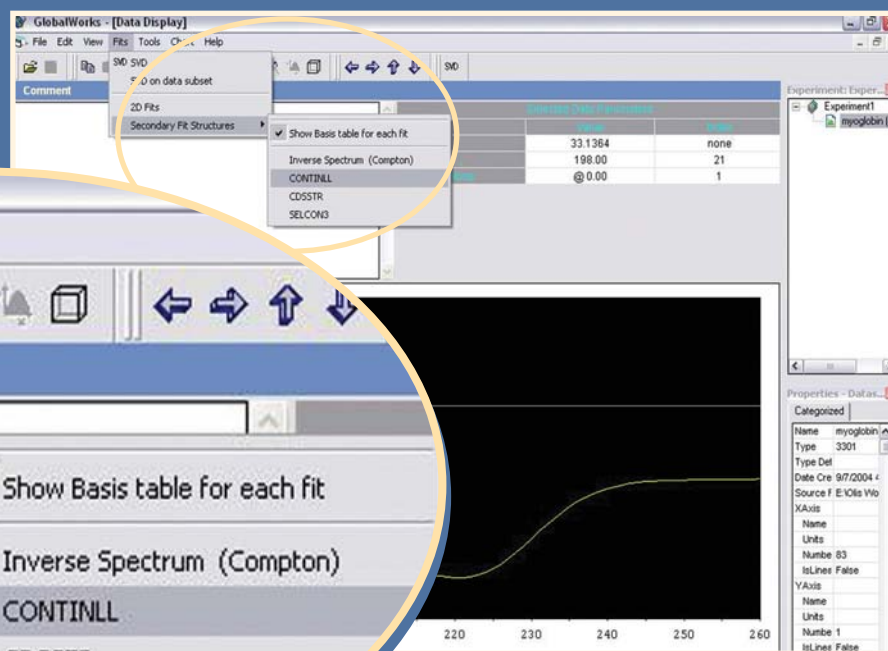


For the CD Spectroscopists:

Three Algorithms for Determination of Protein Secondary Structure

CDSSTR, CONTINLL, and SELCON3 as pull-down fits



These secondary structure determination algorithms are mathematically as their authors wrote them. Olis, Inc. reproduced them from the public literature¹ and public domain site of Narasimha Sreerama, CDPro², and made them easily accessible under the 'Fits' menu of the Olis GlobalWorks and SpectralWorks programs.

All three algorithms use variable selection of known secondary structure to achieve estimation of secondary structure composition from an unknown CD spectrum.

¹ Reviews available in the literature compare these fits (see back page)

² <http://lamar.colostate.edu/~sreeram/CDPro>

One of ten basis sets of known spectra can be chosen for each fit.

The **basis set** is chosen based on how closely it matches the experimental protein in **protein** structure (soluble, denatured, or membrane), **secondary structure** composition and **wavelength** range of the spectrum.

Basis set	Ref. Set	Proteins	Secondary Structures	WaveLength
<input checked="" type="radio"/> 1	SP29	29-Soluble	H1, H2, S1, S2, T, U	178-260
<input type="radio"/> 2	SP23	23-Soluble	H, 3-10, S, T, P2, U	178-260
<input type="radio"/> 3	SP1	29-Soluble	H, S, P2, T, U	178-260
<input type="radio"/> 4	SP37	37-Soluble	H1, H2, S1, S2, T, U	185-240
<input type="radio"/> 5	SP2	37-Soluble	H, S, P2, T, U	185-240
<input type="radio"/> 6	SP42	Sp37+5 Denatrd	H1, H2, S1, S2, T, U	185-240
<input type="radio"/> 7	SMP50	SP37+13 Membrane	H1, H2, S1, S2, T, U	185-240
<input type="radio"/> 8	SP43	43 Soluble	H1, H2, S1, S2, T, U	190-240
<input type="radio"/> 9	SDP48	SP43+5 Denatrd	H1, H2, S1, S2, T, U	190-240
<input type="radio"/> 10	SMP56	SP43+13 Membrane	H1, H2, S1, S2, T, U	190-240
<input type="radio"/> 11	CLSTR	Soluble/Denatrd	H1, H2, S1, S2, T, U	190-240

