The SS Calc Python tool



The SS Calc Python tool calculated the secondary structure content from a CD or IR spectrum.

## Input files

All input files must be in the format of either a tab delimited ASCII or a CSV file with two columns

* first column is the wavelength (CD) / k value (IR)
* second column is Molar Extinction (Δε) for CD or absorbance for IR. For IR spectra, the absorbance will be re-scaled such that the maximum absorbance becomes 1.

For CD data the highest wavelength must be at least 240 nm to fit the reference datasets. The lowest wavelength is determined by the program, and the data are interpolated into 1 nm steps. The wavelength may be in either ascending or descending order

For IR data the interval from 1600 to 1800 cm-1 must be included in the file. All k values must be in ascending order (low to large k values). The data will be interpolated to fit the IR reference data in 1 cm-1 steps.

## Running the code

The Python 3 code to run is:

SSCalcPyGUI.py

The code needs Numpy, MatPlotLib and PyQt5 installed to run (as well as Python of course). It is recommended that the newest version of Anaconda3 (<https://www.anaconda.com/>) is installed to ensure all these libraries are available

With Anaconda3 installed, open Spyder and drag in SSCalcPyGUI.py. Run this file (F5 or press the Run file button ).

If you know where the python.exe is installed, e.g. in C:\ProgramData\Anaconda3, then it is possible to run the program in windows with a batch file. One is included in the program distributed runSSCalcPyGUI.bat:

C:\ProgramData\Anaconda3\python.exe SSCalcPyGUI.py

pause

## Reference data and secondary structure calculation methods

Two methods are implemented in SSCalcPy:

Selcon3

CDSSTR

Both may be chosen for calculations on CD data. For now only Selcon3 is available for calculations on IR data.

The reference data for the calculations included are:

1. The SP175 (71 proteins) and SMP180 (128 proteins) data sets curated at Aarhus University from the PCDDB data bank (SP\_AU-PCDDB and SMP\_AU-PCDDB). Here the DSSP routing is used for secondary structure content extraction with the following scheme:

Alpha Helix: H and G

Beta Sheet: E

Turns: T

Other: All other DSSP groups

The Alpha Helix content is grouped into a distorted group (AlphaD) where the first 2 and last 2 residues in a helix are included and a regular group (AlphaR) where the central residues, not in the distorted group, are included. Residues in Helixes shorter than 4 residues are all counted as distorted.

Similar the Beta sheet content is grouped into distorted (BetaD) and regular (BetaR), but with the difference that only the first and last residue are counted as distorted

1. The IR data set from ULB in Brussels, Belgium, a set of 50 protein spectra from thin films on an ATR crystal. All spectra are normalized such that the largest absorbance is rescaled to 1.

The SP175 and SMP180 datasets and the PCDDB should be referenced when using the SSCalcPy toolkit by the following publications:

SP175:

A reference database for circular dichroism spectroscopy covering fold and secondary structure space.

Jonathan G. Lees, Andrew J. Miles, Frank Wien and B. A. Wallace. Bioinformatics 22:1955-1962 (2006)

<https://doi.org/10.1093/bioinformatics/btl327>

SMP180:

A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy.

Ali Abdul-Gader, Andrew John Miles and B.A. Wallace. Bioinformatics 27:1630-1636 (2011)

<https://doi.org/10.1093/bioinformatics/btr234>

PCDDB:

The PCDDB (protein circular dichroism data bank): A bioinformatics resource for protein characterisations and methods development.

Sergio Gomes Ramalli, Andrew John Miles, Robert W. Janes and B.A. Wallace. Journal of Molecular Biology 434:167441 (2022)

<https://doi.org/10.1016/j.jmb.2022.167441>

The IR ATR dataset should be referenced when using the SSCalcPy toolkit by the following publication:

SOMSpec as a General Purpose Validated Self-Organising Map Tool for Rapid Protein Secondary Structure Prediction From Infrared Absorbance Data.

Marco Pinto Corujo, Adewale Olamoyesan, Anastasiia Tukova, Dale Ang, Erik Goormaghtigh, Jason Peterson, Victor Sharov, Nikola Chmel and Alison Rodger. Front. Chem. 9:784625 (2022)

<https://doi.org/10.3389/fchem.2021.784625>