Published online 14 August 2019 *Nucleic Acids Research, 2020, Vol. 48, Database issue* ***D261–D264***
doi: 10.1093/nar/gkz714

NAR Breakthrough Article

The ABCD database: a repository for chemically defined antibodies

Wanessa C. Lima[1,*], Elisabeth Gasteiger[2], Paolo Marcatili[3], Paula Duek[4], Amos Bairoch[4] and Pierre Cosson[1,*]

[1]Geneva Antibody Facility, Faculty of Medicine, University of Geneva, CH-1211 Geneva, Switzerland, [2]Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, CH-1211 Geneva, Switzerland, [3]Department of Bio and Health Informatics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark and [4]CALIPHO Group, Faculty of Medicine, University of Geneva and SIB Swiss Institute of Bioinformatics, CH-1211 Geneva, Switzerland

Received June 19, 2019; Revised August 01, 2019; Editorial Decision August 03, 2019; Accepted August 05, 2019

ABSTRACT

The ABCD (for AntiBodies Chemically Defined) database is a repository of sequenced antibodies, integrating curated information about the antibody and its antigen with cross-links to standardized databases of chemical and protein entities. It is freely available to the academic community, accessible through the ExPASy server (https://web.expasy.org/abcd/). The ABCD database aims at helping to improve reproducibility in academic research by providing a unique, unambiguous identifier associated to each antibody sequence. It also allows to determine rapidly if a sequenced antibody is available for a given antigen.

INTRODUCTION

Antibodies are one of the most widespread tools used in biological sciences. However, they are currently deemed one of the major culprits in the reproducibility crisis plaguing bio-medical research (1). Problems include batch-to-batch variability, poorly characterized and/or non-validated antibodies that sometimes do not recognize the presumptive target, or recognize more than one target, lack of explicitly described procedures adapted to each antibody, decreasing scrutiny of results by scientists and misleading antibody nomenclature. The 2 million antibodies available on the market might represent as few as 250'000 actual clones (1).

Standardized guidelines for antibody validation have been proposed to reduce reproducibility issues. These guidelines delineate a working framework to define antibody specificity and functionality for different research applications (2). In order to apply these guidelines, it is of course necessary that each antibody is identified easily and unambiguously.

Although the scientific community is well aware of this serious problem, few concerted solutions have appeared until now. The most advanced initiatives for centralizing information of antibodies are probably the portals Antibodypedia (3) and Antibody Registry [http://antibodyregistry.org/]), but both still rely largely on information provided by commercial vendors (such as antibody clone names). They also include an overwhelming majority of unsequenced or polyclonal antibodies, whose identity is difficult to clearly establish.

One of the solutions for this problem is to employ only sequenced antibodies that are unambiguously defined by their primary amino-acid sequence (4,5). In this way, researchers can be sure to be using the exactly same binding reagent. While it seems unlikely that systematic characterization of millions of antibodies will be achieved, for the estimated 20 000 currently described chemically defined (*i.e.* sequenced) monoclonal antibodies, the goal would seem more attainable. The IMGT database (created decades ago by Marie-Paule Lefranc and colleagues (6)) is an invaluable knowledge resource on sequences of immunoglobulins, but it is primarily aimed at studying the diversity of immune molecules, rather than their binding specificity.

Our goal is to provide the academic community with a wider access to recombinant, chemically defined antibodies (7). For this the recently launched ABCD database lists publicly available sequenced antibodies, and provides for each antibody a unique identifier and a link to its antigenic target.

