Electronic Supplementary Information

Bayesian inference assessment of protein secondary structure analysis using circular dichroism data – how much structural information is contained in protein circular dichroism spectra?

Simon E.F. Spencer and Alison Rodger

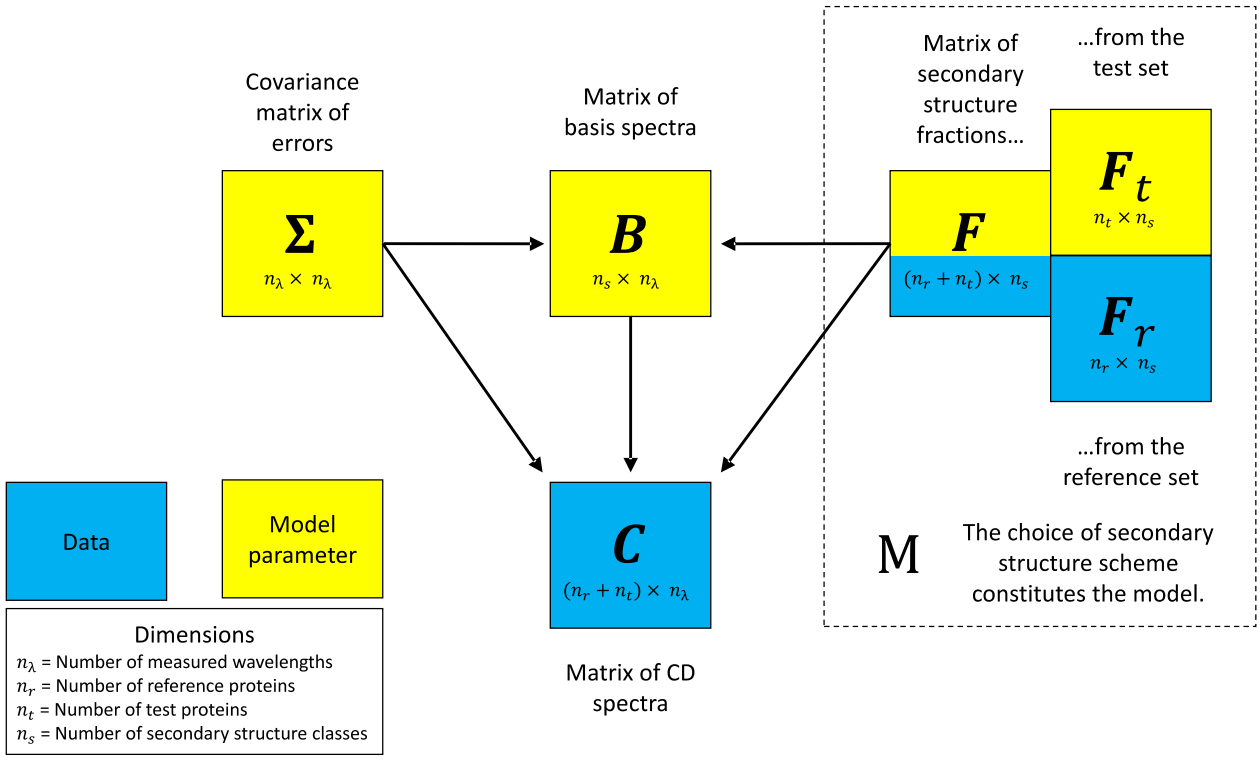

Fig S1: Conditional independence diagram showing the relationships between the model parameters and the data for our Bayesian model.

[a] *Department of Statistics, University of Warwick, Coventry, UK.* *s.e.f.spencer@warwick.ac.uk*
[b] *Author contribution: J. Franco worked on this project in its early stages, but he is unaware that this paper is being submitted for publication as we have no contact details for him. Therefore, he does not take any responsibility for the contents.*
[c] *Department of Molecular Sciences, Macquarie University, NSW, 2109, Australia. alison.rodger@mq.edu.au; here.*
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

Table S1. Cross-validation results for the SP175 proteins with 3 secondary structure classes from DSSP: α-helix, -sheet and Other. Results for competing approaches (SEL-MAT3, PCR, PLS, NN and SVM, taken from reference [31]. The best per-forming approach for each measure is given in bold.

	α-helix			β-sheet			Other		
	δ	ζ	*r*	δ	ζ	*r*	δ	ζ	*r*
Bayesian	0.061	3.48	0.959	0.127	1.25	0.770	0.137	0.77	0.652
SELMAT3	0.063	3.35	0.957	0.083	1.90	0.862	0.078	1.35	0.701
PCR	0.057	3.70	0.966	0.069	2.29	0.906	0.066	1.61	0.796
PLS	**0.053**	**3.98**	**0.971**	0.073	2.17	0.895	0.068	1.56	0.781
NN	0.055	3.84	0.968	**0.067**	**2.36**	**0.912**	**0.062**	**1.71**	**0.816**
SVM	0.057	3.70	0.966	0.069	2.29	0.908	0.066	1.61	0.792

Table S2. Cross-validation results for the SP175 proteins with the SELCON secondary structure scheme (see Section 2.5). Results for competing approaches (SELMAT3 and PCR, see Section 3.2) taken from Lees et al. (2006a). The best performing approach for each measure is given in bold.

	Bayesian			SELMAT3			PLS		
Structure	δ	ζ	*r*	δ	ζ	*r*	δ	ζ	*r*
Regular helix	0.091	1.73	0.836	0.048	3.28	0.956	**0.040**	**3.94**	**0.971**
Distorted helix	0.129	0.46	0.043	**0.035**	**1.70**	**0.809**	0.036	1.66	0.791
Regular β-strand	0.090	1.31	0.695	0.073	1.62	0.792	**0.063**	**1.88**	**0.853**
Distorted β-strand	0.281	0.17	-0.081	**0.020**	**2.41**	**0.913**	0.023	2.10	0.889
Turn	0.201	0.27	0.098	**0.052**	**1.04**	0.325	**0.052**	**1.04**	**0.332**
Other	0.169	0.43	0.278	**0.050**	**1.45**	0.717	**0.050**	**1.45**	**0.720**

Table S3. Cross validation results for the SP175 proteins with the BeStSel secondary structure scheme for different minimum wavelengths. The classification scheme for each number of classes was chosen to be the scheme with the largest marginal likelihood. Values of the normalised measure zeta above one indicate that the predictions are more accurate than a random draw from the reference set. Values below one have been highlighted in orange.

		ζ values					
	Minimum wavelength (nm)	175	180	185	190	195	200
	BeStSel classes						
3 classes	Helix1	3.57	3.55	3.50	3.37	3.45	3.52
	Anti1+Anti2+Parallel	2.02	2.00	2.03	2.02	1.97	1.86
	Helix2+Anti3+Turn+Others	1.84	1.83	1.86	1.90	1.86	1.72
4 classes	Helix1	4.08	4.16	3.99	4.19	4.23	4.14
	Anti1+Anti2	1.37	1.31	1.33	1.30	1.35	1.51
	Parallel	0.54	0.52	0.43	0.50	0.58	0.69
	Helix2+Anti3+Turn+Others	1.32	1.29	1.11	1.33	1.46	1.38
5 classes	Helix1	3.47	3.44	3.49	3.46	3.59	3.67
	Anti1+Anti2	1.42	1.35	1.36	1.35	1.40	1.69
	Helix2+Anti3	0.32	0.32	0.34	0.37	0.34	0.33
	Parallel	0.51	0.50	0.53	0.50	0.65	0.86
	Turn+Others	0.43	0.45	0.48	0.50	0.50	0.52