ADPriboDB: The database of ADP-ribosylated proteins

Christina A. Vivelo¹, Ricky Wat¹, Charul Agrawal¹, Hui Yi Tee¹ and Anthony K. L. Leung¹,²,*

¹Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA and ²Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD 21205, USA

Received June 22, 2016; Revised July 11, 2016; Accepted July 13, 2016

ABSTRACT

ADP-ribosylation refers to the addition of one or more ADP-ribose units onto proteins post-translationally. This protein modification is often added by ADP-ribosyltransferases, commonly known as PARPs, but it can also be added by other enzymes, including sirtuins or bacterial toxins. While past literature has utilized a variety of methods to identify ADP-ribosylated proteins, recent proteomics studies bring the power of mass spectrometry to determine sites of the modification. To appreciate the diverse roles of ADP-ribosylation across the proteome, we have created ADPriboDB – a database of ADP-ribosylated proteins (http://ADPriboDB.leunglab.org). Each entry of ADPriboDB is annotated manually by at least two independent curators from the literature between January 1975 and July 2015. The current database includes over 12 400 protein entries from 459 publications, identifying 2389 unique proteins. Here, we describe the structure and the current state of ADPriboDB as well as the criteria for entry inclusion. Using this aggregate data, we identified a statistically significant enrichment of ADP-ribosylated proteins in non-membranous RNA granules. To our knowledge, ADPriboDB is the first publicly available database encapsulating ADP-ribosylated proteins identified from the past 40 years, with a hope to facilitate the research of both basic scientists and clinicians to better understand ADP-ribosylation at the molecular level.

INTRODUCTION

ADP-ribosylation refers to the addition of one or more ADP-ribose units onto proteins through the transfer of the ADP-ribose group from NAD⁺ to target proteins post-translationally. ADP-ribose groups can be attached singly as mono(ADP-ribose) or in polymeric chains as poly(ADP-ribose) (PAR) by the enzymatically active members of the family of 17 human ADP-ribosyltransferases, commonly known as PARPs (1–3). In addition, mono(ADP-ribosylation) can be facilitated by other enzymes including sirtuins, extracellular membrane-associated ADP-ribosyltransferases and bacterial toxins (4–6). Of note, non-enzymatic addition of ADP-ribose groups onto proteins was also reported in vitro (7,8). ADP-ribosylation can be added onto amino acids of diverse chemistry, including glutamate, aspartate, lysine, arginine and cysteine (9,10). Some of these ADP-ribosyl conjugations have shown to be reversible and their removal is mediated by two broad classes of enzymes that cleave PAR or the bond between ADP-ribose and its conjugated amino acid (11,12). Poly(ADP-ribose) glycohydrolase (PARG) and ADP-ribosylhydrolase 3 (ARH3) cleave the ribose–ribose bonds between ADP-ribose subunits (11,12). While ADP-ribosylhydrolase ARH1 removes single ADP-ribose groups from conjugated arginine residues, ADP-ribosylhydrolases MacroD1, MacroD2 and TARG1 remove the last remaining ADP-ribose groups from poly(ADP-ribose) substrates or single ADP-ribose groups from mono(ADP-ribosyl)ated proteins conjugated at acidic residues (11,12). TARG1 is unique for its additional ability to remove the whole PAR chain specifically at glutamate–ADP-ribose ester bonds (13). Several techniques have been developed to enrich and identify ADP-ribosylated substrates (14). Some techniques use antibodies or protein domains that bind ADP-ribose to pull out the modification from cell lysates and use mass spectrometry to identify ADP-ribosylated substrates (15–17). Other techniques utilize protein microarrays and recombinant PARPs to identify PARP-specific substrates (18,19).

ADP-ribosylation has numerous and diverse effects on protein functions and cellular pathways, including DNA damage, transcription, chromatin organization, stress responses, circadian rhythms, cell cycle regulation and RNA metabolism (2,20–24). The addition of the modification can affect the substrate’s stability and activity, as well as serve as a scaffold to recruit other proteins non-covalently (25–27). ADP-ribosylation activity spans a diverse range of organisms across different kingdoms, from viruses to bacteria to mammals, highlighting the importance of this modi-

*To whom correspondence should be addressed. Tel: +1 410 955 2926; Email: anthony.leung@jhu.edu
Present address: Charul Agrawal, Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology Delhi, India.

© The Author(s) 2016. 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 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
fication in living organisms (28,29). Because of the diverse roles they play in the cell, ADP-ribosylation and ADP-ribosyltransferase/PARP activity have been implicated in a spectrum of disease pathogenesis, such as cancer, inflammatory diseases and neurological disorders (30,31). Notably, PARP inhibitors have already shown clinical benefits in multiple cancers, including Olaparib (LynparzaTM) that has already been approved by the Food and Drug Administration (FDA) in the United States and European Medicines Agency (32,33). Thus, understanding substrate specificity of PARPs and how protein function is affected by ADP-ribosylation are of timely importance and have the potential to improve our understanding of the biology of diseases.

