

Macrodomains: Structure, Function, Evolution and Catalytic activities

Johannes Gregor Matthias Rack¹, Dragutin Perina², and Ivan Ahel¹

Affiliations

¹Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, United Kingdom

²Division of Molecular Biology, Ruđer Bošković Institute, Zagreb 10002, Croatia

emails

johannes.rack@path.ox.ac.uk

Drago.Perina@irb.hr

Ivan.ahel@path.ox.ac.uk

Correspondence

ivan.ahel@path.ox.ac.uk

Keywords

ADP-ribose, PARP family, PARG, NAD, posttranslational modifications

Table of Contents

ABSTRACT	3
EVOLUTION AND DIVERSIFICATION OF MACRODOMAINS.....	5
STRUCTURAL AND CATALYTIC FEATURES OF MACRODOMAINS.....	7
MONO(ADP-RIBOSYL) HYDROLASES	8
POLY(ADP-RIBOSYL) GLYCOHYDROLASES	9
THE FUNCTION OF MACRODOMAINS	10
MACRODOMAINS AS READERS OF PROTEIN ADP-RIBOSYLATION AND SENSORS OF	
METABOLITES	11
MACRODOMAINS AS ERASERS OF PROTEIN ADP-RIBOSYLATION.....	14
MACRODOMAINS IN MICROORGANISMS AND PATHOGENS	18
THE THERAPEUTIC POTENTIAL OF MACRODOMAIN INHIBITORS	21
OUTLOOK.....	22
ACKNOWLEDGMENTS	24
REFERENCES.....	25
ACRONYMS AND DEFINITIONS	30

Abstract

ADP-ribose is a versatile nucleotide moiety employed by various signaling pathways, either in the form of metabolites or as a modification covalently bound to proteins. Recent developments indicate that macrodomains, an ancient and diverse protein domain family, are key players both in the recognition, interpretation and turnover of ADP-ribose signaling. Thereby, macrodomains regulate a wide variety of cellular and organismal processes including DNA damage repair, signal transduction and immune response. Their importance is further stressed by the fact that dysregulation or mutation of macrodomains is associated with several disease including cancer, developmental defects and neurodegeneration. In this review we summarize the current insight into macrodomain evolution and how it influenced the structural and functional diversification. Furthermore, highlight some aspects of their role in pathobiology and as emerging therapeutic targets.

Introduction

Macrodomains are evolutionarily conserved structural modules of 130 – 190 amino acids that are found in proteins with diverse cellular functions, from all domains of life as well as some viruses (1). Genomic sequencing identified the first macrodomains (termed ‘X domain’) as “a domain of considerable conservation” within the genomes of the murine hepatitis virus (MHV) and infectious bronchitis virus (2). Shortly after a homologous domain was identified as part of the rat macroH2A protein, a histone variant that consists of a fusion between histone H2A and a domain of – at that time – unknown function (3). Since macroH2A is the largest histone variant the novel domain was dubbed *macrodomain*.

In contrast to many other modification recognition domains, which are adapted to recognize a single or small number of modification type(s) (4), macrodomains can recognize ADP-ribose (ADPr), both in its free and protein linked form, related ligands, such as *O*-acyl-ADP-ribose (OAADPr), and even ligands unrelated to ADPr (5-10). They exert regulatory influence on inter- and intracellular signaling, transcription, DNA repair pathways, maintenance of genomic stability, telomere dynamics, cell differentiation and proliferation, as well as necrosis and apoptosis (5, 11-19). Two factors further highlight the importance of macrodomains in controlling various cellular processes: first, the number of macrodomain-containing proteins correlates well with the complexity of the organism and second, the macrodomains coevolved most notably with two important NAD⁺ utilizing protein families, poly(ADP-ribose) polymerases (PARPs) and sirtuins (20-22). The majority of sirtuins use NAD⁺ as co-substrate for protein deacylation, releasing OAADPr as byproduct in the reaction (23, 24), but in some instances sirtuins also catalyze protein ADP-ribosylation (25, 26). PARPs catalyze the transfer of a single or multiple ADPr unit(s) from NAD⁺ onto an acceptor site, most prominently a glutamate or aspartate residues, thus resulting in mono or poly(ADP-ribosylation) of the target proteins (20-22, 27, 28). Note that only some members of the

PARP family can catalyze the formation of the *O*-glycosidic ribose-ribose bond and synthesize linear or branched chains of ADPr polymers (27, 29-31).

