







A workflow to define structural classes and classify nucleic acids circular dichroism spectra

Kevin Mosca^{1,2,3} , Søren Vrønning Hoffmann⁴ , Alice Grangier¹ , Frank Wien³ ,
Veronique Arluison^{2,5}  and Sergio Marco⁶ 

Perspective

Cite this article: Mosca K, Hoffmann SV, Grangier A, Wien F, Arluison V, Marco S (2025). A workflow to define structural classes and classify nucleic acids circular dichroism spectra. *QRB Discovery*, 6: e22, 1–7 <https://doi.org/10.1017/qrd.2025.10008>.

Received: 27 January 2025

Revised: 13 May 2025

Accepted: 14 May 2025

Keywords:

Circular Dichroism; Spectroscopy; Nucleic Acids; Spectral Analysis

Corresponding author:

Sergio Marco;

Email: sergio.marco@sanofi.com

Abstract

Circular dichroism (CD) spectroscopy is a widely utilized technique for studying the structures of chiral molecules, including nucleic acids. It is particularly valued for its ability to quickly probe structural changes in these biomolecules. Despite its potential, the prediction of nucleic acid structures by CD has been challenging due to insufficient families' reference spectral data. This study introduces a robust method for defining CD spectra families of nucleic acid structures in solution. Our approach demonstrates high robustness and accuracy in assigning CD spectra to specific nucleic acid folds, facilitating advancements in nucleic acid structure analysis. The algorithm we developed identifies structural classes based on reference spectra, aiding in the assignment of unknown spectra. This method paves the way for creating a comprehensive list of reference spectra for various nucleic acid structures, like those already available for proteins.

Introduction

Circular dichroism (CD) spectroscopy is a widely utilized technique for studying the structures of chiral molecules (Gottarelli et al., 2008; Nordén et al., 2010). While extensively employed in biology to determine the secondary structure of proteins (Greenfield, 2006), CD can also be applied to investigate other chiral biomolecules, such as nucleic acids, which include many forms (Gray et al., 1981; Steely et al., 1986; Johnson, 1990; Gray et al., 1995; Kypr et al., 2009; Del Villar-Guerra et al., 2018). This makes CD suitable for studying their folding patterns as well, which is important for understanding the functions of nucleic acid sequences. The interest of using CD measurements is that it is a fast, non-destructive method to identify nucleic acid folding. Compared to other techniques, such as crystallography, NMR, or cryo-electron microscopy that provide 3D information (Neidle and Sanderson, 2022a, 2022b), CD can be applied to water solutions of nucleic acids without impacting their structure. For protein studies, a circular dichroism structural database (PCDDb) exists (Ramalli et al., 2022), enabling the indexing of unknown structures by comparing their CD spectra to those referenced in the PCDDb and other sources (Micsonai et al., 2022; Nagy et al., 2024). For nucleic acids, a similar database exists called the NACDDb, comprising previous and new spectra (Cappannini et al., 2023). However, due to the flexibility, structural variability, and greater repartition of electronic transitions within a larger distance compared to the peptide bond, the number of possible spectra observed for polynucleotides is extensive compared to proteins. Four major secondary structural types have been assigned in proteins (α -helix, β -sheet, turn, and random coil) (Manavalan and Johnson, 1987; Sreerama and Woody, 1994; Wallace, 2009; Kuril et al., 2024). For protein CD, basis spectra have been correlated to secondary structure elements. They correspond to secondary structure subclasses, distinguishing regular and distorted α -helices, parallel β -sheets (including three distinct twisting patterns) and antiparallel β -sheets, turns, and others (Micsonai et al., 2022; Burastero et al., 2025). Multivariate statistical analysis performed in the NACDDb has revealed a greater number of potential reference spectra compared to proteins due to their numerous sequence specificities and chemical variations (Cappannini et al., 2023).

As a result, there is currently no well-established list of reference spectra for known nucleic acid secondary structures, which hinders the assignment of unknown nucleic acid CD spectra to specific structures.

Classical approaches, such as multivariate statistical analysis or other unsupervised classification methods, as shown in Supplementary Figure 1a,b, are unsuitable for establishing these reference spectra due to the structural heterogeneity and the limited availability of CD spectroscopic data for nucleic acids. One relevant factor to consider is the correct annotation of spectra, as different spectra assigned to the same structural families and a wide spectral range may exist in

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<http://creativecommons.org/licenses/by-nc/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

the literature. Also, while most spectra in the literature have been acquired in the 190 to 300 nm range, it has been demonstrated that spectral extension down to the far UV (170 nm), accessible by synchrotron radiation circular dichroism (SRCD) or with the very latest CD top bench spectrometers, is crucial for discriminating structural families (Gray *et al.*, 1992; Le Brun *et al.*, 2020). Due to all these limitations, there is currently no robust method to classify nucleic acid CD spectra. To address this issue, we have developed a workflow identifying different structural classes and determining their corresponding reference spectra.

