D226–D231 *Nucleic Acids Research, 2023, Vol. 51, Database issue* *Published online 25 October 2022*
https://doi.org/10.1093/nar/gkac829

NAR Breakthrough Article

NACDDB: Nucleic Acid Circular Dichroism Database

Andrea Cappannini[1,†], Kevin Mosca[2,†], Sunandan Mukherjee[1], S. Naeim Moafinejad[1], Richard R. Sinden[3], Veronique Arluison[2,4,*], Janusz Bujnicki[1,*] and Frank Wien[5,*]

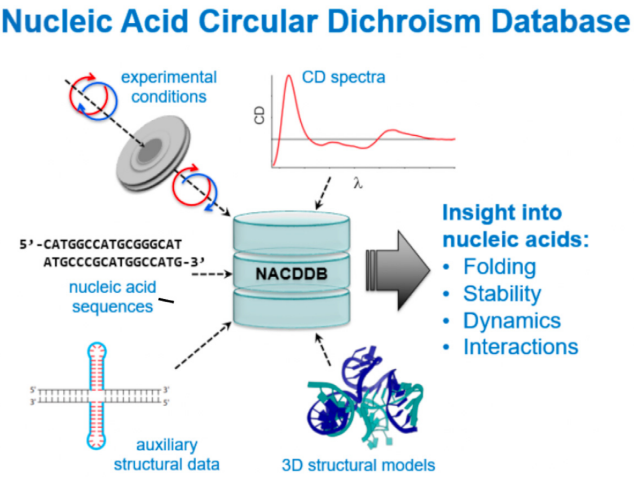
[1]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, ul. Ks. Trojdena 4, PL-02-109 Warsaw, Poland, [2]Université Paris-Saclay, CEA, CNRS, LLB, 91191, Gif-sur-Yvette, France, [3]Department of Chemistry, Biology, and Health Sciences, South Dakota School of Mines and Technology, Rapid City, SD 57701, USA, [4]Université Paris Cité, UFR SDV, 75006 Paris, France and [5]Synchrotron SOLEIL, L'Orme des Merisiers, Saint Aubin BP 48, 91192, Gif-sur-Yvette, France

Received August 19, 2022; Editorial Decision September 04, 2022; Accepted September 15, 2022

ABSTRACT

The Nucleic Acid Circular Dichroism Database (NACDDB) is a public repository that archives and freely distributes circular dichroism (CD) and synchrotron radiation CD (SRCD) spectral data about nucleic acids, and the associated experimental metadata, structural models, and links to literature. NACDDB covers CD data for various nucleic acid molecules, including DNA, RNA, DNA/RNA hybrids, and various nucleic acid derivatives. The entries are linked to primary sequence and experimental structural data, as well as to the literature. Additionally, for all entries, 3D structure models are provided. All entries undergo expert validation and curation procedures to ensure completeness, consistency, and quality of the data included. The NACDDB is open for submission of the CD data for nucleic acids. NACDDB is available at: https://genesilico.pl/nacddb/.

GRAPHICAL ABSTRACT



INTRODUCTION

Circular dichroism (CD) is a sensitive absorption spectroscopy probing the chirality of biological macromolecules such as proteins, nucleic acids, and sugars. Widely used by molecular biologists and biophysics in industrial and academic research, CD measures the differential absorption of circular polarized light by chiral macromolecules. In the ultraviolet (UV) band between 340 and 170 nm electronic transitions of the covalent bonds; mainly n–π, π–π* but also charge transfers between adjacent bonds; are excited and hence absorbed. Sensitive for low and high concentrations of samples in long or short path lengths, respectively, following the Beer–Lambert law, CD is a sample con-

*To whom correspondence should be addressed. Tel: +33 169359695; Email: frank.wien@synchrotron-soleil.fr
Correspondence may also be addressed to Veronique Arluison. Email: veronique.arluison@u-paris.fr
Correspondence may also be addressed to Janusz Bujnicki. Email: janusz@iimcb.gov.pl
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

serving non-destructive spectroscopy. Well-established and fast qualitative and quantitative analysis allows for the determination of macromolecular structure-folding, thermal stability measurements, dynamics as well as inter- and intramolecular interactions to be observed.

