Nucleic Acids Research, 2025, **53**, W158–W168
https://doi.org/10.1093/nar/gkaf350
Advance access publication date: 26 April 2025
Web Server issue

ChiraKit: an online tool for the analysis of circular dichroism spectroscopy data

Osvaldo Burastero [1,2,*], Nykola C. Jones [3], Lucas A. Defelipe [1,2], Uroš Zavrtanik [4], San Hadži [4], Søren Vrønning Hoffmann [3], Maria M. Garcia-Alai [1,2,*]

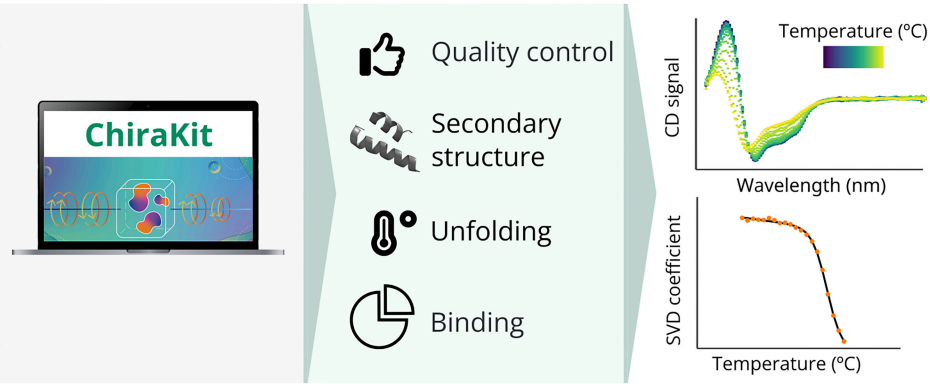
[1]European Molecular Biology Laboratory Hamburg, Notkestrasse 85, 22607 Hamburg, Germany
[2]Centre for Structural Systems Biology, Notkestrasse 85, 22607 Hamburg, Germany
[3]ISA,Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, 8000 Aarhus, Denmark
[4]Department of Physical Chemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia

[*]To whom correspondence should be addressed. Email: maria.garcia@embl-hamburg.de
Correspondence may also be addressed to Osvaldo Burastero. Email: oburastero@embl-hmburg.de

Abstract

Circular dichroism (CD) spectroscopy is an established biophysical technique to study chiral molecules. CD allows investigating conformational changes under varying experimental conditions and has been used to understand secondary structure, folding, and binding of proteins and nucleic acids. Here, we present ChiraKit, a user-friendly, online, and open-source tool to process raw CD data and perform advanced analysis. ChiraKit features include the calculation of protein secondary structure with the SELCON3 and SESCA algorithms, estimation of peptide helicity using the helix-ensemble model, the fitting of thermal/chemical unfolding or user-defined models, and the decomposition of spectra through singular value decomposition or principal component analysis. ChiraKit can be accessed at https://spc.embl-hamburg.de/.

Graphical abstract



Introduction

Circular dichroism (CD) spectroscopy is a biophysical technique for probing the structural properties of chiral biomolecules. It consists of measuring the differential absorption of left- and right-handed circularly polarized light as it passes through a sample, providing insights into the overall conformation of the biomolecules [1]. A common use of CD spectroscopy is for the study of large biological macromolecules such as proteins and oligonucleotides. For nucleic acids, the signal is mostly determined by the geometry of the stacked bases [2–4], while for proteins CD signals arise due to the secondary and tertiary structures made by the residues. In the far-UV (<240 nm) the secondary structure elements, such as α-helices, β-sheets, or turns, are defined by having different angles between the residues, thus giving rise to a CD spectra characteristic of that type of structure [5]. The amount of structural information contained in the CD spectra depends on the wavelength range. Routine benchtop instruments typically operate within 190–340 nm, whereas synchrotron radiation circular dichroism (SRCD) facilities can extend the lower wavelength limit to 170 nm.

CD spectroscopy offers several advantages: it is rapid, cost-effective, requires minimal sample volumes, and is often nondestructive. Therefore, it is particularly well-suited for monitoring conformational changes under varying experimental conditions. For instance, it has been used to characterize protein folding and stability—via thermal or chemical denaturation—to assess DNA–ligand interactions and

Received: January 30, 2025. **Revised:** March 26, 2025. **Editorial Decision:** April 8, 2025. **Accepted:** April 24, 2025

reaction kinetics, and to validate protein mutants [6–10]. The analysis of CD data ranges from straightforward spectral averaging, baseline correction, and normalization, to more advanced modeling approaches. These include secondary structure deconvolution or fitting spectral changes as a function of experimental parameters such as temperature or ligand concentration.

Numerous computational tools have been developed to extract insights from CD data. CDToolX (desktop application) enables standard processing and singular value decomposition (SVD) [11]. DichroWeb and BeStSel (online tools) calculate protein secondary structure content [12, 13]. DichroWeb provides five algorithms (SELCON3, CONTINLL, CDSSTR, VARSLEC, and K2D) [14–18]. BeStSel was optimized for β-sheet-rich and has a module for fold recognition [19]. SESCA (Python package) predicts CD spectra from structures and estimates secondary structure content [20, 21]. CDPal (desktop application) can be used to monitor protein thermal and chemical stability [22]. CalFitter (online tool) offers complex models for thermal unfolding studies [23, 24]. CDNuss (online tool) estimates nucleic acids secondary structure content [25]. Lastly, there are two CD databases: the Protein Circular Dichroism Data Bank (PCDDB) and the Nucleic Acid Circular Dichroism Database (NACDB) [26, 27].

In this work, we present ChiraKit, the latest addition to our eSPC data analysis platform for molecular biophysics [28–30]. ChiraKit is open-source, online, user-friendly, interactive, and free, and provides functions for preprocessing and advanced analysis, including the deconvolution of spectra, thermal and chemical unfolding models, user-defined models, and protein secondary structure calculations. ChiraKit can be accessed at https://spc.embl-hamburg.de/.

Materials and methods

Lysozyme sample

Hen egg-white lysozyme was purchased from Sigma–Aldrich (CAS-Number: 12650-88-3). The sample was prepared by dissolving the powder in MilliQ water (no buffer or salts). The concentration of the sample, 1.1 mg/ml, was determined from the measured absorbance at 205 nm using the sequence from the UniProt ID P00698 (UniProt Consortium, 2023). The sample was prepared from the powder and kept in the fridge until the measurements (same day).

Synchrotron radiation circular dichroism measurements of lysozyme

The CD spectra were recorded on the AU-CD beam line of the ASTRID2 synchrotron radiation light source at the Department of Physics and Astronomy, Aarhus University. Correct operation of the CD spectrometer was confirmed through measurement of a known concentration of camphor sulfonic acid (CSA, CAS number: 3144-16-9). Measurements were carried out with the sample in a nominal 0.1 mm path length quartz suprasil cell (Starna type 31B), with the path length determined to be 0.1106 mm. Measurements were carried out using the fast scanning mode, whereby the monochromator is in constant motion through the wavelength range, allowing data to be collected for a single scan in just under 1 min. One scan was taken at each set temperature step of 2.5°C with an equilibration time of 1 min after the temperature was reached, before starting each measurement.

Expression and purification of the *Homo sapiens* clathrin heavy chain N-terminal domain

The *Homo sapiens* clathrin heavy chain N-terminal domain (CHC-NTD, 1-364) was cloned in a pETM-30 vector, which contains a His-GST tag and a TEV cleavage site. Chemocompetent *E. coli* BL21 (DE3) cells were transformed with 100 ng of plasmidic DNA and grown overnight at 37°C with 30 μg/ml of kanamycin. For protein expression, a 1:100 dilution of the preculture was done in 2xTY (16 g tryptone, 10 g yeast extract, and 5 g NaCl per liter). Cells were grown at 37°C until reaching OD 0.8–1. Then, the temperature was decreased to 18°C, and induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.3 mM. Induction was performed overnight. Cells were harvested at 12 000 × *g* for 15 min at 18°C and stored at −20°C until purification. Cells were resuspended in 5–10 ml of lysis buffer [20 mM sodium phosphate pH 7.5, 200 mM NaCl, 0.05% (w/v) Tween-20, 2 mM $MgCl_2$, 0.5 mM TCEP, +400 U DNAse I, 12.5 mM imidazole, and a tablet of Complete EDTA-free protease inhibitor cocktail, Roche, per 100 ml of buffer) per gram of cells. Lysis of the cells was done with an Emulsiflex C3 (Avestin, Ottawa, Canada) cell disruptor at 15 kPsi three times. A centrifugation at 42 000 × *g* for 1 h at 4°C was performed to clear the lysate.

