D480–D486 *Nucleic Acids Research, 2011, Vol. 39, Database issue* *Published online 11 November 2010*
doi:10.1093/nar/gkq1026

PCDDB: the protein circular dichroism data bank, a repository for circular dichroism spectral and metadata

Lee Whitmore[1], Benjamin Woollett[1], Andrew John Miles[1], D. P. Klose[2], Robert W. Janes[2,]* and B. A. Wallace[1,]*

[1]Department of Crystallography, Institute of Structural and Molecular Biology, Birkbeck College, University of London, London WC1E 7HX and [2]School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, UK

Received September 8, 2010; Revised October 4, 2010; Accepted October 10, 2010

ABSTRACT

The Protein Circular Dichroism Data Bank (PCDDB) is a public repository that archives and freely distributes circular dichroism (CD) and synchrotron radiation CD (SRCD) spectral data and their associated experimental metadata. All entries undergo validation and curation procedures to ensure completeness, consistency and quality of the data included. A web-based interface enables users to browse and query sample types, sample conditions, experimental parameters and provides spectra in both graphical display format and as downloadable text files. The entries are linked, when appropriate, to primary sequence (UniProt) and structural (PDB) databases, as well as to secondary databases such as the Enzyme Commission functional classification database and the CATH fold classification database, as well as to literature citations. The PCDDB is available at: http://pcddb.cryst.bbk.ac.uk.

INTRODUCTION

Circular dichroism (CD) spectroscopy is a technique that has been widely used in structural biology for examining secondary structure, conformational changes, folding and interactions involving protein molecules. It is often employed as a complementary method in conjunction with other structural techniques such as protein crystallography and nuclear magnetic resonance spectroscopy, to elucidate functionally-related environmental effects on macromolecular structures. The popularity of the technique is borne out by the fact that in the last 5 years there have been more than 6000 publications on proteins that have included the use of CD spectroscopy.

Synchrotron radiation CD (SRCD) spectroscopy is a more recent, related technique that, as its name suggests, utilizes synchrotron radiation as its light source. This method can produce improved spectra, including data in the lower (vacuum ultraviolet) wavelength range, and higher signal-to-noise levels than are present in conventional CD spectra.

With so many CD and now SRCD spectra being collected and utilized world-wide, there is a need for a public repository where authors can store their data and make it available to other researchers.

The creation of a Protein Circular Dichroism Data Bank (PCDDB) was first proposed in 2006 (1) as a means of satisfying the needs of both the biological and bioinformatics research communities and funding bodies, by providing ready access to CD and SRCD data. Its development has been subject to extensive public consultations and advice from the structural, spectroscopic and bioinformatics communities. The first public release of the PCDDB (in accession mode only) was in December 2009 (2). The full deposition mode of the resource will become publicly available in early 2011, although authors can currently request deposition access if a journal requires deposition prior to publication.

The PCDDB name arose as an homage to the long-established Protein Data Bank (PDB) (3) which is a widely-used database resource for crystallographic, NMR and electron microscopy structural data and associated metadata. The aim of the PCDDB has been to make CD and SRCD data accessible both to experts within the spectroscopy community and to researchers in the wider biological field, who may be less familiar with the technique and with the types of information that are attainable from it.

The PCDDB is populated by entries deposited by data producers. Ultimately, the aim is to include multiple

*To whom correspondence should be addressed. Tel: +44 207 631 6800; Fax: +44 207 631 6803; Email: b.wallace@mail.cryst.bbk.ac.uk
Correspondence may also be addressed to Robert W. Janes. Tel: +44 207 882 8442; Fax: +44 208 983 0973; Email: r.w.janes@qmul.ac.uk

entries for the same protein that are obtained under different conditions such as pH, ionic strength, with and without ligands, with and without denaturants, thermal scans and titration series, often produced by more than one lab (the latter enabling cross-validation comparisons to be made). In this way the data bank can be mined to enable comparisons that were not possible in the original papers and thus permit new science.

In addition to making data available for other studies, there is increasing recognition within the structural biology and bioinformatics communities and by journal editors and granting bodies that published data should be fully accessible to the public, including all the pertinent details and raw data which may not appear in the original publication. In this way the community can review and re-examine the data upon which experimental conclusions are based. From a researcher's perspective, the long-term archiving of data at a centralized repository not only increases the data usability and visibility but also decreases the risk of data loss.

