

Impact case study (REF3b)

Institution: University College London / Birkbeck College
Unit of Assessment: 5 – Biological Sciences
Title of case study: The DichroWeb Analysis Server and Protein Circular Dichroism Data Bank: analysis tools for structural biology
1. Summary of the impact (indicative maximum 100 words) <p>DICHROWEB is a comprehensive, user-friendly server that provides access to computational tools for the determination of protein secondary structure from data obtained through circular dichroism (CD) and synchrotron radiation (SRCD) spectroscopy. The Protein Circular Dichroism Data Bank (PCDDDB) is a database of spectra obtained using these techniques and allied data. Both resources are widely and increasingly used in many countries and are proving useful in industrial research (for example, in drug discovery) as well as academia and advanced teaching. DICHROWEB currently has over 3,600 registered users and over 375,000 DICHROWEB analyses have been run. Since the launch of PCDDDB in 2009, the database has had over 175,000 unique hits from 41 different countries, and 89,890 downloads.</p>
2. Underpinning research (indicative maximum 500 words) <p>Circular dichroism (CD) spectroscopy is a powerful tool for the structural analysis of peptides and proteins that was first developed in the 1960s. It measures the absorption spectra of molecules in the far ultra-violet (UV) part of the electromagnetic spectrum. As this absorption is influenced by the geometry of the peptide backbone, the absorption spectrum can be used to determine the proportion of each secondary structure type present in a protein. It is a method that is widely used in both academia and industry for the characterisation of proteins. In recent years, technical improvements, particularly the development of synchrotron radiation CD (SRCD) spectroscopy, have enabled lower wavelength data to be collected and thus increased the information content of the spectra. Wallace and her group have been and are still at the forefront of the development of these new techniques.</p> <p>An analysis of the secondary structure of a protein from its CD spectrum is based on comparisons with a set of reference spectra obtained from proteins with structures known to atomic resolution. Many different programs for the analysis of CD data have been made publicly available, although these required input and produced output data in different formats and units, and produced different types of comparison metrics. Wallace and her co-workers developed and continue to update and maintain the DICHROWEB server (http://dichroweb.cryst.bbk.ac.uk) as a user-friendly web-based interface to a wide range of both well-characterised and novel tools for the analysis of SRCD and CD spectroscopy data [1]. DICHROWEB users are able to submit their spectra in any recognised format and have access to many combinations of analysis programs and reference databases. The degree of agreement between the experimental data and a spectrum obtained by back-calculation from the predicted secondary structures is returned as a simple root mean square goodness-of-fit parameter, as well as provided as an easily interpretable graphical comparison of the experimental and calculated data [2].</p> <p>Research by Wallace and her collaborators has enabled a range of new features, input parameters, data types and reference databases to be added to DICHROWEB, taking into account advances in X-ray crystallography, NMR spectroscopy and bioinformatics that have greatly increased the range of atomic-resolution protein structures available and categorised them more precisely into fold families. Among these, new reference data were produced which cover secondary structure of the protein and fold space, using high quality SRCD data, allowing more accurate prediction of secondary structure [3]. Furthermore, new computational and bioinformatics developments have been incorporated by the Wallace group [4]. Techniques for the analysis of the dynamic properties of protein folding and of protein-protein interactions have also been developed [5].</p>

Impact case study (REF3b)

In a parallel development, Wallace created the Protein Circular Dichroism Data Bank (PCDDDB) in collaboration with Dr Robert Janes (Queen Mary University of London) as a publicly available repository for CD and SRCD data and a BBSRC Bioinformatics resource for data sharing. It was first released in December 2009 and is now fully operational for data deposit as well as access [6]. Besides the spectra, which are available for download graphically or as text files, this databank includes metadata such as sample types and experimental conditions. It also contains a set of tools for validating CD spectra which are extremely valuable for the community, particularly as journals do not insist on CD spectra being validated if the main emphasis of the publication is on a different technique. The database is freely available to the community via a web-based interface that runs on all up-to-date platforms and browsers and is user-friendly enough for non-experts to use. Each entry in PCDDDB is linked to a wide range of other structural biology resources and also linked through to DICHROWEB to allow users to run their own analyses on stored datasets. This database is a valuable resource for bioinformatics and experimental studies on proteins; it parallels the Protein Data Bank (PDB) as a repository for crystallographic data and enables comparisons with known proteins [7].