The established workflow enabled us to determine reference spectra for five well-known structures: parallel DNA quadruplexes, DNA triplexes, Z-DNA, DNA, and RNA stem loops (Sinden, 1994; Vanegas *et al.*, 2012). Moreover, the method's robustness was demonstrated by correctly assigning unknown spectra (predicting their structure) to the correct spectral family and reclassifying spectra manually assigned to the incorrect family. This workflow can thus serve as a useful tool to create a list of reference spectra for nucleic acids' various structures, akin to those existing for proteins, and to assign unknown spectra to a defined family. Due to redundancy issues and the limited number of available spectra compared to assigned structures, it is currently not possible to expand the list beyond five structures (basis spectra). However, we are convinced that this number will increase as additional spectra are published or made publicly available. In the future, a complementary approach that allows determining the number or percentage of distinct structures in larger and more complex nucleic acids could be developed.

Methods

CD data sets

The dataset utilized for developing our workflow comprises 118 spectra. Among these, 64 were sourced from the NACDDB (Cappannini *et al.*, 2023), with 59 initially acquired on the DISCO beamline of the Synchrotron SOLEIL and the other 5 originating from the literature (Gray *et al.*, 1981; Steely *et al.*, 1986; Johnson, 1990; Gray *et al.*, 1995; Del Villar-Guerra *et al.*, 2018). Among the remaining 54 spectra not yet included in the NACDDB, 48 have been acquired from the DISCO beamline and 6 from literature (AI Holm *et al.*, 2010; Vanloon *et al.*, 2023). All spectra in the dataset are scaled to differential molar ellipticity ($\Delta\epsilon$) following the formulae:

$$\Delta\epsilon = \frac{\theta \times \text{MRW}}{\text{PL} \times C \times 3298}$$

where θ is the circular dichroism measured in millidegrees (mdeg), MRW is the mean residue weight of the sample ($\text{g.mol}^{-1}.\text{residue}^{-1}$), PL is the path length in centimeters (cm), C is the concentration of the sample in grammes per liter (g L^{-1}), and 3298 is a constant used for unit conversion. In all, the $\Delta\epsilon$ is expressed in $\text{M}^{-1}.\text{cm}^{-1}.\text{residue}^{-1}$.

The spectra retained were the ones including signals between 175 and 300 nm. This range presents characteristic UV absorption maxima and minima corresponding to the absorption by electronic transition of the base pairing, stacking, and overall twisting of the polynucleotide, e.g. $n \rightarrow p$, $p \rightarrow p^*$ as well as $n \rightarrow s^*$ (Miyahara *et al.*, 2012, 2016). Although a clear attribution of electronic transitions within a strand of nucleic acids has not yet been established, there exists a few indications, such as the 260–280 nm CD-absorption band for the base stacking, a band around 190 nm for the backbone conformation, and another one in-between for the twisting of helical nucleic acids (AI Holm *et al.*, 2010).

For workflow validation, a validation subset of 56 spectra, each corresponding to a well-known nucleic acid structure, was established. The list of utilized spectra and their corresponding structures is presented in Supplementary Table 1. It contains 7 families: DNA quadruplexes parallel (3 spectra), DNA triplexes (6 spectra), Z-DNA (3 spectra), DNA loops (3 spectra), RNA loops (15 spectra), DNA loops (6 spectra), and unclassified spectra (20 spectra). The latter group comprises spectra belonging to 11 other structural families but with representative spectra count per family lower than 3.

Statistical tools

Spectra normalization

All scaled spectra used in this work have been normalized to average 0 and standard deviation 1. This normalization ensures that spectra are comparable to a centered normal distribution weighing the contribution of high amplitudes that would otherwise biasing the analysis. To achieve this, we calculated the mean and standard deviation for all wavelengths of each spectrum, then subtracted that mean and divided it by the standard deviation for all wavelengths.

Self-organizing mapping

For classification methods, we employed a simple neural network known as Kohonen self-organizing maps (SOM), which has previously been used to classify nucleic acid CD spectra data (Sathyaseelan *et al.*, 2021). The implementation of the SOM was performed using the MiniSom Python package (<https://github.com/JustGlowing/minisom/>). The neural network was customized using several parameters following recommendations from the MiniSom function package built-in help.

Multivariate statistical analysis

Multivariate analysis was conducted on the entire dataset to group the spectra into families. Initially, hierarchical clustering was employed using the Ward method and Euclidean distances between each pair of spectra. This analysis was conducted using the Python package SciPy (<http://www.scipy.org/>). Additionally, principal component analysis was carried out simultaneously using the SIMCA software (V17) to identify clusters and significant components for class differentiation.