Though the modification was discovered over 50 years ago (34,35), it was only until recently that several proteomics techniques have been developed to identify the ADP-ribosylation sites (9,36). Such site information will provide initial insights to generate hypotheses for testing the function of ADP-ribosylation in modified substrates. To facilitate researchers in appreciating the depth and breadth of ADP-ribosylation’s role across the proteome, we have created ADPriboDB, a database of ADP-ribosylated proteins curated from the literature between 1975 and July 2015 (ADPriboDB.leunglab.org). ADPriboDB is freely accessible online and provides information for each entry on the protein name, gene symbol, UniProt ID, species accessible online and provides information for each entry on the protein name, gene symbol, UniProt ID, species/cell line of origin, experimental details and states the original figure/table from which the entry was curated. Entries from mass spectrometry studies include the modified residues and peptide sequences wherever available. The database curation effort is ongoing and the website also allows single entries or bulk format upload for publication-associated data sets. It is our hope that ADPriboDB will offer insights into the ADP-ribosylated proteome, and may reveal patterns of ADP-ribosylation across substrate proteins, and highlight important interactions and networks, which will aid in our understanding of the myriad roles of ADP-ribosylation played in cells.

Criteria for inclusion in the ADPriboDB

ADPriboDB was compiled through a manual curation of the literature accessed from PubMed, using the following 16 search terms: ‘PARylation’ or ‘PARsylation’ or ‘poly(ADP-ribosylation)’ or ‘poly-ADP-ribosylation’ or ‘poly(ADP-ribosylation)’ or ‘poly(ADP-ribosyl)ation’ or ‘poly-adp-ribosylation’ or ‘poly(ADP-ribosyl)ation’ or ‘poly(ADP-ribosyl)ation’ or ‘poly(ADP-ribosyl)ation’ or ‘poly(ADP-ribosyl)ation’ or ‘poly(ADP-ribosyl)ation’ or ‘poly(ADP-ribosyl)ation’ or ‘poly(ADP-ribosyl)ation’ or ‘poly(ADP-ribosyl)ation’ or ‘poly(ADP-ribosyl)ation’ or ‘poly(ADP-ribosyl)ation’ or ‘MARylation’ or ‘MAR-ribosylation’ or ‘MARsylation’ or ‘MAR-ribosylation’ or ‘MAR-ribosylation’ or ‘MAR-ribosylation’ or ‘MAR-ribosylation’. This parameter was further limited for our search between January 1975 and July 2015 yielding 1456 articles. Each journal article was assessed for the inclusion in ADPriboDB by at least two independent curators.

Diverse techniques have been applied to demonstrate that the entries are ADP-ribosylated (‘Method’; see Supplementary Table S1), including immunoprecipitation with anti-PAR antibodies from cells and autoradiography of in vitro modification of recombinant proteins by ADP-ribosyltransferases with 32P-labeled NAD+. Recent years have seen the application of mass spectrometry approaches for validation of protein modifications and site identification. For the sake of comprehensiveness, ADPriboDB includes all proteins identified from immunoprecipitation/affinity-based approaches as tentatively ADP-ribosylated. The underlying caveat of these approaches using antibodies and ADP-ribose binding domains is that it cannot distinguish whether these proteins are non-covalently bound with ADP-ribose groups or covalently attached to them. However, such distinction can now be made possible with the ability to identify the ADP-ribosylation sites (9,36). As a result, we include modification details based on mass spectrometry analyses in ADPriboDB whenever they are available. For example, CEBPB was identified in an affinity-based proteomics study using the AF1521 ADP-ribose binding macrodomain (15), but the ADP-ribosylation site was subsequently identified in a global site mapping study (37) (Figure 1).

ADPriboDB structure

The database is divided into four sections (Figure 1): ‘Identifiers’, ‘Experimental Details’, ‘Modification Details’ and ‘Literature References’. If a modified protein was identified, the protein name, gene symbol and corresponding UniProt ID were included in the ‘Identifiers’ section. In the circumstance in which a protein was identified but no UniProt IDs and gene symbols were associated in the publication, the curator(s) determined the best assignment of a UniProt ID based on all the information provided and noted that this entry as ‘self-assigned’, along with any other comments in the ‘Curator Comments’ field. If multiple variants of the protein are available on UniProt, the first variant was chosen.