The data bank also includes validation tools that are not only important for ensuring the integrity of the database entries, but also have the advantage of establishing and applying data standards within the field. Not only can authors use these for pre-checking data prior to deposition or publication, they can also be especially useful during the journal review process for reviewers who may not be experts in the technique. Often there is little or no checking of the quality of CD data included in a paper if the major thrust of the study is another technique, for example molecular biology, biochemistry, crystallography or NMR spectroscopy. The availability of the PCDDB and its validation procedures can therefore enable simple checks of the CD data quality even by non-experts.

DATA BANK ENTRIES

Each entry has a unique accession ID of letters (L) and numbers (N) in the format: LLNNNLLNN. These are assigned randomly, and the user cannot request a specific ID code. However, if a user is depositing a series of spectra on the same protein or related proteins, they can pre-reserve a sequential list (up to 100) of entry names which have the same (assigned) initial seven characters.

Each entry includes information on the sample characteristics, including the protein name (and alternative names, if appropriate), the source of the protein (and any changes from the wild-type sequence), its concentration and purity, and buffer and baseline contents (including ligands, if any, and for complexes, their macromolecular binding partners). It includes the CD spectral data plus the HT (high tension or high voltage, or pseudo-absorbance) spectrum and information on the experimental conditions, such as temperature, cell pathlength and type, instrument used, spectral parameters such a minimum and maximum wavelength and wavelength intervals and number of repeat scans or accumulations.

Entries include raw spectral data for both the sample and baseline, as well as information on data processing and the final processed spectral data that are published. There is also optional (but highly recommended) information on instrument calibration standards and their spectra.

Each entry normally includes not only a hyperlink to the relevant UniProt file (4), but also includes a sequence for the protein (in one-letter code) so that differences from the UniProt entry can be identified.

When there is a crystal structure available for the protein, the PDB code ID (3) is included, as are hyperlinks to that entry at all three archiving sites—the RCSB (5), PDBj (6) and PDBe (7) websites. The reason for all three links is that each of the PDB sites has unique and complementary information and means of displaying and querying the data, and we believe easy access to all of them will benefit PCDDB users. In addition, the secondary structures derived from the crystal structure as defined by the DSSP algorithm (8) are also included.

The Enzyme Commission (EC) number (9) and a hyperlink to the associated website is included for enzymes with assigned EC numbers, as a guide to their functional properties. In addition, where available, a link to the entry on the CATH domain fold classification (10) website is also included.

Each entry includes the deposition date and the name of the depositor and the lab they are from, along with the appropriate literature citation and a link to its Medline entry. All download formats include the literature reference, so users of the data can clearly identify and cite the data producers.

These contents have been developed and refined following discussions with members of the user community as a result of extensive public consultations and suggestions from the PCDDB Technical Advisory Board members, who represent manufacturers of all of the commercial CD instruments and scientists from all SRCD beamlines worldwide (see Acknowledgements for list).

The first set of entries were the 71 soluble proteins that comprise the SP175 reference data set (11) currently available for use with the DichroWeb (12) online analysis server. These were followed by the nine proteins in the CRYST175 reference database (13). Following that, additional user-submitted protein spectra have been deposited (14), but their release is embargoed until the release of the full deposition version in early 2011.

DATABASE ORGANIZATION AND WEB INTERFACE

The underlying relational database for the PCDDB is MySQL. The data bank is available via the world-wide web at: http://pcddb.cryst.bbk.ac.uk, with an interface created utilizing the PHP scripting language. It is accessible on all platforms/operating systems with modern web browsers that are JavaScript-enabled. A more in depth description of the software and database design was presented in the pilot phase of the project (15).

The home page includes information on database contents, news and an Email address for contacting the PCDDB team, plus links to terms and conditions, an 'about the PCDDB' page with information on members of the development and curation team and advisory board members listed, an information page containing literature citations, and a glossary defining terms used on the website.

