The Natural History of ADP-Ribosyltransferases and the ADP-Ribosylation System

L. Aravind, Dapeng Zhang, Robson F. de Souza, Swadha Anand and Lakshminarayan M. Iyer

Abstract Catalysis of NAD^+-dependent ADP-ribosylation of proteins, nucleic acids, or small molecules has evolved in at least three structurally unrelated superfamilies of enzymes, namely ADP-ribosyltransferase (ART), the Sirtuins, and probably TM1506. Of these, the ART superfamily is the most diverse in terms of structure, active site residues, and targets that they modify. The primary diversification of the ART superfamily occurred in the context of diverse bacterial conflict systems, wherein ARTs play both offensive and defensive roles. These include toxin–antitoxin systems, virus-host interactions, intraspecific antagonism (polymorphic toxins), symbiont/parasite effectors/toxins, resistance to antibiotics, and repair of RNAs cleaved in conflicts. ARTs evolving in these systems have been repeatedly acquired by lateral transfer throughout eukaryotic evolution, starting from the PARP family, which was acquired prior to the last eukaryotic common ancestor. They were incorporated into eukaryotic regulatory/epigenetic control systems (e.g., PARP family and NEURL4), and also used as defensive (e.g., pierisin and CARP-1 families) or immunity-related proteins (e.g., *Gig2*-like ARTs). The ADP-ribosylation system also includes other domains, such as the Macro, ADP-ribosyl glycohydrolase, NADAR, and ADP-ribosyl cyclase, which appear to have initially diversified in bacterial conflict-related systems. Unlike ARTs, sirtuins appear to have a much smaller presence in conflict-related systems.

Supplementary MaterialA list of Genbank identifier and domain architectures of the proteins discussed in this chapter might be found at: ftp://ftp.ncbi.nih.gov/pub/aravind/ADPR/ADPR.html

L. Aravind (✉) · D. Zhang · S. Anand · L. M. Iyer
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA
e-mail: aravind@ncbi.nlm.nih.gov

R. F. de Souza
Microbiology Department, Biomedical Sciences Institute, University of Sao Paulo, Av. Prof. Lineu Prestes, 1374-Biomédicas II-Sala 250, Cidade Universitária, São Paulo 05508-900, Brazil

Current Topics in Microbiology and Immunology
DOI: 10.1007/82_2014_414
© Springer International Publishing Switzerland 2014

Contents

1 Introduction
2 The Many Faces of ADP-Ribosylation
3 The ART Superfamily
3.1 Biochemical Diversity in the ART Superfamily
3.2 Structural and Catalytic Features of the ART Superfamily
4 Evolutionary History and Diversity of the ART Superfamily
4.1 The H-H-h Clade
4.2 The H-Y-[EDQ] Clade
4.3 The R-S-E Clade
4.4 The Aberrant Clades
5 Sirtuins and TM1506
6 A Synthetic Overview of the Tendencies in the Diversification of the ADP-Ribosylation System
7 Conclusion
References

1 Introduction

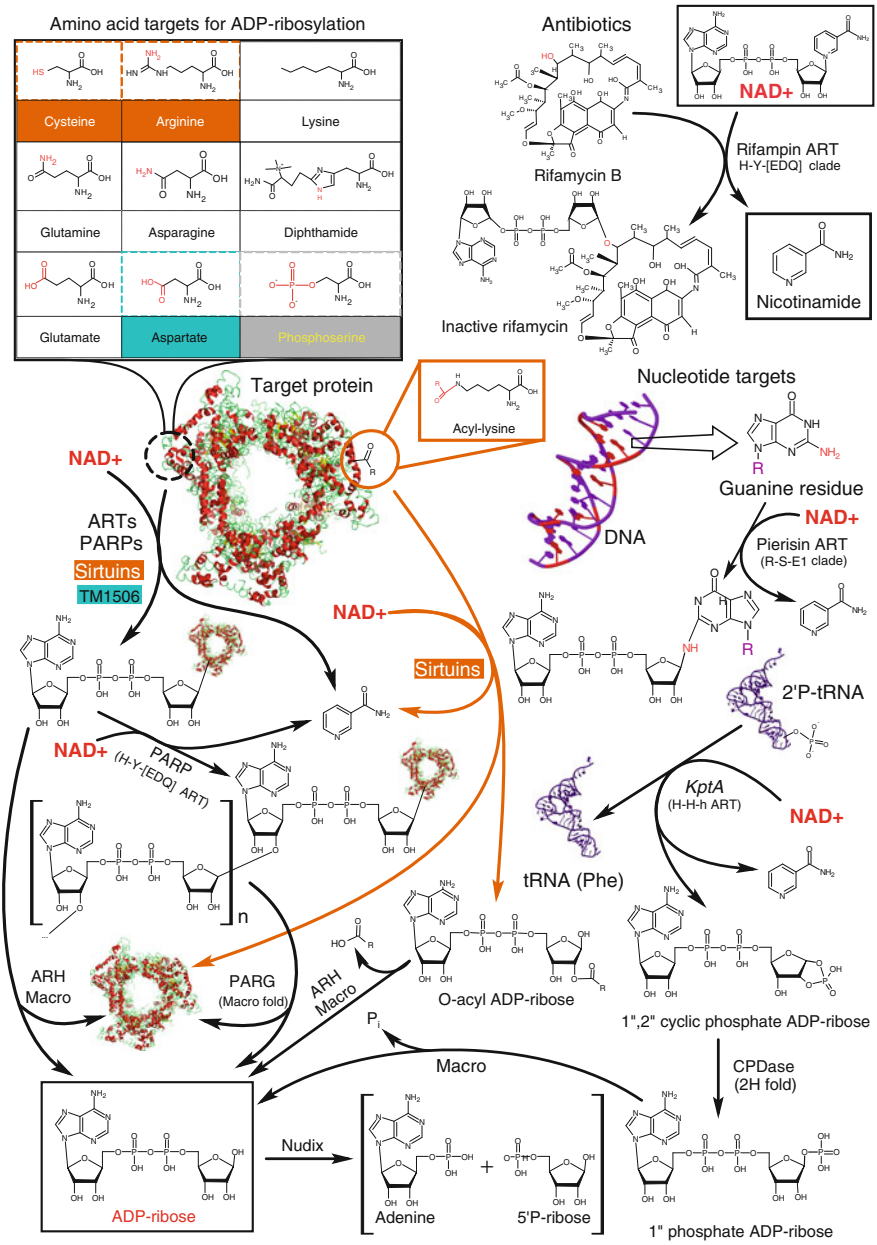
Nicotinamide adenine dinucleotide (NAD^+) is a versatile metabolite that is at the center of a large array of biochemical processes across all domains or superkingdoms of life (Belenky et al. 2007; Gazzaniga et al. 2009; Sorci et al. 2010). Its most widespread and best understood role is that of a redox cofactor (along with its phosphorylated derivative $NADP^+$) for diverse dehydrogenases of the Rossmann fold, which utilizes the facile transition between the hydrogenated and dehydrogenated states of the nicotinamide ring (Berg et al. 2012). The role of NAD^+ as a substrate for diverse protein- and nucleic acid-modifying enzymes has also become increasingly apparent in a striking array of subcellular contexts. In these latter reactions, rather than undergoing reversible oxidation state changes, different fragments of the NAD^+ molecule are transferred to substrate molecules resulting in specific modifications. One of these, catalyzed by the DNA ligase, which is observed in all bacteria, is the transfer of the adenylate moiety from NAD^+ to the 5′ phosphate of DNA (Park et al. 1989). This adenylated DNA serves as an intermediate for the ligation of polynucleotides. Another class of transfer reactions targets the bond between the nicotinamide ring and the ADP-ribose moiety transferring the latter moiety to targets with the release of nicotinamide. Such ADP-ribose transfer reactions are catalyzed by at least three evolutionarily unrelated superfamilies of enzymes, namely the ADP-ribosyltransferase (ART), the Sirtuins, and the obscure TM1506 superfamilies (Corda and Di Girolamo 2003; Hassa et al. 2006; Hawse and Wolberger 2009; Iyer et al. 2011).

2 The Many Faces of ADP-Ribosylation

ADP-ribosylation catalyzed by the ART superfamily first came to light in two independent studies in the 1960s, which respectively identified a novel biopolymer in vertebrate cells (Chambon et al. 1963) and showed the need for NAD^+ for the action of the diphtheria toxin produced by *Corynebacterium diphtheria* (Collier and Pappenheimer 1964). A flurry of investigations in the next 5 years showed that the common theme unifying these disparate activities was the enzymatic conjugation of ADP-ribose (hereinafter ADPR) derived from NAD^+ to proteins and/or other ADPR moieties (Sugimura and Miwa 1994; Ueda and Hayaishi 1985). Over the next 20 years it became firmly established that NAD^+-dependent ADP-ribosylation was a widespread phenomenon across diverse lineages of life as well as viruses (Sugimura and Miwa 1994; Ueda and Hayaishi 1985). Among the key issues that became clear was that this modification came in two basic forms: polyADP-ribosylation, where long-branched polymers of ADP-ribose are generated and added to target proteins, and mono-ADP-ribosylation, which involved transfer of a single ADP-ribose moiety to the target (Sugimura and Miwa 1994; Ueda and Hayaishi 1985). It was also established that this modification targets several distinct amino acid side chains and termini resulting in the ribose being linked via *N*-, *O*- or *S*- glycosidic bonds (Corda and Di Girolamo 2003; Laing et al. 2011) (Fig. 1).

It was also seen that these modifications catalyzed by ARTs were at the center of a wide range of biological phenomena. Of these, the first to be described in detail were the mono-ADP-ribosylation activities of diverse bacterial toxins (Corda and Di Girolamo 2003; Ueda and Hayaishi 1985). Around the same time, it also became apparent that the bacteriophage T4 encodes one or more ARTs, which are deployed to mono-ADP-ribosylate host proteins, such as the RNA polymerase α subunit (Koch and Ruger 1994; Ueda and Hayaishi 1985). In parallel, there was also accumulating evidence for the roles of cellular ARTs. For example, the ART-ADP-ribosyl glycohydrolase pair, DraT and DraG was found to alternately modify or demodify dinitrogenase reductase and regulate its activity (Ludden 1994). In more recent times the roles of such ARTs and polyADP-ribosyltransferases (PARPs/PARTs) have become increasingly evident in diverse eukaryotic processes (Hassa et al. 2006). Modifications of nuclear proteins (e.g., histones) by PARPs, cytoskeletal (e.g., actin) by intracellular ARTs and surface proteins (e.g., nucleotide-gated ion channels and integrins) by ecto-ARTs play important roles in epigenetics, regulation of DNA repair, apoptosis, signaling, and complex biological processes, such as long-term memory in animals (Cohen-Armon et al. 2004; Hassa et al. 2006; Koch-Nolte et al. 2008).