Several key experimental details were collected under the ‘Experimental Details’ section, including the origin of the cell line/tissue (‘Origin’) and the species they are derived from (‘Species’). Additionally, we noted whether the modification is induced/inhibited by certain drug treatment (‘Drug sensitivity’). Under the ‘Modification Details’ section, we noted whether the modification is regulated by certain ADP-ribosyltransferases (‘Enzymes’) and whether the modification is mono(ADP-ribosyl)ation or poly(ADP-ribosylation) (‘Mono/Poly’). In addition, the ADP-ribosylation sites and peptide sequences identified by mass spectrometry were included (‘Sequencing Information’). Under the ‘Literature References’ section, the figure/table number displaying the data for modification and an excerpt of the text describing the data were included under ‘Data Description’. The PubMed ID and year of these publications were included to allow users to easily review or cite the primary sources for the identification of specific ADP-ribosylated proteins.

All entries, given a unique ADPriboDB identification number, were organized in a flat text file database. A user search interface was then created using Xataface, a web application framework for implementing the open-source relational database management system MySQL. ADPriboDB can be searched by Protein Name, Gene Symbol, UniProt ID, PubMed ID and Enzymes (e.g. PARP1). The rest of the webpages to explain the functioning of the database were created through Weebly.
Current state of ADPriboDB

Manual curation of literature between January 1975 and July 2015 identified 12,428 entries, in which 8,867 of them have protein sequencing data available. The current database entries are derived from 459 publications with experimental supports for PARylated and/or MARylated proteins. The database currently includes 2,389 unique proteins and 573 of them are associated with sequencing information. Though 93% of entries are derived from human, over 860 entries are from diverse organisms, which include not only the common model systems, such as mouse, rat, fruitfly and Arabidopsis, but also quail, cow, chicken, sea urchin, octopus, various bacteria and viruses (Figure 2A).

Amongst 459 publications, 58% of them (n = 264) identified single ADP-ribosylated proteins, 5% (n = 25) identified at least 10 substrates and 2% (n = 10) from proteomics studies with more than 50 substrates identified (Figure 2B). Among the 2,389 unique proteins identified, 80% of ADP-ribosylated substrates (1,908) were uniquely identified in these 10 proteomics studies and 17% (402) from the rest of the 449 publications (Figure 2C). Surprisingly, only 79 ADP-ribosylated substrates were overlapped between these two sets, suggesting that we have yet to reach saturation in probing the ADP-ribosylated proteome. Consistently, >71% (1,708 out of 2,389 unique proteins) were identified only once (Figure 2D). ADP-ribosylation of human PARP1 (P09874) has been the most comprehensively analyzed protein with 111 publications, followed by histones, other PARPs and their homologs in mouse and rat (Table 1). Other proteins identified in at least 10 publications, including tumor suppressor p53 and DNA Topoisomerase 1, are also included in Table 1.

ADP-ribosylation and non-membraneous RNA granules

The availability of this aggregate data potentially allows researchers to gain insights into global functions of ADP-ribosylation. Our previous informatics analyses indicated that ADP-ribosylated substrates are significantly enriched with proteins that are enriched for low-complexity regions, which aid self-assembly of non-membranous organelles (25). Using 457 human proteins that have sequencing in-