DATABASE ACCESSION

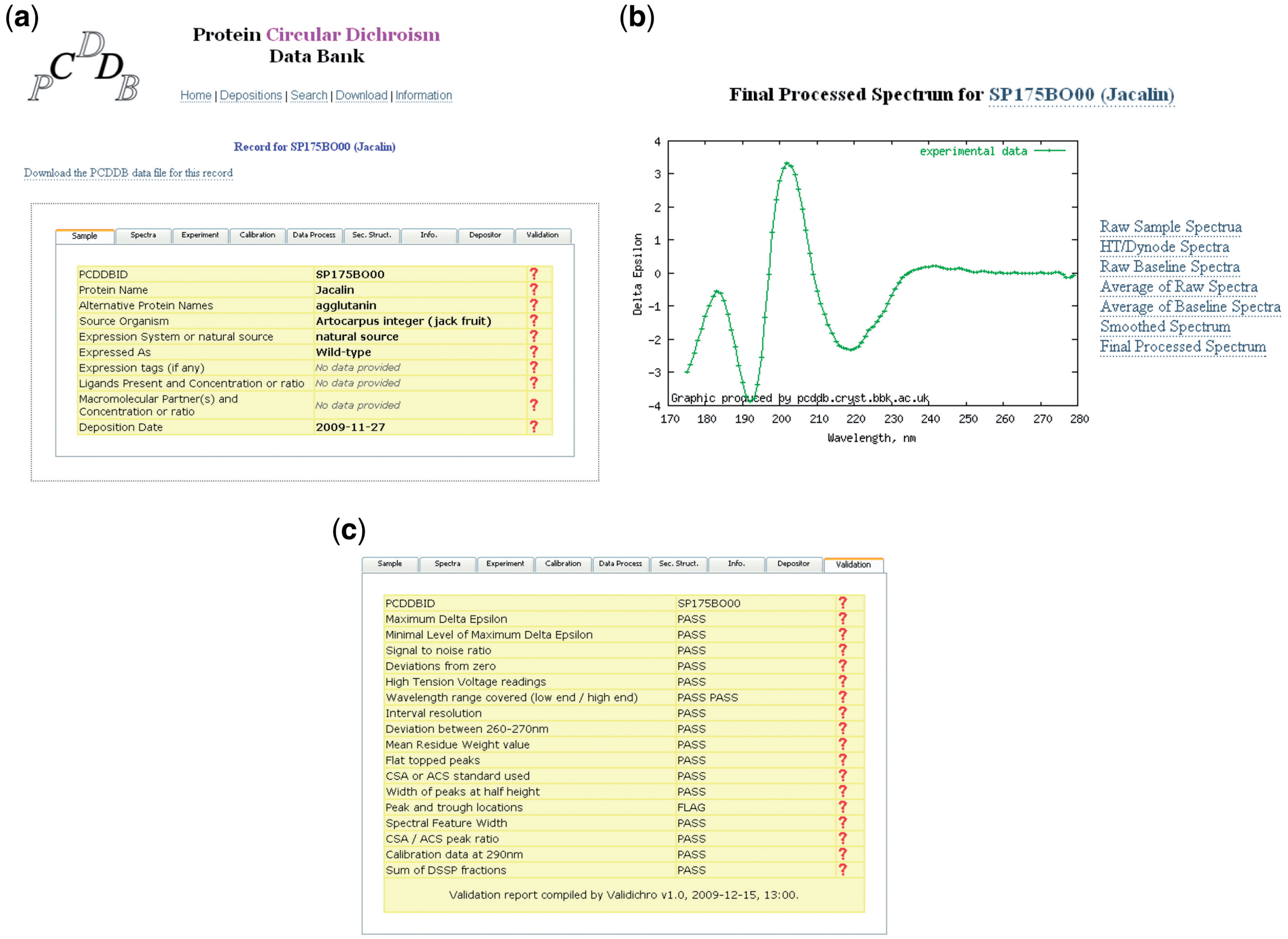
User registration is not necessary for accession. However, in the future additional accession features will be available to registered users such Email alerts for new entries and saving of search queries.

The database holdings may be queried via the 'Search' page. Text searching of the metadata is enabled in a number of categories (Supplementary Table S1), with automatic wildcards in place. For example, the search term 'my' used against the protein name field will match to myoglobin and to amylase. The output of a database search is a list of entry names/IDs which contain the query text.

Once an entry is chosen, the user is presented with a series of tables (Figure 1) in a tabbed interface. The tables include information on: the sample characteristics (Figure 1a), the spectra (Figure 1b) (clickable accession to raw and processed CD data and the associated HT spectrum), experimental conditions, instrument calibration, data processing, DSSP-defined secondary structure (8) (if known from crystal structure), information such as literature reference and links to other databases, depositor information, and a validation report (Figure 1c). Next to each item is a help icon, which displays a definition of the item when the cursor hovers over it.