CD spectra can be acquired with benchtop CD-spectrophotometers or synchrotron radiation CD (SRCD) beamlines. The former being fast and very efficient in the near-UV up to visible light, the latter extending the spectral band down to 170 nm in solutions and to 120 nm in films with a relative humidity of rH 98%. The spectral extension ultimately increases the information content obtainable from a nucleic acid spectrum. SRCD beamlines are accessible at AU-CD (ASTRID2, Denmark), B-23 (Diamond, UK), BL-12 (HISOR, Japan), CEDRO (LNLS, Brazil), DISCO (SOLEIL, France) and 03A1 (NSRRC, Taiwan).

The utility of CD for proteins has been long recognized. The Protein Circular Dichroism Database (PCDDB) has been developed to host the spectra and metadata (1). Various methods and analytical tools have been developed for the interpretation of spectral data and characterization of the targets (2,3). Unfortunately, no comparable resource to archive and maintain CD data for nucleic acids has been developed, beyond individual efforts to curate datasets of spectra for nucleic acids (4).

Nucleic acids, including DNA, RNA, their hybrids and various chemically modified variants, have been studied for the past 50 years by CD (5–9). Nucleotide residues comprising a base connected to a sugar and linked by phosphate groups are asymmetric and chiral, which is at the origin of the CD differential absorption in the UV to vacuum UV (VUV). Each nucleotide unit contains charged, polar, and aromatic groups, which can make diverse types of interactions with other nucleotides as well as with water, ions, small molecules, proteins, and other molecules. Thermodynamically, it is favorable for bases in nucleic acids to stack on one another and form hydrogen bonds with each other. Within the stacking-induced arrangement, base pairs can assemble in helices or other structural elements.

Genomic DNA generally exists as a right-handed double helix formed by two very long complementary strands interacting via canonical Watson-Crick base pairs G-C and A-T, with about 10.4 bp per helical turn, denoted as B-DNA. Under certain conditions such as protein-binding or dehydration, DNA can be driven to form another helical form denoted A-DNA. DNA can also form more complicated structures, including cruciforms, intramolecular triplexes, left-handed Z-DNA helices, slipped strand structures, and four-stranded quadruplexes that all exhibit biological functions (10).

RNA transcripts generally consist of a single chain, which is folded to maximize stacking and favorable hydrogen-bonding interactions. Canonical Watson–Crick base pairing (G–C and A–U) is important to form A-form helices, however, it has to be emphasized that every nucleotide in an RNA chain can favorably interact with every other nucleotide in specific structural contexts, and in most cases, non-canonical pairs are critical for creating the tertiary interactions that stabilize the functional conformations. In well-structured RNAs, canonically-paired A-helical regions serve as geometrically regular segments of the structural framework connected by other structural elements (turns, kinks, stable loops, and various complex motifs) stabilized by non-canonical pairs.

Adding to DNA and RNA structural complexity, nucleotide residues can be chemically modified, which changes their chemical and physical properties and may influence nucleic acid conformation locally and globally (11). Furthermore, depending on the environment (ionic strength, pH, and temperature) nucleic acid molecules may change their structure. CD spectra can capture all these aspects of the structure of nucleic acids, and enable studying the effects of changes of sequence (mutations), chemical modifications, and the environmental effects on the structure that are otherwise hard to detect by other methods (12). Spectral changes based on the excitation of electronic transitions indicate changes in the polynucleotide backbone spatial arrangement, including base-pairing and the helicity of the numerous complex structures of polynucleotides such as those encountered in RNA and DNA molecules (e.g. ssDNA, dsDNA, triplexes and quadruplexes) (13–18).

With the development of nucleic acid therapeutics, including mRNA-based vaccines (19), it has become clear that CD of nucleic acids is a very accessible method for qualitative and quantitative analysis (20,21). This technique stands out in its methodology, being non-destructive, consuming very little amounts of sample, which may be very precious (2–4 μl loading volume for a concentrated sample ∼10 mM expressed in nucleotides or 100 μl of diluted sample ∼ 0.5 mM), and it lends itself very well to fast high throughput microfluidic sampling with short acquisition times (minutes). Apart from using CD for quality control of nucleic acid samples, CD also enables mechanistic studies on the structure and function of nucleic acids. The correlation of the CD spectral changes with the dynamic behavior and equilibria of the chiral polynucleotides is at the heart of current intense research in industry and academia (22).