The lysate was filtered with a 0.45 μm filter and loaded onto a Ni-NTA (Carl Roth, Germany) gravity column equilibrated with buffer A (20 mM sodium phosphate pH 7.5, 500 mM NaCl, and 12.5 mM imidazole). The column was washed with 10 column volumes (CV) of buffer A and eluted with buffer B (20 mM sodium phosphate pH 7.5, 500 mM NaCl, and 250 mM imidazole), collecting 1 ml of fractions. The purity of the protein samples was assessed using Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS–PAGE). Subsequently, selected fractions were pooled and treated with TEV protease at a ratio of 1 mg of TEV per 25 mg of protein. This mixture was then diluted in a 1:1 ratio with a dialysis buffer composed of 20 mM TRIS (pH 7.5), 200 mM NaCl, and 0.5 mM TCEP. Finally, the solution was subjected to overnight dialysis at 4°C with the same buffer.

A reverse Ni-NTA was performed to remove HisTag-TEV and the cleaved HisGST from the solution. A prepacked Ni-NTA column was equilibrated with a dialysis buffer and the flowthrough was collected and concentrated to 2 ml using a 10 kDa MWCO concentrator (Amicon Ultra 15, Millipore). The sample was loaded onto a Superdex 200 HiLoad 16/600 size-exclusion chromatography (SEC) column equilibrated with SEC buffer (50 mM Tris pH 9, 150 mM NaCl, and 0.5 mM TCEP). Fractions were analyzed for purity using SDS-PAGE, pooled, concentrated to 20 mg/ml, flash frozen in liquid nitrogen, and stored at −70°C until used.

CD measurements of CHC-NTD (secondary structure)

The CHC-NTD sample was dialyzed overnight at 4°C against 20 mM sodium phosphate buffer (pH 7.5), containing 150 mM sodium fluoride and 0.5 mM TCEP, using a 10 kDa MWCO Slide-A-Lyzer MINI dialysis device (Thermo Fisher Scientific). The final protein concentration was adjusted to 0.133 mg/ml. CD spectra were acquired in the range of 180–300 nm using a ChiraScan spectrophotometer (Applied Photophysics, Leatherhead, UK), equipped with a Quantum Northwest TC 125 temperature controller (Liberty Lake,

Washington, USA), and set at 20°C. Measurements were performed in a 300 µl quartz cuvette with a 1 mm pathlength (Hellma, Müllheim, Germany). Three scans per spectrum were recorded, each with a step size of 1 nm and a dwell time of 1 s.

CD measurements of CHC-NTD (chemical unfolding)

CHC-NTD solutions at a concentration of 0.41 mg/ml were incubated overnight with size exclusion chromatography (SEC) buffer or SEC buffer supplemented with increasing concentrations of urea, ranging from 0.5 to 6.5 M in 0.5 M increments. CD spectra were recorded between 205 and 260 nm using a ChiraScan spectrophotometer (Applied Photophysics, Leatherhead, UK) equipped with a Quantum Northwest TC 125 temperature controller (Liberty Lake, Washington, USA), maintained at 20°C. Measurements were conducted in a 300 µl quartz cuvette with 1 mm path length (Hellma, Müllheim, Germany). Concentrations from 0.5 to 3.5 M were measured in triplicate, while concentrations from 4 to 6.5 M were measured in duplicate. Three scans per spectrum were recorded, each with a step size of 1 nm and a dwell time of 1 s.

CHC-NTD solvent accessible area surface and *m*-values

The change in solvent-accessible surface area (ΔASA) of CHC-NTD was calculated using the equation located in the inset of Fig. 1 from [31]:

$$\Delta ASA\ \left(^2\right) = -907 + 93\,(n_{res}) \quad (1)$$

where n_{res} is the number of residues (364). The global *m*-value (urea) was estimated according to the equation 29.9 of [32]:

$$m = 0.13\Delta ASA + 243\text{cal/mol/}M \quad (2)$$

CD measurements of intrinsically disordered peptides

Peptides were purchased from Sigma–Aldrich or China Peptides (AK32 peptide) with a minimum purity of 95%. The sequences are listed in Supplementary Table S1. Peptides were dissolved and dialyzed against a 20 mM phosphate buffer (pH 7.0) containing 150 mM NaCl, except for the alanine peptide AK32, which was dialyzed in the same phosphate buffer with a higher salt concentration (1 M NaCl). Prior to each experiment, peptide stock solutions were extensively centrifuged, and their concentrations were determined by measuring absorbance at 280 nm using extinction coefficients of 1.450 M^{-1} cm^{-1} for tyrosine and 5.690 M^{-1} cm^{-1} for tryptophan [33].

CD spectra were recorded using a JASCO J-1500 spectrometer. All peptides were diluted to a final concentration of 60 µM, except for the AK32 peptide, which was prepared at 100 µM. Spectra were measured in a 1 mm (pathlength) cuvette at 25°C. The spectrometer was calibrated prior to each measurement using ammonium-*d*-camphor-10-sulfonate (Sigma–Aldrich). For TFE (trifluoroethanol) induction experiments, TFE from Sigma–Aldrich was used. For thermal melting experiments, ellipticity at 222 nm was monitored at 1°C intervals with an integration time of 8 s per measurement.

Results and discussion

Software description

The usage workflow can be divided into four steps (Fig. 1). The data importing step, preprocessing, the analysis, and export. To start with, the raw CD data are processed, and the final spectra are obtained. Second, the CD data are analyzed as a function of a certain experimental parameter, or the protein secondary structure is calculated. Models for thermal and chemical unfolding are included. A module for the comparison of spectra is also available. Lastly, the final spectra, the fitted parameters, the fitted curves, and a logbook containing all the performed steps can be exported. A comparison between ChiraKit and existing software is shown in Supplementary Table S2.

Input file

ChiraKit accepts several file types, including those generated by "Applied Photophysics" and "JASCO" instruments, generic (.gen) files for PCDDB/NACDDB databases, PCDDB (.pcd) files, SRCD (.dx or .dat) files, and CSV files with column-formatted data. High-tension voltage data, associated with the CD data collected, if available, are also parsed and displayed.

Processing

CD spectra can be added, subtracted, baseline corrected (zeroed), smoothed, scaled, and averaged. Six working units are available: millidegrees, differential absorbance, molar extinction, molar ellipticity, mean unit molar extinction ($\Delta\varepsilon$), and mean unit molar ellipticity ($[\Theta]$). Molar extinction/ellipticity units normalize the CD data to the path length, concentration, and molecular weight. $[\Theta]$ and $\Delta\varepsilon$ are useful for proteins because they take into account the number of chromophores (e.g. peptide bonds) [34]. The wavelength range can be manually adjusted or automatically by using a high-tension voltage threshold.

Analysis options

To calculate a protein's secondary structure content, we provide the SESCA Bayesian method [21] and a Python-based implementation of the SELCON3 algorithm [35]. These methods are well-suited for proteins but not for completely unstructured proteins or peptides. The default reference sets for SELCON are based on the SP175 [36] and SMP180 [37] sets, but custom reference sets can be imported. For peptides adopting only helix or coil conformations, the helicity estimation using the ensemble model by Zavrtanik *et al.* is available [38].