Beyond the ability of macrodomains to bind ADPr derivatives, an early biochemical genomics study identified the yeast macrodomain Poa1 as a phosphatase of ADP-ribose-1''-phosphate (ADPr-1''-p), a by-product of tRNA splicing (32, 33). Since then it became apparent that processing of ADPr derivatives is common amongst certain types of macrodomains. The catalyzed reactions include the hydrolysis of the 2',1''-*O*-glycosidic ribose-ribose bond in poly(ADP-ribose) (PAR), the protein-ADPr ester bond or the acyl-ADPr ester (8, 27, 34-40), as detailed below. Thus macrodomains represent a rare example of a protein fold which is utilized to directly modulate multiple aspects of posttranslational protein modifications, i.e. signal 'reading', 'erasing' and 'interpreting'. The ADPr released in some macrodomain reactions acts both as a precursor for the regeneration of NAD⁺ as well as an important secondary messenger (41). Moreover, reports on GDAP2 and SUD-M type of macrodomains suggest that further ADPr-unrelated binding partners, like nucleic acids, may exist for macrodomains (9, 10, 42).

In this review we highlight some aspects of the macrodomain evolution and the resulting functional and structural diversification. Furthermore, we discuss the role of macrodomains in pathobiology, including the cause of disease via mutation and dysregulation as well as their potential as therapeutic targets.

Evolution and diversification of macrodomains

The intricate connection between macrodomains and NAD⁺-utilizing pathways suggests that by studying their (co)evolution, one could infer the functions and regulation of specific proteins involved in these processes. Phylogenetically macrodomains can be subdivided into six distinct classes (Figure 1). Probably the most diverged classes are PARG and SUD-M since they were recognized as macrodomains only after the first 3-dimensional structure of

their representatives had been determined (10, 40, 43). Thus far Macro2 remains the least understood branch with almost no functional data available. While macrodomains appear less regularly in archaea, most bacteria contain at least one type of macrodomain (Table 1). In eukaryotes, the number of macrodomain proteins, generally increases with the complexity of the organism. Amongst vertebrata, *Actinopterygii* (a class of bony fishes) possess the highest number of macrodomain proteins due to a fish-specific third genome duplication event. A discrepancy is also observed in some animals with accelerated evolution (e.g. nematode *Caenorhabditis elegans* and fruit fly *Drosophila melanogaster*) which possess fewer macrodomain representatives than simple non-bilaterian animals and this correlates well with the reduced number of PARPs in these organisms (20). *Saccharomyces cerevisiae* possesses only two macrodomain proteins and has lost some macrodomains present in the common ancestor of *Opisthokonta*, as well as obvious PARP homologues (20). In human, there are at least 12 different proteins containing 16 macrodomains distributed over four branches: MacroD-type, macroH2A-like, ALC1-like and PARG (Figure 1 and 2 and Table 1) (8, 11).

The phylogenetic classification of macrodomains correlates noticeably with their functional diversification, for example members of the macroH2A-like class show strong binding to ADP-ribosylated proteins, but possess no catalytic activity, whereas degradation of PAR is specifically associated with the PARG class (27, 36, 40) and removal of mono(ADP-ribosylation) (MAR) with the MacroD-type class (34, 38). However, these associations are not exclusive as TARG1, a member of the ALC1-like class, also possess de-MARylation activity, albeit utilizing a different reaction mechanism (39). The emergence of the same molecular function within two distinct phylogenetic branches stresses the importance of ADP-ribosylation reversal and demonstrates the adaptability of macrodomains to exert regulation.

Structural and catalytic features of macrodomains

As revealed by structure determination, macrodomains adopt a globular $\alpha/\beta/\alpha$ sandwich fold composed of a central six-stranded mixed β -sheet and flanked by 5 α -helices (Figure 3a) (6, 44-46). Substrate binding occurs via a deep cleft on the ‘crest’ of the domain. The macrodomain fold shares some resemblance to the DNA binding domain of leucine aminopeptidases as well as the P-loop nucleotide hydrolase fold, and therefore a nucleotide binding or processing activity was originally suggested (44). Indeed, further structural and biochemical characterization showed that ADPr and its derivatives can be accommodated within the cleft (Figure 3b) (6, 45, 46). The interaction between ligand and macrodomain is stabilized by several conserved interactions within the binding pocket: (A) the adenosine moiety readily undergoes π - π stacking with a conserved aromatic residue while its N6 nitrogen is further coordinated by an aspartate residue (Figure 3b-e) (5, 6, 8, 47). Mutagenesis analysis showed that the aspartate residue is crucial for ADPr binding by the macrodomains of human ALC1 (Amplified in Liver Cancer 1) and *Archaeoglobus fulgidus* Af1521. (B) The central part of the cleft stabilizes substrate binding by several side chain/backbone-pyrophosphate contacts, which induce a more closed conformation of the macrodomain (6, 48). (C) The pyrophosphate and distal ribose are accommodated between two substrate binding loops (termed loop 1 and 2). While both loops contribute to substrate specificity, loop 1 harbors the catalytic residues of most macrodomains exhibiting hydrolase activity (for this reason it has been also termed ‘catalytic loop’). The loop 2 provides further coordination of the pyrophosphate, hence also termed ‘diphosphate binding loop’ (Figure 3b-e) (15, 37, 40, 49). The stable interaction between the ligand and the macrodomain can trigger a variety of downstream effects, including recruitment to DNA damages sites, a hotspot of PAR generation, or formation of protein complexes (50).