Sirtuins, whose activities have come to light relatively recently, were first identified as protein ADP-ribosyltransferases (Frye 1999). However, this activity has not been much studied; rather they are best known for their action as protein deacylases—in these reactions they transfer the ADPR from NAD^+ to acylated lysines in proteins (Smith et al. 2002). As a consequence, the acyl group is removed in the form of an *O*-acylated ADP-ribose (OAADPR; Fig. 1). In addition


Amino acid targets for ADP-ribosylation
Cysteine
Arginine
Lysine
Glutamine
Asparagine
Diphthamide
Glutamate
Aspartate
Phosphoserine
Antibiotics
NAD+
Rifampin ART
H-Y-[EDQ] clade
Rifamycin B
Inactive rifamycin
Nicotinamide
Target protein
Acyl-lysine
Nucleotide targets
Guanine residue
NAD+
Pierisin ART
(R-S-E1 clade)
DNA
ARTs
PARPs
Sirtuins
TM1506
NAD+
Sirtuins
2'P-tRNA
NAD+
PARP
(H-Y-[EDQ] ART)
KptA
(H-H-h ART)
NAD+
tRNA (Phe)
ARH
Macro
PARG
(Macro fold)
ARH
Macro
O-acyl ADP-ribose
1",2" cyclic phosphate ADP-ribose
Pi
Macro
CPDase
(2H fold)
ADP-ribose
Nudix
Adenine
5'P-ribose
1" phosphate ADP-ribose

◀ **Fig. 1** Summary of known ADP-ribosylation targets and pathways. Most substrates are tagged by ADP-ribose by different members of the ADP-ribosyltransferase fold (ARTs and PARPs). Cysteine and arginine residues (*orange*) may also serve as substrates for reactions catalyzed by members of the sirtuin family and aspartate (*turquoise*) is possibly a substrate of the TM1506 family. The enzyme responsible for phosphoserine (*gray*) ADP-ribosylation remains unknown. At the bottom of the figure, some key reactions for processing ADP-ribose derivatives are represented. *ART* ADP-ribosyltransferase, *PARP* poly(ADP-ribosyl) polymerase, *CDPase* 2′-cyclic phosphate hydrolase, *KptA/Tpt1* RNA 2′-phosphotransferase (ART fold), *Pierisin* lepidopteran cytotoxin (ART fold), *ARH* ADP-ribosylglycohydrolase, *PARG* poly(ADP-ribosyl)-glycohydrolase, *Macro* ADP-ribose associated Macro domain, *Nudix* "nucleotide diphosphate linked to X" hydrolase domain. Sirtuins might deacylate proteins with several distinct acyl groups (show as R-), such as acetyl, malonyl, crotonyl, succinyl or palmitoyl groups

to ARTs and sirtuins, a variety of other enzymes have been characterized as belonging to the ADPR-centered systems, such as ADP-ribosyl glycohydrolases (ARHs), which deconjugate ADP-ribose from targets (Koch-Nolte et al. 2008), polyADP-ribosyl glycohydrolases (PARGs) which hydrolyze ADPR polymers (Slade et al. 2011), ADP-ribosyl cyclase/cyclic-ADP-ribose hydrolase (e.g., CD38) (Guse and Lee 2008) and binding domains, which recognize ADPR and its derivatives generated by the action of the above enzymes (de Souza and Aravind 2012). In the past two decades, the advent of genome sequencing, structural genomics, and development of sensitive sequence analysis methods have greatly advanced our understanding of these proteins: First, new sequence and structure analysis methods have helped identify new members using sensitive similarity searches (Bazan and Koch-Nolte 1997; de Souza and Aravind 2012; Fieldhouse et al. 2010; Otto et al. 2005; Pallen et al. 2001). Second, they have given us an unprecedented window into the evolution of these enzymes and helped us understand their origins (de Souza and Aravind 2012; Zhang et al. 2012). Finally, they have also opened the door for new approaches to the understanding of their functions in experimentally difficult or less-studied organisms (e.g., algae such as *Emiliania*, apicomplexan parasites like *Plasmodium*, trypanosomes, and bacterial pathogens such *as Mycobacterium tuberculosis*) (Garcia-Salcedo et al. 2003; Merrick and Duraisingh 2007).

In this chapter, we provide an overview of the biochemical diversity and evolutionary history of ADPR-transferring enzymes in light of recent advances in genomics and computational sequence analysis. We also summarize new functional insights and possible new directions offered by these studies.

3 The ART Superfamily

3.1 Biochemical Diversity in the ART Superfamily

Members of the ART superfamily (hereinafter ARTs) can catalyze the transfer of ADPR from NAD^+ to diverse substrates resulting in *N*- *O*- or *S*- glycosidic linkages with the 1″ position of ribose (Fig. 1). This reaction proceeds via the SN1

mechanism: the nicotinamide leaves first with the generation of a reactive oxacarbenium ion as an intermediate, which is presented for nucleophilic attack by the target moiety at the 1″ position (Jorgensen et al. 2008a). The transfer is accompanied by an anomeric inversion of the newly formed glycosidic bond at the 1″ position with respect to the linkage of nicotinamide in NAD^+ (Ueda and Hayaishi 1985). The reaction mechanism, wherein nicotinamide leaves first, is consistent with the proposal that some members of the ART superfamily such as the halovibirins and the streptococcal NADase (SPN) toxins act primarily as NAD^+-glycohydrolases (Smith et al. 2011; Stabb et al. 2001). However, in the former case the released ADPR is noncatalytically conjugated to arginines in proteins (Stabb et al. 2001), suggesting that even the latter enzymes could potentially use the same mechanism to generate products effectively same as *bona fide* ARTs. In contrast to enzyme superfamilies catalyzing most other protein modifications, ARTs modify a striking variety of side chains and terminal positions in proteins (Fig. 1) (Corda and Di Girolamo 2003; Hassa et al. 2006; Koch-Nolte et al. 2001; Laing et al. 2011; Ueda and Hayaishi 1985). To date, ARTs have been shown to modify: (1) the terminal NH_2 of the guanidino group of arginine (e.g., the bacterial DraT which ADP-ribosylates the dinitrogen reductase, the T4 proteins Alt, ModA and ModB, several bacterial toxins directed at eukaryotic hosts such as the cholera, *Bacillus sphaericus* mosquitocidal toxin (MTX), *Bacillus cereus* VIP2, *Clostridium difficile* CDT and *C. perfringens* iota toxins, as well as eukaryotic mono-ARTs (ARTC1-ARTC5) modifying endogenous proteins such as actins, integrins, and nucleotide-gated ion channels); (2) The intracyclic NH in diphthamide, which is a highly modified amino acid derived from histidine in eukaryotic EF2 (e.g., the diphtheria and *P. aeruginosa* toxin A); (3) The NH_2 of the amide group of the glutamine side chain (e.g., the *Photorhabdus luminescens* ART toxin); (4) The NH_2 of the amide group of the asparagine side chain (e.g., the *Clostridium botulinum* C3 toxin); (5) The SH group in the side chain of cysteine (e.g., the pertussis toxin which targets cysteines in multiple vertebrate Gα proteins); (6) The COOH group of glutamate [e.g., PARP10/ARTD10 ADP-ribosylates itself on glutamate 882 (Kleine et al. 2008)]. The COOH group of the aspartate side chain has also been proposed as a potential mono-ART substrate in eukaryotes, although to date the responsible enzyme remains unidentified (Hassa et al. 2006). Early studies also detected ADP-ribosylation of phosphoserine in Histone H1 but again the enzyme remains unknown (Smith and Stocken 1974). Furthermore, given this diversity it is conceivable that there might be ART domains which modify side chains of serine or threonine, especially among versions found in bacterial toxins.

In the case of PARPs, successive ADPR chains might be added via *O*-glycosidic linkages between the 2′ or 2″ positions and 1″ positions of successive ADPR units. As each unit has two ribose moieties successive links to either of them can result in large branched molecules with up to 200 units that can be visualized via electron microscopy (Hassa et al. 2006; Ueda and Hayaishi 1985). In eukaryotes, these poly ADPR (pADPR) molecules apparently modify a wide range of nuclear proteins including core and linker histones, chromatin proteins like HMGA/B, transcription factors like p53, RNA-binding proteins such hnRNPs and DNA

repair/replication factors like topoisomerases and PCNA (Hassa et al. 2006; Ueda and Hayaishi 1985). It is widely believed that in these substrate proteins the side chain COOH of glutamate, the terminal COOH and/or perhaps aspartate are modified by pADPR. However, recent studies have shown that the automodification of PARP1 (ARTD1) occurs via lysine epsilon-NH_2 groups (Altmeyer et al. 2009).

In addition to proteins, ARTs modify nucleic acids: cytotoxic ARTs of lepidopterans and molluscs modify the exo-cyclic amino group at the second position of guanine in DNA to induce apoptosis (Carpusca et al. 2006; Nakano et al. 2006). Another nucleic acid-modifying ART is the RNA 2′-phosphotransferase KptA/Tpt1 (ARTD18), an enzyme that repairs the 2′-phosphate generated as a result of tRNA splicing and RNA ligase action (Spinelli et al. 1999). Here, the ADPR is added to 2′ phosphate which then leaves as ADP-ribose 1″-2″ cyclic diphosphate (Appr>p). ARTs also modify low molecular weight substrates, such as the rifamycin antibiotics, which are inactivated by ADP-ribosylation of an OH group on carbon 23 (Baysarowich et al. 2008) (Fig. 1).

3.2 Structural and Catalytic Features of the ART Superfamily

The ART superfamily is characterized by extreme divergence at the sequence level often making it difficult to identify new members through sequence analysis alone (Bazan and Koch-Nolte 1997; de Souza and Aravind 2012; Fieldhouse et al. 2010). However, at the structural level all members of the superfamily show a strongly conserved core fold (Figs. 2 and 3) (Bazan and Koch-Nolte 1997; de Souza and Aravind 2012). This core consists of a split β-sheet formed by two distinct units of three strands each, with most strands being alternately interleaved between the two units (i.e., strand order in the sheet is 4-5-2|1-3-6, where "|" represents the split in the sheet; Fig. 2 bottom panel). Typically, most of these strands are followed by a downstream helical element suggesting that the entire fold emerged through a series of duplications of a strand-helix element with the strands being distributed alternately between two units of the split sheet. In most members of the family the key active site residues emerge from the "lower" surface of strand 1, and the "upper" surface of strands 2 and 5 (Fig. 3). As a result the NAD^+ molecule is sterically constrained between the lower and upper surfaces respectively of the two units of the split sheet with the central diphosphate being wedged in the central split in the sheet (Jorgensen et al. 2008a). The characteristic loop between helix 1 and strand 2 is observed throughout the fold and forms a "wall" of the NAD^+-binding pocket (Fig. 3). The active site residues from strand 1 typically contact the ribose linked to the adenine and also the nicotinamide moiety. Those from strands 2 and 5 contact the ribose linked to the nicotinamide. This conformational constraining of NAD^+ between the two elements of the split sheet appears to be critical for catalysis by these enzymes for it appears to strain the substrate to facilitate its cleavage.