3. References to the research (indicative maximum of six references)

- [1] Lobley A, Whitmore L, Wallace BA. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*. 2002 Jan;18(1):211-2. <http://dx.doi.org/10.1093/bioinformatics/18.1.211>
- [2] Whitmore L, Wallace BA. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res*. 2004 Jul 1;32(Web Server issue):W668-73. <http://dx.doi.org/10.1093/nar/gkh371>
- [3] Lees JG, Miles AJ, Wien F, Wallace BA. A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics*. 2006 Aug 15;22(16):1955-62. http://people.cryst.bbk.ac.uk/~ubcg25a/BI_reprint.pdf
- [4] Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*. 2008 May;89(5):392-400. <http://dx.doi.org/10.1002/bip.20853>
- [5] Wallace BA, Janes RW. Synchrotron radiation circular dichroism (SRCD) spectroscopy: an enhanced method for examining protein conformations and protein interactions. *Biochem Soc Trans*. 2010 Aug;38(4):861-73. <http://dx.doi.org/10.1042/BST0380861>
- [6] Whitmore L, Woollett B, Miles AJ, Janes RW, Wallace BA. The protein circular dichroism data bank, a Web-based site for access to circular dichroism spectroscopic data. *Structure*. 2010 Oct 13;18(10):1267-9. <http://dx.doi.org/10.1016/j.str.2010.08.008>
- [7] Whitmore L, Woollett B, Miles AJ, Klose DP, Janes RW, Wallace BA. PCDDDB: the Protein Circular Dichroism Data Bank, a repository for circular dichroism spectral and metadata. *Nucleic Acids Res*. 2011 Jan;39(Database issue):D480-6. <http://dx.doi.org/10.1093/nar/gkq1026>

Grants supporting this work:

BBSRC BB/J019135/1. Bioinformatics Resources for Circular Dichroism Spectroscopy. 1/1/2013-31/12/2017. £856,528.

BBSRC F010346: Support for the Protein Circular Dichroism Data Bank and the Dichroweb Analysis Server. 01/01/2008-31/12/2012. £545,279.

BBSRC B02959. Bioinformatics tools for CD spectroscopy: validation and processing software and creation of a CD deposition data bank. 01/04/04-31/03/06. £235,446.

BBSRC 27/B13586. Bioinformatics for CD spectroscopy. Apr 01-Mar 04. £152,256.

4. Details of the impact (indicative maximum 750 words)

Both DICHROWEB and the PCDDDB are widely used by researchers in industry as well as for academic research and advanced teaching. DICHROWEB currently has over 3,600 registered users. Over 375,000 DICHROWEB analyses have been run and the service has been cited in over 1,000 publications [a]. This resource is free for academic and non-profit research use but industrial subscribers each pay a small annual fee for use of the service. The list of past and present industrial users includes representatives of the pharmaceutical, biotechnology and food sectors, and ranges from large multinational companies to the SME sector. Fees from industrial users provide an income of £2-5,000/year which is used as a contribution to running costs. No formal registration is required to access PCDDDB, but users can register to facilitate multiple queries or updates of contents. Currently registered industrial users include a number of large pharmaceutical companies and life science SMEs, plus about a dozen biomedical product and clinical diagnostic suppliers, amongst others. Since its launch in 2009, the database has had over 175,000 unique hits from 41 different countries, and 89,890 downloads (as at March 2012) [b].

Analysis of citations of our key publications reveals that a wide range of companies have obtained valuable results from using the tools that have been made accessible through DICHROWEB (many of which were developed by Wallace's group). For example, three studies cited below [c-e] illustrate typical recent cases in which the analysis tools within DICHROWEB have been applied to analysis of the stability and solution properties of proteins being developed as biological products, particularly vaccine candidates and tools for biomedical investigation. A complete list of citations with industrial co-authors is available [f]. This indicates use of DICHROWEB by a wide range of large and small companies from the pharmaceutical, vaccine, biotech, food and instrumentation sectors. A review of the patent literature has shown 11 patent applications from major corporations referring to the underpinning research papers, further indicating the commercial relevance of our work. The list of assignees that hold these patents include Roche Holding AG, Pfizer Inc. and Collplant Ltd [g].