Singular value decomposition

Singular value decomposition (SVD) was performed using the NumPy Python package (Harris *et al.*, 2020) to identify the initial references during workflow initialization.

Normalized correlation coefficient and normalized mutual information

A normalized correlation coefficient (NCC) was used to measure the linear resemblance of a family's reference spectrum compared to all spectra of our dataset, whereas normalized mutual information (NMI) was to measure non-linear resemblance.

The NCC is defined as

$$\text{NCC}(X, Y) = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}}$$

where X and Y correspond to spectra value vector, x and y to spectra value at a wavelength and m_x and m_y to the means of the spectra.

These coefficients were computed using the Python package SciPy (<http://www.scipy.org/>).

Each NMI was computed using the mutual information coefficient (I) defined as

$$I(X, Y) = \sum P(X \cap Y) \log \frac{P(X \cap Y)}{P(X)P(Y)}$$

where X and Y corresponding to spectra value vectors, $P(X)$ and $P(Y)$ the probability for the spectra to reach a certain value and $P(X \cap Y)$ the probability for both spectra to reach the same value at the same wavelength. Probability values are calculated from integer rounded spectral intensities.

Entropy (H) is defined as

$$H(X) = - \sum P(X) \log P(X)$$

the NMI was then calculated by applying the following equation:

$$NMI(X, Y) = \frac{I(X, Y)}{[H(X) + H(Y)]/2}$$

as implemented in the scikit-learn Python package.

Workflow initialization

The workflow was initiated by manually defining structural families based on the theoretical understanding of their structures.

The assignment of a spectrum to a particular family relies on the anticipated structure of an oligonucleotide sequence and the characteristics of the spectrum, including the position and intensity of its peaks when normalized. Once a structural family accumulates at least four normalized spectra (heuristically determined value), an SVD is performed on it to define an initial reference spectrum (first eigen-vector) for the family. Subsequently, the initial reference spectrum is validated by ensuring that the spectra forming the basis of the reference exhibit an NCC and an NMI whose product exceeds a threshold value. This threshold value is determined by identifying the first peak above the baseline in the derivative of this product.

Workflow validation

To assess the robustness of the workflow, three metrics were calculated: sensitivity, specificity, and similarity (Jaccard index). Each of the seven families within the validation subset was individually evaluated against the entire validation subset by running the workflow. For each run, the number of true positives, false positives, and false negatives was determined. The total count for each category was then calculated by summing the results obtained for each family. These cumulative totals were utilized to compute the value of each figure of merit.

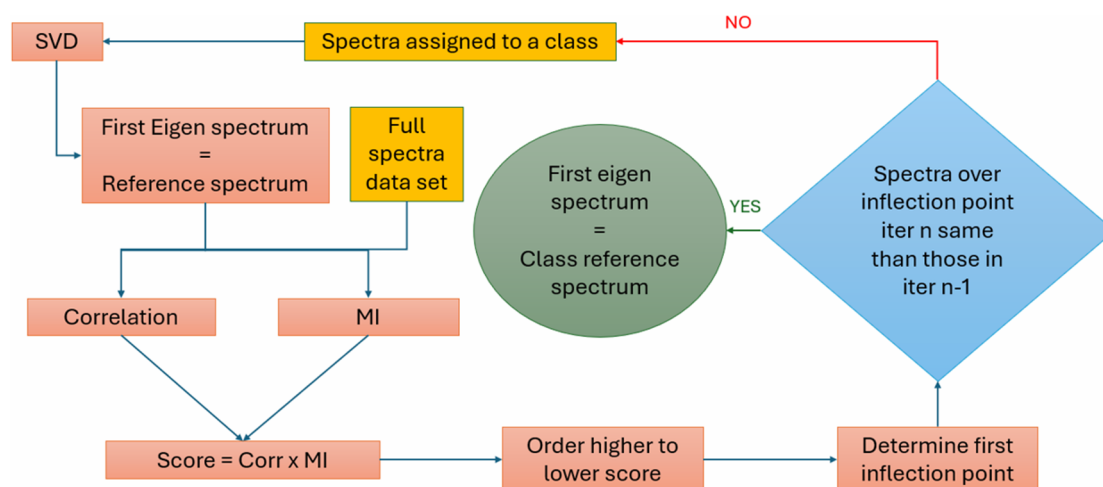


Figure 1. Graphical diagram of the iterative workflow. In yellow the point where data are selected, red are mathematical operation, blue the decision point and green the output.