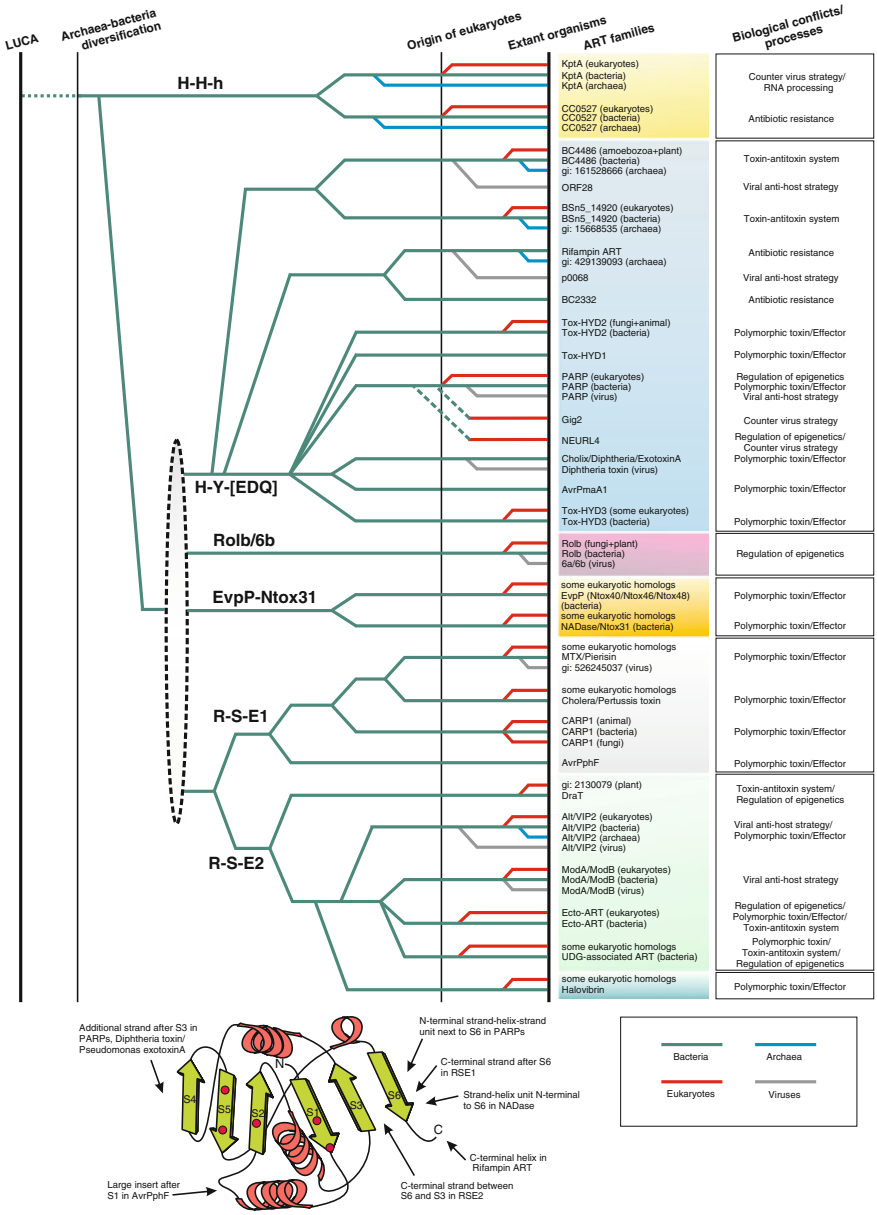


Fig. 2 Reconstructed evolutionary history for the different members of ART superfamily. Horizontal lines are colored according to their observed phyletic distributions (key: *bottom right*). *Dashes* indicate uncertainty in terms of the origins of a family, while the *gray ellipse* indicates that the RolB/6b clade likely underwent rapid divergence from either the H-Y-[EDQ] or R-S-E clade. The H-H-h and H-Y-[EDQ] include the ARTD proteins, while the R-S-E clade includes the ARTC proteins in the previously presented nomenclature for ART domains (Hottiger et al. 2010). For each family the known or inferred functional role in biological conflicts or processes is indicated (*right column*). Representatives of each family are shown in Figs. 4 and 5. In the bottom panel shows the idealized topology of the ART fold with markup indicating specific structural features associated with certain ART families and positions of active site residues

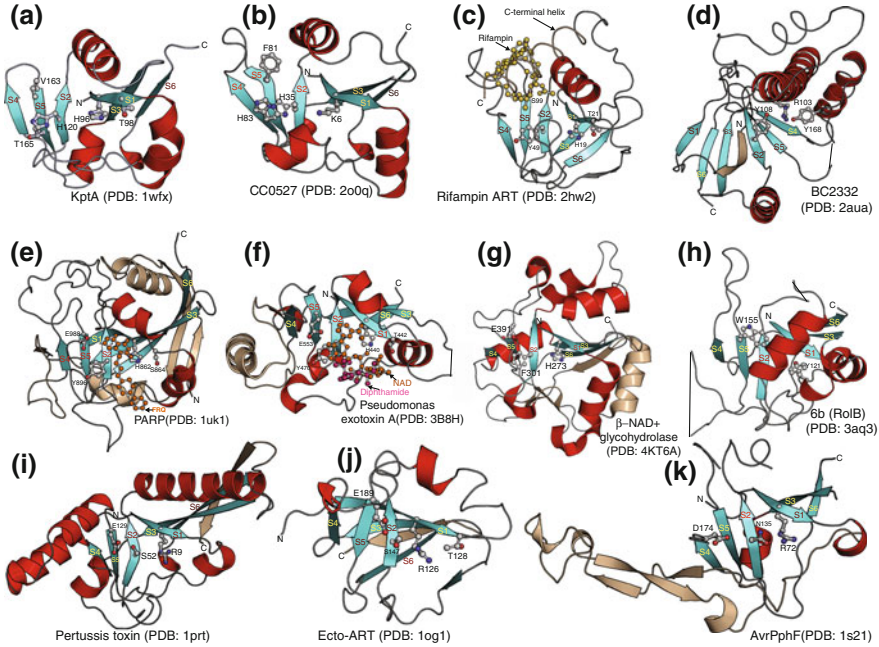


Fig. 3 Cartoon structures of diverse members of the ART superfamily. Conserved strands and helices of the core fold are shown in *aquamarine* and *red*, respectively. Additional strands that pack with the sheets and inserts are shown in *wheat color*. Conserved strands of the core are labeled, as also active site residues and ligands. **a** KptA (PDB: 1wfx). **b** CC0527 (PDB: 2o0q). **c** Rifampin ART (PDB: 2hw2). **d** BC2332 (PDB: 2aua). **e** PARP (PDB: 1uk1). **f** Pseudomonas exotoxin A (PDB: 3B8H). **g** β-NAD$^+$ glycohydrolyase (PDB: 4KT6A). **h** 6b (RolB) (PDB: 3aq3). **i** Pertussis toxin (PDB: 1prt). **j** Ecto-ART (PDB: log1). **k** AvrPphF (PDB: 1s21)

Examination of the active site residues reveals three widespread configurations which are described in terms of the consensus conserved residue supplied by strands 1, 2, and 5 of the core:

(1) *H-H-h*: This is seen primarily in two families, namely the KptA (ARTD18) or tRNA 2′ phosphotransferase family and the CC0527 family (Fig. 3a, b), which has been largely obscure prior to this chapter (see below). Here, the first strand supplies a conserved H, which is usually in a HX[ST] motif, where both the H and the downstream alcoholic residue might make polar contacts with the 2′ and 3′ OH of ribose groups. Strand-2 supplies a histidine and strand-5 has just a hydrophobic residue (h), which contacts the ribose linked to the nicotinamide. In these ARTs the hydrophobic residue in strand 5 is followed by an additional polar residue two residues downstream, typically [TS] in KptA or H in CC0527, which is also likely to contribute to the active site (Fig. 3a, b).

(2) *H-Y-[QED]*: This is seen in several families such as diphtheria toxin-like ARTs, PARPs, and related mono-ARTs (Figs. 2 and 3c–f). Domains with this

active site configuration and those with the above configuration have been previously designated as ARTD (Hottiger et al. 2010). Here, the strand 1 histidine occurs in the same motif as observed in the H-H-h configuration. The residue from strand 2 is an aromatic residue which is typically tyrosine and less frequently phenylalanine. That from the beginning of strand 5 is acidic or in some cases a glutamine and it typically makes a polar contact with the 2″ OH (Jorgensen et al. 2008a). It often occurs as the downstream residue in a [QE]X[QED] motif, with the first [QE] position playing a role in recognition of the target moiety that is ADP-ribosylated (Baysarowich et al. 2008; Jorgensen et al. 2008a, b). A variation on this theme is seen in the rifamycin ART family, where the strand 5 displays a TXSXR motif, where the TXS maps to the [QE]X[QED] in the above motif (Fig. 3c). Similarly, in a related novel family of ARTs prototyped by the version from the *Acinetobacter* phage ZZ1 (p0068; gi: 570033484) the acidic residue is often absent, being replaced by a tyrosine.

(3) *R-[ST]-E*: This configuration is seen in several bacterial toxins, like the heat labile toxin of *E. coli*, the cholera toxin, bacteriophage ARTs, like Alt, ModA and ModB, and certain eukaryotic intracellular (e.g., molluscan CARP-1) and extracellular ARTs (Fig. 3). Domains with this active site configuration have been previously designated as ARTC (Hottiger et al. 2010). In this configuration, the arginine plays the equivalent role as the histidine in strand 1 and the [ST] plays a role comparable to the aromatic residue (Y) in strand 2. The E from strand 5 is identical in function to the equivalent residue in those showing the H-Y-E configuration (Figs. 2 and 3).

(4) *Aberrant configurations*: Several of members of the ART superfamily show highly derived active site configurations that do not match any of the above patterns. The previously unrecognized clade of ARTs typified by the *Edwardsiella tarda* EvpP toxin (Zheng and Leung 2007) and the SPN toxins (Smith et al. 2011) (see below) displays either an arginine (rarely histidine; e.g., NADase) or no specific conserved residue in strand 1, an aromatic position (usually Y or F) equivalent to the Y or [ST] in strand 2 and a glutamate/glutamine at the beginning of strand-5. Thus this clade displays a hybrid between the H-Y-[QED] and R-[ST]-E configurations. The RolB family of ARTs prototyped by the *Agrobacterium* Ti plasmid (T-DNA) 6b-related proteins (Wang et al. 2011) represents another aberrant configuration (Fig. 3h), where most residues corresponding to any of the above configurations are not conserved. The only exception is a highly conserved aromatic residue from strand-5 that corresponds to the hydrophobic residue in the H-H-h configuration (Fig. 3h). This residue along with another residue (typically tyrosine or arginine) from the helical insert upstream of strand 3 is likely to constitute a novel active site in this family. Another novel family, the BC2332 (Pfam: DUF2441) family has undergone an even more drastic modification, both in structural and sequence terms (Fig. 3d). It appears to have emerged via triplication of the ancestral ART fold followed by reconstitution of a domain by elements drawn from each of the three repeats. As a

consequence, the strand order is re-organized, now adopting a 4-1-2-6-3-5 linear sequence from the N- to C-terminus, where the numbers stand for the strand numbers in the typical ART domain. This is concomitantly accompanied by the emergence of a completely new active site that is unrelated to all other members of the fold (Fig. 3d). This pervasive plasticity in the active site of the ARTs lends further support to the proposal that the conformation in which NAD^+ is "wedged" in the split sheet is probably more critical for the catalytic mechanism than the strict presence of particular residues.

The minimal version of the fold with just six strands and no extended inserts between the elements is seen only in the KptA and CC0527 families (Hottiger et al. 2010). The core split β-sheet is often augmented by N-terminal (e.g., RolB family) (Wang et al. 2011) or C-terminal extensions (e.g., diphtheria toxin-like ARTs) or internal insertions (e.g.,an inserted strand after strand 3 in PARPs) (Hottiger et al. 2010). A subset of ARTs showing the R-S-E configuration shows an additional C-terminal strand 7, which forms a hairpin with core strand-6 and stacks against strand 3 in a parallel configuration (Fig. 3j). The rifamycin ART family shows a C-terminal helix which plays a key role in substrate recognition by forming a "cap" over the active site (Baysarowich et al. 2008). Additionally, several versions might show inserted elements. The characteristic loop between helix 1 and strand 2 is a frequent site for independent insertions in different families and some of these inserts play a role in substrate recognition (e.g., rifamycin and diphtheria toxin-like ART families). Some PARPs display an insert of a Zn-cluster just N-terminal to strand 3, which might also play a role in substrate recognition (Hottiger et al. 2010). The dramatic structural reorganization in the BC2332 family has resulted in the emergence of unprecedented large inserts between some of the elements drawn from the different copies of the above mentioned triplication. Thus, emergence of lineage-specific inserts appears to have played an important role in the origin of distinct substrate specificities in the superfamily.

4 Evolutionary History and Diversity of the ART Superfamily

Availability of multiple structures from distant members of the ART superfamily, sensitive sequence profile searches to identify novel versions, and determination of phyletic patterns have resulted in a new and detailed understanding of the evolutionary history of the ART superfamily (Bazan and Koch-Nolte 1997; de Souza and Aravind 2012; Finn et al. 2010; Schaffer et al. 2001; Soding et al. 2005). In particular the above-described structural features and active site configurations help establish the higher-order relationships between distantly related families (Fig. 2). Within families, higher levels of sequence conservation allow establishment of relationships based on conventional phylogenetic analysis (Price et al. 2010).

By combining these, we present below an evolutionary reconstruction reflecting the current state of the data (Fig. 2).

Broadly, the ART superfamily can be divided into three major higher-order clades that more-or-less reflect the above-described three basic active site configurations. The most primitive or earliest branching clade appears to be the H-H-h clade followed by a crown group comprised of the H-Y-[EDQ] and the R-S-E clades (Fig. 2). Thus, the presence of an active site H in the first strand appears to be the ancestral feature of the ART superfamily, with the R in the R-S-E clade being a derived feature. The acidic residue at the beginning of the strand 5 appears to be a shared derived character unifying the crown clades. Below, we briefly describe the various constituent families of within each of these major clades.