The Nucleic Acid Circular Dichroism Database (NACDDB) aims to fill the gap between the need for systematic information on the CD of nucleic acids and the dispersed information on that topic. NACDDB provides the research community, including academia and industry, with a resource containing the first extensive repertoire of CD experiments standardized, calibrated, and normalized for various nucleic acid molecules, including DNA, RNA, DNA/RNA hybrids and several nucleic acid derivatives. NACDDB provides a collection of CD data sets for many different types of nucleic acid molecules, with detailed descriptions of experimental conditions, along with models of 3D structure, associated data, and literature references. We also encourage and are open to accommodating the CD data for nucleic acids from all sources.

DATABASE CONTENT

The current (as of 2022/09/21) version of NACDDB contains 135 entries, and this dataset is expected to grow with the planned updates. Among these entries deposited in the NACDDB, 63 references spectra were deposited using SRCD at DISCO beamline (Synchrotron SOLEIL) and

72 were deposited once digitized from the literature using WebPlotDigitizer (https://automeris.io/WebPlotDigitizer/). In addition Eigenspectra representing the most significant contributions (15 spectra currently) are retrievable and updated following novel CD spectra submissions

Each entry in NACDDB provides information on:

- CD absorption versus wavelength in delta epsilon (Δε) versus nanometer (nm) units
- the nucleic acid sequence (for heterogeneous samples sequence variability is indicated)
- spectral band coverage
- wavelength measurement interval

For SRCD spectra additional information is provided, including:

- High tension (HT) of the detector, in volt units
- Absorbance spectra or pseudo-absorbance based on the HT, in optical density units (OD)
- Dwell time at each datapoint in seconds (s)
- Cell and pathlength (mm) used for sample loading
- Spectral zeroing band

In addition to CD data, NACDDB provides information on the nucleic acid structure obtained with other methods, including experimental and computational approaches. In cases where the 3D structure was determined experimentally and is available from the RCSB database (23), the link to the RCSB structural model is provided. In cases where experimental data were available in RCSB for a related but non-identical sequence, comparative modeling was carried out with ModeRNA (24). In the majority of cases where no experimentally determined high-resolution structure was available, we generated a predictive model using SimRNA and its variants (25,26). All computational models had their geometry optimized by QRNAS (27). For CD experiments involving heterogeneous nucleic acid sequences (e.g. repetitive polymers with variable length), a single representative sequence was selected for 3D structure modeling. The source of each 3D structural model is indicated, and in the case of computational predictions, a warning is included for the users that the model is tentative and should be considered with caution.

NACDDB also hosts spectra of nucleic acid sequences with modified residues. The modified nucleotides are represented with the alphabet used by the MODOMICS database (11). Each modified residue serves as a link to the MODOMICS page describing the modified residue.

In terms of information content obtainable from the NACDDB, we have chosen the 63 SRCD spectra representing a mix of characteristic polynucleotide sequences and folds. From these 63 spectra (Figure 1) a decomposition into distinct spectra using single value decomposition (SVD) revealed 15 relevant spectra. The statistical correlation of the distinct spectra (Eigenspectra) was $r = 0.04$ with a standard error of sr = 0.08. Based on the very low correlation of the eigenspectra we assume a nonlinear correlation. Surely this is not excessive but leads us to the hypothesis that it will be possible to identify characteristic polynucleotide folding patterns similar to the secondary structure elements

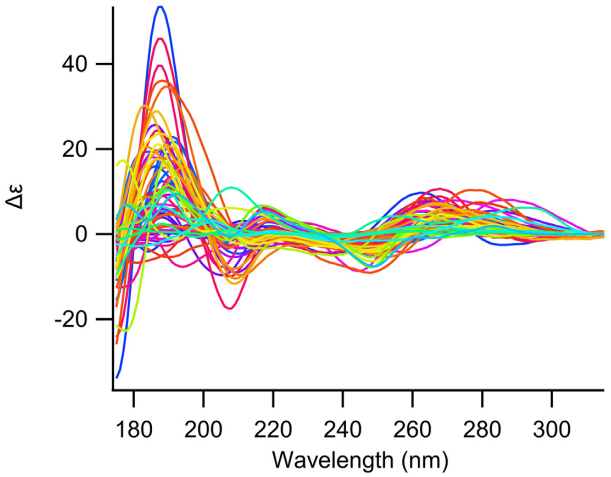


Figure 1. Presentation of 63 spectra collected with SRCD normalized to Δε representing the average nucleotide contribution of a polynucleotide sequence.

found in proteins. Indeed, 8 eigenspectra contribute to the deconvolution of protein CD spectra according to eight secondary structures. It is evident that the electronic transitions within a polynucleotide (e.g. between to base-paired nucleotides) span a much larger distance than observed in the peptide bond, ±17.5Å versus 2.25Å, respectively. Eigenspectra are included in NACDDB as special entries (eigenspectra), which can be queried separately and compared pairwise with experimental spectra.