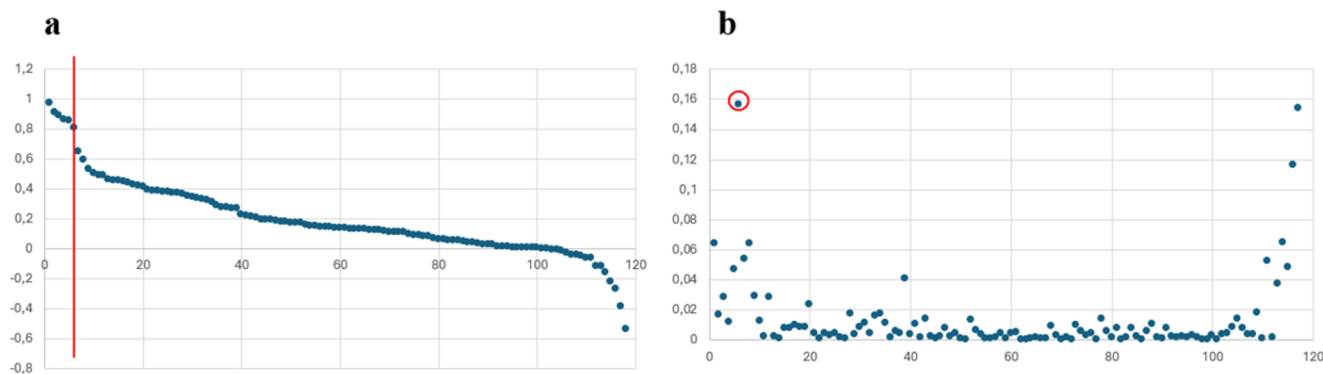


Figure 2. Example of threshold for class assignment. (a) Plot of the correlation multiplied by the mutual information ordered from higher to lower values. Red line depicts the inflexion point where score above are spectra kept for the class. (b) First derivative of data shown in (a). Red circle evidences the position of the inflection point used to defining the threshold shown in (a). Abscises correspond to the spectrum position ordered from higher to lower NCC and NMI product. Ordinates has no unit as it corresponds to coefficient product value.