4.1 The H-H-h Clade

This is the most primitive or basal clade of the ART superfamily as indicated by an unelaborated structure and active site lacking an acidic position in strand 5 (Figs. 2 and 3a, b). This clade contains only two families the KptA and CC0527 (Pfam: DUF952) which are unified by above-described active site features associated with strand 5. Of these, KptA represents the most widely distributed lineage of the ART superfamily and is typically present in a single copy in organisms that encode it (de Souza and Aravind 2012). This family is seen in most eukaryotes barring the basal-most branches (parabasalids and diplomonads), several euryarchaeal and crenarchaeal clades and representatives of most major bacterial lineages albeit in a somewhat patchy distribution. This widespread phyletic pattern suggests that KptA could have potentially been present in the last universal common ancestor (LUCA). However, phylogenetic analysis suggests that both the archaeal and eukaryotic branches are nested within the bacterial radiation making it possible that it actually emerged in bacteria and was independently transferred to the former two lineages (Fig. 2). Nevertheless, its wide distribution and structural simplicity suggest that it is close to the ancestral version of the entire ART superfamily and that other ARTs were likely derived from it by extensive and rapid divergence. In eukaryotes and archaea, it appears to be part of the core RNA-processing mechanism required for tRNA maturation following removal of the intron (Spinelli et al. 1999). Interestingly, knockout of the sole mouse representative of this family Trpt1 did not apparently result in RNA-processing defects (Harding et al. 2008), suggesting that despite its strong conservation in eukaryotes it might be a nonessential gene in mammals. However, its function in bacteria is less understood. Its presence in operons where it is combined with the RNA repair $5' \rightarrow 3'$ polymerase Thg1 suggests that it might be involved in RNA repair, perhaps as strategy against bacteriophage or toxin attack on tRNAs (de Souza and Aravind 2012) (Fig. 4). However, in several bacteria it occurs as part of large NAD^+/ADPR metabolism operon combined with genes for several other enzymes which use these metabolites, such as sirtuins, Macro, PARG (also belonging to Macro

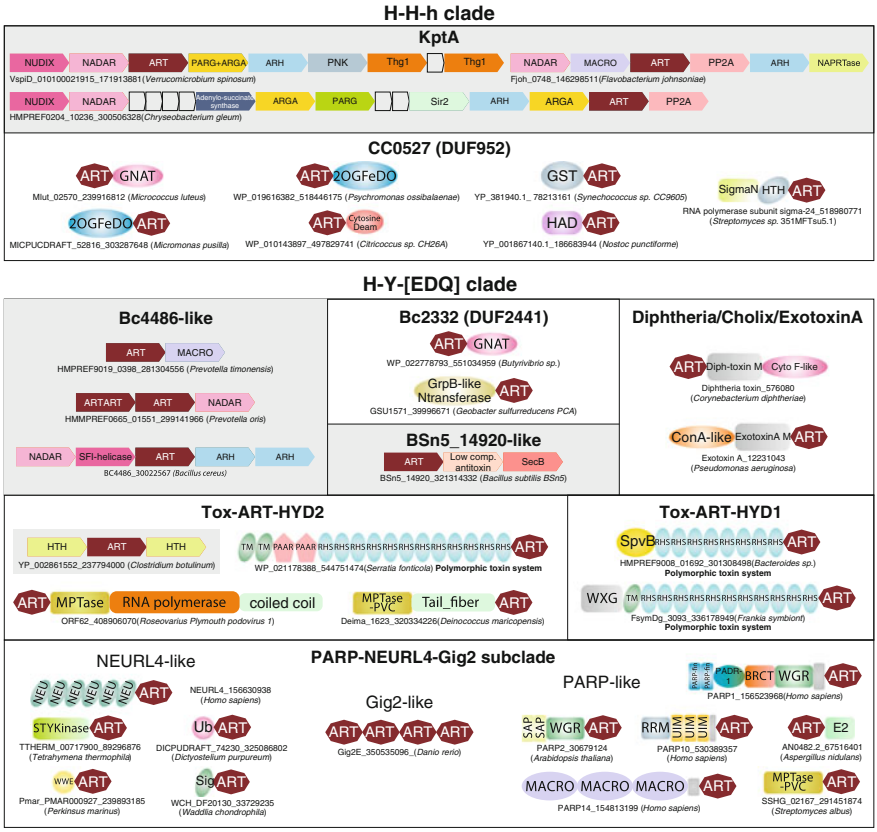


Fig. 4 Domain architectures and gene neighborhoods of ARTs of the H-H-h and H-Y-[EDQ] clades. Gene neighborhoods and domain architectures are labeled with the gene name, Genbank index (gi) number and the species name, which is shown in brackets. For domain architectures, the label corresponds to the ART containing gene. Gene neighborhoods are shown as *box-arrows* with the *arrowhead* pointing to the gene in the 3′ direction of the coding strand. These are further shaded *gray*. The architectures are grouped based on the type of ART. Standard domain abbreviations are used to label genes and domains

superfamily), NADAR, and ARHs but no RNA-related genes (de Souza and Aravind 2012) (Fig. 4). This raises the possibility that a subset of the bacterial KptAs might modify substrates other than RNA.

The CC0527 family is widely but patchily distributed across the bacterial superkingdom. It appears to have also been independently transferred to several eukaryotes such as fungi, viridiplantae, stramenopile algae *Emiliania*, and *Naegleria* (Fig. 2). While its structure is available (PDB: 2o0q), its function has remained obscure. Examination of its domain architectures reveals fusions to glutathione S-transferases, GCN5-like NH_2 acetyltransferases (GNAT superfamily), cytosine deaminases, or 2 oxoglutarate/Fe-dependent dioxygenases (Fig. 4).

All these associated enzymes have been previously implicated in modification and detoxification of antibiotics by respectively adding, glutathione and acetyl groups, deaminating cytosine in Blasticidin S-like compounds and hydroxylation (Ramirez and Tolmasky 2010; Walsh 2003). Thus, by the principle of "guilt by association" (Aravind 2000) it is likely that the CC0527 ADP-ribosylates antibiotics or other toxic compounds in a reaction comparable to the rifamycin ARTs.

4.2 The H-Y-[EDQ] Clade

The primary and explosive diversification of this clade appears to have occurred in the context of several prokaryotic conflict systems (Zhang et al. 2012). These include intragenomic conflict systems such as type-II Toxin–Antitoxin (T–A) systems, intergenomic conflicts between bacteriophages and their host genome, olecule toxins or antibiotics, and interorganismal conflict using proteinaceous effectors (Aravind et al. 2012; Zhang et al. 2012). Indeed, in each of these conflict systems ARTs of this clade have been extensively deployed as toxins to target proteins (Fig. 2). Three families belonging to this clade are encountered in T–A systems that are widely distributed across the bacterial tree (Fig. 4). The first of these is the recently described BC4486 family (prototyped by BC4486 from *Bacillus cereus*), which is typically found in a two gene T–A system, where the second gene encodes either a member of the NADAR superfamily (BC4488) or a member of the Macro superfamily (de Souza and Aravind 2012). These two genes follow a strict gene order with that encoding the BC4486 family ART always occurring as the first gene in the operon (Fig. 4). Based on this, it has been suggested that the BC4486 protein functions as an ADP-ribosylating toxin while the Macro or the NADAR protein function as antitoxins that counter the former's action by removing the ADP-ribose moiety. Thus, they might be seen as analogous to the DraG-DraT system (see below). Also belonging to this family is a subfamily of ARTs prototyped by ORF28 (gi: 46309426) of *Agrotis segetum* granulovirus, which is encoded by certain arthropod DNA viruses such as baculoviruses, ascoviruses, and iridoviruses (Fig. 2).

The second of the T–A-associated families, the BSn5_14920 family (prototyped by the eponymous protein from *Bacillus subtilis* BSn5) is encoded by a 3-gene T–A wherein the ART is again encoded by the first gene (Fig. 4). This is followed by a gene for the antitoxin which assumes the form of a low complexity or intrinsically disordered protein with two highly conserved hydrophobic residues at the C-terminus and the third gene encoding the chaperone SecB (Sala et al. 2013). In these systems, the intrinsically disordered antitoxin is postulated to bind SecB by the C-terminal hydrophobic residues, while SecB stabilizes it, allowing it to inhibit the toxin ART. If under protein unfolding stress conditions the SecB is drawn away toward other unfolded proteins, then the antitoxin is destabilized (Sala et al. 2013), unleashing the action of the ART, which presumably induces dormancy by modifying certain targets. The third T–A-associated family is the

Tox-ART-HYD2 family prototyped by a version from a *Clostridium botulinum* T–A system (gi: 237794000; Fig. 4). In these, the postulated toxin ART is typically accompanied by two predicted antitoxins, which are DNA-binding helix-turn-helix proteins (HTH) proteins, encoded by flanking genes (Fig. 4). Outside TA systems, ARTs of this family are also encountered in bacterial polymorphic toxin systems with variable C-terminal domains that are likely to be used in intraspecific conflicts by bacteria (e.g., gi: 544751474 from *Serratia fonticola* (Zhang et al. 2012)). Here, the ART occurs as one of the variable C-terminal toxin tips preceded by a long stretch of RHS repeats which help present the toxin on cell-surfaces of bacteria. Other members occur in phages such as *Roseovarius* Plymouth podovirus 1(gi: 408906070) where they are fused to the viral RNA polymerase and might help regulate transcription in course of infection (Fig. 4). Members of this family have also been transferred to and expanded in eukaryotes such as *Acanthamoeba* (gi: 470494435), choanoflagellates, and the fungal plant pathogen *Puccinia*. The Tox-ART-HYD1 family is another ART family with comparable features, which is found in several polymorphic toxin systems from several bacterial lineages (Fig. 4) (Zhang et al. 2012).

The family typified by the p0068 protein of the *Acinetobacter* phage ZZ1 appears to be involved in phage-host conflicts and is found in several bacteriophage genomes, including a lineage-specific expansion in *Bacillus* phage G (10 paralogs) (de Souza and Aravind 2012). These phage versions are predicted to act similarly to the T4 and N4 phage ARTs (see below for details) in modifying host proteins to facilitate viral replication (Koch and Ruger 1994; Tiemann et al. 2004). Cellular versions are found in cyanobacteria and the nanohaloarchaeon *Halorediviyus*. This family is further related to another family, rifamycin ART (Figs. 2 and 3c), which is implicated as a defensive element in interorganismal conflict by neutralizing rifamycin antibiotics (Baysarowich et al. 2008). The spread of this family across bacteria through lateral transfer appears to have been a key determinant of resistance toward this class of antibiotics. The structurally highly modified BC2332 family is likely to have been derived from this assemblage and its domain fusions and operonic associations (Fig. 4) are also suggestive of a role comparable to the rifamycin ART family in detoxification of small molecules. Most families discussed thus far are characterized by more or less a minimal ART domain with few or no additional structural inserts. This suggests that they are likely to represent the basal-most lineages within the H-Y-[EDQ] clade (Fig. 2).

