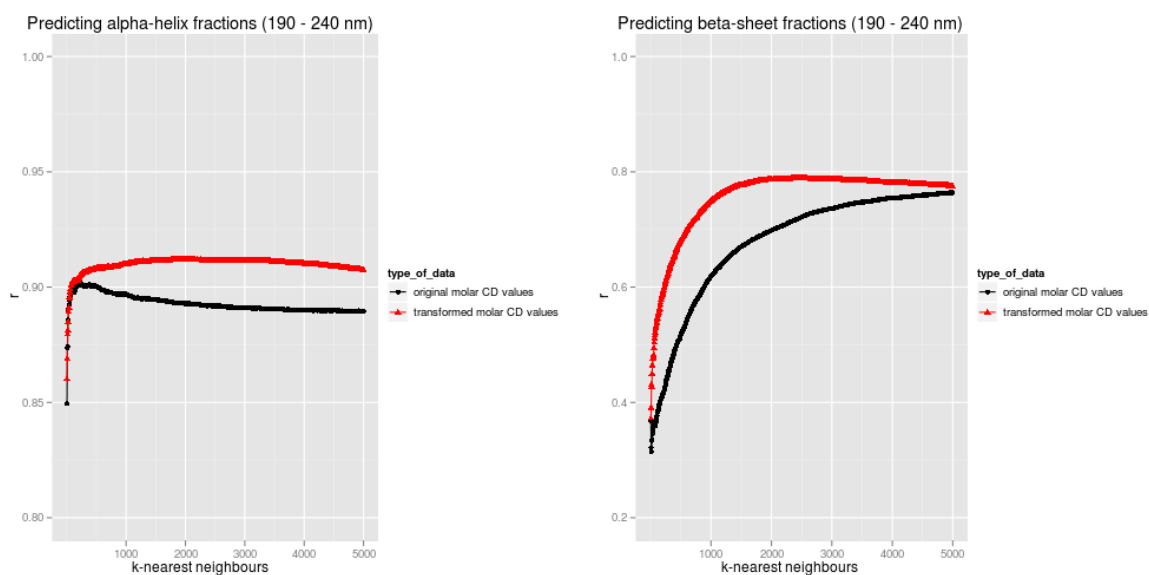
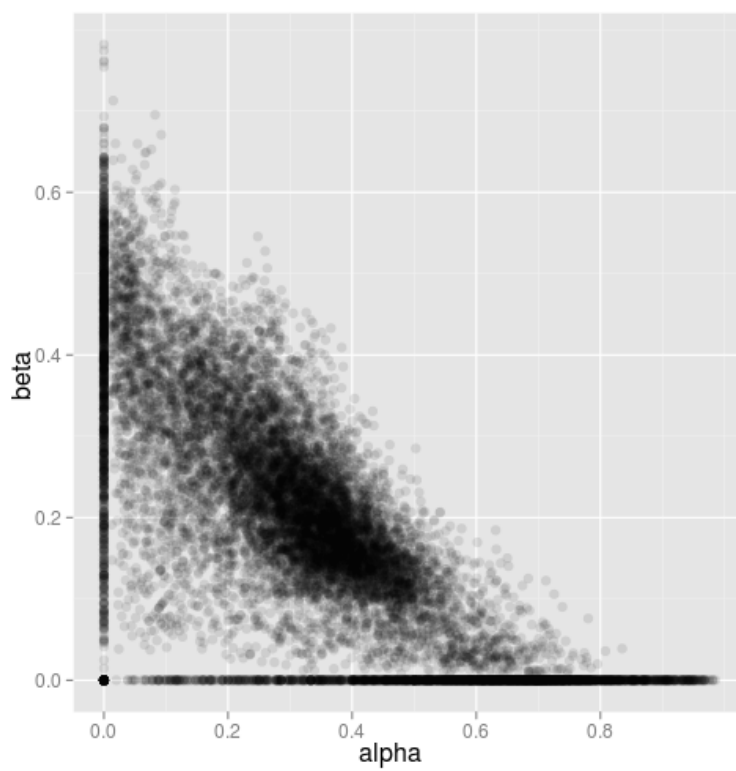


## Supplementary data

**Figure S1:** Improvement of the predictions in terms of Pearson's correlation coefficient ( $r$ ) when the spectra output from DichroCalc are further calibrated using actual spectral data.