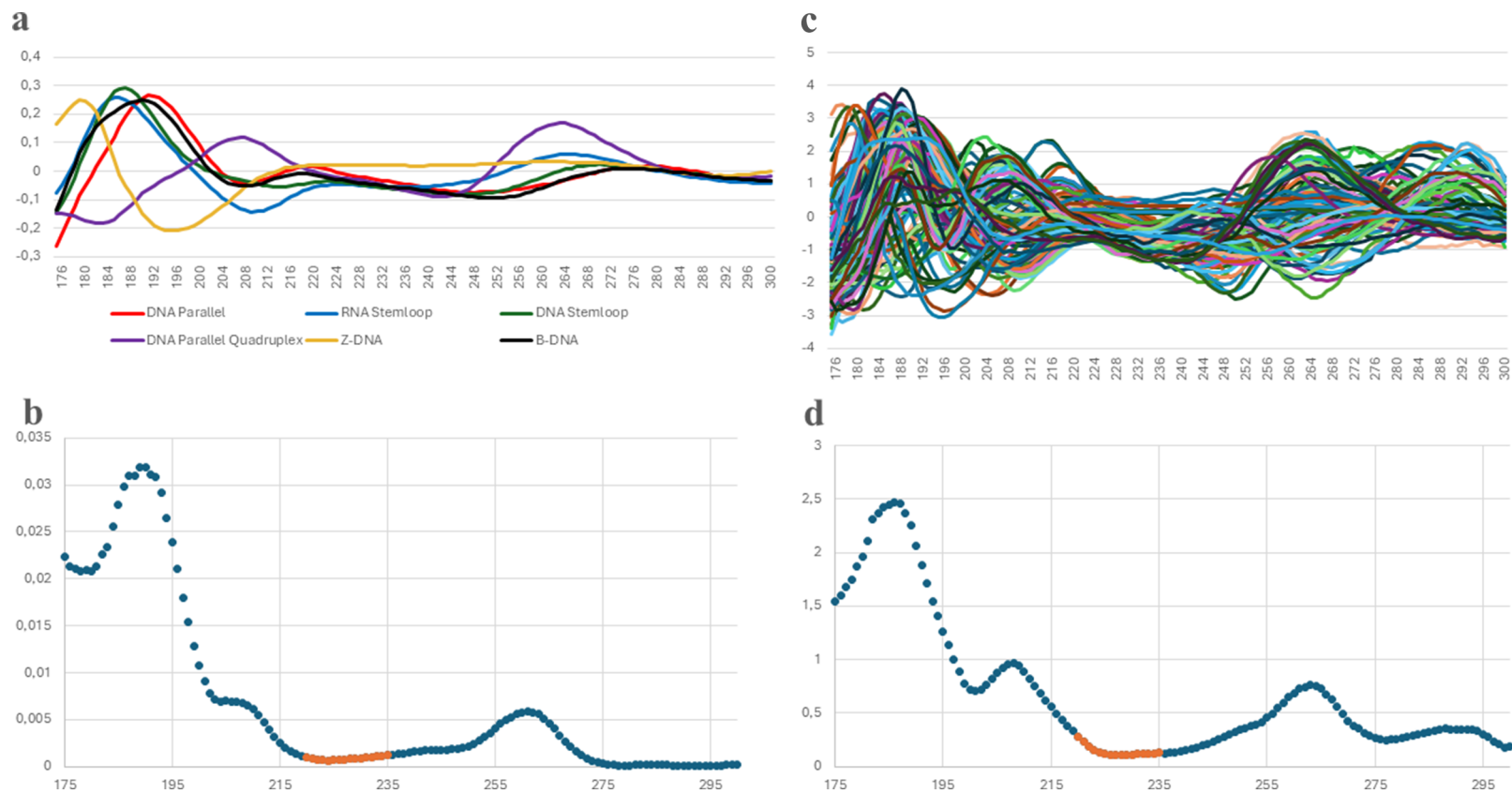


Figure 3. Comparison of the reference spectra and the dataset used. (a) reference spectra obtained after spectra value decomposition and (b) the variance of this data set at each wavelength. (c) The whole dataset showed normalized to have comparable intensities and (d) its' associated variance at each wavelength. The orange points are the ones where chirality is invariant in the two datasets. Abscises correspond to the wavelength in nanometers (nm). Ordinates has no unit as it corresponds to normalized data or variances.

Results

Due to the large structural heterogeneity and the yet limited availability of relevant CD spectroscopic data, the use of multivariate statistical analysis and neural networks does not produce relevant and reproducible results. Specifically, the first eigen-vectors explaining the highest percentage of variance by principal component analysis do not correspond to any spectrum having physical significance. Moreover, hierarchical classification (Supplementary Figure 1) merges spectra belonging to different structural families. Equivalent results, with inconsistent family assignments, were observed for self-organizing mapping. Therefore, we chose to combine approaches targeting two different types of information: shape similarity (by using the NCC) and probability of value occurrence (NMI).

Workflow allowing to define nucleic acids structural classes from CDs spectra

Based on NCC and NMI, we have established an iterative workflow (Figure 1) to determine the reference spectrum for each structural family. The workflow is applied to each manually defined family determined during the initialization process as follows:

- (1) MCC and NMI values are calculated between every normalized $N(0,1)$ spectrum from our dataset and the reference spectrum for the family.
- (2) The product of these values ($\text{Score} = \text{NCC} \times \text{NMI}$) is ordered from highest to lowest, thus defining the order as an abscissa (X_n) and the result of the product as an ordinate (Y_n).
- (3) The first derivative of the (X_n, Y_n) array is computed to determine the position of the first inflection point.
- (4) The coordinates of the first inflection point are used as a family belonging threshold (Figure 2).

- (5) Spectra whose Score are above the Score at the inflection point are included in the family, regardless of whether they were part of the initial group used for the family definition.
- (6) SVD is computed from all spectra of a family and the first component is used as the new reference spectrum for that family.
- (7) The process is repeated from (1) until the included spectra are constant (convergence of the iterative workflow).

Once convergence is reached, the first component of SVD computed from a family's normalized spectra is set as the reference spectrum for that structural family.

Evaluation of the workflow

Once the five CD reference spectra have been determined, the robustness and accuracy of the workflow was evaluated by using a data set of 56 manually assigned spectra and standardized figures of merit as described in materials and methods. Sensibility, specificity, and Jaccard (similarity) values were 1, 0.94, and 0.94, respectively. This confirms that the workflow is robust enough to assign unknown spectra to one of the defined families. Other workflows previously described in the literature appear to be less accurate with 87.33%, 85.33%, and 78.66% for the XGBoost algorithm, neural network, and Kohonen approaches, respectively (Sathyaseelan et al., 2021).