Multiple spectra can be reduced to a set of basis spectra. The basis spectra can be linearly combined to reconstruct all the original spectra. Hence, we take advantage of the information provided by the full spectrum rather than a single wavelength. Two methods are provided: SVD and PCA (without *Z*-score normalization). PCA is equivalent to SVD after centering the data [39]. The number of relevant basis spectra is determined by the user, and each basis spectrum will be associated with as many coefficients as the original spectra (e.g. one coefficient per temperature). To help interpreting the results, the spectra can be inverted and/or rotated. The inversion operation changes the sign, while the rotation operation consists of creating a new basis set with one basis spectrum that resembles the first measured spectrum (e.g. lowest temperature) [40].

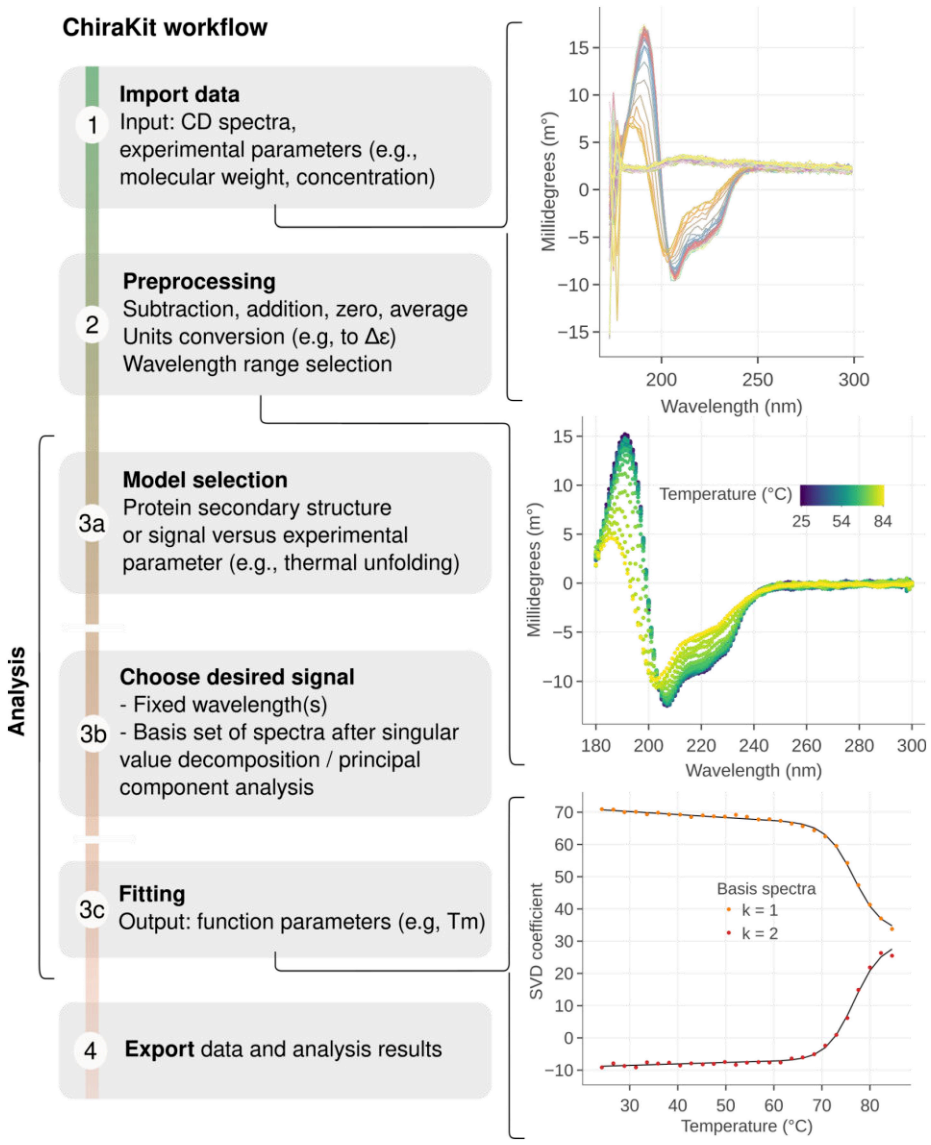


Figure 1. Usage workflow of ChiraKit. First, the data file is imported and the experimental parameters, such as path length, are inputted (1). Subsequently, the CD units are selected (e.g. molar extinction) and the spectra are preprocessed (2). Third, the CD spectra are analyzed. For proteins, the secondary structure can be estimated, or, in the case of thermal/chemical unfolding experiments, two or three-state models can be applied. For other types of biomolecules (or experiments), the CD signal can be investigated by employing user-defined models. Moreover, the analysis can focus on single wavelengths or use a set of basis spectra derived from SVD/PCA (3b). After selecting a model, the fitting is done and the relevant parameters such as the temperature of melting (T_m) are obtained (3c). Lastly, the finalized spectra, fitted parameters and fitted curves can be exported, in text format and/or as figures (if applicable) (4).

For protein unfolding analysis, the user can fit two- and three-state reversible unfolding models for chemical and thermal unfolding [41]. In the former case, the models are based on the empirical linear extrapolation model (LEM) and depend on the parameters *D50* and *m*, where *D50* is the denaturant concentration at which the free energy equals zero, and *m* is a parameter that correlates with the solvent accessible area [31, 42]. In the latter case, the two important thermodynamic parameters are the enthalpy of unfolding (ΔH) and the temperature where the free energy equals zero (temperature of melting T_m or $T_{1/2}$).

ChiraKit can fit many curves simultaneously with shared thermodynamic parameters. For instance, the CD signal monitored at wavelengths 198 and 222 nm must have the same T_m but may have different baselines. The global fitting can be done at the protein concentration level, improving the estimation of dimer, trimer, and tetramer unfolding models. Furthermore, two- and three-state unfolding models that lead to an irreversibly denatured state (Lumry–Eyring) are available for melting curves of monomers [43]. The two additional important thermodynamic parameters are the energy of activation (E_a) and the temperature where the rate constant equals one (T_f).

The CD signal can be also studied as a function of any desired experimental parameter. The user needs to associate each spectra with a certain value of the experimental parameter and input a fitting function as a string. This approach could be relevant for other types of CD datasets such as 2,2,2-trifluoroethanol (TFE) titrations or DNA–ligand binding [8, 44, 45].

Lastly, ChiraKit has a module to compare samples (e.g. wild-type versus mutant) or the same sample at different conditions. Given two or more groups of spectra, the averages and associated standard deviations are computed, together with the "difference" spectrum (of the averages). To assess the importance of the differences, the intra and inter groups euclidean distances are provided. To compare the shapes of spectra and discard the influence of differences in absorbance intensity, spectra can be normalized to have unit length of one (L2 normalization) [46].

Case studies

Secondary structure and thermal unfolding of hen egg-white lysozyme

Lysozyme, an enzyme known for hydrolyzing bacterial cell wall peptidoglycan, is a model protein for studies on stability and unfolding. Hen egg-white lysozyme solubilized in water was measured using SRCD. Six scans were performed for both lysozyme in water and water alone. Following data acquisition, the averaged water spectrum was subtracted from the averaged sample spectrum. The CD signal was zeroed using the mean signal in the 296–300 nm range, and the lower wavelength limit was set to 176 nm, based on a voltage threshold indicating low levels of light. The final CD spectrum was normalized to $\Delta\varepsilon$ units. The processing was done with the ChiraKit Processing submodule (under 1. Import data).

The secondary structure content was assessed using the SELCON3 method with the "AU-SP175" reference set [47] (ChiraKit module "2c. Protein Secondary Structure"). The calculated values confirm lysozyme's predominantly α-helical structure (40 versus 41.4%) (Supplementary Table S3). This outcome is expected, given that lysozyme's spectrum is included in the Selcon3 reference set. The β-sheet content differs by a factor of two, highlighting the deviations of the method (12 versus 6.3%). Furthermore, we applied the SESCA method with the "DSSP-TSC1" basis set, yielding values consistent with the crystallographic structure (Supplementary Table S4), as anticipated because lysozyme is part of the training set. An advantage of SESCA is that it provides error estimates for each secondary structure element.