**Figure S2:** Coverage on alpha-helix and beta-strand percentages by the proteins in the reference dataset.



**Table S1:** Proteins of BENCH83 coming from SMP56 and PCDDDB.

| <b>PDB Id</b> | <b>Protein names</b>                                  | <b>length</b> | <b><math>\alpha</math>-helix</b> | <b><math>\beta</math>-sheet</b> | <b>source</b> |
|---------------|-------------------------------------------------------|---------------|----------------------------------|---------------------------------|---------------|
| 2DHQ          | 3-dehydroquinate dehydratase                          | 146           | 0.338                            | 0.184                           | PCDDDB        |
| 3ADK_A        | adenylate kinase                                      | 195           | 0.5464                           | 0.1289                          | SPM56         |
| 8ADH_A        | alcohol dehydrogenase                                 | 374           | 0.2246                           | 0.2273                          | SPM56         |
| 1ADO          | Aldolase                                              | 363           | 0.425                            | 0.138                           | PCDDDB        |
| 1ED9          | Alkaline phosphatase                                  | 450           | 0.278                            | 0.183                           | PCDDDB        |
| 1VJS          | Alpha amylase                                         | 482           | 0.241                            | 0.205                           | PCDDDB        |
| 2CGA_A        | alpha chymotripsinogen                                | 245           | 0.0735                           | 0.3224                          | SPM56         |
| 2ABX_A        | alpha-bungarotoxin                                    | 74            | 0                                | 0.0541                          | SPM56         |
| 1RAV          | Avidin                                                | 128           | 0                                | 0.504                           | PCDDDB        |
| 1AZU_A        | azurin                                                | 128           | 0.1129                           | 0.2581                          | SPM56         |
| 1REI_A        | bence jones protein                                   | 107           | 0                                | 0.4766                          | SPM56         |
| 1FA2          | Beta amylase                                          | 498           | 0.311                            | 0.114                           | PCDDDB        |
| 1BGL          | Beta galactosidase                                    | 1023          | 0.103                            | 0.364                           | PCDDDB        |
| 1BEB_A        | beta-lactoglobulin                                    | 162           | 0.0962                           | 0.4231                          | SPM56         |
| 5PTI_A        | bovin pancreatic trypsin inhibitor                    | 58            | 0.1379                           | 0.2414                          | SPM56         |
| 1HA7          | c-Phycocyanin                                         | 334           | 0.743                            | 0                               | PCDDDB        |
| 1SU4_A        | Ca <sup>2+</sup> ATPase muscle sarcoplasmic reticulum | 994           | 0.4266                           | 0.1549                          | SPM56         |
| 1LIN          | Calmodulin                                            | 153           | 0.575                            | 0.055                           | PCDDDB        |
| 1CA2_A        | carbonic anhydrase                                    | 259           | 0.082                            | 0.2891                          | SPM56         |
| 5CPA_A        | carboxypeptidase A                                    | 307           | 0.3616                           | 0.1629                          | SPM56         |
| 1DGF          | Catalase                                              | 526           | 0.282                            | 0.169                           | PCDDDB        |
| 1KCW          | Ceruloplasmin                                         | 1046          | 0.087                            | 0.343                           | PCDDDB        |
| 2CTS          | Citrate synthase                                      | 437           | 0.588                            | 0.014                           | PCDDDB        |
| 1COL_A        | colicin A                                             | 204           | 0.7563                           | 0                               | SPM56         |
| 2CTV_A        | concanavalin A                                        | 237           | 0                                | 0.4641                          | SPM56         |
| 5CYT_R        | cytochrome C                                          | 104           | 0.4078                           | 0                               | SPM56         |
| 1ERI_A        | EcoR1 endonuclease                                    | 276           | 0.2835                           | 0.1877                          | SPM56         |
| 3EST_A        | elastase                                              | 240           | 0.0542                           | 0.3417                          | SPM56         |
| 2FDN          | Ferredoxin                                            | 55            | 0.073                            | 0.182                           | PCDDDB        |
| 1FX1_A        | flavodoxin                                            | 148           | 0.2925                           | 0.2177                          | SPM56         |
| 4GCR_A        | gamma-crystallin                                      | 174           | 0.0287                           | 0.4598                          | SPM56         |
| 1CF3          | Glucose oxidase                                       | 583           | 0.27                             | 0.193                           | PCDDDB        |
| 3MW9          | Glutamate dehydrogenase I                             | 501           | 0.454                            | 0.125                           | PCDDDB        |
| 3GRS_A        | glutathione reductase                                 | 478           | 0.2863                           | 0.2408                          | SPM56         |
| 3GPD_R        | glyceraldehyde 3-p dehydrogenase                      | 334           | 0.2545                           | 0.2096                          | SPM56         |
| 1GPB          | Glycogen phosphorylase-b                              | 842           | 0.451                            | 0.153                           | PCDDDB        |
| 1EMA_A        | green fluorescent protein                             | 236           | 0.0356                           | 0.5067                          | SPM56         |
| 1BN6          | Haloalkane dehalogenase                               | 294           | 0.371                            | 0.186                           | PCDDDB        |
| 2HMZ_A        | hemerythrin                                           | 113           | 0.646                            | 0                               | SPM56         |
| 1IGT          | Immunoglobulin G                                      | 658           | 0.04                             | 0.451                           | PCDDDB        |
| 1KU8          | Jacalin                                               | 133           | 0                                | 0.634                           | PCDDDB        |
| 6LDH_A        | lactate dehydrogenase                                 | 330           | 0.3982                           | 0.1611                          | SPM56         |
| 1BLF          | Lactoferrin                                           | 689           | 0.298                            | 0.177                           | PCDDDB        |
| 1LES          | Lectin (lentil)                                       | 232           | 0.018                            | 0.48                            | PCDDDB        |
| 1AX8          | Leptin                                                | 146           | 0.631                            | 0                               | PCDDDB        |