Individual records can be downloaded via a link that appears on the entry record page. They can be produced in either PCDDB data format (as a .pcd ASCII file) or in a CDTools- (16) compatible format (as a .gen ASCII file). The .pcd output format contains all of the PCDDB metadata fields arranged in key-value pairs, followed by a spectral section arranged as a six column list consisting of wavelength, final processed CD spectrum, HT spectrum, smoothed data, averaged sample and averaged baseline data. If available, these are followed by a two column (wavelength, CD value) listing of the relevant instrument calibration spectrum. The .gen output format includes a limited amount of metadata as well as multicolumn spectral data (wavelength, final processed CD spectrum, HT spectrum). Both download formats include the original literature citation. The individual raw spectral data can be separately accessed via the 'Spectra' table.

Although the web interface is primarily designed to be viewed interactively, it is also possible to use a spider



Record for SP175BO00 (Jacalin)

Download the PCDDB data file for this record

PCDDBID	SP175BO00	?
Protein Name	Jacalin	?
Alternative Protein Names	agglutanin	?
Source Organism	Artocarpus integer (jack fruit)	?
Expression System or natural source	natural source	?
Expressed As	Wild-type	?
Expression tags (if any)	*No data provided*	?
Ligands Present and Concentration or ratio	*No data provided*	?
Macromolecular Partner(s) and Concentration or ratio	*No data provided*	?
Deposition Date	2009-11-27	?

Final Processed Spectrum for SP175BO00 (Jacalin)

PCDDBID	SP175BO00	?
Maximum Delta Epsilon	PASS	?
Minimal Level of Maximum Delta Epsilon	PASS	?
Signal to noise ratio	PASS	?
Deviations from zero	PASS	?
High Tension Voltage readings	PASS	?
Wavelength range covered (low end / high end)	PASS PASS	?
Interval resolution	PASS	?
Deviation between 260-270nm	PASS	?
Mean Residue Weight value	PASS	?
Flat topped peaks	PASS	?
CSA or ACS standard used	PASS	?
Width of peaks at half height	PASS	?
Peak and trough locations	FLAG	?
Spectral Feature Width	PASS	?
CSA / ACS peak ratio	PASS	?
Calibration data at 290nm	PASS	?
Sum of DSSP fractions	PASS	?

Validation report compiled by Validichro v1.0, 2009-12-15, 13:00.

Figure 1. Screenshots of entry tables for the protein jacalin (PCDDB ID: SP175BO00): (**a**) sample information page, (**b**) final processed spectrum, (**c**) validation page.

script written in any scripting language that is able to request URLs such as perl or python to browse the entire database.

To obtain the complete released holdings of the databank, the 'Download' page offers compressed archive files in .zip (windows compatible) and .tar.gz (Linux or similar and Macintosh compatible) formats.

DEPOSITIONS

Depositors must register their details with the PCDDB via a simple online form. Registration enables curation and validation feedback to the depositor and a depositor can leave a partially completed submission and return to it later.

Our aim has been to produce simple deposition procedures that minimize the burden of submission, whilst producing full and accurate documentation not only of the spectral data, but also of the sample conditions, contents and purity and the experimental parameters used to collect the data.

The deposition interface is very similar to the accession one, with a table-based design. Users can upload spectral files that have rich headers containing data collection details that are mined for information required on the deposition form. To date, this is possible for files in the .gen format output by the CD Tools processing programme (16), but in the future it will be possible to do so for all commercial instrument ASCII format files, as well as for all SRCD beamline files.

Some of the additional information required (for instance the buffer contents) is in free-form text format to enable the submitter to describe the conditions as fully as possible. Other fields, such as instrument type/manufacturer, are available as pull-down menus that trigger further boxes for the appropriate parameters specific to that instrument. This is possible as there are a limited number of types of CD instruments available: only four conventional CD manufacturers and about a dozen SRCD beamlines worldwide. Other examples of pull-down boxes are the names of the calibration standards or units of measurement. The pull downs are used as opposed to free texts for these items in order to ensure consistency in terminology and spelling between entries, thus aiding in search functions. In some fields, such as the expression system, it is not possible to anticipate every possible answer, so the option of selecting 'other' from the pull-down menu and typing in a value is also provided.

Some journals do/will require accession numbers before acceptance of a paper for publication. Depositors can make the depositions, have them validation-checked (see below) and receive ID codes before submitting the paper, but the files will not be released until publication, at which time the literature citation and links must be added to the entry. An embargoed file can be downloaded by the author for disclosure to reviewers; it will contain the validation information as a quality-control step in the review process.