Accept Cancel Don't show again

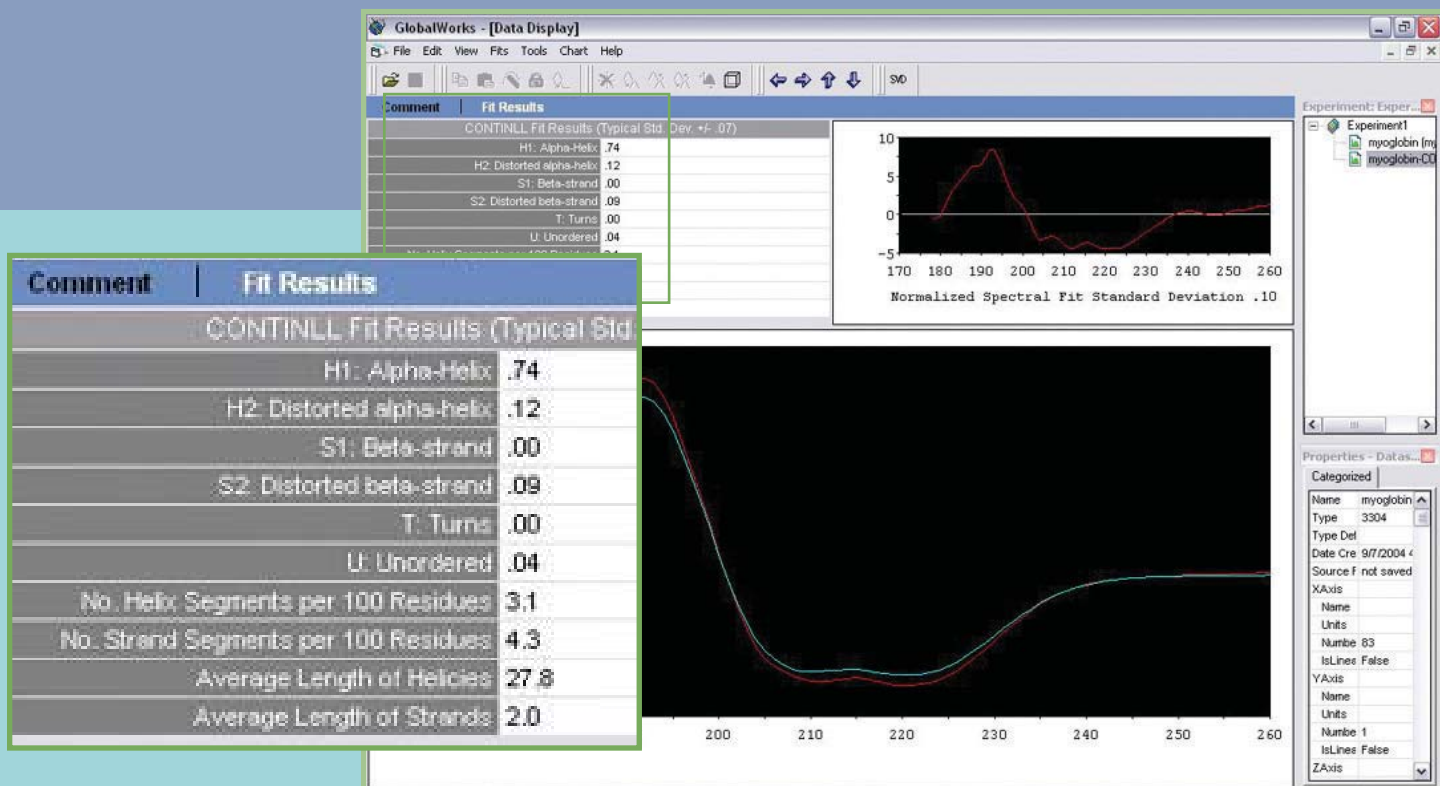
Click in circle next to 1 – 10 and then press "Accept" or "Cancel." Click "Don't Show again" if you want to continue using this basis set.

Basis Sets

All of the secondary structure fits require a basis set, which contains CD spectra of proteins of known secondary structure. Secondary structures of these model proteins were obtained from X-ray crystal structures. The data are acquired at 1 nm intervals.

Secondary Structure Results

The results of a protein secondary structure determination are presented as fractional composition of secondary structure motifs. These include alpha-helix, beta-strand, turns, proline turns, and unordered segments. The list of motifs included in the results will depend on the basis set used to produce the fit. The typical standard deviation for this fit is 0.06. The sum of these compositions should be close to 1.0. A graphical display of the fit shows how closely the calculated spectrum represents the data. A residual plot, which displays the difference between the calculated spectrum and the original CD spectrum, is also displayed with the results.