Mono(ADP-ribosyl) hydrolases

Thus far two classes of macrodomains were shown to contain mono(ADP-ribosyl) hydrolases as members: MacroD-type and ALC1-like. Biochemical and structural work on MacroD1, MacroD2 and TARG1, the human representatives of the classes, revealed that both classes utilize fundamentally different catalytic mechanisms.

In MacroD1/2 the distal ribose is bound in a constrained conformation, bending it towards the ADPr α -phosphate group (Figure 3c) (8, 34). This orientation is forced through the presence of a highly conserved aromatic residue that is part of the bipartite, MacroD-type signature motif Nx(6)GG[V/L/I] and G[V/I/A][Y/F]G located in loop 1 and 2, respectively (34). A major difference between ‘reading’ and ‘erasing’ macrodomains is the presence of a groove in the pyrophosphate binding site in which a structural water molecule is coordinated. The exact mechanism of catalysis is as yet not determined, but two possible reaction sequences were put forward: first, a substrate assisted mechanism in which the structural water molecule, positioned between the α -phosphate and distal ribose, becomes activated through the α -phosphate group and carries out a nucleophilic attack on the protein-ADPr ester bond (34, 35). While the constrained conformation of the substrate appears to be crucial for the catalysis, it is still controversial whether the low pK_a of the α -phosphate is sufficient to activate the water molecule (34, 35). Second, a conserved aspartate residue in the active site acts as general base for the activation of a water molecule, which in turn carries out a nucleophilic attack on the C1' atom of the distal ribose (8, 38). Recent quantum mechanics simulation and kinetic isotope effect measurements of the hydrolysis of OAADPr by MacroD1 support the latter mechanism and suggest a concerted mechanism involving simultaneous nucleophile attack and ester bond breakage (51). However, it is important to point out, while conserved in MacroD1 and MacroD2, not all catalytically active MacroD-type enzymes possess an isostructural aspartate residue (34).

In contrast to the MacroD-type enzymes, the reaction catalyzed by TARG1 progresses via a conserved lysine residue (Lys84 in human TARG1; **Figure 3d**). The lysine nucleophilically attacks the C1' atom of the distal ribose, leading to the formation of a lysyl-ADPr intermediate and release of the demodified glutamate/aspartate residue (39). The lysine residue is restored by resolving the intermediate via a proximal catalytic aspartate residue (Asp125 in human TARG1), thus releasing ADPr from TARG1. A second difference that sets TARG1 apart from MacroD1/2 is a higher solvent accessibility at the adenosine ribose 2''-OH position, which suggests that TARG1 can bind to PARylated target protein and release PAR from its substrates. This possibility was experimentally demonstrated *in vitro*, but whether it occurs *in vivo* remains an open question (35, 39).

Poly(ADP-ribosyl) glycohydrolases

In mammals, PARG is composed of three domains: an N-terminal putative regulatory region, the PARG accessory domain and the macrodomain (Figure 3e) (40, 52-54). Together, the latter two form the catalytic core found in all canonical PARG homologues. The PARG macrodomain is composed of a seven-stranded, mixed β -sheet accompanied by five α -helices and contains the PARG-specific catalytic motif GGGx(6-8)QEE within substrate loop 1 (40, 52, 54-56). Most of the PARG-substrate contacts are established via the macrodomain and the function of the accessory domain is less well understood. It was proposed that the latter acts indirectly by stabilizing the macrodomain-fold (19). Structural insights from bacterial and mammalian PARGs show the positioning of the 2',1''-*O*-glycosidic ribose-ribose bond in immediate spatial proximity to the second, catalytic glutamate (Glu756 in human). The catalytic mechanism is distinct from both MacroD1/2 and TARG1 (reviewed in detail in 19, 27, 49) and is initiated through an acid/base protonation of the ribose 2'-OH in the proximal ADPr unit. The resulting oxocarbenium intermediate is stabilized by spatial constraints imposed by a phenylalanine (Phe875 in human PARG) present in substrate loop 2 (PARG