IMPLEMENTATION

NACDDB was implemented using Django 3.0.5 (https://www.djangoproject.com/) and HTTP server Nginx (https://www.nginx.com/). The database uses a SQLite relational database (https://www.sqlite.org/index.html) and leverages several open-source Javascript libraries to serve its content. DataTables (http://datatables.net) was implemented to allow sorting and searching in the experiment and publication tables, allowing users to search through the data. Additionally, we leverage the open source Mol* Viewer project (28), which enables web-based molecular visualization of the required molecule. In addition to these, Chart.js (https://www.chartjs.org/) is used to display spectra (Figure 2). NACDDB provides spectra as downloadable files, using SweetAlert2 (https:/sweetalert2.github.io/) based endpointAPI, in the original '*.gen*' format. This service is also available for tertiary structure visualization, provided '*.pdb*' formatted files do exist. The website is HTTPS-enabled, which means that the data exchange between the user and the NACDDB server is secured by an encrypted connection. The website is mobile-friendly, adapting to the user's screen size and device, making NACDDB accessible e.g. from tablets and smartphones.

The NACDDB home page provides a brief introduction regarding circular dichroism (CD) and describes the purpose of the database. The menu bar provides access to all the pages available through which the database can be navigated. The experiment results are provided under the Ex-

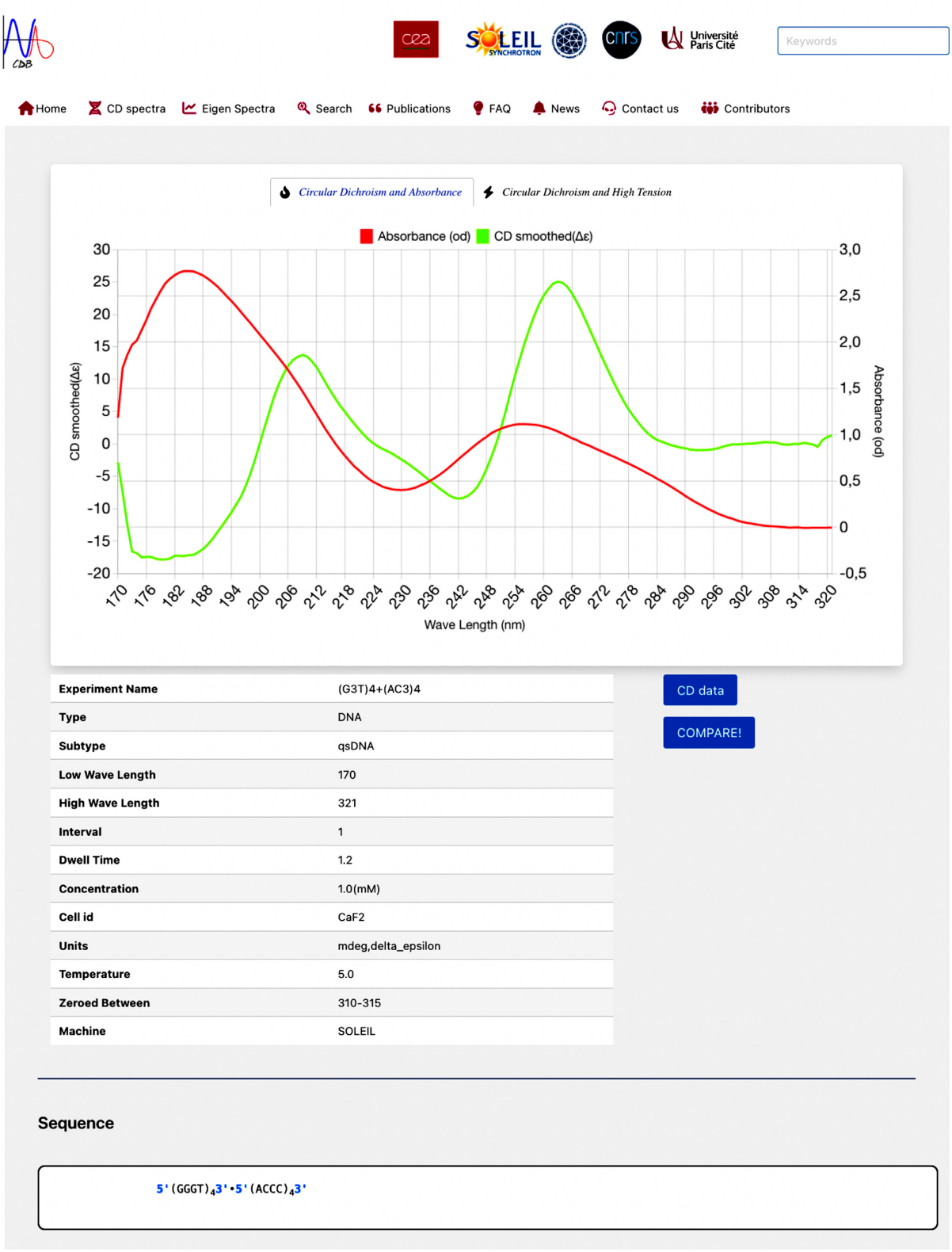


Figure 2. Example entry in NACDDB: SRCD spectrum for the (G3T)4+(AC3)4 quadruplex DNA (green), absorption (red) and metadata including sequence information.

periment tab. The toolbar displays all the database's pages and a search bar to query the database contents.

The database can be queried through a keyword search function and will therefore reveal not only the results related to the searched keyword but also the contextual matches to it. In the same section, NACDDB enables the comparison of CD spectra, including the representative eigenspectra. The search returns a ranking list of spectra most similar to the query spectrum, ranked according to the p-value of the two-sample Kolmogorov–Smirnov (KS) test. Spectral data, including raw circular dichroism data and the associated wavelength vector can be entered by a user. Then the Empirical Cumulative Distribution Function (ECDF) of that spectrum and spectra in the NACDDB are calculated. Finally, the two-sample KS test is calculated and results, including p-values, are displayed in a table. The first column contains the ID of each spectrum that is linked to comparison charts. A scatter plot is presented in which the ECDFs of the two spectra provide a visual representation of the KS test. In addition, for further comparison, a scatter plot of both spectra is provided. We also provide the best fit for the database-spectra and the one provided by the user, for further comparison. The polynomial used to represent the curves is chosen from six possible ones. The best fit returned is the one that returns the largest r-squared value.