Applications and workflow limits

Based on the robustness of the workflow, we have successfully defined reference spectra for five families using 118 normalized spectra, with or without initial manual assignment. These families are DNA quadruplexes parallel, DNA triplexes, Z-DNA, DNA loops, and RNA loops. The superposition of the five references

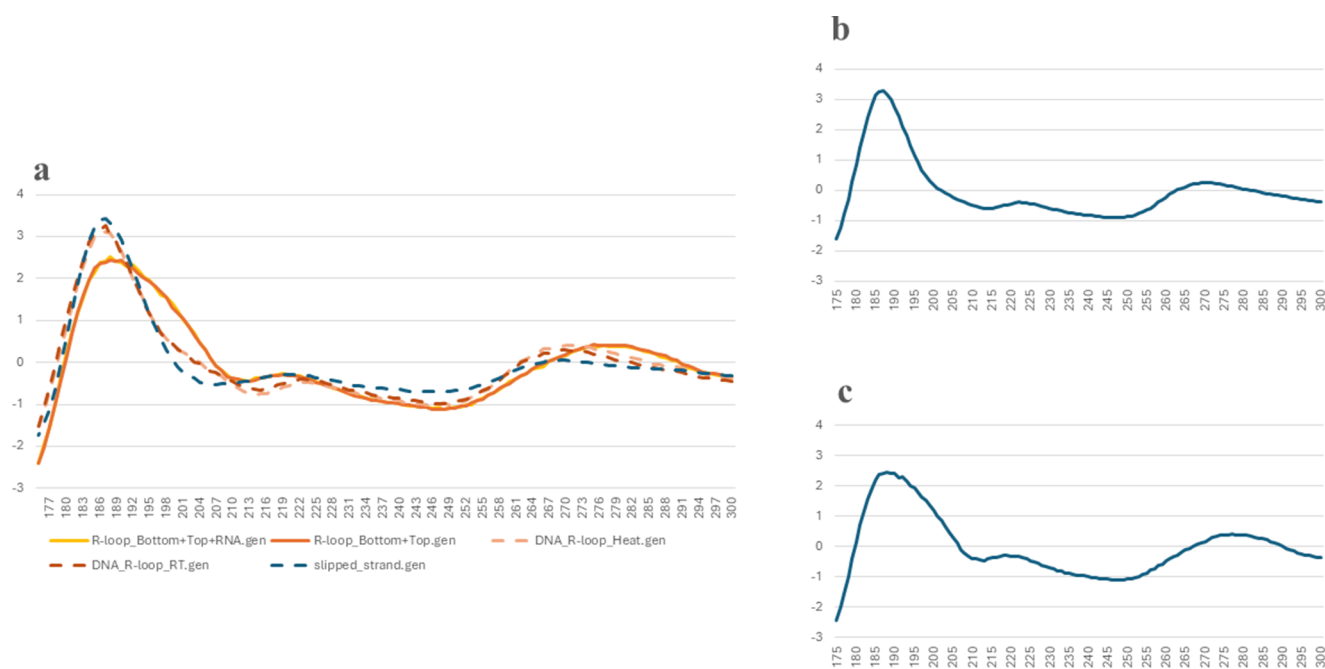


Figure 4. Spectra of the used for the definition of the DNA loop family. (a) The normalized spectra used to define the family. (b) The spectra of the newly defined DNA loop family. (c) The spectra defined for the R-loop family. Abscises correspond to the wavelength in nanometers (nm). Ordinates has no unit as it corresponds to normalized data or variances.

(Figure 3a) allows us to identify regions (between 220 and 250 nm and between 275 and 300 nm), where the CD signal remains invariant (orange in Figure 3b). This observation holds even when the normalized spectra of the entire dataset are superposed (Figure 3c,d). It is noteworthy that, due to the limited number of available spectra in databases or published structures, we opted to apply the workflow without any discontinuity in the wavelength.

Furthermore, several spectra not initially assigned to any family were identified as belonging to one in coherence with their biological characteristics. For instance, the classification of the DNA sequence TT(GGGT)₄, predicted by the workflow to belong to the quadruplex family, was confirmed by NMR (ref unpublished results, personal communication). Interestingly, of the four spectra that had been manually assigned as R-loops (orange lines in Figure 4a), two of them (dashed light and dark orange lines in Figure 4a) were rejected from that family due to the NCC and NMI product being below the determined threshold. As these 2 spectra have a similar shape, and as their corresponding sequences are compatible with DNA loops, a new reference spectrum was generated for the DNA-loops family. By running the workflow with the DNA-loops family reference across the entire dataset, we identified an additional spectrum (dashed blue line in Figure 4a), which provided validation of this new family. In summary, running the workflow allowed us to define a new reference spectrum for both the DNA loops (Figure 4b) and the R-loops (Figure 4c) families.

Originally designed to determine reference CD spectra, the workflow described here can also predict yet unknown secondary structures. However, it is limited to identifying elementary reference spectra from sequences with a single secondary structure. It cannot be used to determine the percentage of different structures in complex spectra from sequences with multiple secondary structures.