|        |                                           |     |        |        |        |
|--------|-------------------------------------------|-----|--------|--------|--------|
| 1LYS_A | lysozyme                                  | 129 | 0.3178 | 0.062  | SPM56  |
| 1AF6_A | maltoporin                                | 421 | 0      | 0.5962 | SPM56  |
| 1MOL   | Monellin                                  | 94  | 0.17   | 0.521  | PCDDDB |
| 4MBN_A | myoglobin                                 | 153 | 0.7451 | 0      | SPM56  |
| 1K6J   | Nitrogen metabolite repression regulator  | 352 | 0.37   | 0.174  | PCDDDB |
| 1OVA   | Ovalbumin                                 | 385 | 0.29   | 0.32   | PCDDDB |
| 1DOT   | Ovotransferrin                            | 686 | 0.261  | 0.166  | PCDDDB |
| 9PAP_A | papain                                    | 212 | 0.2311 | 0.1698 | SPM56  |
| 5CPV_A | parvalbumin                               | 109 | 0.5    | 0.037  | SPM56  |
| 1AIR   | Pectate lyase C                           | 353 | 0.088  | 0.321  | PCDDDB |
| 2PSG_A | pepsinogen                                | 370 | 0.1081 | 0.3865 | SPM56  |
| 7ATJ   | Peroxidase C1                             | 308 | 0.452  | 0.02   | PCDDDB |
| 1HNN   | Phenylethanolamine N-methyltransferase    | 289 | 0.33   | 0.244  | PCDDDB |
| 3PMG   | Phosphoglucosmutase 1                     | 561 | 0.33   | 0.25   | PCDDDB |
| 3PGK_A | phosphoglycerate kinase                   | 416 | 0.3446 | 0.1108 | SPM56  |
| 1UNE   | Phospholipase A2                          | 123 | 0.447  | 0.065  | PCDDDB |
| 1PHO_A | phosphoporin                              | 330 | 0.0121 | 0.5485 | SPM56  |
| 2OMF_A | porin OmpF from E coli                    | 340 | 0.0353 | 0.5794 | SPM56  |
| 2POR_A | porin Rhodobacter capsulatus              | 301 | 0.0565 | 0.5681 | SPM56  |
| 2PAB_A | prealbumin                                | 127 | 0.0702 | 0.4912 | SPM56  |
| 1QHJ_A | purple membrane Halobacterium halobium    | 248 | 0.7456 | 0.0526 | SPM56  |
| 1A49   | Pyruvate kinase                           | 530 | 0.362  | 0.193  | PCDDDB |
| 1IFC_A | rat intestinal fatty acid binding protein | 132 | 0.1145 | 0.5878 | SPM56  |
| 1RHD_A | rhodanase                                 | 293 | 0.2765 | 0.1092 | SPM56  |
| 1F88_A | rhodopsin bovine retina                   | 348 | 0.6006 | 0.0355 | SPM56  |
| 3RN3_A | ribonuclease A                            | 124 | 0.1774 | 0.3306 | SPM56  |
| 1R0I   | Rubredoxin                                | 54  | 0      | 0.151  | PCDDDB |
| 2SNS_A | staphylococcal nuclease                   | 149 | 0.1844 | 0.1986 | SPM56  |
| 1STP   | Streptavidin                              | 159 | 0.033  | 0.554  | PCDDDB |
| 1SBT_A | subtilisin BPN                            | 275 | 0.3018 | 0.1782 | SPM56  |
| 2SBT_A | subtilisin novo                           | 275 | 0.2145 | 0.1382 | SPM56  |
| 2SOD_O | superoxide dismutase                      | 152 | 0      | 0.3841 | SPM56  |
| 2LZM_A | T4 lysozyme                               | 164 | 0.6646 | 0.0854 | SPM56  |
| 1THW   | Thaumatococcus I                          | 207 | 0.106  | 0.357  | PCDDDB |
| 8TLN_E | thermolysin                               | 316 | 0.3711 | 0.1635 | SPM56  |
| 3TIM_A | trios phosphate isomerase                 | 250 | 0.3896 | 0.1526 | SPM56  |
| 1TNF_A | tumor necrosis factor                     | 157 | 0      | 0.4276 | SPM56  |
| 1UBI   | Ubiquitin                                 | 76  | 0.158  | 0.316  | PCDDDB |