*To whom correspondence should be addressed. Tel: +41 22 379 5294; Email: wanessa.delima@unige.ch
Correspondence may also be addressed to Pierre Cosson. Tel: +41 22 379 5293; Fax: +41 223 795 260; Email: Pierre.Cosson@unige.ch

A

ABCD_AI179 in the ABCD (AntiBodies Chemically Defined) Database

Antigen information	
Target type	Protein
Target link	UniProt: P01106 Homo sapiens (Human)
Target name	Proto-oncogene c-Myc, Myc proto-oncogene protein, MYC
Epitope	Myc-tag, Myc tag (EQKLISEEDLN)
Antibody information	
Antibody name	anti-cMyc-9E10
Comments	Use: Protein tag
Applications	X-ray crystallography
Cross-references	PDB: 2OR9 Cellosaurus: CVCL_L708
Publications	DOI: 10.24450/journals/abrep.2019.e27 PMID: 18473392
Antibody sequence	
If you want to have the protein sequence of this antibody, please check the Publications and Cross-references links. If you have trouble finding it, just send us an email using the contact form.	
Choices of Fc for production at the Geneva Antibody facility - more information.	
Mouse - Human - Rabbit - Other options	

B

ABCD_AF583 in the ABCD (AntiBodies Chemically Defined) Database

Antigen information	
Target type	Chemical
Target link	ChEBI: 17347
Target name	Testosterone
Epitope	Testosterone conjugated via the 3-O position
Antibody information	
Antibody name	ARK17-9
Applications	ELISA
Publications	Patent: WO2008009960
Antibody sequence	
If you want to have the protein sequence of this antibody, please check the Publications and Cross-references links. If you have trouble finding it, just send us an email using the contact form.	
Choices of Fc for production at the Geneva Antibody facility - more information.	
Mouse - Human - Rabbit - Other options	

Figure 1. Examples of antibody entries. Each entry has a unique identifier with the format ABCD_[A-Z][A-Z][0–9][0–9][0–9]. The Antibody table contains names and synonyms of the antibody, a published reference (with a link to PubMed, in case of scientific papers, or to the WIPO database, in case of a patent), and technical applications. The antibody sequence (see Figure 2 legend) is available on the Cross-references and Publications links provided, or upon request. (**A**) Target is a *Protein:* the Antigen table contains the name and species of target, a link to the UniProtKB UID, and information on the epitope when available. (**B**) Target is a *Chemical*: the Antigen table contains the target name, a link to the ChEBI UID, and information on the epitope when available.

OVERVIEW OF DATABASE CONTENT

The ABCD database is, to our knowledge, the first effort to provide freely accessible, curated information on chemically defined antibodies (*i.e.* antibodies with a known primary amino-acid sequence) connected with their antigenic target, which can be either a protein (linked to an UniProtKB unique identifier (UID) [(8), https://www.uniprot.org/]) or a chemical entity (linked to a ChEBI UID [(9), https://www.ebi.ac.uk/chebi/]).

Each ABCD entry corresponds to a unique primary amino-acid sequence, defined by a unique ABCD identifier. For each entry, information about the antigen and about the antibody are provided (Figure 1).

Regarding the *antibody*, in addition to its ABCD identifier, the following information is given:

(i) recommended name (most frequently, the name provided in the referenced publication) and a list of synonyms;
(ii) technical applications for which the antibody has been used (by no means an exhaustive inventory, as it lists only the applications described on the referenced publications);
(iii) at least one bibliographic reference (either a published scientific article—with a PubMed UID or a Digital Object Identifier (DOI)—or a patent, with a link to the WIPO database) in which the antibody sequence is provided. Note that this is not meant to be a comprehensive list of all the publications describing a given antibody;
(iv) cross-references to other databases (listed in Table 1).

Regarding the *antigen*, the following is given:

(i) type of target (if a protein or a chemical);
(ii) name of the antigen (and, in the case of a protein, also the species against which the antibody was produced);
(iii) link to UniProtKB (for a protein) or ChEBI (for a chemical) databases;
(iv) when available, information about the epitope recognized (for example, a domain or a specific amino-acid subsequence).

The antibody amino-acid sequence can be obtained in the links to the publications and the databases used as source (this is extensively explained on our FAQ section, with links and examples on how to obtain any given sequence). Alternatively, the information is also available upon request by email (via our Contact form). The stored information corresponds to the sequence of the variable region of both the heavy and light chains (or, in the case of camelid antibodies or nanobodies, the sequence of the unique variable chain) (Figure 2). When needed, definition of heavy and light chain boundaries, based on alignment with germline sequences, was done using the VBASE2 server (10).