CHECKING/VALIDATION

In order to ensure the integrity of the data in the database, all entries undergo a mixture of automated and manual curation checks for completeness, consistency and quality.

The first types of checks are for completeness and are of two types: (i) Whether all of the required items have been included in the file. Not all fields are required, but the required ones are indicated with an asterisk on the right-hand side of the listing on each table. When all of the required items (either metadata information or spectral files) are completed on a given table, then the table displays a green tick instead of a red cross on its tab. When all tabs have a green tick, this indicates the minimum set of required data has been included, and the entry may then be submitted for further checking and validation. (ii) Whether the spectral data are complete. For example, the submitted files are subjected to a test to identify if there are any missing data points.

The second types of checks are for self-consistency and accuracy within the metadata fields. These include numerical and textual cross-checking. An example of the numerical checking is whether the value for mean residue weight (MRW) given by the authors (used in processing calculations to convert the spectral magnitudes to standard units) is consistent with the protein molecular weight, sequence and number of residues reported. An example of textual cross-checking is whether the name of the protein given is found in the associated UniProt and PDB files. If not, this flags a potential mismatch that may be as simple as a typing error for the code, or an alternative name may need to be included in the name field.

The third, and most significant types of checking, are the validation tests which examine the quality of the data. These types of checks are philosophically similar to the validation checks associated with the PDB, which include MolProbity (17), PROCHECK (18), What_Check (19). They are done to ensure the quality of the data in the database entries. The PCDDB quality checks are based on existing standards established within the community for CD (20,21) and SRCD (22) spectra, and further validation parameters developed through feedback from the researcher, bioinformatics and instrument communities and the PCDDB Technical Advisory Board. They include such criteria as whether the signal-to-noise level is adequate for the measurement and whether the instrument calibration spectrum is correct. The help icons on the validation report page provide information/explanations associated with each criterion. It is anticipated that the validation checks may ultimately lead to an improvement of data standards within the field, as happened when such checking was initiated for PDB entries.

There are four possible outcomes for any of the validation tests: pass, flag, fail or data not available (the latter only in the case of the few optional deposition parameters, such as certain calibration files). The validation test report is forwarded to the depositor, who is given the opportunity to change the file or add a comment as to why the reported value is acceptable. [For example in Figure 1c, there is a flag noted for the 'Peak and Trough Locations' criterion. This is because this protein has an unusual spectrum (Figure 1b) relative to spectra seen in typical β-sheet rich proteins. The spectrum is perfectly correct, but the flag indicates this is an unusual feature. Such flags may ultimately be especially interesting for users, alerting them to interesting and novel features present in the spectra of some proteins]. The depositor's reply comments will be visible along with the full validation report once the file has been released. Note the PCDDB will not prohibit inclusion if an entry contains flags or fails, but any record with one or more 'fails' will only be available in an optional 'all entries' search, whereas the default searches will be for files that do not have any of their test statuses as 'fail'.

Entries must be approved by the head of the supervising lab before they will be publicly released. Authors may request that their entries be embargoed until publication, but they are required to notify the PCDDB team of the citation details at the time of publication.

The validation report is included as the final tab of the entry table. Entries are stamped with respect to the date and version of the validation protocol that has been applied. The reason for this is that as the validation criteria are refined/enhanced, users can see which version they were checked with (the criteria for each version will be documented in the help section), and can chose to run a new validation report produced with later criteria.

OTHER RESOURCES

The PCDDB includes links to other related resources for CD spectroscopy, including the DichroWeb analysis server for calculating secondary structures based on CD data (12) located at: http://dichroweb.cryst.bbk.ac.uk and the 2Struc server (23), for comparing different methods of secondary structure calculations/classifications based on crystallographic data, located at: http://2struc.cryst.bbk.ac.uk. There is also a link to the website for the CDTools processing software (19), which includes information on the .gen format. In the future, it will include links to other types of CD analysis/visualization software. Authors of such software can contact the PCDDB (at pcddb@mail.cryst.bbk.ac.uk) in order to request linkage.

We have produced exemplar depositions (consisting of components found in a number of popular CD analysis reference datasets), not only so potential depositors can see what format and information is required of them, but also because we believed that in a new endeavour such as this one, if data-producers are to embrace the new culture of open access, they need to see examples of a lab that is willing to do so. We felt the most expeditious avenue, before asking others to share their data, was that we share ours publicly as examples. The first version of the database includes 71 spectra of soluble proteins for which there are cognate crystallographic structures (11). In this way potential depositors and accessors/users can see the full breadth of the links and information available.