The remaining families of this clade appear to constitute a monophyletic assemblage, which appears to have initially diversified in toxin systems that are used as effectors by bacteria in interorganismal conflicts (Fig. 2). Of these, the PARP-NEURL4-*Gig2* subclade is a major monophyletic assemblage that includes several bacterial effectors, such as one encoded by the intracellular chlamydia-like pathogen of animals and amoebae, *Waddlia chondrophila* (gi: 337292305) (de Souza and Aravind 2012). PARP family includes both *bona fide* PARPs (e.g., human PARP1-3/ARTD1-3, vPARP/ARTD4, and the tankyrases/ARTD5-6) and related mono-ARTS (ARTD7-17; also called Pl-MARTs) (Citarelli et al. 2010; Hassa et al. 2006; Otto et al. 2005). The eukaryotic versions are closely related to

effectors encoded by bacteria such as *Legionella drancourtii* (gi: 363538754) and *Vibrio caribbenthicus* (gi: 497289311) and appear to have been derived from such precursors (Zhang et al. 2012). At least two PARP family members, including the histone-modifying PARP1 can be reconstructed as being present in the last eukaryotic common ancestor LECA (Citarelli et al. 2010). In course of eukaryotic evolution they expanded extensively to spawn several distinct lineages such as the telomere-protein-modifying tankyrases and the vPARP, which is a subunit of the vault, an organelle associated with small noncoding RNAs. Thus, the PARP family was incorporated into a diverse set of eukaryotic systems ranging from epigenetic regulation to control of RNA-processing and chromosome dynamics (Hottiger et al. 2010). This was accompanied by fusion to several distinct domains such as the PARP-Zn-finger, WGR (also of bacterial origin) and SAP domains, all involved in DNA binding, and the phosphopeptide binding BRCT domain in *bona fide* PARPs and domains such as the E2 ubiqutin ligase, peptide-binding WWE, ubiquitin-binding UIM and RNA-binding RRM domains in the case of the mono-ARTs from this family (Fig. 4). The eukaryotic *Gig2* family (Jiang et al. 2009) shows lineage-specific expansions in fishes, *Oxytricha trifallax*, and the haptophyte alga *Emiliania huxleyi* and includes proteins which may possess multiple tandem ART domains (Fig. 4). In fishes, it has been shown to be induced by interferon as part of the antiviral response (Jiang et al. 2009; Krasnov et al. 2011). Hence, these ARTs might have been recruited by eukaryotes to counter viruses by modifying their proteins. The NEURL4 family typified by the eponymous human protein is found in most animal lineages and several more basal eukaryotic lineages (de Souza and Aravind 2012). They are typically fused to other domains such as neuralized repeats, ubiquitin, WWE, and MORN-repeats and have undergone independent lineage-specific expansions in the fungus *Rhizophagus irregularis*, slime molds, sponges, crustaceans, amphioxus, and ciliates (Fig. 4). While NEURL4 itself has been implicated in centrosomal assembly (Al-Hakim et al. 2012), these expanded versions might have roles in antiviral responses similar to the *Gig2* family. Another more divergent family in this subclade is the Tox-ART-HYD3 family, which is present as the toxin tip of a recently identified class of bacterial cell surface effectors (Fig. 2; DZ, LMI, LA unpublished).

The large assembly of effectors deployed by both animal and plant-pathogenic bacteria comprises the final major family in the H-Y-[EDQ] clade (Fig. 2) (Corda and Di Girolamo 2003; Hassa et al. 2006; Koch-Nolte et al. 2001). The precise relationships in this family are difficult to discern due to their rapid sequence divergence as a consequence of adaptation against host immune responses and the selection for targeting different host proteins. One distinct monophyletic lineage within this group includes the diphthamide-targeting ARTs such as diphtheria toxin, the *Vibrio cholerae* cholix toxin, and the *P. aeruginosa* Exotoxin A. A more divergent version of this group is the ART domain of the AvrPm1 of the plant pathogen *Pseudomonas syringae* (Rohmer et al. 2003). This assemblage of toxins might have originated from phage ARTs used in intergenomic conflicts as suggested by the fact that several are encoded by temperate phages (Boyd 2012).

4.3 The R-S-E Clade

The diversification of the R-S-E clade closely parallels that of the above clade in that it primarily occurred in the context of the same bacterial conflict systems as those discussed above. On structural grounds the R-S-E clade can be divided into two major subclades (Figs. 2 and 3i–k). The first retains the ancestral state of the split β-sheet of the ARTs, with the core 6 strands. The second subclade is characterized by the above-described additional seventh strand that is spatially inserted between the core strand 3 and strand 6 (the additional strand subclade).

The first subclade contains five families of bacterial effectors deployed both in intraspecific conflict (polymorphic toxins) and against more distantly related organisms and hosts (Fig. 2). The first of these families is the MTX/Pierisin family, which is typified by the arthropod-killing toxin deployed by *Bacillus sphaericus* and the apoptotic pierisin-like cytotoxin from lepidopterans (Carpusca et al. 2006; Nakano et al. 2006; Takahashi-Nakaguchi et al. 2013). Interestingly, this family has evolved a wide range of substrate specificities which include both protein and DNA. The close relationship between lepidopteran cytotoxins like pierisin and the insecticidal effectors encoded *Paenibacillus larvae* temperate bacteriophage phiIBB_Pl23 (gi: 526245037; Fig. 5) suggest they were acquired by insects from their bacterial symbionts/pathogens and incorporated in their cellular systems as mediators of apoptosis or as defensive toxins against parasitoids (Takahashi-Nakaguchi et al. 2013). Similar to the lateral transfers to insects, members of this family have also been transferred to several plants such as the dicot *Medicago truncatula* and the grass *Aegilops tauschii*, where they appear to be seed-associated albumins which might play a role in defense against seed-eating animals (Fig. 5). Members of this family are also found as toxin tips of polymorphic toxin-like proteins such as a toxin from *Pseudomonas fluorescens* (gi: 515906858), where it occurs C-terminal to series of RHS repeats (Fig. 5). Two other closely related families in this subclade are the cholera toxin-heat-labile enterotoxin A (CTXA/LTA) and the pertussis toxin (PTX) families, which are prototyped by the eponymous proteins (Fig. 2) (Corda and Di Girolamo 2003; Laing et al. 2011). The former family, widely disseminated across pathogenic bacteria, includes effectors of forms infecting both animals and plants (e.g., *Xanthomonas fuscans* gi: 495244509), and symbionts such as *Sodalis* (gi: 573023525) which colonizes the tsetse fly (Snyder and Rio 2013). Interestingly, members of the CTXA/LTA family also appear to have been transferred to the plant *Selaginella* (gi: 302792354), and several insect- and plant-pathogenic fungi (e.g., *Metarhizium* and *Colletotrichum* respectively), where they have undergone lineage-specific expansion (Fig. 5). These latter ART domains are on occasions fused to a C-terminal conserved, apparently enzymatic domain, which is also found fused to ricin-like RNA-targeting toxins of these fungi and in several plant proteins (e.g., gi: 629706939); we predict that they might function as secreted effectors deployed in defensive interactions (Fig. 5). PTX family shows a similar distribution in bacteria but is mainly limited to animal pathogens.

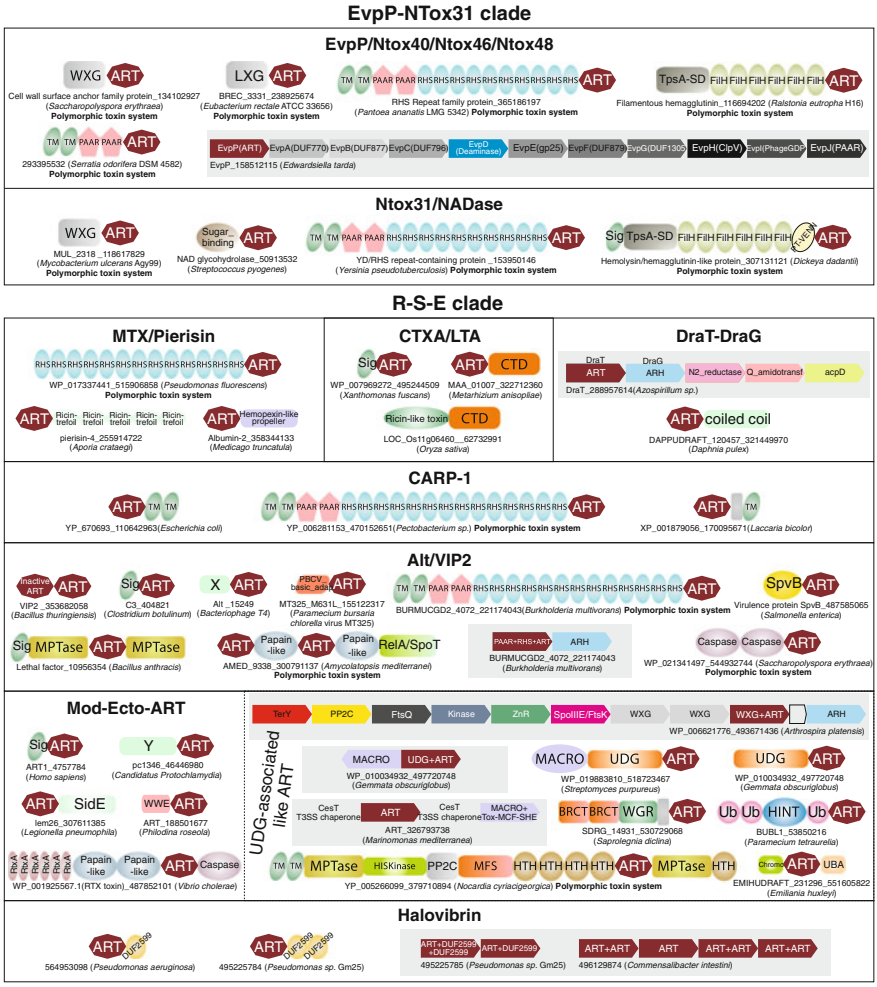


Fig. 5 Domain architectures and gene neighborhoods of ARTs of the R-S-E clade. The labeling scheme and grouping of proteins and gene neighborhoods are as described in Fig. 4. Standard domain abbreviations are used to label genes and domains. Additionally, X and Y are uncharacterized domains

The fourth family in the first subclade is typified by the molluscan DNA-modifying toxin CARP-1 from the clam *Meretrix lamarckii* (Nakano et al. 2006). We also found four members of this family in the oyster *Crassostrea gigas*, suggesting a more widespread presence of these proteins in bivalves. Members of the CARP-1 family are conserved in most filamentous fungi of both ascomycete and basidiomycete lineages (Fig. 5) and certain other eukaryotes such as the choanoflagellate *Monosiga* and the stramenopile *Phytophthora*. In bacteria, they are found as toxin tips of polymorphic toxins (Zhang et al. 2012) (e.g., gi:

470152651 from *Pectobacterium*; Fig. 5) or certain membrane-associated proteins (e.g., gi: 110642963 from *Escherichia coli* 536). The eukaryotic versions appear to have been acquired from the bacterial version by lateral transfer on multiple occasions. Interestingly, DNA-modifying activity has emerged in this family independently of the pierisin family. A divergent and distinctive family within the first subclade is the distinctive AvrPphF family that is prototyped by the eponymous effector of the plant pathogen *P. syringae* (Singer et al. 2004). This family is typified by a large insert after the first strand-helix unit (Fig. 3k). Moreover, other than the arginine from the first strand the remaining active site residues have undergone drastic modification: the S in strand 2 is replaced by a polar residue often asparagine or arginine while the acidic residue from strand 5 is entirely lost. Instead its place appears to be taken by a similarly positioned aspartate from the end of strand 4.

