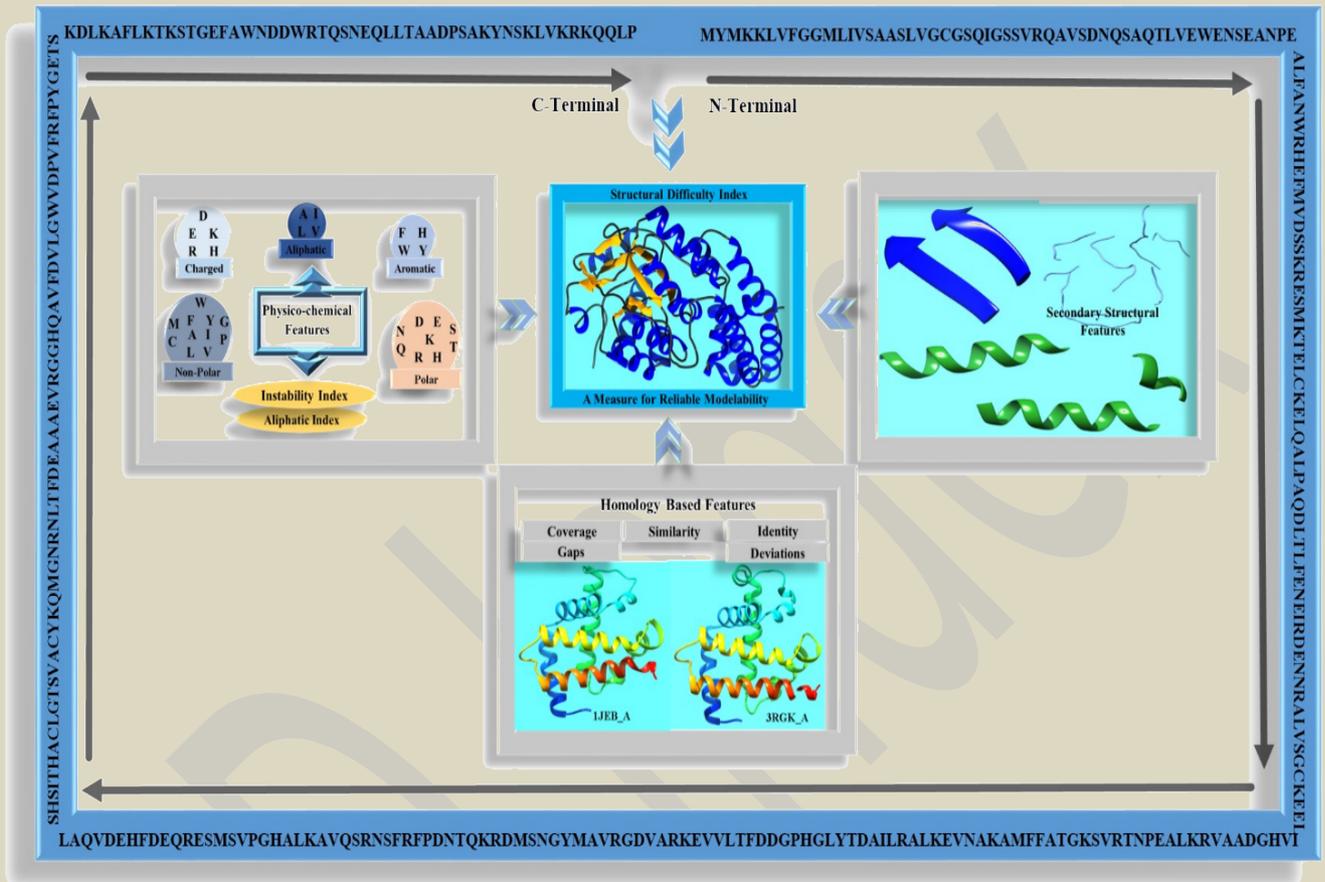
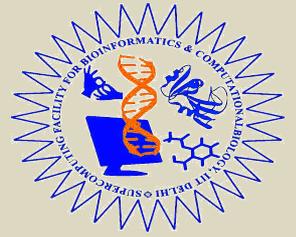


