The SS Calc Python tool



The SS Calc Python tool calculates the secondary structure content from a CD or IR spectrum. When using this tool, please reference the following paper:

Hoffmann SV, Jones NC, Rodger A.  
Protein secondary structure determined from independent and integrated infra-red absorbance and circular dichroism data using the algorithm SELCON.   
QRB Discovery. Published online 2025.  
<https://doi.org/10.1017/qrd.2025.4>

This readme file contains the following sections:

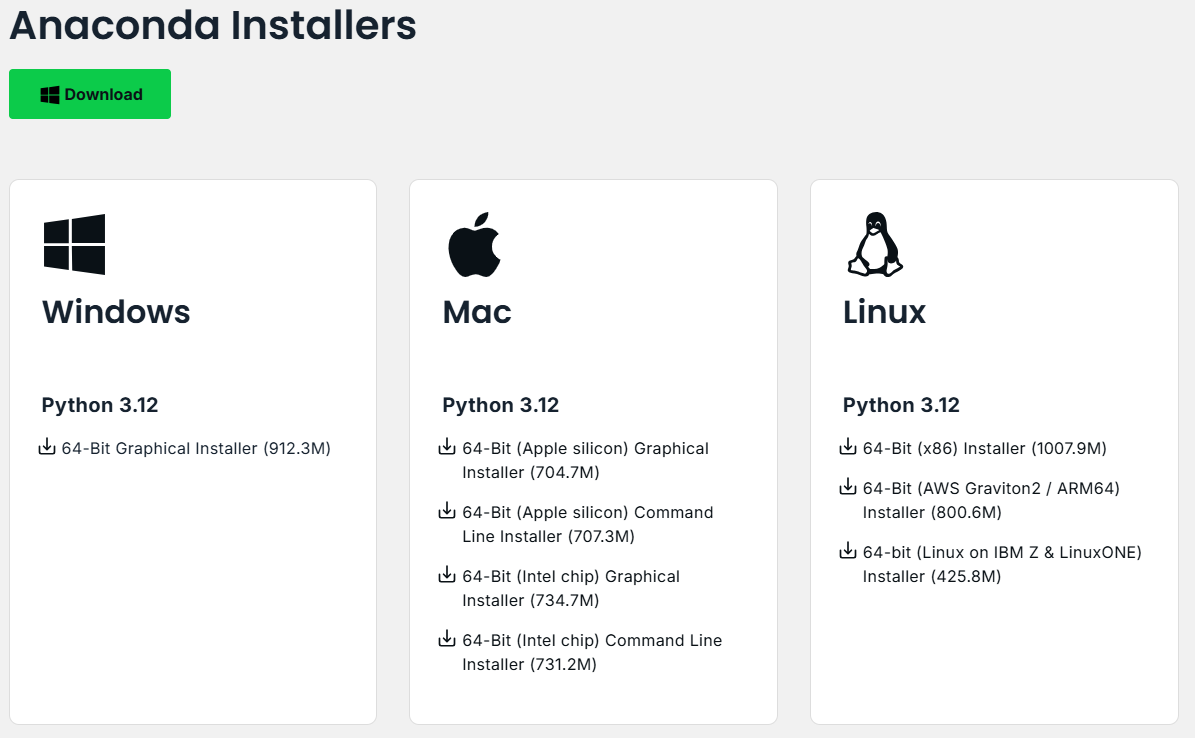
* Prerequisites
* Running the code for SSCalc
* GUI interface to the code
* Input files
* Reference data and secondary structure calculation methods

## Prerequisites

To be able to use the SSCalc Python tool, you need to be able to run Python code with several different libraries (Numpy, MatPlotLib and PyQt5) installed.

If you do not already have Python installed, we recommend that you download and install the latest version of Anaconda3 from <https://www.anaconda.com/download>.

The Anaconda software is free, but you may need to create an account to be able to download it (there is currently an option to "Skip registration", but may change in the future). Installation files are available for several operating systems.



Once this is installed on your computer you should have all of the necessary libraries to run the SSCalc Python tool, as well as "Spyder" a console to be able to edit and run the code if needed.