**Table S2:** Slope and intercepts for the regressions of DichroCalc on real data for BENCH83 proteins for each wavelength together with the coefficient of determination (adjusted\_R\_squared) and Pearson's correlation coefficient (correlation\_coefficient).

| #wavelength | slope  | intercept | adjusted_R_squared | correlation_coefficient |
|-------------|--------|-----------|--------------------|-------------------------|
| 190         | 0.9966 | -1.1891   | 0.7105             | 0.8429                  |
| 191         | 1.0255 | -1.6196   | 0.6938             | 0.833                   |
| 192         | 1.052  | -1.9761   | 0.6647             | 0.8153                  |
| 193         | 1.0841 | -2.3476   | 0.6241             | 0.79                    |
| 194         | 1.1131 | -2.6197   | 0.5683             | 0.7538                  |
| 195         | 1.1303 | -2.7354   | 0.4972             | 0.7051                  |
| 196         | 1.1132 | -2.573    | 0.4018             | 0.6339                  |
| 197         | 1.0248 | -1.9859   | 0.2845             | 0.5334                  |
| 198         | 0.8078 | -0.8376   | 0.1558             | 0.3947                  |
| 199         | 0.5053 | 0.3211    | 0.0595             | 0.2439                  |
| 200         | 0.2888 | 0.6535    | 0.0198             | 0.1408                  |
| 201         | 0.246  | 0.1714    | 0.0237             | 0.1538                  |
| 202         | 0.2936 | -0.486    | 0.0658             | 0.2565                  |
| 203         | 0.3582 | -1.0003   | 0.1515             | 0.3893                  |
| 204         | 0.4114 | -1.3367   | 0.2626             | 0.5125                  |
| 205         | 0.4565 | -1.5435   | 0.3804             | 0.6167                  |
| 206         | 0.4905 | -1.6541   | 0.4832             | 0.6951                  |
| 207         | 0.5148 | -1.6951   | 0.566              | 0.7523                  |
| 208         | 0.5271 | -1.6869   | 0.6239             | 0.7899                  |
| 209         | 0.5314 | -1.6545   | 0.6658             | 0.816                   |
| 210         | 0.5312 | -1.6052   | 0.6904             | 0.8309                  |
| 211         | 0.5325 | -1.5568   | 0.7065             | 0.8405                  |
| 212         | 0.5398 | -1.5096   | 0.7156             | 0.846                   |
| 213         | 0.5582 | -1.4609   | 0.7174             | 0.847                   |
| 214         | 0.5875 | -1.4056   | 0.7176             | 0.8471                  |
| 215         | 0.6312 | -1.3323   | 0.7182             | 0.8475                  |
| 216         | 0.6915 | -1.2243   | 0.7236             | 0.8507                  |