# Structural Difficulty (SD) Index: A measure for modelability of protein structures



The SD Index web utility can be freely accessed <http://www.scfbio-iitd.res.in/SDIndex/>

## **INSTALLATION AND USER GUIDE FOR SD INDEX ON LOCAL MACHINE**

### **(LINUX COMPATIBLE ONLY)**

“Structural Difficulty (SD)” index, which is derived from secondary structures, homology and physico-chemical features of protein sequences, reflects the capability of predicting good quality structures with some of the best methodologies available currently. The SD index also helps to assess the plausibility for developing proteome level structural databases for various organisms.

SD index calculation has a few dependencies which are required to be installed in mentioned folders. These are freely (or on request) available for download in public domains. The dependencies along with their links and details for download are provided below:

**I. NCBI BLAST:** BLAST+ is a suite of command-line tools to run BLAST. It can be downloaded as source code or executables from <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>. If downloaded as source code then user needs to compile as per the instructions given in user manual else the downloaded executables (according to user's machine architecture) may be copied to SDINDEX/SOURCE/BLAST/ folder directly.

**II. PSIPRED:** PSIPRED is one of the most widely used software for secondary structure prediction of proteins. It is also freely available for download in public domains. It can be downloaded from [http://bioinf.cs.ucl.ac.uk/software\\_downloads/](http://bioinf.cs.ucl.ac.uk/software_downloads/). After extracting the downloaded version, copy the entire content to SDINDEX/SOURCE/PSIPRED/.

**III. NR Database:** Non-redundant database of protein sequences can be downloaded from <ftp://ftp.ncbi.nlm.nih.gov/blast/db/>. The downloaded files after extracting should be copied to SDINDEX/SOURCE/NRDB. Warning: These files will be very large in size (~8 GB).

**IV. PDB Database:** These files can be directly downloaded from RCSB PDB website in Download section. This downloaded file can be used as input for generating database using makeblastdb tool of NCBI BLAST. This database contains all the protein sequences with their protein databank identifier and some additional information. After downloading and creating the database, copy the resultant files to SDINDEX/SOURCE/PDBs/. In the provided downloadable version already a working version of database is present but user needs to replace it with updated files to get best results.

Once all the dependencies are successfully implemented. For testing the installation run SDINDEX.sh script on TEST folder as follows:

```
sh SDINDEX.sh TEST
```

## Structural Difficulty (SD) Index: A measure for modelability of protein structures

The successful execution should generate two new files in SDINDEX/UPLOAD/TEST/ folder named as TEST.RES (detailed result file) and TEST.SDI (per residue SD Index scores).

### How to Run SDIndex Calculation

The SDIndex calculations can be performed in two modes

- i). When template information is not known
- ii). When template information is known (User defined template)

i). When the template information is not known (or no user defined template) then user need to make a folder in UPLOAD folder (e.g. UPLOAD/TEST/). This folder should have protein sequence file in fasta/plane format (e.g. UPLOAD/TEST/TEST.seq). Now user can run the main script as follows:

**sh SDINDEX.sh <Name of Folder containing sequence file> (e.g. sh SDINDEX.sh TEST)**

ii). When the template information is known (or with user defined template) then user need to make a folder in UPLOAD folder (e.g. UPLOAD/TEST/). This folder should have protein sequence file in fasta/plane format (e.g. UPLOAD/TEST/TEST.seq) and user defined template information file as shown in TEST.info. Now user can run the main script as follows:

**sh SDI\_TEMPLATE.sh <Name of Folder> (e.g. sh SDI\_TEMPLATE.sh TEST)**

In case of any problem please contact "[rahul@scfbio-iitd.res.in](mailto:rahul@scfbio-iitd.res.in)"

Please cite us if you find the SD Index server/standalone useful for your work. You can cite SD Index as

“Rahul Kaushik and B. Jayaram, Structural difficulty index: A reliable measure for modelability of protein tertiary structures, Protein Engineering Design and Selection, 2016. [doi: 10.1093/protein/gzw025](https://doi.org/10.1093/protein/gzw025)”