The use of both DICHROWEB and the PCDDDB is promoted and encouraged by all the major commercial manufacturers of equipment for CD spectroscopy, who have established links from their own company websites either to DICHROWEB alone or to both DICHROWEB and the PCDDDB [h]. The well-established and prestigious RCSB Protein Data Bank Knowledge Base website, which is a primary access point for publicly available atomic-resolution protein structures, also links relevant entries to the appropriate entries in the PCDDDB; both sites are also linked from a number of widely used bioinformatics websites and are now being promoted via social media.

Wallace's work has been recognised in several awards and in invitations to present her work at international conferences. Her presentation with the prestigious AstraZeneca Award by the Biochemical Society in 2010 recognised her "*outstanding work which, through biomedical advances, leads to development of a new reagent or method*" [i]. In 2011 she was a keynote lecturer at the CASSS Higher Order Structure meeting on spectroscopy for industry (Rockville, MD, USA) [j].

Both DICHROWEB and the PCDDDB are widely used in advanced teaching both in the UK outside Birkbeck College and internationally. These include workshops for researchers such as an annual EU/UK summer school, which regularly involves delegates from industry, and university courses; DICHROWEB is mentioned in course notes and curricula from the Universities of Texas and Pittsburgh [k]. The use of both databases as resources for undergraduate teaching of CD spectroscopy has recently been recommended [l].

5. Sources to corroborate the impact (indicative maximum of 10 references)

[a] Exact figures from DICHROWEB Annual Report (March 2012): 3,459 users; 1,079 publications.

Impact case study (REF3b)

Report available on request.

- [b] Exact figure from PCDDDB Annual Report (March 2012): 175,527. Report available on request.
- [c] Salnikova MS, Joshi SB, Rytting JH, Warny M, Middaugh CR. Physical characterization of clostridium difficile toxins and toxoids: effect of the formaldehyde crosslinking on thermal stability. J Pharm Sci. 2008 Sep;97(9):3735-52. <http://dx.doi.org/10.1002/jps.21261>.
- [d] Martino A, Magagnoli C, De Conciliis G, D'Ascenzi S, Forster MJ, Allen L, Brookes C, Taylor S, Bai X, Findlow J, Feavers IM, Rodger A, Bolgiano B. Structural characterisation, stability and antibody recognition of chimeric NHBA-GNA1030: an investigational vaccine component against Neisseria meningitidis. Vaccine. 2012 Feb 8;30(7):1330-42. <http://dx.doi.org/10.1016/j.vaccine.2011.12.066>.
- [e] Paoletti F, Malerba F, Konarev PV, Visintin M, Scardigli R, Fasulo L, Lamba D, Svergun DI, Cattaneo A. Direct intracellular selection and biochemical characterization of a recombinant anti-proNGF single chain antibody fragment. Arch Biochem Biophys. 2012 Jun 1;522(1):26-36. <http://dx.doi.org/10.1016/j.abb.2012.04.003>.
- [f] A complete list of citations with industrial collaborators available on request and summarized below:
- Reference [1]: 385 citations in total (July 2013); 11 papers with industrial co-authors
- Reference [2]: 462 citations in total (July 2013); 13 papers with industrial co-authors
- Reference [4]: 363 citations in total (July 2013); 14 papers with industrial co-authors
- [g] Report by Cambridge IP Ltd available on request.
- [h] Links on commercial websites include
- http://www.avivbiomedical.com/circular_links.php (Aviv Biomedical)
 - <http://www.olisweb.com/Olis-Links.html> (Olis)
 - <http://www.photophysics.com/tutorials/cdnn-secondary-structure-analysis/3-more-analysis-programs> (Applied Photophysics, also links to other Wallace Lab resources)
 - <http://www.jasco.co.uk/dichroweb.php> (Jasco UK Ltd)
- [i] Astra Zeneca award notification: <http://www.biochemistry.org/Awards/AstraZenecaAward.aspx>
- [j] Programme available at http://www.casss.org/associations/9165/files/HOS2011_FINALProgram_091211.pdf
- [k] University of Texas: <http://www.biomachina.org/courses/structures/052.pdf>
University of Pittsburgh: www.cs.cmu.edu/~pittMB3/Lectures/CDLecture.ppt
- [l] DICROWEB and the PCDDDB are cited in: Urbach AR. Circular dichroism in the undergraduate curriculum. J. Chem. Ed. 2010;87(9):891-3. <http://dx.doi.org/10.1021/ed1005954> (see Refs 14 & 15).