POTENTIAL USES

The PCDDB will increase in utility as the number and diversity of entries grows. A few examples of the potential uses for the data are for: (i) development and testing of new algorithms, both empirical ones and ones based on first principles, for calculating the secondary structures of proteins and their homologues, (ii) development of methods for identifying protein folds based on spectral characteristics, and for deducing activities based on detection of spectral nearest neighbours with known functions, (iii) standards used for testing the design and output of new CD and SRCD instruments, (iv) protein identification purposes, (v) eliminating duplications of effort by obviating the need to produce proteins or peptides in order to re-collect spectra of already published samples for direct comparisons, (vi) enabling comparisons of a known native protein with a new mutant or homologue, looking for correct folding and conformational differences, (vii) examination of the effects of environment on the protein structure, including the effects of solvent dielectric constant on spectral peak positions, (viii) comparison of the structures of isolated proteins and as parts of complexes, (ix) examination of natively unfolded proteins, a category of biologically-important structures for which crystallographic and NMR structures are often not available, (x) comparisons of series of spectra (i.e. thermal and denaturant unfolding) to enable analysis of cooperativity of unfolding of an individual protein and comparison of folding processes across a range of similar and different types of protein folds, (xi) production of specialist secondary structure reference data sets, or broader ones including more diverse structural types (e.g. membrane proteins), which will further aid in the analyses of protein structures by CD (24) and (xii) data standard creation for regulatory bodies (25).

Furthermore, just as it was found after the establishment of the PDB, many additional unforeseen uses are likely to be conceived with the public availability of this data.

FUTURE DEVELOPMENTS

It is anticipated that the number of entries will grow rapidly since CD spectroscopy is used as a major tool in structural biology, not only in its own right, but also as a complementary technique to other structural biology methods. As both granting and regulatory bodies and journals expand their requirements for documentation, validation and data sharing, the availability of the PCDDB as a public repository for CD and SRCD data is expected to become a significant resource for the structural biology and bioinformatics communities.

In the future, we plan to continue developing tools for automated deposition and curation and new tools for analyses, searches and interpretations, as well as creating international mirror sites located at synchrotrons that have SRCD beamlines and at major bioinformatics centres. New software currently in development includes DichroMatch, for identifying spectral nearest neighbours present in the PCDDB for a query protein spectrum, PDB2CD, which back calculates a CD spectrum based on a protein's PDB coordinates, and RDBCreate, which will create new specialized and general reference data sets for secondary structure analyses using existing algorithms (21). We also aim to create a direct link to the DichroWeb analysis server from a given PCDDB protein entry to enable the user to undertake whatever secondary structure analyses they prefer using a range of algorithms and reference data sets that may not have been used in the original publications.

Additional tools are envisioned for data mining (especially searching by spectral features) and visualization, which are aimed at enhancing the convenience of use for both expert and non-expert users, thereby promoting the main-stream and regular usage of CD data by structural biologists.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to express our thanks to the PCDDB Technical Advisory Board members for their advice and guidance: Dr Peter Baumgärtel (BESSY-II Synchrotron, Germany), Dr Paul Boxrud (Olis Inc., USA), Mr Steven Cave (Jasco UK Ltd, UK), Dr Lindsay Cole (Applied Photophysics Ltd, UK), Prof. Kunihiko Gekko (Hiroshima University, Japan), Dr Søren Vrønning Hoffmann (Institute for Storage Ring Facilities, Denmark), Dr Yi-hung Lin (National Synchrotron Radiation Research Centre, Taiwan), Dr David Moss (ANKA Synchrotron, Germany), Dr Glen Ramsay (Aviv Biomedical Inc, USA), Dr Giuliano Siligardi (Diamond Light Source, UK), Prof. John Sutherland (East Carolina University and Brookhaven National Laboratories, USA), Dr Ye Tao (Institute of High Energy Physics Synchrotron, China), and Dr Frank Wien (SOLEIL Synchrotron, France).

FUNDING

UK Biotechnology and Biological Sciences Research Council (BBSRC) (grant numbers F010346 and F010362 from the Bioinformatics and Biological Resources Fund to B.A.W. and R.W.J., respectively); BBSRC (grant number G023476 from the Tools and Resources Development Fund to B.A.W.); UK Medical Research Council DTA Studentship (to B.W.). Funding for open access charge: BBSRC.