The ABCD database is populated with data coming from (see Table 1 for a list of source databases): (i) sequences published in scientific articles or patents; (ii) 3D structural data; (iii) a few publications and repositories of large-scale phage display or hybridoma sequencing projects (11–15). We only include sequenced antibodies with a known and defined target. However, the source of such information is of variable quality, and we encourage users to verify (and to publish) the reactivity of each antibody that they use.

DATABASE DESIGN AND IMPLEMENTATION

The ABCD database is developed by the Geneva Antibody Facility team (https://www.unige.ch/medecine/antibodies/), in collaboration with the CALIPHO and Swiss-Prot groups at the Swiss Institute of Bioinformatics (https://www.sib.swiss/). The database is available at the ExPASy web server (https://web.expasy.org/abcd/).

Data is indexed for full text search using the Apache Lucy search engine library in PERL (https://lucy.apache.org/). This is a 'loose C' port of the Apache Lucene™ search engine library for Java. The query interface and entry display is implemented on the ExPASy server using PERL CGI scripts.

The ABCD database website consists of a simple, user-friendly interface. Each antibody page is dynamically linked to external resources and databases (see Table 1). Entries can be searched by antibody name, antigen name, antigen species, UniProtKB or ChEBI UIDs, epitope information and reference UID (PubMed, DOI or Patent), via a full-text search field.

The current release (v 4.0) contains 10′525 entries, referencing 9′076 proteins (1′642 unique UniProtKB UIDs) and 1′203 chemicals (261 unique ChEBI UIDs).

Table 1. List of databases and websites used as source of information or cross-reference

Database	Link	Data use	Ref.
Abysis	www.bioinf.org.uk/abysis2.7/	Source for Kabat sequences	(16)
Addgene	www.addgene.org	Source for antibody sequences inside vectors	(17)
Cellosaurus	web.expasy.org/cellosaurus/	X-ref for hybridomas	(18)
ChEBI	www.ebi.ac.uk/chebi/	X-ref for chemical targets	(9)
DigIt	circe.med.uniroma1.it/digit/	Source for sequences of annotated variable domains	(19)
IMGT/mAb-DB	imgt.org/mAb-DB/	Source for therapeutic antibody sequences	(6)
InterPro	www.ebi.ac.uk/interpro/	X-ref for domains	(20)
NCBI Taxonomy	www.ncbi.nlm.nih.gov/Taxonomy/	X-ref for species taxonomy	(21)
PROSITE	prosite.expasy.org	X-ref for domains	(22)
PubMed	www.ncbi.nlm.nih.gov/pubmed/	X-ref for publications Source for published sequences	(21)
RAN	recombinant-antibodies.org	Source for Recombinant Antibody Network antibodies	(12)
RCSB/PDB	www.rcsb.org/pdb/	X-ref for 3D structures Source for published sequences	(23)
UniProt	www.uniprot.org	X-ref for protein targets	(8)
WIPO Patents	patentscope.wipo.int	X-ref for patent publications	—

