha-helices, TCO is near +1, for beta-sheets TCO is near -1. Not used for structure definition.

KAPPA

virtual bond angle (bend angle) defined by the three C-alpha atoms of residues I-2,I,I+2. Used to define bend (structure code ’S’).

ALPHA

virtual torsion angle (dihedral angle) defined by the four C-alpha atoms of residues I-1,I,I+1,I+2.Used to define chirality (structure code ’+’ or ’-’).

PHI PSI

IUPAC peptide backbone torsion angles

X-CA Y-CA Z-CA

echo of C-alpha atom coordinates

Warnings

The values for solvent exposure may not mean what you think!

- effects leading to larger than expected values: solvent exposure calculation ignores unusual residues, like ACE, or residues with incomplete backbone, like ALA 1 of data set 1CPA. it also ignores HETATOMS, like a heme or metal ligands. Also, side chains may be incomplete (an error message is written).
- effects leading to smaller than expected values: if you apply this program to protein data bank data sets containing oligomers, solvent exposure is for the entire assembly, not for the monomer. Also, atom OXT of c-terminal residues is treated like a side chain atom if it is listed as part of the last residue. also, peptide substrates, when listed as atoms rather than hetatoms, are treated as part of the protein, e.g. residues 499 s and 500 s in 1CPA.

Unknown or unusual residues are named X on output and are not checked for standard number of sidechain atoms. All explicit water molecules, like other hetatoms, are ignored.

Input file

Coordinate file in PDB format.