In summary, the workflow introduced here, with its Python code available in the [supplementary data](https://doi.org/10.1017/qrd.2025.10008), helps identify elementary reference CD spectra for nucleic acids. Currently limited to five families, this number is expected to grow as more nucleic acid spectra are added to public reference datasets. This advancement lays the groundwork for an online tool to determine the percentage of structures in complex CD spectra, similar to existing tools for proteins. The next steps include designing this tool and developing accurate algorithms for deconvoluting complex spectra.

Open peer review. To view the open peer review materials for this article, please visit <http://doi.org/10.1017/qrd.2025.10008>.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/qrd.2025.10008>.

Data availability statement. Python implemented script of the workflow can be downloaded from <https://github.com/Sanofi-Public/CD-spectra-classification>. License conditions apply, including limitation to non-commercial uses only.

Acknowledgments. We thank our Sanofi colleagues Marc François-Heude and Frédéric Greco for useful discussions and Jean-Sébastien Bolduc for critically reviewing the manuscript. We are thankful to Synchrotron SOLEIL (France) for attribution of SRCD beamtime (proposals 20201304, 20210819, and 20240053).

Author contribution. K.M. was responsible for designing and coding the workflow, compiling and acquiring spectra, evaluating spectra classification, performing critical results analysis, writing the first draft, and revising the manuscript following authors' comments. S.V.H. proposed analytical methods implemented in the workflow, performed critical results analysis, and contributed to manuscript revisions. A.G. prepared and managed biological materials, coordinated biological aspects at Sanofi, and participated in scientific

discussions. F.W. acquired spectra data, evaluated spectra classification, performed critical results analysis, and contributed to manuscript revisions. V.A. designed nucleic acid sequences for data acquisition, evaluated spectra classification, performed critical results analysis and contributed to manuscript revisions. S.M. designed the workflow, evaluated spectra classification, performed critical results analysis, wrote the first draft, and revised the manuscript following authors' comments.

Financial support. This work was funded by Sanofi.

Competing interests. K.M., A.G., and S.M. are Sanofi employees and may hold shares and/or stock options in the company. The remaining authors declare no competing interests exist.

References

- Burastero, O, Jones, NC, Defelipe, LA, Zavrtanik, U, Hadži, S, Hoffmann, SV, Garcia-Alai, MM (2025) *ChiraKit: an online tool for the analysis of circular dichroism spectroscopy data*. Nucleic Acids Research (in press). doi: [10.1093/nar/gkaf350](https://doi.org/10.1093/nar/gkaf350).
- Cappannini A, Mosca K, Mukherjee S, Moafinejad SN, Sinden RR, Arluison V, Bujnicki J and Wien F (2023) NACDDB: Nucleic acid circular Dichroism database. *Nucleic Acids Research* **51**, D226–D231. <https://doi.org/10.1093/nar/gkac829>.
- Del Villar-Guerra R, Trent JO and Chaires JB (2018) G-Quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angewandte Chemie (International ed in English)* **57**(24), 7171–7175.
- Gottarelli G, Lena S, Masiero S, Pieraccini S and Spada GP (2008) The use of circular dichroism spectroscopy for studying the chiral molecular self-assembly: An overview. *Chirality* **20**(3–4), 471–485. <https://doi.org/10.1002/chir.20459>.
- Gray DM, Hung SH and Johnson KH (1995) Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes. *Methods in Enzymology* **246**, 19–34. [https://doi.org/10.1016/0076-6879\(95\)46005-5](https://doi.org/10.1016/0076-6879(95)46005-5).
- Gray DM, Liu JJ, Ratliff RL and Allen FS (1981) Sequence dependence of the circular dichroism of synthetic double-stranded RNAs. *Biopolymers* **20**, 1337–1382.
- Gray DM, Ratliff RL and Vaughan MR (1992) Circular dichroism spectroscopy of DNA. *Methods in Enzymology* **211**, 389–406.
- Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**(6), 2876–2890. <https://doi.org/10.1038/nprot.2006.202>.
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, Del Rio JF, Wiebe M, Peterson P, Gerard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C and Oliphant TE (2020) Array programming with NumPy. *Nature* **585**(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
- Holm AI, Nielsen LM, Hoffmann SV and Nielsen SB (2010) Vacuum-ultraviolet circular dichroism spectroscopy of DNA: A valuable tool to elucidate topology and electronic coupling in DNA. *Physical Chemistry Chemical Physics* **12**(33), 9581–9596. <https://doi.org/10.1039/c003446k>.
- Johnson WC (1990) Electronic circular dichroism spectroscopy (CD) spectroscopic of nucleic acids. In *Biophysics Berlin/Heidelberg*: Springer Verlag, 2275–2280.
- Kuril AK, Vashi A and Subbappa PK (2024) A comprehensive guide for secondary structure and tertiary structure determination in peptides and proteins by circular dichroism spectrometer. *Journal of Peptide Science*, e3648. <https://doi.org/10.1002/psc.3648>.
- Kypr J, Kejnovska I, Rencuk D and Vorlickova M (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Research* **37**(6), 1713–1725. <https://doi.org/10.1093/nar/gkp026>.
- Le Brun E, Arluison V and Wien F (2020) Application of synchrotron radiation circular dichroism for RNA structural analysis. *Methods in Molecular Biology* **2113**, 135–148.
- Manavalan P and Johnson WC (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra.