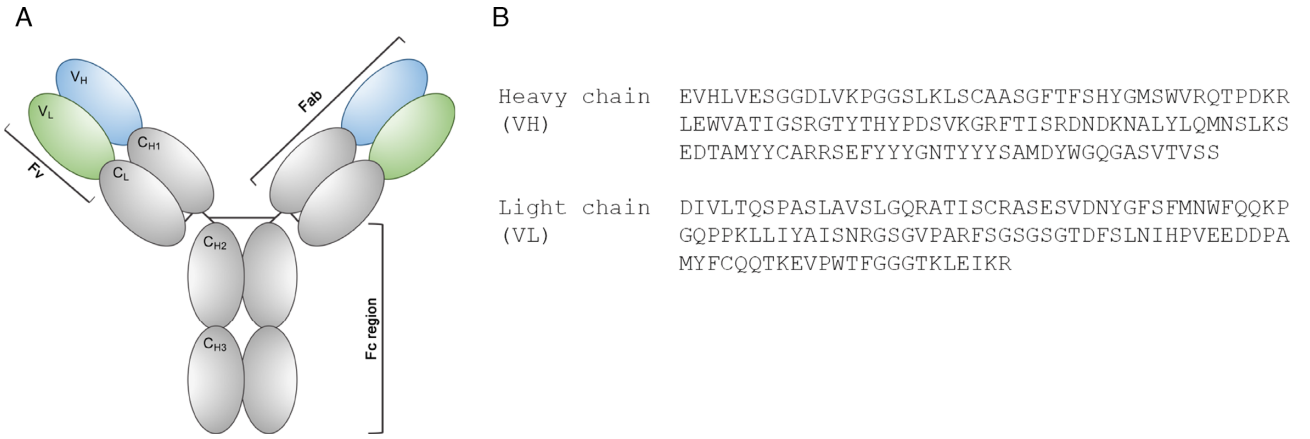

Figure 2. Antibody sequence information. (**A**) An immunoglobulin consists of constant (C, in gray) and variable (V, in blue and green) chains. The paratope (or specific binding site) of an antibody is located at the variable moiety of the light (V_L) and heavy (V_H) chains. (**B**) The ABCD database stores as sequence information the amino-acid sequence of both V_L and V_H chains (the example given corresponds to sequence of entry ABCD_AI179, the anti-cMyc 9E10 clone).

CONCLUSION AND PERSPECTIVES

We believe that this initiative is a valuable step in setting up a centralized repository of sequenced antibodies, allowing the unique and unambiguous identification of binding reagents for research and publication purposes.

Depositing or publishing the sequence information of any given antibody should be a required step during any antibody characterization procedure; careful and thorough validation is still obligatory, but knowing the precise identity of a given reagent would allow others to repeat the exact same experiment.

All entries in the ABCD database are manually curated and, hence, the database growth is linear and slow. Using computational approaches is not a desirable strategy: defining the identity of a given antibody targets is a cumbersome process, involving extensive literature mining, a process that is not easily automatized. One approach to allow for a faster inclusion of entries is to promote the submission of sequences by colleagues around the world, originating from large-scale discovery projects or sequencing of hybridomas or purified antibodies.

FUNDING

ProCare Foundation. Funding for open access charge: Swiss National Science Foundation (31003A-172951).
Conflict of interest statement. None declared.

REFERENCES

1. Baker,M. (2015) Reproducibility crisis: Blame it on the antibodies. *Nature*, **521**, 274–276.
2. Uhlen,M., Bandrowski,A., Carr,S., Edwards,A., Ellenberg,J., Lundberg,E., Rimm,D.L., Rodriguez,H., Hiltke,T., Snyder,M. *et al.* (2016) A proposal for validation of antibodies. *Nat. Methods*, **13**, 823–827.
3. Bjorling,E. and Uhlen,M. (2008) Antibodypedia, a portal for sharing antibody and antigen validation data. *Mol. Cell Proteomics*, **7**, 2028–2037.
4. Bradbury,A. and Pluckthun,A. (2015) Reproducibility: standardize antibodies used in research. *Nature*, **518**, 27–29.