contains the same signature motif as MacroD1/2 in loop 2, see above). The intermediate is resolved through a nucleophilic attack onto the oxocarbenium ion by a water molecule, followed by the release of ADPr and the remaining PAR chain (40, 52, 54, 55). *In vitro* data show that PARG has a preference for chain termini, hence is predominantly an exo-glycohydrolase, with only minor contributions of endo-cleavage. In addition, structural constraints limit the possibility that PARG can cleave at PAR branch points (40, 54, 55). Since endo-cleavage remains a possibility under condition of extreme cellular stress and PARP1 over-activation, PARG could, similarly to TARG1, release oligo- or poly(ADPr) fragments, which in turn were proposed to induce apoptotic cell death (19, 35, 57).

The function of macrodomains

A driving factor for the evolution of macrodomains may lie in increasing complexity of NAD⁺ signaling/consumption processes, regulating DNA repair, redox defense, chromatin architecture, protein acylation and response to viral infection amongst others (58, 59). Support for the latter idea comes from studies of the evolution of PARP genes as well as vertebrate NAD metabolism (20, 60, 61), which suggest that the macrodomain evolution, at least in macroPARPs, is still ongoing (58, 61). While the majority of identified macrodomains, especially in bacteria and archaea, are single-domain proteins, the protein families database (Pfam) accounts for more than 180 different domain architectures containing at least one macrodomain (pfam.xfam.org). Amongst these are homologues of human macrodomain-containing proteins as well as distinct domain combinations in lower organisms (Figure 2). The presence of these proteins suggests a complex interplay between ADPr signaling and large number of different cellular processes. For example, in plants a distinct macrodomain type is fused to the DNA-processing enzymes polynucleotide kinase and Aprataxin suggesting a role in DNA repair (62) (Figure 2). Furthermore, association of macrodomains with NUDIX (nucleoside diphosphate-linked moiety X) domains is commonly

observed, suggesting involvement in the ADPr recycling pathways. NUDIX pyrophosphatases can cleave ADPr into AMP and ribose-5-phosphate for resynthesis of NAD (63, 64) as well as other nucleotides and the human family member NUDT16 was recently shown to act on ADP-ribosylated proteins *in vitro* (65). Further domain compositions suggest the cross-talk of macrodomains with other posttranslational modification, including methylation, phosphorylation and ubiquitination (Figure 2).

Macrodomains as readers of protein ADP-ribosylation and sensors of metabolites

macroH2A

MacroH2A contains an N-terminal histone-fold, allowing integration into chromatin and the C-terminal macrodomain (Figure 2) (66). *Capsaspora owczarzaki*, a *Filasterea*, occupies one of several unicellular sister-groups to Metazoa and is the simplest organism in which macroH2A can be identified. As consequence of a whole genome duplication event in the early vertebrate lineage two *macroH2A* genes, *macroH2A1* and 2, can be identified (Figure 1). Verified in mammals is the further functional diversification of macroH2A through alternative splicing (into isoforms macroH2A1.1 and A1.2), which leads to the exchange of a short stretch of amino acids within the macrodomain (67). Binding studies showed that macroH2A1.1 interacts with ADPr, OAADPr and PAR, while neither macroH2A1.2 nor macroH2A2 can do so (46). Since these two macroH2A variants contain a binding cleft, albeit altered, it is intriguing to speculate whether these proteins have developed an unknown specificity beyond the usual ADPr metabolites.

MacroH2A has been associated with several cellular processes, including cell differentiation and proliferation, transcription repression as well as DNA repair (66, 68, 69). Studies investigating the function of macroH2A during embryonic stem cell (ESC) development indicate that macroH2A1 plays an important role during the early stages of differentiation by

repressing pluripotency genes and facilitating activation of developmental genes, while macroH2A2 stabilizes the cell fate by ‘locking down’ genes of a different cell fate (70, 71). *In vitro* studies on reconstituted nucleosomes suggest that macroH2A1.2 facilitates gene repression by suppression of p300-mediated histone acetylation and blockage of nucleosome sliding and chromatin remodeling (72). *In vivo*, macroH2A1.1 was shown to localize to sites of PARP1 activation in a PAR-dependent manner and to facilitate chromatin compaction (15). The authors hypothesize that PARP activation induces transient macroH2A1.1-dependent chromatin changes.