The second subclade represents another extensive radiation of the R-S-E clade and includes four major families (Fig. 2). The first, the DraT family, contains the structurally and architecturally least elaborated versions. These are represented by T–A systems prototyped by the DraT-DraG system, which itself appears to be a "domesticated" T–A system exapted for dinitrogen reductase regulation (de Souza and Aravind 2012; Ludden 1994). Here, the ART (DraT) acts as the protein-modifying toxin, while the ARH (DraG) acts as the demodifying antitoxin that reverses the action of the former. A member of this family has been horizontally transferred to the crustacean *Daphnia pulex* (gi: 321449970; Fig. 5), where it appears to have been incorporated into a much larger polypeptide. The second family is typified by the secreted halovibrins (HvnA and HvnB) of *Vibrio fischeri*, the extracellular symbiont that confers bioluminescence to the squid *Euprymna scolopes* by residing in its light organ (Stabb et al. 2001). This family is marked by a large central insert in the ART domain and conserved cysteine residues that are predicted to form two disulfide bridges. In addition to symbiotic and pathogenic bacteria (e.g., *Erwinia amylovora*, *P. aeruginosa*, and *P. syringae*) it also found in *Perkinsus marinus* the eukaryotic parasite of oysters and clams (Fig. 5), suggesting that modification via action of halovibrins might play a role in host interactions of multiple unrelated pathogens. The third family in this subclade is the large Alt/VIP2 family (Corda and Di Girolamo 2003; Han et al. 1999; Koch-Nolte et al. 2001; Laing et al. 2011). This includes phage T4 Alt which is packaged into the virion and deployed upon infection to hijack the host transcription machinery by modifying the RNA polymerase α subunit (Koch and Ruger 1994). Interestingly, certain eukaryotic DNA viruses, such as the *Paramecium bursaria Chlorella* virus (PBCV1_A092/093L) also encode an ART belonging to this family, which is packaged into the virion (Dunigan et al. 2012), suggesting that it might be deployed right after infection just as with the bacteriophage versions. Other members are toxins deployed by various animal pathogenic bacteria such as *B. cereus* VIP2, *C. perfringens* iota, *C. botulinum* C2 and C3, *B. anthracis* lethal factor, *Aeromonas hydrophila* VahC and *Salmonella* SpvB (Corda and Di Girolamo 2003; Fieldhouse et al. 2010; Shniffer et al. 2012; Ueda and Hayaishi 1985). These proteins frequently target specific host proteins such as actin and the small GTPase

Rho. This family also contains several polymorphic toxins, which are typically distinguished by N-terminal RHS repeats as in the case of other versions found in polymorphic toxins (e.g., gi: 221174043 from *Burkholderia multivorans*; Fig. 5) (Zhang et al. 2012). Interestingly, the immunity proteins for these toxins are ARHs (e.g., gi: 221174044) suggesting that their action is reversed by hydrolytic removal of the ADPR modification by the latter. A related group of toxins from actinobacteria (e.g., gi: 300791137) contains several tandem copies of the ART domain combined with other enzymatic domains such as metallopeptidases, papain-like and caspase-like thiol peptidases, and RelA/SpoT-like nucleotidyltransferases (Zhang et al. 2012) (Fig. 5).

The fourth, the Mod-Ecto-ART family is also large and like the previous one includes both phage versions and those deployed as toxins in interorganismal conflicts. The phage-encoded ARTs in this family are prototyped by the T4 Mod proteins (ModA and ModB), which are expressed post-infection and catalyze host RNA polymerase subunit modifications to facilitate transcriptional switching to the late genes of T4 (Tiemann et al. 2004). The effectors belonging to this family include a novel effector (pc1346; gi: 46446980) encoded by the endosymbiont *Protochlamydia,* which resides in amoebae, and various type-III secretion system effectors of *Pseudomonas syringae*, namely hopO1-1/2/3, and HopU1, *Xanthomonas axonopodis* effector XopA1 and a *Legionella pneumophila* T4SS effector (gi: 307611385) (de Souza and Aravind 2012; Fu et al. 2007; Zhang et al. 2012). Certain RTX toxins (Linhartova et al. 2010), which are secreted by means of ABC ATPases, by pathogenic bacteria like *V. cholerae* also contain ART domains of this family (e.g., gi: 487852101). Another interesting member of this family predicted to be involved in intragenomic conflicts in bacteria is fused to a DNA glycosylase of the UDG superfamily (e.g., gi: 497720748; Fig. 5). These are encoded in operons along with a gene for a Macro protein and in some cases an ARH. These are predicted to modify DNA and introduce abasic sites in conjunction with the fused DNA glycosylase domain, whereas the Macro and ARH might counter these modifications. Thus, they could function as a potential DNA-modifying restriction system. Related ART proteins are also encoded as part of a mobile novel T7SS and might be deployed as toxins exported by them (Anantharaman et al. 2012). Yet other related ART domains also occur as part of large actinobacterial proteins with several other enzymatic and helix-turn-helix domains (e.g., gi: 379710894; *Nocardia cyriacigeorgica*; Fig. 5); these proteins might represent a multipronged defensive and signaling strategy against antibiotics. Members of this lineage have been transferred to eukaryotes on several occasions, and include a version from oomycetes (e.g., *Saprolegnia diclina* gi: 530729068), where they have displaced the PARP family catalytic domain within a PARP-like polypeptide (Fig. 5). Other members include the BUBL1-like proteins fused to ubiquitin and HINT autopeptidase domains in ciliates (Dassa et al. 2004) and versions fused to methylated peptide-binding chromodomains in *Emiliania* (Fig. 5).

Also included in the above family are the ecto-ARTs of eukaryotes which are secreted proteins that add a single ADPR moiety to extracellular proteins

(Glowacki et al. 2002; Pallen et al. 2001). These ecto-ARTs show a patchy distribution in eukaryotes being found in the animal lineage as well as groups such as the haptophyte alga *Emiliania* and the rhizarian *Reticulomyxa filosa*. In the latter two taxa they show massive lineage-specific expansions. Certain mammalian versions (ARTC2) function as "toxins" for T cells by inducing their apoptosis (Adriouch et al. 2001; Hassa et al. 2006; Seman et al. 2003). This raises the possibility that the expanded version in the more basal eukaryotic lineages might have a role in defense or cell surface organization by modifying extracellular targets. Rotifers display an expansion of intracellular members of this clade, including versions which appear to have displaced the PARP-like H-Y-[EDQ] clade catalytic domain to associate with the N-terminal WWE domain (gi: 188501623; Fig. 5). All these eukaryotic ARTs are nested within the radiation of bacterial effectors, being particularly close to certain versions such as the *P. syringae* effectors hopO1-1/2/3, suggesting that they were acquired by lateral transfer from bacterial symbionts or parasites (Glowacki et al. 2002; Pallen et al. 2001).

4.4 The Aberrant Clades

The most extensive of these is the EvpP-NTox31 clade, which contains secreted ARTs with a "hybrid" active site. Hence, it is unclear if they are derived versions of either the H-Y-[EDQ] or R-S-E clades, or represent a distinct transitional clade (Fig. 2). This clade includes two distinct sub-clades, of which the EvpP sub-clade is constituted by three families of toxin tips of polymorphic toxins (NTox40, NTox46, and NTox48) (Zhang et al. 2012), and the effectors of pathogenic bacteria, such as EvpP of *E. tarda* (Zheng and Leung 2007). Those occurring with polymorphic toxins are characterized by neighboring immunity protein genes and are associated with several distinct secretory systems such as T7SS and T5SS in Gram-positive and proteobacteria, respectively (Fig. 5). Several of the host-targeting effectors, such as EvpP, are associated with T6SS (Zheng and Leung 2007). The second subclade includes Ntox31 that is found in polymorphic toxin systems and SPNs. SPNs from *Streptococcus pyogenes* and related species appear to be derived from polymorphic toxin systems and are reused as host-directed toxins. Polymorphic tips of systems featuring EvpP-NTox31 domains often display toxin cassettes coding for multiple distinct members of this clade, classical R-S-E clade ARTs or ADP-ribosyl cyclases (Fig. 5). Extensive sequence divergence also precludes precise phylogenetic placement of the RolB family (Fig. 2).

5 Sirtuins and TM1506

The ADP-ribosylation activity of the sirtuins is generally believed to be much weaker than that of the ART superfamily (Frye 1999; Hawse and Wolberger 2009). ADP-ribosylation by sirtuins has been shown to be specific for arginine in

nuclear proteins like histones and perhaps tubulin α/β in the cytoskeleton (Hassa et al. 2006; Hawse and Wolberger 2009). Mitochondrial sirtuins (SIRT3, SIRT4, and SIRT5 in humans) have been implicated in mono-ADP-ribosylation of cysteine-119 in glutamate dehydrogenase (Haigis et al. 2006). The conservation of ADP-ribosylation by sirtuins across eukaryotes (confirmed in animals, apicomplexans, and kinetoplastids) (Garcia-Salcedo et al. 2003; Merrick and Duraisingh 2007) and the coupling of bacterial sirtuins in operons with ADP-ribosyl glycohydrolases (de Souza and Aravind 2012) suggests that this activity might indeed be widely functionally relevant, even if less understood. The ADPR transfer-associated deacylation of lysines by sirtuins is conserved in both prokaryotes and eukaryotes and targets the entire spectrum of acyl modifications ranging from the simple acetyl group, through mid-sized propionyl, crotonyl, malonyl, and succinyl groups, all the way to long-chained moieties like the myristoyl group (Belenky et al. 2007; Jiang et al. 2013; Smith et al. 2002).

Sirtuins display a classical α/β Rossmann fold that had diversified prior to the LUCA into two distinct assemblages of enzymes (de Souza and Aravind 2012; Smith et al. 2002). The first of these includes all classical $NAD^+/NADP^+$-dependent dehydrogenases (Andreeva et al. 2008; de Souza and Aravind 2012; Ronimus and Morgan 2003). The sirtuins however belong to the second assemblage along with the deoxyhypusine synthase and the NAD^+–$NADP^+$ transhydrogenases (Andreeva et al. 2008; de Souza and Aravind 2012; Smith et al. 2002). While topologically similar to the former, this version of the Rossmann fold binds NAD+ in the reverse configuration. Moreover, sirtuins have evolved an additional set of catalytic residues (usually a conserved histidine), which are absent in the others dehydrogenases, that have allowed the sirtuins to utilize NAD^+ as a substrate to transfer ADP-ribose to target moieties rather than as a redox cofactor. Sirtuins are present in all the three superkingdoms of life, with phylogenetic analysis suggesting that majority of archaeal sirtuins from euryarchaea, crenarchaea, and korarchaea form a slow-evolving monophyletic clade (de Souza and Aravind 2012). They are also widespread in bacteria where they have undergone a notable diversification. Hence, the ancestral sirtuin was likely to have been present in the LUCA (de Souza and Aravind 2012). While eukaryotes might have inherited at least one sirtuin from their archaeal progenitor, several of the eukaryotic sirtuins appear to have been acquired via lateral transfer from bacteria (Iyer et al. 2008). In particular, sirtuin 4, 5, and 6 appear to have been independently acquired relatively early in eukaryotic evolution from bacterial precursors. Like the PARP family, eukaryotic sirtuins also developed fusions to several distinct domains such as the ubiquitin-binding Ubp-ZnF domain, tetratricopeptide and kelch repeats, and the Macro domain (Iyer et al. 2008), which might play a role in further processing of OAADPR or allosteric regulation by it (Peterson et al. 2011; Timinszky et al. 2009). In some bacteria, sirtuins occur as part of large operons with several other NAD^+-utilizing and ADPR-processing enzymes such as KptA, ARHs, and Macro (de Souza and Aravind 2012) (Fig. 4). This suggests that NAD^+ utilization and ADPR metabolism by a diverse array of enzymes might be coordinately regulated in these organisms. Several bacteria also possess a functionally enigmatic family

of sirtuins, which occur in operons for diverse restriction systems in place of the restriction DNases (Burroughs et al. 2013; Iyer et al. 2004). This suggests that they are functionally equivalent to the restriction enzymes raising the possibility that they target DNA, perhaps via ADP-ribosylation of DNA bases or have undergone a functional shift to catalyze an unrelated reaction, such as DNase activity.

The TM1506 family is prototyped by the eponymous protein from *Thermotoga maritima* (Iyer et al. 2011). This protein was shown to be ADP-ribosylated at aspartate-56 (Xu et al. 2008). Analysis of the structure of this protein showed that it contains a deaminase-like fold and possesses the characteristic active site pocket of this fold with a distinctive set of conserved residues that are likely to support catalytic activity. The crystal structure of TM1506 also showed diffraction density for a ligand in this active site that is consistent with an ADP-ribose linked to D56 or its precursor NAD^+ (Xu et al. 2008). This suggests that it might be a novel ADP-ribosylating enzyme with auto-activity. Thus far, this family is only found in bacteria. TM1506-like genes are often linked in predicted operons to genes encoding a Rossmann fold aldo/keto reductase fused to a rubredoxin-like zinc ribbon and a 5TM protein that is predicted to form a transmembrane channel (Iyer et al. 2011). In bacteroidetes, the domain is also fused to a TonB-like receptor, which is usually involved in the trafficking of small molecules, such as siderophores and peptide antibiotics (Noinaj et al. 2010). Thus, it is conceivable that the activity of the TM1506 family regulates transmembrane trafficking of molecules by the products of associated genes.