To investigate the thermal unfolding of lysozyme, a thermal ramp from 24 to 85°C at an average heating rate of 1°C/min was performed. Single scans were taken at each temperature step and spectra of water measured under the same conditions as the sample was used as a baseline reference at each temperature. The CD spectra generally reduce in magnitude with increasing temperature, with a change in shape, whereby the peak maxima shifts to lower wavelengths at higher temperature (Fig. 2A). The presence of an isodichroic point between 203 and 204 nm suggests that the unfolding process involves two states. If that is the case, fitting the CD signal at any particular wavelength should yield similar thermodynamic parameters.

We applied a global fitting at different wavelengths (190–230 nm, every 10 nm) and obtained a T_m of 76.6 ± 0.2 °C and ΔH of 86.8 kcal/mol ± 3.5 kcal/mol (Fig. 2B), in agreement with previous estimations [48, 49] (ChiraKit module 2a. Thermal unfolding). We recommend considering the relative errors of the fitted parameters as a criterion for discarding a model. These optimistic errors, as in almost all other nonlinear fitting software, are calculated from the covariance matrix of the fitted parameters and assume a linear model.

An improved approach to fully use the entire wavelength range information is spectral decomposition. Reconstructing the data post-SVD decomposition requires two basis spectra (Fig. 2C). Global fitting of the SVD-associated coefficients yields highly consistent estimates: a T_m of 76.6 ± 0.1°C and a ΔH of 88.1 ± 3.5 kcal/mol (Fig. 2D).

Secondary structure and chemical unfolding of the clathrin heavy chain N-terminal domain

The CHC-NTD serves as the primary contact point between clathrin and adaptor proteins, linking the clathrin cage to the membrane through the binding of short linear motifs in the adaptor proteins [50]. It consists of seven β-transducin (WD) repeats that form a β-propeller, resulting in a predominantly β-sheet structure with two short α helices (see Supplementary Fig. S1). The CD spectrum of the CHC-NTD solubilized in sodium phosphate buffer with sodium fluoride salt was measured using a CD benchtop instrument. The lowest, noiseless, wavelength was 188 nm, and the highest wavelength was 300 nm. After averaging the scans and subtracting the baseline, the CD spectrum was converted to $\Delta\varepsilon$ (mean unit molar extinction), and the secondary structure was calculated with the Selcon method. The expected values are in agreement with the mixed β/α structure, although the absolute percentages of β and α deviate almost 1.6-fold from the values obtained via the Dictionary of Secondary Structure in Proteins (DSSP) Analysis of the crystal structure (Table 1).

Then, CD spectra of the CHC-NTD solubilized in Tris buffer with sodium chloride salt were acquired. The urea concentration ranged from 0.5 to 6.5 M, with steps of 0.5 M. Scans taken at the same urea concentration were averaged. The spectra were recorded up to 260 nm and the buffer spectrum was used as a baseline reference. To remove noisy data, the lower wavelength was set to 212 nm. Concentrations of urea ranging from 0.5 to 3.5 were measured in triplicate, while concentrations from 4 to 6.5 M were measured in duplicate, yielding 33 finalized spectra (Fig. 3A).

A visual inspection of the spectra reveals a complex unfolding pattern, with the CD signal initially reducing in magnitude with an increasing concentration of urea up to 2 M. Then, the CD signal increases again to a maximum ∼3–3.5 M, before decreasing again to higher concentrations. This effect can be appreciated in Fig. 3B, where single wavelength data are shown. Considering the two transitions, one between 2 and 3 M urea, and the other between 3.5 and 5 M urea, we applied a three-state reversible unfolding model (N ⇋ I ⇋ U) (ChiraKit module 2b. Chemical unfolding).

We analyzed both the SVD coefficients and single wavelength data (Fig. 3). The inclusion of the pre- and post-transition slopes resulted in a better fitting. The estimated values of $D50_1$ and $D50_2$ were 2.5 and 3.9 M (Supplementary Tables S5 and S6). These parameters are the urea concentrations (M) at which ΔG_1 (N ⇋ I) and ΔG_2 (I ⇋ U) equal zero, respectively. The *m*-value of the first transition could not be reliably estimated, while the second *m*-value was ≈ 4 kcal/mol/M. The *m*-values are linearly related to the difference in solvent accessible surface area between the native and the unfolded states [31, 32, 51]. Indeed, the predicted *m*-value according to the number of residues for the global unfolding process would be ≈4.5 kcal/mol/M. The second *m*-value is similar to the global *m*-value, suggesting that the most significant structural changes occur during this step.