| 217         | 0.7671 | -1.0889   | 0.7349             | 0.8572                  |
| 218         | 0.858  | -0.9192   | 0.7505             | 0.8663                  |
| 219         | 0.9621 | -0.7274   | 0.7638             | 0.874                   |
| 220         | 1.0757 | -0.5201   | 0.776              | 0.8809                  |
| 221         | 1.1965 | -0.3122   | 0.787              | 0.8871                  |
| 222         | 1.324  | -0.1025   | 0.7948             | 0.8915                  |
| 223         | 1.4511 | 0.0904    | 0.7991             | 0.8939                  |
| 224         | 1.5778 | 0.2678    | 0.7998             | 0.8943                  |
| 225         | 1.7022 | 0.4178    | 0.7989             | 0.8938                  |
| 226         | 1.8203 | 0.5349    | 0.7925             | 0.8902                  |
| 227         | 1.9318 | 0.6211    | 0.7847             | 0.8858                  |
| 228         | 2.0334 | 0.6729    | 0.7736             | 0.8796                  |
| 229         | 2.1153 | 0.681     | 0.7607             | 0.8722                  |
| 230         | 2.1935 | 0.6783    | 0.7466             | 0.8641                  |

|     |        |        |        |        |
|-----|--------|--------|--------|--------|
| 231 | 2.2464 | 0.6361 | 0.7309 | 0.8549 |
| 232 | 2.2951 | 0.583  | 0.7134 | 0.8446 |
| 233 | 2.3361 | 0.5174 | 0.7008 | 0.8371 |
| 234 | 2.372  | 0.4449 | 0.6897 | 0.8305 |
| 235 | 2.3958 | 0.3665 | 0.673  | 0.8203 |
| 236 | 2.4241 | 0.3024 | 0.657  | 0.8106 |
| 237 | 2.3948 | 0.2261 | 0.6242 | 0.79   |
| 238 | 2.325  | 0.1548 | 0.5739 | 0.7576 |
| 239 | 2.281  | 0.1036 | 0.5271 | 0.726  |
| 240 | 2.0957 | 0.0437 | 0.4388 | 0.6624 |

**Table S3:** Performance comparisons of K2D2 and K2D3 on the PCDDDB subset of BENCH83.  $r$  and RMSD values for beta strand and alpha-helix predictions. In the case of the 200 to 240 nm interval, a t-test on the individual differences in absolute errors for both alpha-helix and beta-strand predictions, with the null hypothesis that K2D2 predictions are equal of better than K2D3 predictions, was only mildly significant for alpha-helix (p-value = 0.0287) but highly significant for beta strands (p-value = 0.0098). The equivalent tests for 190 to 240 nm were not significant.

|               |      | K2D3     |         | K2D2     |         |
|---------------|------|----------|---------|----------|---------|
|               |      | $\alpha$ | $\beta$ | $\alpha$ | $\beta$ |
| 200 to 240 nm | $r$  | 0.95     | 0.87    | 0.79     | 0.56    |
|               | RMSD | 0.07     | 0.09    | 0.14     | 0.15    |
| 190 to 240 nm | $r$  | 0.93     | 0.84    | 0.92     | 0.81    |
|               | RMSD | 0.07     | 0.1     | 0.08     | 0.1     |