The sequence search functionality utilizes the NCBI BLAST 2.12.0+ (BLASTN) tool (29,30), which allows to search the sequences of nucleic acids in NACDDB in FASTA or RAW format. The search can be fine-tuned by adjusting the e-value threshold (default is 1e-02). Matches are listed in a table. The first column shows NACDDB IDs for which BLAST revealed a significant alignment, the second, the degree of similarity between the two nucleic acid sequences, and the third, the alignment e-value. BLAST results can be downloaded in either a tabular format (outfmt 6) or as alignments in '.afa' format. With new CD spectra deposited in the NACDDB, the database of sequences to be searched will be updated accordingly

DISCUSSION

The nucleic acid structural biology community needs a database that archives and organizes available CD and SRCD data, accepts public deposition and distribution of spectral references, including the metadata that describes experimental conditions, and enables comparative analyses of the data. NACDDB is dedicated to the CD of nucleic acids and it offers a single entry point for data that so far could only be obtained by meticulously analyzing many different sources, often difficult to browse, such as supplementary materials of published papers. This database provides users with an easy-to-use interface with the flexibility to browse from lists of CD experiments and publications, filter the results according to user-defined criteria, and choose the order in which the results are displayed. The database can be queried by keywords and database entries can be searched according to the similarities of nucleic acid sequence or CD spectra obtained.

NACDDB will allow the researchers in the fields of molecular biology and biophysics as well as bioinformatics to systematically study the CD spectra that provide information about nucleic acid structures and polynucleotide-protein interactions that are complementary to other sources of structural information. With this collection of CD spectra, NACDDB will enable comparative analyses and the development of new predictive tools, e.g. for the identification of specific structural features from new spectra.

NACDDB is open for user feedback ensuring transparency and quality. Users will be able to submit their experimental data following a standardization and normalization procedure to conform with the currently deposited spectra. This will allow growth of the spectral content in the NACDDB, thriving to a comprehensible, ordered and structured approach, which ultimately shall allow the prediction of polynucleotide acid folding based on CD spectra. We are open to include data of novel types of molecular systems including nucleic acids, such as RNAs folded in membrane-confined environments used in the novel mRNA vaccines for COVID and in the future for cancer treatments (31,32).