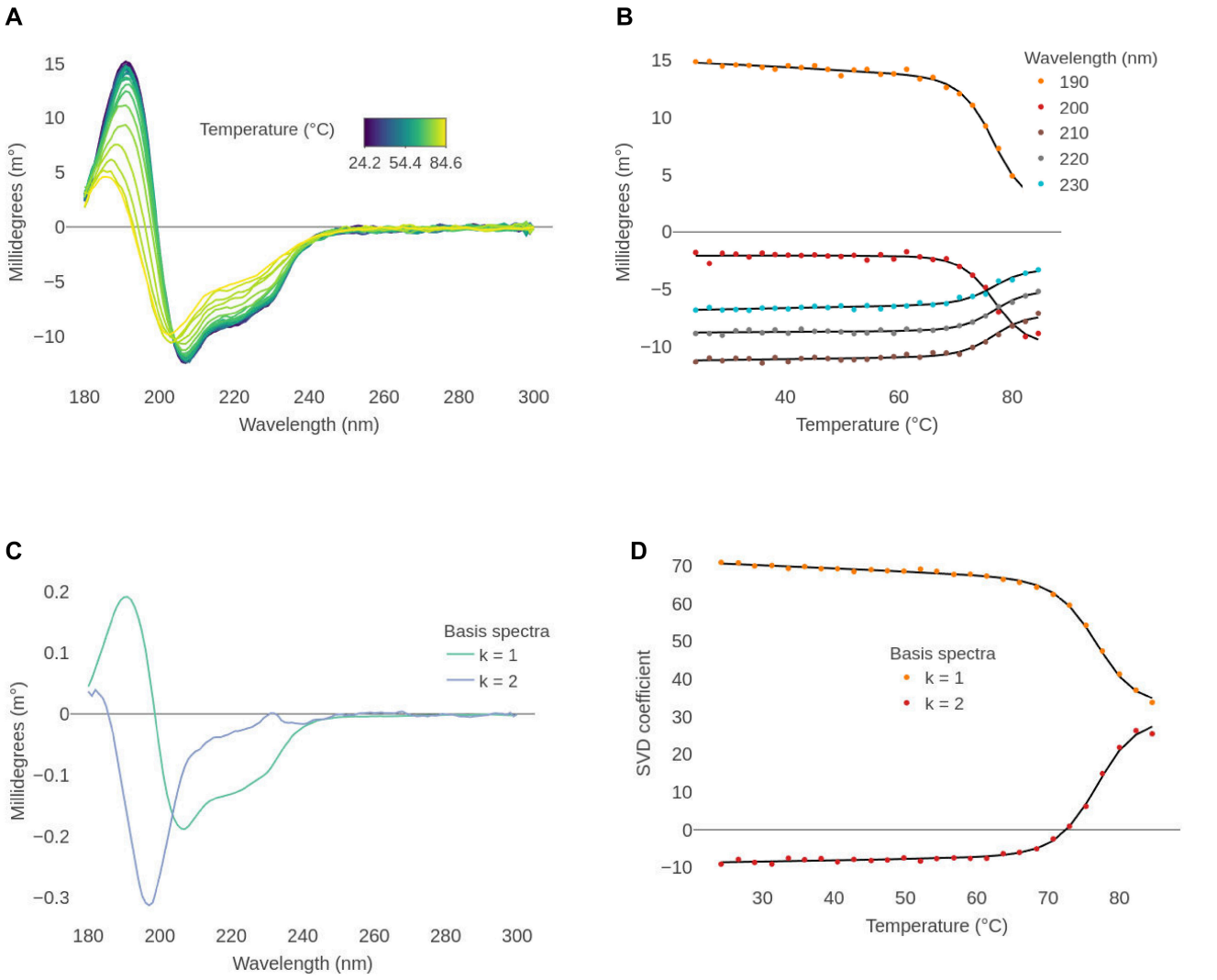


Figure 2. Two-state thermal unfolding analysis of lysozyme. (**A**) Finalized spectra after baseline subtraction and wavelength range correction. (**B**) Points: CD signal versus temperature, at five selected wavelengths. Lines: global fitted values based on a two-state reversible unfolding model. Baseline and slope (folded state only) values vary freely for each curve, while the parameters T_m and ΔH are shared. ΔCp was set to 2.3 kcal/K/mol. (**C**) SVD derived basis spectra required to reconstruct the original spectra. The cumulative explained variances are 97% and 99.93%. A rotation was performed such that the first basis spectrum resembles the spectrum measured at the lowest temperature. (**D**) Points: SVD coefficients versus temperature, for the two basis spectra. Lines: global fitted values based on a two-state reversible unfolding model. Baseline and slope (folded state only) values vary freely for each curve, while the parameters T_m and ΔH are shared. ΔCp was set to 2.3 kcal/K/mol.

Table 1. Comparison of CHC-NTD secondary structure composition as determined by the Selcon3 method versus structural data

Method	α	β	Turns	Other
SELCON	14	33	15	36
DSSP (PDBid 5m5r chain A)	8.3	51.4	15.6	24.7

Analysis of transient helicity in intrinsically disordered proteins

Intrinsically disordered proteins (IDPs) lack a fixed three-dimensional structure under physiological conditions. Nevertheless, they can transiently sample elements of secondary structure or acquire structure upon binding their partner proteins [52]. Common interaction motifs are the helical binding motifs, short IDP segments that fold into an α-helix upon binding [53]. When these regions are unbound, the helical structure is partially formed, and overlaps with the target-binding motif [54]. CD spectra enable quantification of helix content in IDP peptides undergoing folding-upon-binding. Since peptides exist as conformational ensembles, helicity cannot be estimated using algorithms such as SESCA. A traditional method is based on a simple algebraic relation that normalizes the measured signal at 222 nm to the ellipticity of maximal helix [55]. The drawback is that, since ellipticity depends on the length of the helix segment [56], it performs well only for peptides whose ensembles consist of conformers with long helical segments, approaching maximal helix length. In contrast, IDP ensembles typically consist of conformers with short or intermediate helices due to their lower helix propensity, leading to underestimation of helix content. The developed helix ensemble model [38], integrated in ChiraKit, estimates the average peptide helicity by summing together the individual length-corrected contributions of all conformers in the peptide ensemble.

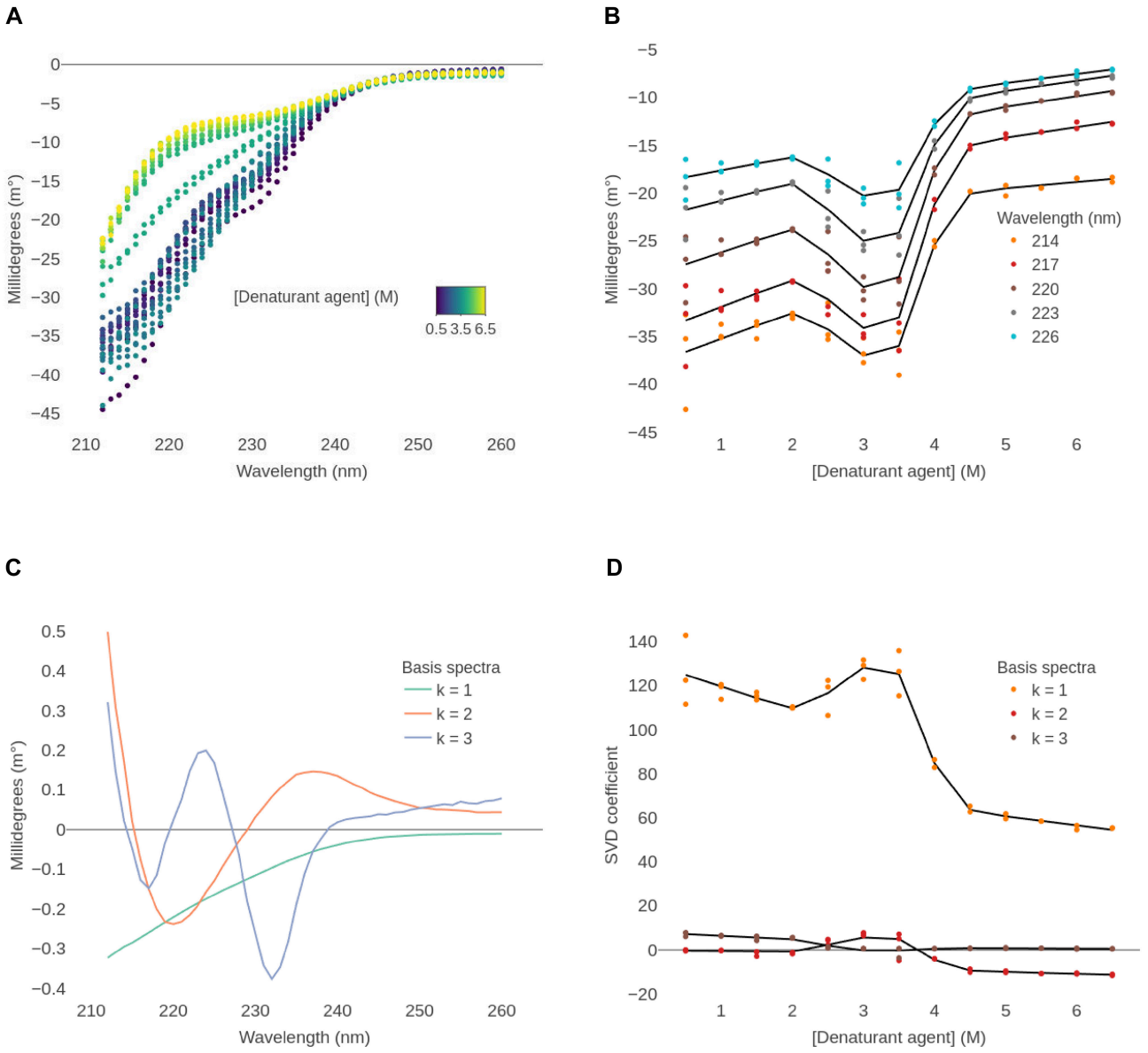


Figure 3. Three-state chemical unfolding analysis of the CHC-NTD. (**A**) Finalized spectra after baseline subtraction and wavelength range correction. (**B**) Points: CD signal versus urea concentration, at five selected wavelengths. Lines: global fitted values based on a three-state reversible unfolding model. Baseline and slope values vary freely for each curve, while the parameters $D50_1$, $D50_2$, m_1, and m_2 are shared. (**C**) SVD derived basis spectra required to reconstruct the original spectra. The cumulative explained variances are 99.43%, 99.84%, and 99.97%. A linear rotation was performed such that the first basis spectrum resembles the spectrum measured at the lowest urea concentration. (**D**) Points: SVD coefficients versus urea concentration for the three basis spectra. Lines: global fitted values based on a three-state reversible unfolding model. Baseline and slope values vary freely for each curve, while the parameters $D50_1$, $D50_2$, m_1, and m_2 are shared.

