MANUAL for: Protein structure quality assessment based on the electrostatic profile of consecutive Cβ atoms

Sandeep Chakraborty[1], Ravindra Venkatramani [2], Basuthkar J. Rao[1], Bjarni Ásgeirsson[3], Abhaya M. Dandekar[4],

1. Department of Biological Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India.
2. Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India.
3. Science Institute, Department of Biochemistry, University of Iceland, Dunhaga 3, IS-107 Reykjavik, Iceland
4. Plant Sciences Department, University of California, Davis, CA 95616.

Installation

1. You have to use the tcsh shell (bash shell wont work).

2. You have to install the cpan perl packages from http://www.cpan.org/. The list depends on the program you use, just run the main program - and it will prompt you about the missing packages. Go ahead and install those.

3. You need to set some environment variables (SRC, BIOPERLHOME, etc). The complete list is in a file called "setup.csh" - you need to change this file and source it. This changes the $path variable too.

Preprocessing

APBS has to be run on the proteins - and the results stored in a directory which can be accessed by the variable $PDBDIR.

Running the program

1. $SRC/ALIGN/escapist.pl -outf out -con -lis list -score pd.CB.score.full

The file "list" should contain the list of PBDs which needs to be checked. These PDB files are to be kept in a directory, and the variable PDBDIR needs to be set to this directory.

Input file: Electrostatic potential difference between different pairs of amino acids barring those having glycine (which does not have a Cβ atom)

Pair–Mean——————-SD——————-numbers
DF -110.837361981036 29.84767599511064110948 907
DY -109.703122535081 30.65716717118522351879 807
DH -109.22915950426 31.50032237470258017677 439
DW -106.80439650934 29.44442391065002650445 379
EH -98.1755990328704 30.89410249532408733962 432
EY -97.159062961348 28.75101641377281812914 727
EF -94.1944678994964 28.51519207145185967702 834
EW -91.7275528184616 33.95685828822522267705 325

DR -71.3102449867486 29.27341191623873556291 935
DN -69.8036967714474 29.24535788357812180733 760
AH -67.8937378531901 29.12599002823296305470 605
AY -67.3964078153199 26.24149152913231416650 912
DP -66.5917248623077 26.04710362402731717592 520
LY -66.2911790780705 26.23124805970472690501 964
IY -66.1664280414583 26.86756754228026501120 576
VY -65.7003282594777 28.32229672982512553635 785
AW -64.4776152830808 25.80816828729334677034 396
AF -63.8736953255319 26.45118497093355173228 987
DK -63.6122738742628 28.84593765801676969297 1119
LW -63.1409032441038 25.15647035708849199605 424
VW -62.8182297987055 28.12226557481797198028 309
...
...
...
...

Output

The output gives the PDScore for each protein in the file "escapist.out". The native protein (1ASH) has the minimum score. 1ASH 19.9
1ASH_1BAB-A_R 29.3
1ASH_1BAB-B_R 29.6
1ASH_1COL-A_R 30.6
1ASH_1CPC-A_R 28.7
1ASH_1ECD_R 24.8
1ASH_1EMY_R 19.8
1ASH_1FLP_R 22.4
1ASH_1GDM_R 27.5
1ASH_1HBG_R 26.5
1ASH_1HBH-A_R 28.1
1ASH_1HBH-B_R 27
1ASH_1HDA-A_R 36.7
1ASH_1HDA-B_R 31.5
1ASH_1HLB_R 21.9
1ASH_1HLM_R 29.3
1ASH_1HSY_R 22
1ASH_1ITH-A_R 19
1ASH_1LHT_R 27.6
1ASH_1MBA_R 22.8
1ASH_1MBS_R 27.9
1ASH_1MYG-A_R 20.5
1ASH_1MYJ-A_R 29.5
1ASH_1MYT_R 26.5
1ASH_2DHB-A_R 29.9
1ASH_2DHB-B_R 27.2
1ASH_2LHB_R 23.3
1ASH_2PGH-A_R 33.2

1ASH_2PGH-B_R 27.7
1ASH_4SDH-A_R 26.1