6 A Synthetic Overview of the Tendencies in the Diversification of the ADP-Ribosylation System

It is currently not clear if there were any member of the ART superfamily in the LUCA or whether they arose first in bacteria (Fig. 2). Nevertheless, KptA appears closest to the ancestral version of this superfamily suggesting that it first arose in the context of RNA repair. From that point on the ART superfamily diverged rapidly and explosively in the bacterial superkingdom in terms of sequence, structure, and active site residues, and evolved a wide range of substrates specificities, albeit without any notable diversification in terms of the basic reaction type they catalyze. Second, as noted in the above discussion the primary diversification of ARTs appears to have taken place in bacteria among diverse systems involved in defensive and offensive strategies in intragenomic, intergenomic, and intraorganismal conflicts (Fig. 2) (Fieldhouse et al. 2010; Glowacki et al. 2002; Pallen et al. 2001; Zhang et al. 2012). This suggests it took place under the combination of the intense selective pressures arising from development of resistance to their action and defensive counter strategies directed against them in such conflict systems, and relaxation of selection for specific interactions with multiple cellular molecules beyond substrates and immunity proteins (Aravind et al. 2012). Strikingly,

eukaryotes have repeatedly (more than 20 distinct occasions) acquired ARTs from these prokaryotic conflict systems (Fig. 2). While some of these like the PARP family were acquired prior to the LECA (Citarelli et al. 2010), others seems to have been sporadically acquired throughout eukaryotic evolution. These ARTs appear to have been deployed in two distinct modes: (1) they appear to more or less retain their ancestral toxin function upon transfer to the eukaryotes. Thus, they appear to be utilized as mediators of apoptosis as seen in the case of pierisin, CARP-1, and certain ecto-ARTs. (2) They are incorporated as components of core regulatory systems, where they modify proteins as part of quintessentially eukaryotic regulatory and epigenetic processes. This process was usually accompanied by fusion to several distinct domains, some of which are unique to eukaryotes (e.g., WWE and PARP-ZnF; Fig. 4), in domain architectures that are unparalleled in prokaryotic counterparts (Hottiger et al. 2010). Another, unique aspect of the evolution of eukaryotic ARTs is the lineage-specific expansions that have occurred on several independent occasions. Given the possible role of unrelated lineage-specifically expanded families in anti-pathogen defenses in eukaryotes (Jiang et al. 2009; Krasnov et al. 2011; Rosado et al. 2013), it is conceivable that ARTs have repeatedly been deployed as part of anti-pathogen strategies of eukaryotes.

The above-described evolutionary tendencies of ART superfamily are in contrast with those observed for the sirtuins (de Souza and Aravind 2012). Sirtuins show only a limited presence in the above-described conflict-related systems. The only such versions are those occurring in restriction systems, wherein they take the place of restriction endoDNases. However, they are currently not known in secreted effectors involved in interorganismal conflict. Most conserved versions of sirtuins appear to be forms that are involved in regulation of intracellular proteins by deacylation or perhaps ADP-ribosylation (Garcia-Salcedo et al. 2003; Haigis et al. 2006; Hassa et al. 2006; Merrick and Duraisingh 2007; Smith et al. 2002). Consistent with this, sirtuins show greater conservation at the sequence level and little variation in their active site residues. Notably, unlike the ARTs they also show little diversity in their domain architectures. Like the ARTs, eukaryotes have acquired sirtuins on multiple occasions from bacteria (Iyer et al. 2008); however, there is no evidence that any of these were acquired from conflict-related systems. Moreover, unlike the ARTs, sirtuins show little propensity for lineage-specific expansions in eukaryotes. These differences suggest that the ART fold is probably more suited for structural diversification; thus it tended to be selected repeatedly as a scaffold for evolving different ADP-ribosylation activities.

However, four other superfamilies of domains in the ADP-ribosylation system, namely the ADP-ribosyl cyclase, Macro, NADAR, and ARH appear to have evolutionary histories comparable to that of the ART superfamily (de Souza and Aravind 2012; Zhang et al. 2012). Of these, Macro NADAR and ARH appear to have emerged in type-II T–A and polymorphic toxin systems as enzymatic antitoxins or immunity proteins removing ADP-ribosyl modifications. Moreover, Macro, NADAR and the ARH domain also appear to have diversified in bacterial systems to cleave ADPR modifications from macromolecules and process other ADPR derivatives such as OAADPR (generated by sirtuins during deacylation)

and ADP-ribose 1″-phosphate (Appr1″p) generated during RNA repair (Guse and Lee 2008; Hofmann et al. 2000; Karras et al. 2005; Koch-Nolte et al. 2008; Slade et al. 2011). All these versions were incorporated into eukaryotic systems upon being acquired by lateral transfer. De-ADP-ribosylating Macro domains were recruited as the PARGs as well as mono-ADP-ribosyl hydrolases of the eukaryotes (de Souza and Aravind 2012; Slade et al. 2011). Similarly, the ARG domain was recruited both in de-ADP-ribosylation and processing of OAADPRs in eukaryotes (Koch-Nolte et al. 2008; Ono et al. 2006). The ADP-ribosyl cyclase (ARC) has previously only been characterized in animals and generates cyclic-ADP-ribose (cADPr) and nicotinic acid adenine dinucleotide phosphate (NAADP), respectively from NAD^+ and NADP (Guse and Lee 2008). cADPr and NAADP have been shown to function as potent inducers of calcium influx via the ryanodine receptors. At the same time by channeling NAD^+, ARCs might also affect protein deacylation by sirtuins and other processes requiring NAD^+ (Guse and Lee 2008). Polymorphic toxins with ARC domains occur in free-living bacteria and are predicted to be delivered via T5SS, T6SS, and T7SS (Zhang et al. 2012). This suggests that they are used in intraspecific conflict rather than against eukaryotes. It remains to be seen if some of these ARC domains might function as NADases or ADP ribosyltransferases. Bacterial ARCs appear to have been transferred to eukaryotes on at least two occasions, namely to animals and fungi, where they function as membrane-associated enzymes (Zhang et al. 2012). Thus, majority of enzymes of the eukaryotic ADPR system might have originally emerged in the context of bacterial conflict systems and were then acquired via lateral transfer by eukaryotes. This pattern resembles another large superfamily of proteins namely the deaminases, which likewise diversified in comparable bacterial contexts and were incorporated into eukaryotic nucleic acid modification and anti-pathogen defense systems on several occasions (Iyer et al. 2011). More generally, it adds further evidence to the recently proposed hypothesis that the rapid evolution of novel biochemical capabilities in bacterial conflict systems has been a major supplier of innovations in the evolution of eukaryotic regulatory systems (Aravind et al. 2012).

7 Conclusion

Little over 50 years after its discovery we have a fairly detailed view of the evolutionary history of the ADP-ribosylation system and the functioning of its various components at the molecular level. While considerable progress has been made in terms of the biochemistry and biology of this system, several new questions have been raised by recent studies. First, identification of new ARTs through sequence and structure analysis has emphasized the diversity of the ART active site. This calls for new understanding of the ART reaction mechanism, especially in cases like RolB or BC2332 where the active site is completely remodeled. Second, newly identified ARTs like NEURL4, plant pierisin-like,

fungal CARP-1-like ARTs, and the various lineage-specifically expanded versions in eukaryotes are interesting targets for discovery of new biological roles for ARTs in both model and nonmodel organisms. Finally, bacterial versions of ART, sirtuin, Macro, NADAR, and ARH domains in diverse conflict systems offer the potential for understanding both the nature of these conflict systems and the true extent of the diversity of biochemical activities in the ADPR system.

Acknowledgments Work by the authors is supported by the Intramural Research Program of the National Library of Medicine, the National Institutes of Health, USA.

References

Adriouch S, Ohlrogge W, Haag F, Koch-Nolte F, Seman M (2001) Rapid induction of naive T cell apoptosis by ecto-nicotinamide adenine dinucleotide: requirement for mono(ADP-ribosyl)transferase 2 and a downstream effector. J Immunol 167(1):196–203

Al-Hakim AK, Bashkurov M, Gingras AC, Durocher D, Pelletier L (2012) Interaction proteomics identify NEURL4 and the HECT E3 ligase HERC2 as novel modulators of centrosome architecture. Mol Cell Proteomics 11(6):M111 014233

Altmeyer M, Messner S, Hassa PO, Fey M, Hottiger MO (2009) Molecular mechanism of poly(ADP-ribosyl)ation by PARP1 and identification of lysine residues as ADP-ribose acceptor sites. Nucleic Acids Res 37(11):3723–3738

Anantharaman V, Iyer LM, Aravind L (2012) Ter-dependent stress response systems: novel pathways related to metal sensing, production of a nucleoside-like metabolite, and DNA-processing. Mol Biosyst 8(12):3142–3165

Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJ, Chothia C, Murzin AG (2008) Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 36(Database issue):D419–D425

Aravind L (2000) Guilt by association: contextual information in genome analysis. Genome Res 10(8):1074–1077

Aravind L, Anantharaman V, Zhang D, de Souza RF, Iyer LM (2012) Gene flow and biological conflict systems in the origin and evolution of eukaryotes. Front Cell Infect Microbiol 2:89

Baysarowich J, Koteva K, Hughes DW et al (2008) Rifamycin antibiotic resistance by ADP-ribosylation: structure and diversity of Arr. Proc Natl Acad Sci U S A 105(12):4886–4891

Bazan JF, Koch-Nolte F (1997) Sequence and structural links between distant ADP-ribosyltransferase families. Adv Exp Med Biol 419:99–107

Belenky P, Bogan KL, Brenner C (2007) NAD^+ metabolism in health and disease. Trends Biochem Sci 32(1):12–19

Berg JM, Tymoczko JL, Stryer L (2012) Biochemistry, 7th edn. W.H. Freeman, New York

Boyd EF (2012) Bacteriophage-encoded bacterial virulence factors and phage-pathogenicity island interactions. Adv Virus Res 82:91–118

Burroughs AM, Ando Y, Aravind L (2013) New perspectives on the diversification of the RNA interference system: insights from comparative genomics and small RNA sequencing. Wiley Interdiscip Rev RNA 5(2):141–181

Carpusca I, Jank T, Aktories K (2006) *Bacillus sphaericus* mosquitocidal toxin (MTX) and pierisin: the enigmatic offspring from the family of ADP-ribosyltransferases. Mol Microbiol 62(3):621–630

Chambon P, Weil J, Mandel P (1963) Nicotinamide mononucleotide activation of a new DNA-dependent polyadenylic acid synthesizing nuclear enzyme. Biochem Biophys Res Commun 11:39–43

Citarelli M, Teotia S, Lamb RS (2010) Evolutionary history of the poly(ADP-ribose) polymerase gene family in eukaryotes. BMC Evol Biol 10:308

Cohen-Armon M, Visochek L, Katzoff A et al (2004) Long-term memory requires polyADP-ribosylation. Science 304(5678):1820–1822

Collier R, Pappenheimer A (1964) Studies on the mode of action of diphtheria toxin. II. Effect of toxin on amino acid incorporation in cell-free systems. J Exp Med 120:1019–1039

Corda D, Di Girolamo M (2003) Functional aspects of protein mono-ADP-ribosylation. EMBO J 22(9):1953–1958

Dassa B, Yanai I, Pietrokovski S (2004) New type of polyubiquitin-like genes with intein-like autoprocessing domains. Trends Genet 20(11):538–542

de Souza RF, Aravind L (2012) Identification of novel components of NAD-utilizing metabolic pathways and prediction of their biochemical functions. Mol Biosyst 8(6):1661–1677

Dunigan DD, Cerny RL, Bauman AT et al (2012) *Paramecium bursaria* chlorella virus 1 proteome reveals novel architectural and regulatory features of a giant virus. J Virol 86(16):8821–8834

Fieldhouse RJ, Turgeon Z, White D, Merrill AR (2010) Cholera- and anthrax-like toxins are among several new ADP-ribosyltransferases. PLoS Comput Biol 6(12):e1001029

Finn RD, Mistry J, Tate J, et al (2010) The Pfam protein families database. Nucleic Acids Res 38(Database issue):D211–D222

Frye RA (1999) Characterization of five human cDNAs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity. Biochem Biophys Res Commun 260(1):273–279