The CD spectra of a set of IDP peptides (Supplementary Table S1) that fold-upon-binding was measured using a benchtop instrument in phosphate buffer (Fig. 4A). Prior to analysis, spectra were buffer subtracted, and CD intensity was calibrated using camphor sulfonic acid. The intensity at 222 nm was then used to estimate fractional helicity (f_H) content, with obtained f_H values ranging between 0.1 to 0.3 (Supplementary Table S7). For comparison, we show that helicity estimated using the conventional approach strongly underestimates peptide helicity for IDPs (Supplementary Table S7, CcdA, error >200%) but performs rather well for peptides with high helix tendency (Lifson–Roig $w > 1.2$ [57]), such as alanine peptides (AK32, error ~5%). Note that the helix ensemble model assumes only an equilibrium between coil and helix conformations; therefore, it is not suitable for analysis of peptides exhibiting β-sheet propensity.

Thermal denaturation can provide insights into forces stabilizing helix structure and help with the energetic dissection of folding and binding contributions of IDP–target interactions. For the PaaA2 peptide, with considerable helicity ($f_H > 0.2$), we measured intensity at 222 nm as a function of temperature (Fig. 4B). Differently from proteins, where thermal denaturation induces highly cooperative transitions, in the case of peptides, loss of helix content occurs with low cooperativity. We used the helix ensemble model to analyze PaaA2 data available under "2.a. Thermal unfolding"

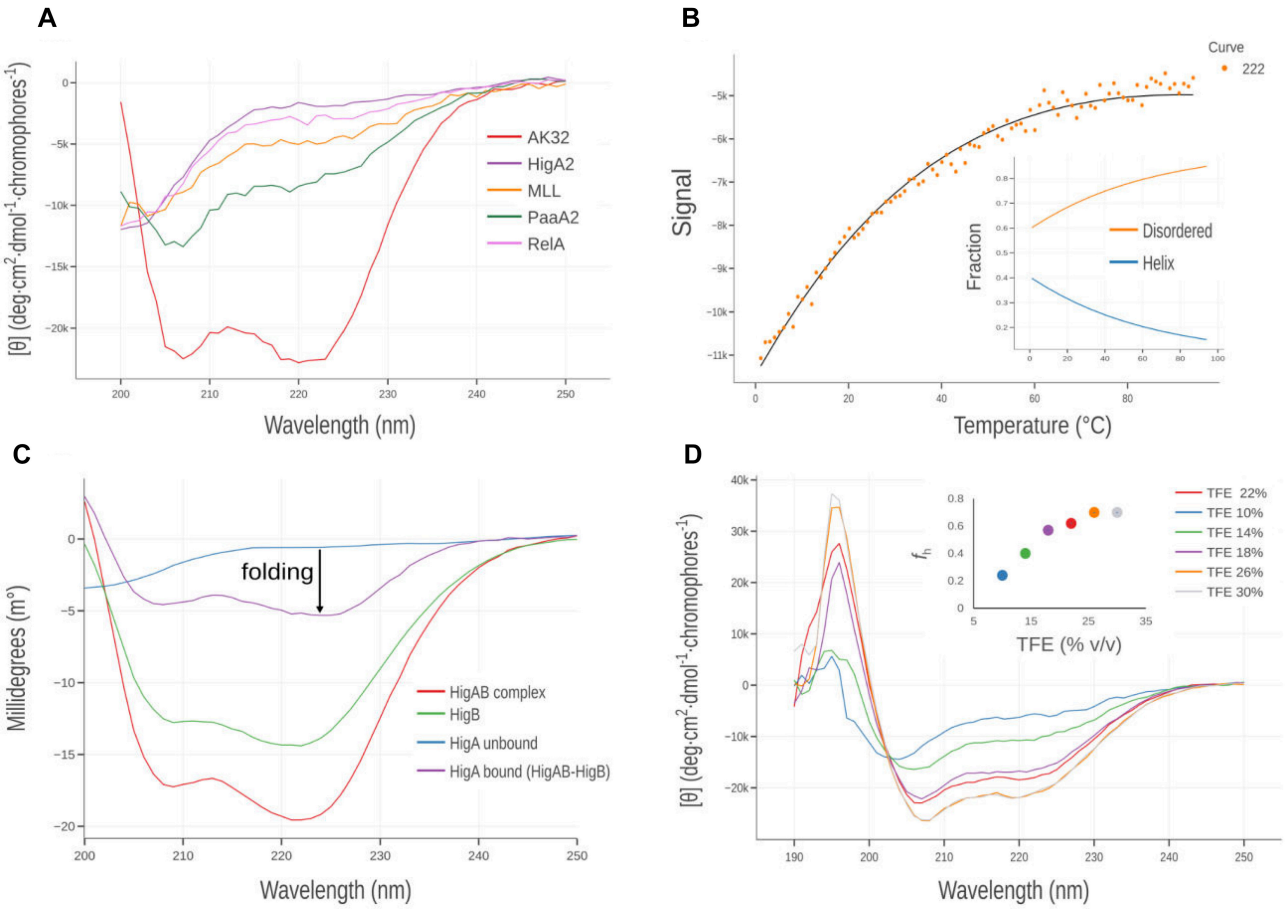


Figure 4. (**A**) Far-UV CD spectra of intrinsically disordered peptides (HigA2, MLL, PaaA2 and RelA) and alanine-rich peptide (AK32). Mean residue ellipticity at 222 nm and the corresponding helix fraction (f_H) estimated helicity are reported in Supplementary Table S7. (**B**) Thermal denaturation of the PaaA2 peptide monitored by CD intensity at 222 nm. Points represent experimental data, and the solid line shows the model fit using the ensemble model. The inset displays the ensemble model-derived overall helicity as a function of temperature. (**C**) Folding-upon-binding analysis of the intrinsically disordered antitoxin HigA upon binding to its globular target toxin HigB. The difference spectrum of the target-bound antitoxin (HigA) was obtained by subtracting the spectrum of the free toxin (HigB) from that of the HigBA complex. (**D**) TFE induction experiment of the disordered HigA peptide. The inset shows helicity as a function of the volume percentage of TFE.

module and obtained thermodynamic parameters of transition $\Delta G = -0.07$ and $\Delta H = -0.3$ kcal mol^{-1} pep. bond^{-1} (Fig. 4B). Assuming that 15 residues of PaaA2 fold into α-helix upon binding its target [58], the associated $\Delta H_{folding}$ for PaaA2 can be estimated to be ∼5 kcal mol^{-1}.

CD spectroscopy is an excellent tool to study IDP folding-upon-binding processes. For example, intrinsically disordered HigA folds into an α-helix upon binding the globular target HigB [59]. This process can be studied by measuring CD spectra of unbound HigA2 and HigB2 and a 1:1 HigA2–HigB2 complex (Fig. 4C). The ChiraKit Processing submodule (under 1. Import data) can then be used to obtain a difference spectrum (i.e. complex-HigB2), corresponding to folded HigA2, assuming no structural change in HigB2. Comparison with the unbound state (Supplementary Table S7) then shows that helix content in HigA2 increased from $f_H = 0.12$ to about $f_H = 0.7$ upon binding HigB2. Finally, in some cases where no binding partner is available, the potential of IDPs to adopt a helix conformation can be assessed using helix-inducing cosolvents. For example, titration with TFE of HigA2 peptide shows considerable gain in helicity which stabilizes ∼25% TFE, where it reaches $f_H = 0.7$, as calculated using ensemble model (Fig. 4D).

Re-analysis of published data

To further validate ChiraKit, we reanalyzed three published datasets with known parameters. These include one protein–DNA binding affinity measurement (ChiraKit module 2e. Custom analysis), two-state unfolding of a protein dimer (module 2c. Chemical unfolding) and three-state unfolding of a protein dimer (module 2c. Chemical unfolding). The three cases are presented in the Supporting Information (Supplementary Figs S2–S4).