<table>
<thead>
<tr>
<th>UniProt Entry</th>
<th>Protein Name</th>
<th>Number of Times Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARP1_HUMAN</td>
<td>Poly(ADP-ribose) polymerase 1</td>
<td>111</td>
</tr>
<tr>
<td>H11_HUMAN</td>
<td>Histone H1.1</td>
<td>27</td>
</tr>
<tr>
<td>H11_MOUSE</td>
<td>Histone H1.1</td>
<td>25</td>
</tr>
<tr>
<td>PARP1_MOUSE</td>
<td>Poly(ADP-ribose) polymerase 1</td>
<td>24</td>
</tr>
<tr>
<td>PARP1_RAT</td>
<td>Poly(ADP-ribose) polymerase 1</td>
<td>21</td>
</tr>
<tr>
<td>P53_HUMAN</td>
<td>Tumor suppressor p53</td>
<td>18</td>
</tr>
<tr>
<td>H31_HUMAN</td>
<td>Histone H3.1</td>
<td>18</td>
</tr>
<tr>
<td>PARP1_BOVIN</td>
<td>Poly(ADP-ribose) polymerase 1</td>
<td>17</td>
</tr>
<tr>
<td>H4_HUMAN</td>
<td>Histone H4</td>
<td>17</td>
</tr>
<tr>
<td>H2A1_HUMAN</td>
<td>Histone H2A.1</td>
<td>16</td>
</tr>
<tr>
<td>H2B1C_HUMAN</td>
<td>Histone H2B type 1-C/E/F/G/I</td>
<td>15</td>
</tr>
<tr>
<td>H11_RAT</td>
<td>Histone H1.1</td>
<td>14</td>
</tr>
<tr>
<td>TNKS1_HUMAN</td>
<td>Tankyrase-1</td>
<td>13</td>
</tr>
<tr>
<td>H10_HUMAN</td>
<td>Histone H1.0</td>
<td>12</td>
</tr>
<tr>
<td>H2A2A_MOUSE</td>
<td>Histone H2A.2</td>
<td>12</td>
</tr>
<tr>
<td>TOP1_HUMAN</td>
<td>DNA topoisomerase 1</td>
<td>11</td>
</tr>
<tr>
<td>PARP2_HUMAN</td>
<td>Poly(ADP-ribose) polymerase 2</td>
<td>11</td>
</tr>
<tr>
<td>H2B1A_RAT</td>
<td>Histone H2B type 1-A</td>
<td>10</td>
</tr>
</tbody>
</table>
formation in ADPriboDB, we tested whether these substrates are enriched in particular subcellular compartments. ADP-ribosylated substrates are significantly enriched in the proteome of non-membraneous RNA granules, including the nucleolus (38) and stress granules (39) (P-value = 1.85 \times 10^{-34} and 5.00 \times 10^{-28} respectively, Fisher’s Exact Test; Figure 3A). These data are consistent with the identification of the ADP-ribosyltransferases in these non-membraneous organelles—PARP1 and PARP2 in the nucleolus (40) and PARP5a, PARP12, PARP13, PARP14 and PARP15 in the stress granules (41). Consistent with a previous analysis (42), both the proteomes of the nucleolus and stress granules are enriched for low complexity domain-containing proteins (Figure 3A). In contrast, similar analyses for N-glycosylated substrates are enriched in the endoplasmic reticulum, golgi, lysosome and membrane fractions and these substrates are not enriched for low complexity domain-containing proteins (Figure 3A and B). Such a statistically significant association between ADP-ribosylated substrates and non-membraneous organelles is in agreement with our informatics-led hypothesis (25) and recent experimental evidence (43,44) indicating that PAR can help seed low complexity domain-containing proteins to form non-membraneous structures.

Future directions

ADPriboDB is built with the anticipation of several major developments in the ADP-ribosylation field—site identification, enzyme–substrate specificity and substrate/site sensitivity to PARP inhibitors. First, given the development of various site identification methods (9,14,36), we are expecting that a series of proteome-wide analyses will soon be available for a global survey of ADP-ribosylated proteins in cells and tissues. The first generation of ADPriboDB will therefore help connect these newly acquired proteomics data with the existing literature between 1975 and 2015. In addition, ADP-ribosylation sites can be used to correlate with existing databases on other post-translational modifications, such as phosphorylation, acetylation and methylation (45,46). Second, 19% of ADPriboDB entries (2419) include the information regarding the enzyme responsible for the modification of specific substrates, where 69% (1659) are mediated by PARP1. However, we expect this trend will be changed soon with the coming-of-age technologies to investigate enzyme–substrate specificity of 17...
human ADP-ribosyltransferases. These include proteome arrays (18,19), the development of analog-sensitive ADP-ribosyltransferases (47–49) as well as the improvement in performing genetic knockdown by RNA interference and knockout using CRISPR in mammalian cell culture. Finally, PARP inhibition is actively used in clinical settings. Therefore, future development of these PARP-specific inhibitors for clinical use should couple with our basic scientific knowledge in how the endogenous ADP-ribosylated proteome is affected (9,50). As a result, ADPriboDB starts to collect information under the field ‘Drug sensitivity’ to report whether certain substrates/sites are sensitive to particular PARP inhibitors (37). With ADPriboDB, we aim to provide a one-stop portal for the community to search ADP-ribosylation substrates and site information, where basic scientists and clinicians can use this cumulative information to understand basic biology of ADP-ribosylation, generate hypotheses in examining ADP-ribosylation functions and explain clinical benefits and side effects observed in patients at the molecular level.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT
The authors thank Jerome Yu for the initial curation work, Diego Miranda-Saavedra for the discussion of ADPriboDB and the Leung lab members for critical discussions.

FUNDING
Allegheny Health Network–Johns Hopkins Cancer Research Fund; Johns Hopkins Catalyst Award; American Cancer Society Research Scholar Award [129539-RSG-16-062-01-RMC]; National Institute of Health [R01-GM104135]. The trainee was funded by an NCI training grant [5T32CA009110 to C.A.V]; Khorana Program to C.A. Funding for open access charge: Allegheny Health Network–Johns Hopkins Cancer Research Fund.

Conflict of interest statement. None declared.

REFERENCES