Using Protein Secondary Structure Fits in GlobalWorks Software

To fit a CD spectrum, open the desired file and click on the spectrum to select the curve. It will turn yellow and a cursor will appear. Select the desired fit from the list of 'Secondary Structure Fits' in the 'Fits' menu. A screen will appear showing the ten available basis sets to use (facing page). Choose the appropriate basis set for the analysis, and click on 'Apply' to initiate the fit. The original (red) and calculated (blue) spectra are shown on the results page along with the resulting compositions and residuals.

CDSSTR, from W. C. Johnston (1999)

CDSSTR is a modification of VARSIC which uses all possible combinations of a fixed number of proteins in the reference set. This method is generally the most accurate of the four, but can take minutes, not seconds, to perform all the necessary calculations.

VARSIC, or Variable Selection, from Johnson et al, 1999, fits a random subset of eight basis spectra using the variable selection method first described by Mosteller and Tukey (1977). The resulting fits are analyzed using four criteria: (1) The sum of the secondary structures is close to unity, (2) No fraction of the secondary structure be less than -0.03, (3) The reconstituted CD spectrum should fit the original spectrum with only a small error, (4). The fraction of alpha-helix should be similar to that obtained using all the proteins as the basis set. The subset that fits these criteria best is used as the basis set. Basis set 1 is most commonly used (see below for more details on basis sets), although GlobalWorks allows for use of any of the eight available basis sets.

CONTINLL, from Provencher and Glockner (1981)

Developed by Provencher and Glockner (1981), this algorithm uses ridge regression¹ to fit a linear combination of spectra of known composition to match a spectrum of unknown composition. CONTIN3 uses this algorithm, with a locally linearized model (van Sukkom et al., 1990). In this model, basis set proteins are screened for those with the lowest RMS deviation from the spectrum of unknown composition, such that only spectra which closely match the unknown spectrum are used.

SELCON3, or Self Consistent, from Woody et al (1993)

Developed in the laboratory of Robert W. Woody (Sreerama and Woody, 1993), SELCON incorporates the self-consistent method in which successive iterations achieve an RMS noise of 0.0025 to predict secondary structure. The algorithm first assigns an initial guess at the composition using the singular value decomposition (SVD) method of Hennessey and Johnson (1981). Then, in the second step, the initial starting compositions are iterated using the SVD calculation until a convergent solution is obtained. The results are then constrained to the sum of components being close to 1 and no composition less than -0.05. Finally, the results are further constrained by the helix limit theorem.

¹ Tikhonov and Arsenin 1974

References

1. Johnson, WC. (1999) *PROTEINS: Structure, Function, and Genetics* 35, 307-312.
2. Mosteller, F. and Tukey, JW (1977) in *Data Analysis and Regression*, Reading, MA, Addison-Wesley, p528.
3. Greenfield, NJ (1996) *Anal. Biochem.* 235, 1-10.
4. Johnson, WC (1988) *Annu. Rev. Biophys. Chem.* 17, 145-166.
5. Sreerama, N. and Woody, RW (1993) *Anal. Biochem.* 209, 32-44.
6. Hennessey, JP and Johnson, WC (1981) *Biochemistry* 20, 1085-1094.
7. Provencher, SW and Glockner, J. (1981) *Biochemistry* 20, 33-37.
8. van Stokkum, IHM, Spoelder, H.JW, Bloemendal, M., van Grondelle, R., and Groen, F. C. A. (1990) 191, 110-118.
9. Venyaminov, SY and Yang, JT (1996) in *Circular Dichroism and the Conformational Analysis of Biomolecules*, Fasman, G. D. ed., Plenum Press, New York, pp. 69-107.
10. Tikhonov, AN and Arsenin, VY. (1974) *Methods of Solving Ill-Posed Problems [In Russian]*, Nauka, Moscow.
11. King, SM and WC Johnson (1999) *PROTEINS: Structure, Function, and Genetics* 35, 313-320.
12. Park, K., Percel, A., and Fasman, GD (1992) *Protein Sci.* 1, 1032-1049.
13. Yang, JT, Wu, C.-S., and Martinez, HM. (1986) *Methods Enzymol.* 130, 208-269.
14. Sreerama, N., Venyaminov, S. Y., and Woody, RW (1999) *Anal. Biochem.* 287, 243-251.
15. Sreerama, N., Venyaminov, SY, and Woody, RW (2001) *Anal. Biochem.* 299, 271-274.
16. Sreerama, N., Venyaminov, SY, and Woody, RW (1999) *Protein Sci.* 8, 370-380.
17. Pancoska & Keiderling (1991). Systematic comparison of statistical analysis of electronic and vibrational circular dichroism for secondary structure prediction of selected proteins. *Biochemistry*, 30, 6885-6895.