Conclusions

In this work, we presented and evaluated ChiraKit, an open-source, free, multi-purpose tool for processing and analyzing CD data. ChiraKit supports multiple input file formats, multiple CD units and preprocessing functions. It does not require installation, nor registration and it provides a user-friendly, interactive interface. ChiraKit provides modules for the comparison of spectra, estimation of peptide helicity, estimation of protein secondary structure, and model-based analysis of thermal and chemical unfolding.

Regarding the case studies, the thermal unfolding of hen egg-white lysozyme could be explained with a two-state

reversible unfolding model. The chemical unfolding of the CHC-NTD follows a three-state process, with major structural changes occuring between the intermediate and final states. The secondary structure estimation of CHC shows an overall agreement with the values extracted from the PDB, with deviations of up to 2-fold.

IDPs may adopt secondary structures, such as α-helices, upon binding to partner proteins. The ChiraKit helix-coil ensemble model provides accurate helicity estimates by considering the contributions of all conformers in peptide ensemble, thus enabling estimation of helical content of unbound IDP and the analysis of the folding-upon-binding process.

In the presented cases, the whole pipeline, from the pre-preprocessing of the CD raw data to the fitting and figure generation, was done in ChiraKit. We expect that CD users find this tool useful and look forward to new developments.

Acknowledgements

We acknowledge technical support by the Sample Preparation and Characterization (SPC) facility at EMBL Hamburg, Germany, and the Synchrotron Radiation Circular Dichroism Facility at ASTRID2 (AU-SRCD), Aarhus University, Denmark. We thank the users of the SPC and AU-SRCD for constructive feedback. ChatGPT (https://chat.openai.com/) was used as an aid to correct written text. ChatGPT and Github Copilot (https://github.com/features/copilot) were used for coding assistance. The authors take full responsibility for the manuscript content and code.

Author contributions: Conceptualization: O.B. and M.M.G. Methodology: O.B., N.C.J., L.A.D., U.Z., S.H., S.V.H. and M.M.G. Investigation: N.C.J., L.A.D. and U.Z. Formal analysis: O.B. and U.Z. Resources: S.H., S.V.H. and M.M.G. Funding acquisition: S.H., S.V.H. and M.M.G. Project administration: M.M.G. Supervision: S.H., S.V.H. and M.M.G. Visualization: O.B. and U.Z. Data curation: O.B. Validation: All authors. Software: O.B., N.C.J. and U.Z. Writing–original draft: O.B. and M.M.G. Writing–review and editing: All authors.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 945405. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004806 (MOSBRI). This work was supported by grants from Slovenian Research and Innovation Agency (ARIS) with core funding P1-0201 and grants J1-50026 to San Hadži. Osvaldo Burastero is funded by the ARISE fellowship (EMBL and the Marie Skłodowska-Curie Actions). Funding to pay the Open Access publication charges for this article was provided by EMBL.

Data availability

ChiraKit is available at https://spc.embl-hamburg.de/app/chirakit. This website is free and open to all users and there is no login requirement. The source code has been deposited at Zenodo (https://doi.org/10.5281/zenodo.15187609). A containerized version that can be run as a desktop application is available at https://hub.docker.com/repository/docker/oburastero/chirakit_circular_dichroism/.

References

1. Woody RW. Theory of circular dichroism of proteins. In *Circular Dichroism and the Conformational Analysis of Biomolecules*. Springer US, Boston, MA, 1996; pp.25–67. https://doi.org/10.1007/978-1-4757-2508-7
2. Kypr J, Kejnovská I, Renciuk D *et al.* Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res* 2009;**37**:1713–25. https://doi.org/10.1093/nar/gkp026
3. Kypr J, Kejnovská I, Bednářová K *et al.* Circular dichroism spectroscopy of nucleic acids. In *Comprehensive Chiroptical Spectroscopy*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012; pp. 575–86. https://doi.org/10.1002/9781118120392
4. del Villar-Guerra R, Trent JO, Chaires JB. G-quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew Chem* 2018;**130**:7289–93. https://doi.org/10.1002/ange.201709184
5. Toumadje A, Alcorn SW, Johnson WC Jr. Extending CD spectra of proteins to 168 nm improves the analysis for secondary structures. *Anal Biochem* 1992;**200**:321–31. https://doi.org/10.1016/0003-2697(92)90473-K
6. Seelig J, Schönfeld H-J. Thermal protein unfolding by differential scanning calorimetry and circular dichroism spectroscopy two-state model versus sequential unfolding. *Quart Rev Biophys* 2016;**49**:e9. https://doi.org/10.1017/S0033583516000044
7. Yadav S, Gupta S, Saxena JK. Monitoring thermal and chemical unfolding of Brugia malayi calreticulin using fluorescence and circular dichroism spectroscopy. *Int J Biol Macromol* 2017;**102**:986–95. https://doi.org/10.1016/j.ijbiomac.2017.04.053
8. Garbett NC, Ragazzon PA, Chaires JB. Circular dichroism to determine binding mode and affinity of ligand–DNA interactions. *Nat Protoc* 2007;**2**:3166–72. https://doi.org/10.1038/nprot.2007.475
9. Burastero O, Defelipe LA, Gola G *et al.* Cosolvent sites-based discovery of protein kinase G inhibitors. *J Med Chem* 2022;**65**:9691–705. https://doi.org/10.1021/acs.jmedchem.1c02012
10. Madsen MM, Jones NC, Nielsen SB *et al.* On the wavelength dependence of UV induced thymine photolesions: a synchrotron radiation circular dichroism study. *Phys Chem Chem Phys* 2016;**18**:30436–43. https://doi.org/10.1039/C6CP05980E
11. Miles AJ, Wallace BA. CDtoolX, a downloadable software package for processing and analyses of circular dichroism spectroscopic data. *Protein Sci* 2018;**27**:1717–22. https://doi.org/10.1002/pro.3474
12. Miles AJ, Ramalli SG, Wallace BA. DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data. *Protein Sci* 2022;**31**:37–46. https://doi.org/10.1002/pro.4153
13. Micsonai A, Moussong É, Wien F *et al.* BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic Acids Res* 2022;**50**:W90–8. https://doi.org/10.1093/nar/gkac345
14. Sreerama N, Woody RW. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem* 1993;**209**:32–44. https://doi.org/10.1006/abio.1993.1079