Future updates of the NACDDB database will include the development of a structure prediction engine, linking the query CD profile to nucleic acids with similar spectra and potentially similar structural features. We will also include the Applications page, with published or patented examples of CD-based practical applications.

The database is continuously updated as new papers on CD experiments on nucleic acids become available. NACDDB users are encouraged to use the contact form under the 'Contact us' tab to provide us with CD spectra not yet included, with the scope of making NACDDB as complete as possible. For the new submissions, we will add the relevant links, 3D structure models, and other types of metadata. We also encourage users to use the contact form under the Contact us tab in the toolbar to report any errors or omissions.

DATA AVAILABILITY

The web interface to the database is available at https://iimcb.genesilico.pl/nacddb/. This website is free for academic (non-commercial) applications, open to all users and no login or password is required.

ACKNOWLEDGEMENTS

We are indebted to the authors of primary databases and services, whose content was used in or linked to NACDDB. We are particularly grateful to Penzhi Wu (ETH, Zurich) and Elsa Baduzzi (LLB, CEA Saclay) for their help at an early stage of this work. We also thank Farhang Jaryani (IIMCB, Warsaw) for help in secondary structure analysis in analyses of the auxiliary structural data. We thank all members of the Arluison, Wien, and Bujnicki research groups for fruitful discussions. The computational modeling of nucleic acid structures for the NACDDB database was supported in part by PLGrid Infrastructure (grant id: *plgsimcry*, *plgsimrnp*).

FUNDING

European Commission [H2020-MSCA-ITN-2020, GA No. 956810, CA16120]; Polish National Science Center [NCN, 2020/37/B/NZ2/02456 to J.M.B. and 2017/26/A/NZ1/01083 to J.M.B. and S.M.]; IIMCB [statutory funds to J.M.B.]. Funding for open access charge: European Commission [H2020-MSCA-ITN-2020, GA No. 956810]. SRCD experiments were performed on 'DISCO' beamline at SOLEIL Synchrotron, France [proposals 20201304, 20200074, 20210819].

Conflict of interest statement. Janusz M. Bujnicki is an Executive Editor of *Nucleic Acids Research.*