Conflict of interest statement. None declared.

REFERENCES

1. Wallace,B.A., Whitmore,L. and Janes,R.W. (2006) The protein circular dichroism data bank (PCDDB): a bioinformatics and spectroscopic resource. *Proteins: Struct. Funct. Bioinform.*, **62**, 1–3.
2. Whitmore,L., Woollett,B., Miles,A.J., Janes,R.W. and Wallace,B.A. (2010) The protein circular dichroism data bank, a web-based site for access to circular dichroism spectroscopic data. *Structure*, **18**, 1267–1269.
3. Berman,H., Henrick,K., Nakamura,H. and Markley,J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **35**, D301–D303.
4. Consortium UniProt. (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.*, **37**, D169–D174.
5. Kouranov,A., Xie,L., de la Cruz,J., Chen,L., Westbrook,J., Bourne,P.E. and Berman,H.M. (2006) The RCSB PDB information portal for structural genomics. *Nucleic Acids Res.*, **34**, D302–D305.
6. Standley,D.M., Kinjo,A.R., Kinoshita,K. and Nakamura,H. (2008) Protein structure databases with new web services for structural biology and biomedical research. *Brief. Bioinformatics*, **9**, 276–285.
7. Velankar,S., Best,C., Beuth,B., Boutselakis,C.H., Cobley,N., Sousa Da Silva,A.W., Dimitropoulos,D., Golovin,A., Hirshberg,M., John,M. *et al*. (2010) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.*, **38**, D308–D317.
8. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
9. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.
10. Greene,L.H., Lewis,T.E., Addou,S., Cuff,A., Dallman,T., Dibley,M., Redfern,O., Pearl,F., Nambudiry,R., Reid,A. *et al*. (2007) The CATH domain structure database: New protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res.*, **35**, D291–D297.
11. Lees,J.G., Miles,A.J., Wien,F. and Wallace,B.A. (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics*, **22**, 1955–1962.
12. Whitmore,L. and Wallace,B.A. (2004) DichroWeb, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.*, **32**, W668–W673.
13. Evans,P., Bateman,O.A., Slingsby,C. and Wallace,B.A. (2007) A reference dataset for circular dichroism spectroscopy tailored for the βγ-crystallin lens proteins. *Experiment. Eye Res.*, **84**, 1001–1008.
14. Powl,A.M., O'Reilly,A.O., Miles,A.J. and Wallace,B.A. (2010) Synchrotron radiation circular dichroism spectroscopy-defined structure of the C-terminal domain of NaChBac and its role in channel assembly. *Proc. Natl Acad. Sci. USA*, **107**, 14064–14069.
15. Whitmore,L., Janes,R.W. and Wallace,B.A. (2006) Protein circular dichroism data bank (PCDDB): data bank and website design. *Chirality*, **18**, 426–429.
16. Lees,J.G., Smith,B.R., Wien,F., Miles,A.J. and Wallace,B.A. (2004) CDtool – an integrated software package for circular dichroism spectroscopic data processing, analysis and archiving. *Anal. Biochem.*, **332**, 285–289.
17. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B., Snoeyink,J., Richardson,J.S. *et al*. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
18. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283–291.
19. Hooft,R.W.W., Vriend,G., Sander,C. and Abola,E.E. (1996) WHAT_CHECK. Errors in protein structure. *Nature*, **381**, 272.
20. Jones,C., Schiffmann,D., Knight,A. and Windsor,S. (2004) Val-CiD best practice guide: CD spectroscopy for the quality control of biopharmaceuticals. *Natl Phys. Lab Report*, DQL-AS 008.
21. Kelly,S.M., Jess,T.J. and Price,N.C. (2005) How to study proteins by circular dichroism. *Biochim. Biophys. Acta*, **1751**, 119–139.

22. Miles,A.J. and Wallace,B.A. (2006) Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics. *Chem. Soc. Rev.*, **35**, 39–51.
23. Klose,D.P., Wallace,B.A. and Janes,R.W. (2010) 2Struc:The secondary structure server. *Bioinformatics*, **26**, 2624–2625.
24. Whitmore,L. and Wallace,B.A. (2006) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, **89**, 392–400.
25. Food and Drug Administration (1999) Guideline Q6B: test procedures and acceptance criteria for biotechnological/biological products. Federal Register 64, 44928–44935.