The ABCD database: a repository for chemically defined antibodies

Lima, Wanessa C.; Gasteiger, Elisabeth; Marcatili, Paolo; Duek, Paula; Bairoch, Amos; Cosson, Pierre

Published in:
Nucleic acids research

Link to article, DOI:
10.1093/nar/gkz714

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
NAR Breakthrough Article

The ABCD database: a repository for chemically defined antibodies

Wanessa C. Lima1,*, Elisabeth Gasteiger2, Paolo Marcatili3, Paula Duek4, Amos Bairoch4 and Pierre Cosson1,*

1Geneva Antibody Facility, Faculty of Medicine, University of Geneva, CH-1211 Geneva, Switzerland, 2Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, CH-1211 Geneva, Switzerland, 3Department of Bio and Health Informatics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark and 4CALIPHO 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 22 379 5260; Email: Pierre.Cosson@unige.ch

© The Author(s) 2019. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
TheABCDDatabaseisdevelopedbytheGenevaAntibody
Facilityteam([https://www.unige.ch/medecine/antibodies/](https://www.unige.ch/medecine/antibodies/)),
in collaboration with the CALIPHO and Swiss-Prot groups at the Swiss Institute of Bioinformatics([https://www.sib.

Data is indexed for full-text search using the Apache Lucene search engine library in PERL ([https://lucy.apache.org/](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,UniProtKBorChEBlUIDsand 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

<table>
<thead>
<tr>
<th>Database</th>
<th>Link</th>
<th>Data use</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abysis</td>
<td><a href="http://www.bioinf.org.uk/abysis2.7/">www.bioinf.org.uk/abysis2.7/</a></td>
<td>Source for Kabat sequences</td>
<td>(16)</td>
</tr>
<tr>
<td>Addgene</td>
<td><a href="http://www.addgene.org">www.addgene.org</a></td>
<td>Source for antibody sequences inside</td>
<td>(17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>vectors</td>
<td></td>
</tr>
<tr>
<td>Cellosaurus</td>
<td>web.expasy.org/cellosaurus/</td>
<td>X-ref for hybridomas</td>
<td>(18)</td>
</tr>
<tr>
<td>ChEBI</td>
<td><a href="http://www.ebi.ac.uk/chebi/">www.ebi.ac.uk/chebi/</a></td>
<td>X-ref for chemical targets</td>
<td>(9)</td>
</tr>
<tr>
<td>DigIt</td>
<td>circe.med.uniroma1.it/digit/</td>
<td>Source for sequences of annotated variable domains</td>
<td>(19)</td>
</tr>
<tr>
<td>IMGT/mAb-DB</td>
<td>imgt.org/mAb-DB/</td>
<td>Source for therapeutic antibody sequences</td>
<td>(6)</td>
</tr>
<tr>
<td>InterPro</td>
<td><a href="http://www.ebi.ac.uk/interpro/">www.ebi.ac.uk/interpro/</a></td>
<td>X-ref for domains</td>
<td>(20)</td>
</tr>
<tr>
<td>PROSITE</td>
<td>prosite.expasy.org</td>
<td>X-ref for domains</td>
<td>(22)</td>
</tr>
<tr>
<td>RAN</td>
<td>recombinant-antibodies.org</td>
<td>Source for Recombinant Antibody</td>
<td>(12)</td>
</tr>
<tr>
<td>RCSB/PDB</td>
<td><a href="http://www.rcsb.org/pdb/">www.rcsb.org/pdb/</a></td>
<td>Network antibodies</td>
<td></td>
</tr>
<tr>
<td>UniProt</td>
<td><a href="http://www.uniprot.org">www.uniprot.org</a></td>
<td>X-ref for protein targets</td>
<td>(8)</td>
</tr>
<tr>
<td>WIPO Patents</td>
<td>patentscope.wipo.int</td>
<td>X-ref for patent publications</td>
<td>—</td>
</tr>
</tbody>
</table>

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, originat-