- Analytical Biochemistry* **167**(1), 76–85. [https://doi.org/10.1016/0003-2697\(87\)90135-7](https://doi.org/10.1016/0003-2697(87)90135-7).
- Micsonai A, Moussong E, Wien F, Boros E, Vadaszi H, Murvai N, Lee YH, Molnar T, Refregiers M, Goto Y, Tantos A and Kardos J (2022) BeStSel: Webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic Acids Research* **50**(W1), W90–W98. <https://doi.org/10.1093/nar/gkac345>.
- Miyahara T, Nakatsuji H and Sugiyama H (2012) Helical structure and circular dichroism spectra of DNA: A theoretical study. *The Journal of Physical Chemistry. A* **117**, 42–55. <https://doi.org/10.1021/jp3085556>.
- Miyahara T, Nakatsuji H and Sugiyama H (2016) Similarities and differences between RNA and DNA double-helical structures in circular dichroism spectroscopy: A SAC-CI study. *The Journal of Physical Chemistry. A* **120**(45), 9008–9018. <https://doi.org/10.1021/acs.jpca.6b08023>.
- Nagy G, Hoffmann SV, Jones NC and Grubmuller H (2024) Reference data set for circular dichroism spectroscopy comprised of validated intrinsically disordered protein models. *Applied Spectroscopy* **78**(9), 897–911. <https://doi.org/10.1177/00037028241239977>.
- Neidle S and Sanderson M (2022a) DNA structure as observed in fibres and crystals. In book *Principles of Nucleic Acid Structure*, Elsevier, pp. 53–108, <https://doi.org/10.1016/b978-0-12-819677-9.00007-x>
- Neidle S and Sanderson M (2022b) RNA structures and their diversity. In book *Principles of Nucleic Acid Structure*, Elsevier, pp. 287–346, <https://doi.org/10.1016/b978-0-12-819677-9.00002-0>
- Nordén B, Rodger A and Daffron T (2010) *Linear Dichroism and Circular Dichroism a Textbook on Polarized-Light Spectroscopy*. Cambridge: RCS Publishing 317–370.
- Ramalli SG, Miles AJ, Janes RW and Wallace BA (2022) The PCDDb (protein circular Dichroism data Bank): A bioinformatics resource for protein characterisations and methods development. *Journal of Molecular Biology* **434**(11), 167441. <https://doi.org/10.1016/j.jmb.2022.167441>.
- Sathyaseelan C, Vijayakumar V and Rathinavelan T (2021) 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. *Journal of Molecular Biology* **433**(11), 166629. <https://doi.org/10.1016/j.jmb.2020.08.014>.
- Sinden RR (1994) *DNA Structure and Function*. Academic Press, Cambridge, 11–12.
- Sreerama N and Woody RW (1994) Protein secondary structure from circular dichroism spectroscopy. Combining variable selection principle and cluster analysis with neural network, ridge regression and self-consistent methods. *Journal of Molecular Biology* **242**(4), 497–507. <https://doi.org/10.1006/jmbi.1994.1597>.
- Steely HT, Gray DM and Ratliff RL (1986) CD of homopolymer DNA-RNA hybrid duplexes and triplexes containing A-T or A-U base pairs. *Nucleic Acids Research* **14**(24), 10071–10090.
- Vanegas PL, Hudson GA, Davis AR, Kelly SC, Kirkpatrick CC and Znosko BM (2012) RNA CoSSMos: Characterization of secondary structure motifs--a searchable database of secondary structure motifs in RNA three-dimensional structures. *Nucleic Acids Research* **40**(Database issue), D439–444. <https://doi.org/10.1093/nar/gkr943>.
- Vanloon J, Bennett HA, Martin A, Wien F, Harroun T and Yan H (2023) Synchrotron radiation circular dichroism spectroscopy of oligonucleotides at millimolar concentrations. *Bioorganic & Medicinal Chemistry Letters* **92**, 129376. <https://doi.org/10.1016/j.bmcl.2023.129376>.
- Wallace BA (2009) Protein characterisation by synchrotron radiation circular dichroism spectroscopy. *Quarterly Reviews of Biophysics* **42**(4), 317–370. <https://doi.org/10.1017/S003358351000003X>.