## Running the code for SSCalc

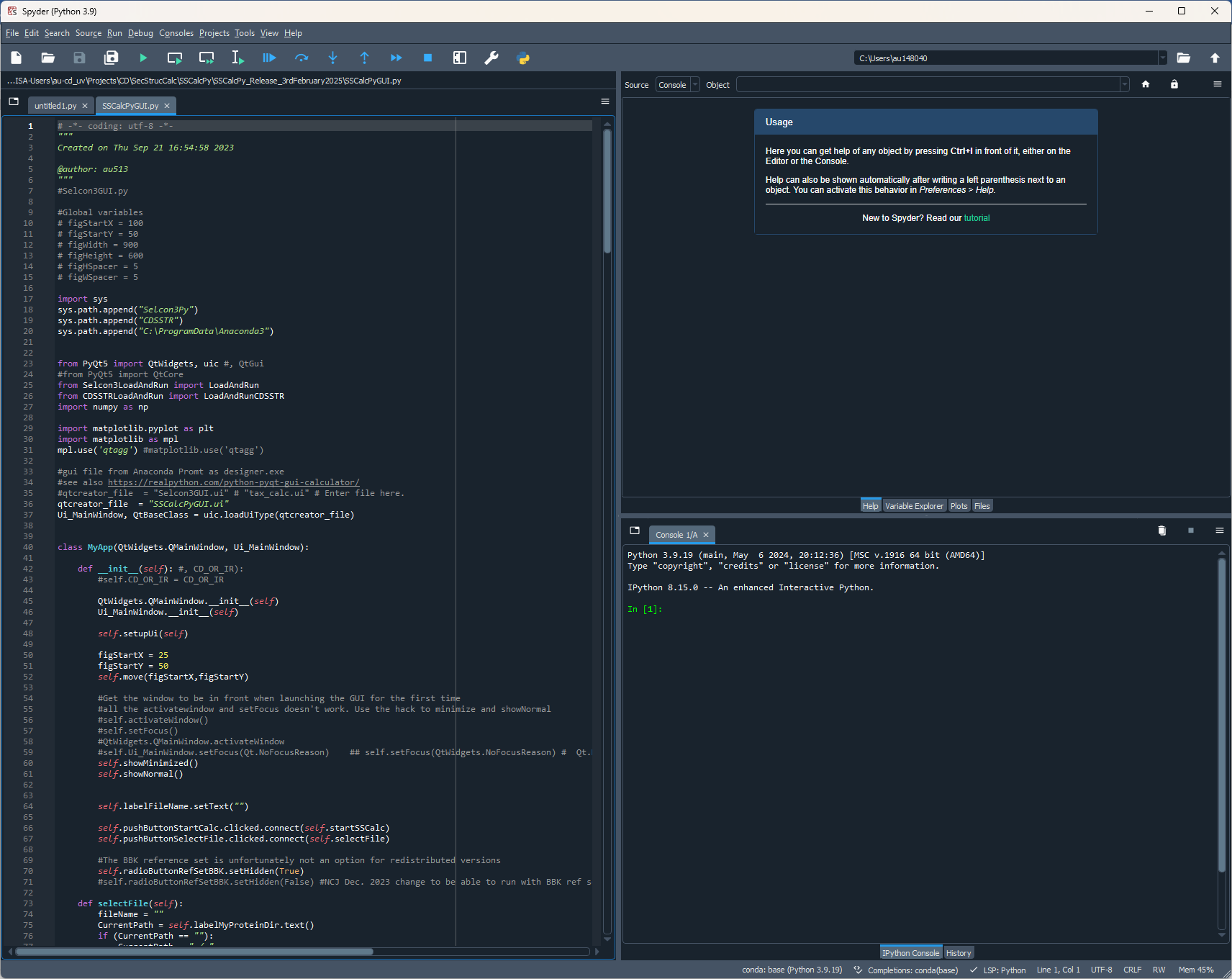
In the SSCalc folder there is a batch file runSSCalcPyGUI.bat. If you carried out installation of Anaconda under Windows to the standard location i.e. in folder C:\ProgramData\Anaconda3\, then double clicking this bat file should open the SSCalc python GUI interface and you are ready to start fitting data (see below).

However, if this is not the path where "python.exe" is found, then in order to use a bat file, it needs to be updated so that the underlined portion of the path (below) is correct for your installation.

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

pause

Alternatively, you can open "Spyder", which should have been installed with Anaconda.



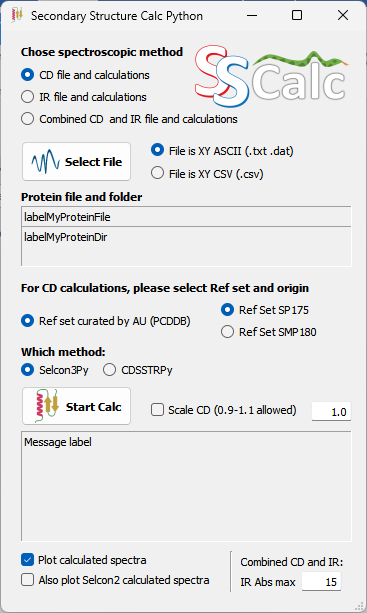
From Spyder the SSCalc Python 3 code file to run is:

SSCalcPyGUI.py

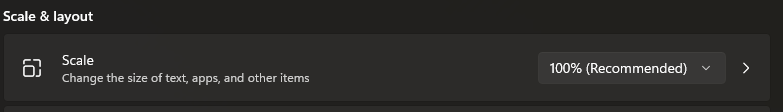
Drag in or browse to the folder containing all of the SSCalc files to open the file SSCalcPyGUI.py. Run this file (F5 or press the Run file button ).

## GUI interface to the code

With successful opening of the SSCalc Py tool, you should see the following window.



To view the window and the text correctly, it is recommended that you set the scaling for viewing texts and graphics on your computer to 100% (e.g. the "Scale and Layout" option under settings in Windows as below). If scaling is set higher, then the text fields will not show all of the text.



## 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.

The input files can contain header information, but should have ";" (a semi-colon without the speech marks) at the beginning of each header line to indicate this, so that they are ignored by the programme.

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.

## Reference data and secondary structure calculation methods

Two methods are implemented in SSCalcPy:

Selcon3

CDSSTR

Either 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 Helices shorter than 4 residues are all counted as distorted.

Similarly the Beta sheet content is grouped into distorted (BetaD) and regular (BetaR), but with the difference that only the first and last residues 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>