15. Provencher SW, Glöckner J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 1981;**20**:33–7. https://doi.org/10.1021/bi00504a006
16. Compton LA, Johnson WC Jr. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem* 1986;**155**:155–67. https://doi.org/10.1016/0003-2697(86)90241-1
17. Manavalan P, Johnson WC Jr. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal Biochem* 1987;**167**:76–85. https://doi.org/10.1016/0003-2697(87)90135-7
18. Andrade MA, Chacón P, Merelo JJ *et al.* Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng Des Sel* 1993;**6**:383–90. https://doi.org/10.1093/protein/6.4.383
19. Micsonai A, Wien F, Kernya L *et al.* Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 2015;**112**:E3095–103. https://doi.org/10.1073/pnas.1500851112
20. Nagy G, Igaev M, Jones NC *et al.* SESCA: predicting circular dichroism spectra from protein molecular structures. *J Chem Theory Comput* 2019;**15**:5087–102. https://doi.org/10.1021/acs.jctc.9b00203
21. Nagy G, Grubmuller H. Implementation of a Bayesian secondary structure estimation method for the SESCA circular dichroism analysis package. *Comput Phys Commun* 2021;**266**:108022. https://doi.org/10.1016/j.cpc.2021.108022
22. Niklasson M, Andresen C, Helander S *et al.* Robust and convenient analysis of protein thermal and chemical stability. *Protein Sci* 2015;**24**:2055–62. https://doi.org/10.1002/pro.2809
23. Kunka A, Lacko D, Stourac J *et al.* CalFitter 2.0: leveraging the power of singular value decomposition to analyse protein thermostability. *Nucleic Acids Res* 2022;**50**:W145–51. https://doi.org/10.1093/nar/gkac378
24. Mazurenko S, Stourac J, Kunka A *et al.* CalFitter: a web server for analysis of protein thermal denaturation data. *Nucleic Acids Res* 2018;**46**:W344–9. https://doi.org/10.1093/nar/gky358
25. Sathyaseelan C, Vijayakumar V, Rathinavelan T. CD-NuSS: a web server for the automated secondary structural characterization of the nucleic acids from circular dichroism spectra using extreme gradient boosting decision-tree, neural network and Kohonen algorithms. *J Mol Biol* 2021;**433**:166629. https://doi.org/10.1016/j.jmb.2020.08.014
26. Ramalli SG, Miles AJ, Janes RW *et al.* The PCDDB (Protein Circular Dichroism Data Bank): a bioinformatics resource for protein characterisations and methods development. *J Mol Biol* 2022;**434**:167441. https://doi.org/10.1016/j.jmb.2022.167441
27. Cappannini A, Mosca K, Mukherjee S *et al.* NACDDB: nucleic acid circular dichroism database. *Nucleic Acids Res* 2023;**51**:D226–31. https://doi.org/10.1093/nar/gkac829
28. Burastero O, Niebling S, Defelipe LA *et al.* eSPC: an online data-analysis platform for molecular biophysics. *Acta Crystallogr D Struct Biol* 2021;**77**:1241–50. https://doi.org/10.1107/S2059798321008998
29. Burastero O, Draper-Barr G, Raynal B *et al.* Raynals, an online tool for the analysis of dynamic light scattering. *Acta Crystallogr D Struct Biol* 2023;**79**:673–83. https://doi.org/10.1107/S2059798323004862
30. Niebling S, Burastero O, García-Alai M. Biophysical characterization of membrane proteins. *Methods Mol Biol* 2023;**2652**:215–30.
31. Myers JK, Pace CN, Scholtz JM. Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 1995;**4**:2138–48. https://doi.org/10.1002/pro.5560041020
32. Scholtz JM, Grimsley GR, Pace CN. Chapter 23 solvent denaturation of proteins and interpretations of the *m* value. In *Methods in Enzymology*. Elsevier, 2009, **466**, 549–65. https://doi.org/10.1016/S0076-6879(09)66023-7
33. Gill SC, Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 1989;**182**:319–26. https://doi.org/10.1016/0003-2697(89)90602-7
34. Rogers DM, Jasim SB, Dyer NT *et al.* Electronic circular dichroism spectroscopy of proteins. *Chem* 2019;**5**:2751–74. https://doi.org/10.1016/j.chempr.2019.07.008
35. Sreerama N, Venyaminov SY, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal Biochem* 2000;**287**:243–51. https://doi.org/10.1006/abio.2000.4879
36. Lees JG, Miles AJ, Wien F *et al.* A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* 2006;**22**:1955–62. https://doi.org/10.1093/bioinformatics/btl327
37. Abdul-Gader A, Miles AJ, Wallace BA. A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. *Bioinformatics* 2011;**27**:1630–6. https://doi.org/10.1093/bioinformatics/btr234
38. Zavrtanik U, Lah J, Hadži S. Estimation of peptide helicity from circular dichroism using the ensemble model. *J Phys Chem B* 2024;**128**:2652–63. https://doi.org/10.1021/acs.jpcb.3c07511
39. Kim D, You K. PCA, SVD, and Centering of Data. arXiv, https://doi.org/10.48550/arXiv.2307.15213, 27 July 2023,preprint: not peer reviewed.
40. Baggesen LM, Hoffmann SV, Nielsen SB. On the formation of thymine photodimers in thymine single strands and calf thymus DNA. *Photochem Photobiol* 2014;**90**:99–106. https://doi.org/10.1111/php.12183
41. Seelig J, Seelig A. Protein unfolding-thermodynamic perspectives and unfolding models. *Int J Mol Sci* 2023;**24**:5457. https://doi.org/10.3390/ijms24065457
42. Pace CN, Hermans J. The stability of globular protein. *CRC Crit Rev Biochem* 1975;**3**:1–43. https://doi.org/10.3109/10409237509102551
43. Mazurenko S, Kunka A, Beerens K *et al.* Exploration of protein unfolding by modelling calorimetry data from reheating. *Sci Rep* 2017;**7**:16321. https://doi.org/10.1038/s41598-017-16360-y
44. Jasanoff A, Fersht AR. Quantitative determination of helical propensities from trifluoroethanol titration curves. *Biochemistry* 1994;**33**:2129–35. https://doi.org/10.1021/bi00174a020
45. de Prat-Gay G. Conformational preferences of a peptide corresponding to the major antigenic determinant of foot-and-mouth disease virus: implications for peptide-vaccine approaches. *Arch Biochem Biophys* 1997;**341**:360–9. https://doi.org/10.1006/abbi.1997.9982
46. Oyama T, Suzuki S, Horiguchi Y *et al.* Performance comparison of spectral distance calculation methods. *Appl Spectrosc* 2022;**76**:1482–93. https://doi.org/10.1177/00037028221121687
47. 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* 2025;**6**:e10. https://doi.org/10.1017/qrd.2025.4
48. Knubovets T, Osterhout JJ, Connolly PJ *et al.* Structure, thermostability, and conformational flexibility of hen egg-white lysozyme dissolved in glycerol. *Proc Natl Acad Sci USA* 1999;**96**:1262–7. https://doi.org/10.1073/pnas.96.4.1262
49. Li-Blatter X, Seelig J. Thermal and chemical unfolding of lysozyme. Multistate Zimm-Bragg theory versus two-State model. *J Phys Chem B* 2019;**123**:10181–91. https://doi.org/10.1021/acs.jpcb.9b08816
50. Defelipe LA, Veith K, Burastero O *et al.* Subtleties in clathrin heavy chain binding boxes provide selectivity among adaptor

proteins of budding yeast. *Nat Commun* 2024;**15**:9655. https://doi.org/10.1038/s41467-024-54037-z
51. Schellman JA. Solvent denaturation. *Biopolymers* 1978;**17**:1305–22. https://doi.org/10.1002/bip.1978.360170515
52. van der Lee R, Buljan M, Lang B *et al.* Classification of intrinsically disordered regions and proteins. *Chem Rev* 2014;**114**:6589–631. https://doi.org/10.1021/cr400525m
53. Zavrtanik U, Hadži S, Lah J. Unraveling the thermodynamics of ultra-tight binding of intrinsically disordered proteins. *Front Mol Biosci* 2021;**8**:726824. https://doi.org/10.3389/fmolb.2021.726824
54. Zavrtanik U, Medved T, Purič S *et al.* Leucine motifs stabilize residual helical structure in disordered proteins. *J Mol Biol* 2024;**436**:168444. https://doi.org/10.1016/j.jmb.2024.168444
55. Scholtz JM, Qian H, York EJ *et al.* Parameters of helix–coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers* 1991;**31**:1463–70. https://doi.org/10.1002/bip.360311304
56. Chen YH, Yang JT, Chau KH. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* 1974;**13**:3350–9. https://doi.org/10.1021/bi00713a027
57. Lifson S, Roig A. (06) On the theory of helix–coil transition in polypeptides. *J Chem Phys* 1961;**34**:1963c74. https://doi.org/10.1063/1.1731802
58. Sterckx YG-J, Jové T, Shkumatov AV *et al.* A unique hetero-hexadecameric architecture displayed by the *Escherichia coli* O157 PaaA2-ParE2 antitoxin–toxin complex. *J Mol Biol* 2016;**428**:1589–603. https://doi.org/10.1016/j.jmb.2016.03.007
59. Hadži S, Garcia-Pino A, Haesaerts S *et al.* Ribosome-dependent *Vibrio cholerae* mRNAse HigB2 is regulated by a β-strand sliding mechanism. *Nucleic Acids Res* 2017;**45**:4972–83. https://doi.org/10.1093/nar/gkx138

Received: January 30, 2025. **Revised:** March 26, 2025. **Editorial Decision:** April 8, 2025. **Accepted:** April 24, 2025