Fu ZQ, Guo M, Jeong BR et al (2007) A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity. Nature 447(7142):284–288

Garcia-Salcedo JA, Gijon P, Nolan DP, Tebabi P, Pays E (2003) A chromosomal SIR2 homologue with both histone NAD-dependent ADP-ribosyltransferase and deacetylase activities is involved in DNA repair in *Trypanosoma brucei*. EMBO J 22(21):5851–5862

Gazzaniga F, Stebbins R, Chang SZ, McPeek MA, Brenner C (2009) Microbial NAD metabolism: lessons from comparative genomics. Microbiol Mol Biol Rev 73(3):529–541 (Table of Contents)

Glowacki G, Braren R, Firner K et al (2002) The family of toxin-related ecto-ADP-ribosyltransferases in humans and the mouse. Protein Sci 11(7):1657–1670

Guse AH, Lee HC (2008) NAADP: a universal Ca^{2+} trigger. Sci Signal 1(44):re10

Haigis MC, Mostoslavsky R, Haigis KM et al (2006) SIRT4 inhibits glutamate dehydrogenase and opposes the effects of calorie restriction in pancreatic beta cells. Cell 126(5):941–954

Han S, Craig JA, Putnam CD, Carozzi NB, Tainer JA (1999) Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex. Nat Struct Biol 6(10):932–936

Harding HP, Lackey JG, Hsu HC et al (2008) An intact unfolded protein response in Trpt1 knockout mice reveals phylogenic divergence in pathways for RNA ligation. RNA 14(2):225–232

Hassa PO, Haenni SS, Elser M, Hottiger MO (2006) Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? Microbiol Mol Biol Rev 70(3):789–829

Hawse WF, Wolberger C (2009) Structure-based mechanism of ADP-ribosylation by sirtuins. J Biol Chem 284(48):33654–33661

Hofmann A, Zdanov A, Genschik P, Ruvinov S, Filipowicz W, Wlodawer A (2000) Structure and mechanism of activity of the cyclic phosphodiesterase of Appr>p, a product of the tRNA splicing reaction. EMBO J 19(22):6207–6217

Hottiger MO, Hassa PO, Luscher B, Schuler H, Koch-Nolte F (2010) Toward a unified nomenclature for mammalian ADP-ribosyltransferases. Trends Biochem Sci 35(4):208–219

Iyer LM, Makarova KS, Koonin EV, Aravind L (2004) Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res 32(17):5260–5279

Iyer LM, Anantharaman V, Wolf MY, Aravind L (2008) Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int J Parasitol 38(1):1–31

Iyer LM, Zhang D, Rogozin IB, Aravind L (2011) Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic Acids Res 39(22):9473–9497

Jiang H, Khan S, Wang Y et al (2013) SIRT6 regulates TNF-alpha secretion through hydrolysis of long-chain fatty acyl lysine. Nature 496(7443):110–113

Jiang J, Zhang YB, Li S, Yu FF, Sun F, Gui JF (2009) Expression regulation and functional characterization of a novel interferon inducible gene *Gig2* and its promoter. Mol Immunol 46(15):3131–3140

Jorgensen R, Wang Y, Visschedyk D, Merrill AR (2008a) The nature and character of the transition state for the ADP-ribosyltransferase reaction. EMBO Rep 9(8):802–809

Jorgensen R, Purdy AE, Fieldhouse RJ, Kimber MS, Bartlett DH, Merrill AR (2008b) Cholix toxin, a novel ADP-ribosylating factor from *Vibrio cholerae*. J Biol Chem 283(16): 10671–10678

Karras GI, Kustatscher G, Buhecha HR et al (2005) The macro domain is an ADP-ribose binding module. EMBO J 24(11):1911–1920

Kleine H, Poreba E, Lesniewicz K et al (2008) Substrate-assisted catalysis by PARP10 limits its activity to mono-ADP-ribosylation. Mol Cell 32(1):57–69

Koch-Nolte F, Reche P, Haag F, Bazan F (2001) ADP-ribosyltransferases: plastic tools for inactivating protein and small molecular weight targets. J Biotechnol 92(2):81–87

Koch-Nolte F, Kernstock S, Mueller-Dieckmann C, Weiss MS, Haag F (2008) Mammalian ADP-ribosyltransferases and ADP-ribosylhydrolases. Front Biosci 13:6716–6729

Koch T, Ruger W (1994) The ADP-ribosyltransferases (gpAlt) of bacteriophages T2, T4, and T6: sequencing of the genes and comparison of their products. Virology 203(2):294–298

Krasnov A, Timmerhaus G, Schiotz BL et al (2011) Genomic survey of early responses to viruses in Atlantic salmon, *Salmo salar* L. Mol Immunol 49(1–2):163–174

Laing S, Unger M, Koch-Nolte F, Haag F (2011) ADP-ribosylation of arginine. Amino Acids 41(2):257–269

Linhartova I, Bumba L, Masin J et al (2010) RTX proteins: a highly diverse family secreted by a common mechanism. FEMS Microbiol Rev 34(6):1076–1112

Ludden PW (1994) Reversible ADP-ribosylation as a mechanism of enzyme regulation in prokaryotes. Mol Cell Biochem 138(1–2):123–129

Merrick CJ, Duraisingh MT (2007) *Plasmodium falciparum* Sir2: an unusual sirtuin with dual histone deacetylase and ADP-ribosyltransferase activity. Eukaryot Cell 6(11):2081–2091

Nakano T, Matsushima-Hibiya Y, Yamamoto M et al (2006) Purification and molecular cloning of a DNA ADP-ribosylating protein, CARP-1, from the edible clam *Meretrix lamarckii*. Proc Natl Acad Sci U S A 103(37):13652–13657

Noinaj N, Guillier M, Barnard TJ, Buchanan SK (2010) TonB-dependent transporters: regulation, structure, and function. Annu Rev Microbiol 64:43–60

Ono T, Kasamatsu A, Oka S, Moss J (2006) The 39-kDa poly(ADP-ribose) glycohydrolase ARH3 hydrolyzes O-acetyl-ADP-ribose, a product of the Sir2 family of acetyl-histone deacetylases. Proc Natl Acad Sci U S A 103(45):16687–16691

Otto H, Reche PA, Bazan F, Dittmar K, Haag F, Koch-Nolte F (2005) In silico characterization of the family of PARP-like poly(ADP-ribosyl)transferases (pARTs). BMC Genomics 6:139

Pallen MJ, Lam AC, Loman NJ, McBride A (2001) An abundance of bacterial ADP-ribosyltransferases—implications for the origin of exotoxins and their human homologues. Trends Microbiol 9(7):302–307 (discussion 308)

Park UE, Olivera BM, Hughes KT, Roth JR, Hillyard DR (1989) DNA ligase and the pyridine nucleotide cycle in *Salmonella typhimurium*. J Bacteriol 171(4):2173–2180

Peterson FC, Chen D, Lytle BL, Rossi MN, Ahel I, Denu JM, Volkman BF (2011) Orphan macrodomain protein (human C6orf130) is an O-acyl-ADP-ribose deacylase: solution structure and catalytic properties. J Biol Chem 286(41):35955–35965

Price MN, Dehal PS, Arkin AP (2010) FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One 5(3):e9490

Ramirez MS, Tolmasky ME (2010) Aminoglycoside modifying enzymes. Drug Resist Updat 13(6):151–171

Rohmer L, Kjemtrup S, Marchesini P, Dangl JL (2003) Nucleotide sequence, functional characterization and evolution of pFKN, a virulence plasmid in *Pseudomonas syringae* pathovar maculicola. Mol Microbiol 47(6):1545–1562

Ronimus RS, Morgan HW (2003) Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. Archaea 1(3):199–221

Rosado MM, Bennici E, Novelli F, Pioli C (2013) Beyond DNA repair, the immunological role of PARP-1 and its siblings. Immunology 139(4):428–437

Sala A, Calderon V, Bordes P, Genevaux P (2013) TAC from *Mycobacterium tuberculosis*: a paradigm for stress-responsive toxin-antitoxin systems controlled by SecB-like chaperones. Cell Stress Chaperones 18(2):129–135

Schaffer AA, Aravind L, Madden TL et al (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res 29(14):2994–3005

Seman M, Adriouch S, Scheuplein F et al (2003) NAD-induced T cell death: ADP-ribosylation of cell surface proteins by ART2 activates the cytolytic P2X7 purinoceptor. Immunity 19(4):571–582

Shniffer A, Visschedyk DD, Ravulapalli R et al (2012) Characterization of an actin-targeting ADP-ribosyltransferase from *Aeromonas hydrophila*. J Biol Chem 287(44):37030–37041

Singer AU, Desveaux D, Betts L et al (2004) Crystal structures of the type III effector protein AvrPphF and its chaperone reveal residues required for plant pathogenesis. Structure 12(9): 1669–1681

Slade D, Dunstan MS, Barkauskaite E et al (2011) The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase. Nature 477(7366):616–620

Smith CL, Ghosh J, Elam JS, Pinkner JS, Hultgren SJ, Caparon MG, Ellenberger T (2011) Structural basis of *Streptococcus pyogenes* immunity to its NAD^+ glycohydrolase toxin. Structure 19(2):192–202

Smith JA, Stocken LA (1974) Chemical and metabolic properties of adenosine diphosphate ribose derivatives of nuclear proteins. Biochem J 147:523–529

Smith JS, Avalos J, Celic I, Muhammad S, Wolberger C, Boeke JD (2002) SIR2 family of NAD(+)-dependent protein deacetylases. Methods Enzymol 353:282–300

Snyder AK, Rio RV (2013) Interwoven biology of the tsetse holobiont. J Bacteriol 195(19):4322–4330

Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33(Web Server issue):W244–W248

Sorci L, Kurnasov OV, Rodionov DA, Osterman A (2010) Genomics and enzymology of NAD biosynthesis. In: Mander LN, Lui HW (eds) Comprehensive natural products II : chemistry and biology, vol 7. Elsevier, Amsterdam, pp 213–257

Spinelli SL, Kierzek R, Turner DH, Phizicky EM (1999) Transient ADP-ribosylation of a 2′-phosphate implicated in its removal from ligated tRNA during splicing in yeast. J Biol Chem 274(5):2637–2644

Stabb EV, Reich KA, Ruby EG (2001) *Vibrio fischeri* genes *hvnA* and *hvnB* encode secreted NAD(+)-glycohydrolases. J Bacteriol 183(1):309–317

Sugimura T, Miwa M (1994) Poly(ADP-ribose): historical perspective. Mol Cell Biochem 138(1–2):5–12

Takahashi-Nakaguchi A, Matsumoto Y, Yamamoto M, Iwabuchi K, Totsuka Y, Sugimura T, Wakabayashi K (2013) Demonstration of cytotoxicity against wasps by pierisin-1: a possible defense factor in the cabbage white butterfly. PLoS One 8(4):e60539

Tiemann B, Depping R, Gineikiene E, Kaliniene L, Nivinskas R, Ruger W (2004) ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. J Bacteriol 186(21):7262–7272

Timinszky G, Till S, Hassa PO et al (2009) A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. Nat Struct Mol Biol 16(9):923–929

Ueda K, Hayaishi O (1985) ADP-ribosylation. Annu Rev Biochem 54:73–100

Walsh C (2003) Antibiotics : actions, origins, resistance. ASM Press, Washington, D.C.

Wang M, Soyano T, Machida S, Yang JY, Jung C, Chua NH, Yuan YA (2011) Molecular insights into plant cell proliferation disturbance by *Agrobacterium* protein 6b. Genes Dev 25(1):64–76

Xu Q, Kozbial P, McMullan D et al (2008) Crystal structure of an ADP-ribosylated protein with a cytidine deaminase-like fold, but unknown function (TM1506), from *Thermotoga maritima* at 2.70 A resolution. Proteins 71(3):1546–1552

Zhang D, de Souza RF, Anantharaman V, Iyer LM, Aravind L (2012) Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. Biol Direct 7:18

Zheng J, Leung KY (2007) Dissection of a type VI secretion system in *Edwardsiella tarda*. Mol Microbiol 66(5):1192–1206