5. Weller,M.G. (2016) Quality issues of research antibodies. *Anal. Chem. Insights*, **11**, 21–27.
6. Lefranc,M.P., Giudicelli,V., Duroux,P., Jabado-Michaloud,J., Folch,G., Aouinti,S., Carillon,E., Duvergey,H., Houles,A., Paysan-Lafosse,T. *et al.* (2015) IMGT(R), the international ImMunoGeneTics information system(R) 25 years on. *Nucleic Acids Res.*, **43**, D413–D422.
7. Cosson,P. and Hartley,O. (2016) Recombinant antibodies for Academia: A practical approach. *Chimia (Aarau)*, **70**, 893–897.
8. UniProt Consortium, T. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169
9. Hastings,J., Owen,G., Dekker,A., Ennis,M., Kale,N., Muthukrishnan,V., Turner,S., Swainston,N., Mendes,P. and Steinbeck,C. (2016) ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.*, **44**, D1214–D1219.
10. Retter,I., Althaus,H.H., Munch,R. and Muller,W. (2005) VBASE2, an integrative V gene database. *Nucleic Acids Res.*, **33**, D671–D674.
11. Andrews,N.P., Boeckman,J.X., Manning,C.F., Nguyen,J.T., Bechtold,H., Dumitras,C., Gong,B., Nguyen,K., van der List,D., Murray,K.D. *et al.* (2019) A toolbox of IgG subclass-switched recombinant monoclonal antibodies for enhanced multiplex immunolabeling of brain. *Elife*, **8**, e43322
12. Hornsby,M., Paduch,M., Miersch,S., Saaf,A., Matsuguchi,T., Lee,B., Wypisniak,K., Doak,A., King,D., Usatyuk,S. *et al.* (2015) A high Through-put platform for recombinant antibodies to folded proteins. *Mol. Cell Proteomics*, **14**, 2833–2847.
13. Jain,T., Sun,T., Durand,S., Hall,A., Houston,N.R., Nett,J.H., Sharkey,B., Bobrowicz,B., Caffry,I., Yu,Y. *et al.* (2017) Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 944–949.
14. Schoenherr,R.M., Saul,R.G., Whiteaker,J.R., Yan,P., Whiteley,G.R. and Paulovich,A.G. (2015) Anti-peptide monoclonal antibodies generated for immuno-multiple reaction monitoring-mass spectrometry assays have a high probability of supporting Western blot and ELISA. *Mol. Cell Proteomics*, **14**, 382–398.
15. Schofield,D.J., Pope,A.R., Clementel,V., Buckell,J., Chapple,S., Clarke,K.F., Conquer,J.S., Crofts,A.M., Crowther,S.R., Dyson,M.R. *et al.* (2007) Application of phage display to high throughput antibody generation and characterization. *Genome Biol.*, **8**, R254.
16. Swindells,M.B., Porter,C.T., Couch,M., Hurst,J., Abhinandan,K.R., Nielsen,J.H., Macindoe,G., Hetherington,J. and Martin,A.C. (2017) abYsis: Integrated antibody sequence and Structure-Management, analysis, and prediction. *J. Mol. Biol.*, **429**, 356–364.
17. Kamens,J. (2015) The Addgene repository: an international nonprofit plasmid and data resource. *Nucleic Acids Res.*, **43**, D1152–D1157.
18. Bairoch,A. (2018) The cellosaurus, a Cell-Line knowledge resource. *J. Biomol. Tech.*, **29**, 25–38.
19. Chailyan,A., Tramontano,A. and Marcatili,P. (2012) A database of immunoglobulins with integrated tools: DIGIT. *Nucleic Acids Res.*, **40**, D1230–D1234.
20. Mitchell,A.L., Attwood,T.K., Babbitt,P.C., Blum,M., Bork,P., Bridge,A., Brown,S.D., Chang,H.Y., El-Gebali,S., Fraser,M.I. *et al.* (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.*, **47**, D351–D360.
21. Sayers,E.W., Agarwala,R., Bolton,E.E., Brister,J.R., Canese,K., Clark,K., Connor,R., Fiorini,N., Funk,K., Hefferon,T. *et al.* (2019) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **47**, D23–D28.
22. Sigrist,C.J., de Castro,E., Cerutti,L., Cuche,B.A., Hulo,N., Bridge,A., Bougueleret,L. and Xenarios,I. (2013) New and continuing developments at PROSITE. *Nucleic Acids Res.*, **41**, D344–D347.
23. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.