REFERENCES

1. Ramalli,S.G., Miles,A.J., Janes,R.W. and Wallace,B.A. (2022) The PCDDB (Protein circular dichroism data bank): a bioinformatics resource for protein characterisations and methods development. *J. Mol. Biol.*, **434**, 167441.
2. Miles,A.J., Ramalli,S.G. and Wallace,B.A. (2022) DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data. *Protein Sci.*, **31**, 37–46.
3. Miles,A.J. and Wallace,B.A. (2018) CDtoolX, a downloadable software package for processing and analyses of circular dichroism spectroscopic data. *Protein Sci.*, **27**, 1717–1722.
4. 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. *J. Mol. Biol.*, **433**, 166629.
5. Bishop,G.R. and Chaires,J.B. (2003) Characterization of DNA structures by circular dichroism. *Curr. Protoc. Nucleic Acid Chem.*, **Chapter 7**, 7.11.1–7.11.18.
6. Jaumot,J., Eritja,R., Navea,S. and Gargallo,R. (2009) Classification of nucleic acids structures by means of the chemometric analysis of circular dichroism spectra. *Anal. Chim. Acta*, **642**, 117–126.
7. Figueiredo,J., Miranda,A., Lopes-Nunes,J., Carvalho,J., Alexandre,D., Valente,S., Mergny,J.L. and Cruz,C. (2021) Targeting nucleolin by RNA G-quadruplex-forming motif. *Biochem. Pharmacol.*, **189**, 114418.
8. Vorlickova,M., Kejnovska,I., Bednarova,K., Renciuk,D. and Kypr,J. (2012) Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality*, **24**, 691–698.
9. Vanloon,J., Harroun,T. and Yan,H. (2021) Circular dichroism spectroscopy of DNA duplexes at near-biological concentrations. *Bioorg. Med. Chem. Lett.*, **43**, 128053.
10. Sinden,R.R. (1994) In: *DNA Structure and Function*. Academic Press.
11. Boccaletto,P., Stefaniak,F., Ray,A., Cappannini,A., Mukherjee,S., Purta,E., Kurkowska,M., Shirvanizadeh,N., Destefanis,E., Groza,P. *et al.* (2022) MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res.*, **50**, D231–D235.
12. Le Brun,E., Arluison,V. and Wien,F. (2020) Application of synchrotron radiation circular dichroism for RNA structural analysis. *Methods Mol. Biol.*, **2113**, 135–148.
13. Del Villar-Guerra,R., Trent,J.O. and Chaires,J.B. (2018) G-Quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew. Chem. Int. Ed Engl.*, **57**, 7171–7175.
14. Gray,D.M., Hung,S.-H. and Johnson,K.H. (1995) *Methods in Enzymology*. Elsevier, pp. 19–34.
15. Holm,A.I., Nielsen,L.M., Hoffmann,S.V. and Nielsen,S.B. (2010) Vacuum-ultraviolet circular dichroism spectroscopy of DNA: a valuable tool to elucidate topology and electronic coupling in DNA. *Phys. Chem. Chem. Phys.*, **12**, 9581–9596.
16. Cech,C.L. and Tinoco,I. Jr (1977) Circular dichroism calculations for double-stranded polynucleotides of repeating sequence. *Biopolymers*, **16**, 43–65.
17. Randazzo,A., Spada,G.P. and da Silva,M.W. (2013) Circular dichroism of quadruplex structures. *Top. Curr. Chem.*, **330**, 67–86.
18. Kypr,J., Kejnovska,I., Renciuk,D. and Vorlickova,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.
19. Kulkarni,J.A., Witzigmann,D., Thomson,S.B., Chen,S., Leavitt,B.R., Cullis,P.R. and van der Meel,R. (2021) The current landscape of nucleic acid therapeutics. *Nat. Nanotechnol.*, **16**, 630–643.
20. Kloczewiak,M., Banks,J.M., Jin,L. and Brader,M.L. (2022) A biopharmaceutical perspective on higher-order structure and thermal stability of mRNA vaccines. *Mol Pharm*, **19**, 2022–2031.
21. DeLong,R.K., Huber,H., Aparicio-Lopez,C., Bhatti,A., Swanson,R., Shrestha,T.B. and Gaudreault,N.N. (2022) Enzyme nanoscale interactions with manganese zinc sulfide give insight into potential antiviral mechanisms and SARS-CoV-2 inhibition. *ACS Pharmacol. Transl. Sci.*, **5**, 449–457.
22. Eriksson,M. and Norden,B. (2001) Linear and circular dichroism of drug-nucleic acid complexes. *Methods Enzymol.*, **340**, 68–98.
23. Burley,S.K., Bhikadiya,C., Bi,C., Bittrich,S., Chen,L., Crichlow,G.V., Christie,C.H., Dalenberg,K., Di Costanzo,L., Duarte,J.M. *et al.* (2021) RCSB protein data bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.*, **49**, D437–D451.
24. Rother,M., Rother,K., Puton,T. and Bujnicki,J.M. (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.*, **39**, 4007–4022.
25. Boniecki,M.J., Lach,G., Dawson,W.K., Tomala,K., Lukasz,P., Soltysinski,T., Rother,K.M. and Bujnicki,J.M. (2016) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.*, **44**, e63.
26. Wirecki,T.K., Nithin,C., Mukherjee,S., Bujnicki,J.M. and Boniecki,M.J. (2020) Modeling of three-dimensional RNA structures using SimRNA. *Methods Mol. Biol.*, **2165**, 103–125.
27. Stasiewicz,J., Mukherjee,S., Nithin,C. and Bujnicki,J.M. (2019) QRNAS: software tool for refinement of nucleic acid structures. *BMC Struct. Biol.*, **19**, 5.
28. Sehnal,D., Bittrich,S., Deshpande,M., Svobodova,R., Berka,K., Bazgier,V., Velankar,S., Burley,S.K., Koca,J. and Rose,A.S. (2021) Mol* viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.*, **49**, W431–W437.
29. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
30. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinf.*, **10**, 421.
31. Pardi,N., Hogan,M.J., Porter,F.W. and Weissman,D. (2018) mRNA vaccines - a new era in vaccinology. *Nat Rev Drug Discov*, **17**, 261–279.
32. Hou,X., Zaks,T., Langer,R. and Dong,Y. (2021) Lipid nanoparticles for mRNA delivery. *Nat. Rev. Mater.*, **6**, 1078–1094.