Reduced macroH2A expression was observed in several cancer types, including breast and lung cancer, and has been associated with increased tumor proliferation and metastatic potential (reviewed in 73). The underlying mechanism is complex and involves loss of transcriptional control over cancer genes (e.g. *CDK8* and *c-Fos*), a dysregulation of the cell cycle and promotion of differentiation. Taken together the available data support a role of macroH2A as epigenetic tumor suppressor.

ALC1

ALC1 (also known as CHD1L) homologues are present in representatives from three eukaryotic supergroups and were probably also present in last eukaryotic common ancestor. They belong to the SNF2 family of DNA helicases and possess two helicase-like domains (ATP binding and C-terminal domain) followed by nuclear localization sequence and a distinct type of macrodomain at the C-terminus (Figure 1 and 2) (5, 47, 74). In mammals ALC1 is involved in DNA damage repair, gene regulation, cell proliferation, development and p53-independent apoptosis (5, 17, 75-77). ALC1 is recruited to sites of DNA damage via its macrodomain by sensing PARP1 generated PAR (5, 47). Recognition of PAR leads both to the formation of a stable ALC1-nucleosome-PARylated PARP1 complex and an increase in the chromatin remodeling activity of ALC1 (5, 47, 78). The function of ALC1 as oncogene

came into focus, through several studies reporting on its involvement in tumorigenesis. ALC1 is frequently amplified in certain cancer types, including hepatocellular carcinoma and bladder cancer (79). Ectopic expression of ALC1 in mice increased the rate of spontaneous tumor formation as well as promoted tumor susceptibility in response to a model of inducible hepatocyte lesion (80). Its overexpression was shown to sensitize cells to DNA damage most likely due to its increased chromatin retention at the damage site (5). In addition, missense mutations near or proximal to the macrodomain, which reduce its affinity for PAR, were identified in patients with congenital anomalies of the kidney and urinal tract (81). Based on the correlation, immunohistology and gene expression analysis it seems that ALC1 plays an important role in kidney development most likely by impairment of its chromatin remodeling function.

macroPARPs

Three human members of the PARP family (9, 14 and 15) contain multiple macrodomains, in addition to their PARP catalytic domain (Figure 2) (11, 27). The macrodomain-PARP ancestor probably arose early in *Unikonts*' (predecessors of *Amebozoa*, *Opisthokonta* and *Apusozoa*) evolution and was most similar to the recent PARP14 homologues (Figure 1) (20). While no PARP activity has as yet been detected for PARP9, both PARP14 and 15 show robust auto-ADP-ribosylation activity (82). All the macroPARPs in humans are encoded within ~200 kbp in the chromosomal 3q20 region that is associated with multiple haematological malignancies (82-84). Likewise, the highest expression of PARP9 and 14 was found in lymphatic tissue (85). In response to DNA damage, PARP9 (also called BAL1) and its binding partner BBAP are recruited to the damage site via the PARP9 macrodomain 2 in a PARP1/PAR-dependent manner. Subsequent, BBAP-mediated histone ubiquitylation serves to recruit further repair factors, including 53BP1 and BRCA1 (86). Experimental data show that PARP14 localizes to the end of actin stress fibers (28). Loss-of-function

experiments showed that, while the focal adhesion assembly appeared to be normal, the fiber turnover was reduced, resulting in elongated cellular protrusions and increased adherence. Hence, PARP14 appears to have an important function in cell morphology and motility.

Dysregulation of both PARP9 and 14 are often associated with lymphoma. For example, amplification of PARP9 in diffuse large B cell lymphomas is associated with increased cell migration and poor prognosis (83, 84). The increase in PARP9 leads to the formation of an ADP-ribosylation/macrodomein-dependent complex between PARP9 and the transcription repressor STAT1 β , followed by translocation into the nucleus and inhibition of the expression of the tumor suppressor IRF1 (87).

Even though no functional data relating to PARP15 cellular activity are available, it is interesting to note that a recent high-throughput genomic study investigating cancer-related gene inactivation identified PARP15 as novel tumor suppressor in head and neck squamous cell carcinoma (88).

GDAP2: A macrodomain protein of unknown function

GDAP2 (ganglioside induced differentiation associated protein 2) is a highly conserved protein of unknown function found from plants to humans. Although its macrodomain is similar to human MacroD1/2 proteins (Figure 1), it does not bind to derivatives of ADPr, but instead appears to have specificity for poly(A) (42). In addition, GDAP2 proteins possess a lipid-binding SEC14 domain at the C-terminus (Figure 2), which is known to act in signal transduction, transport and organelle biology (89), thus a function in the integration of ADPr- and lipid-mediated signaling appears possible.

Macrodomains as erasers of protein ADP-ribosylation

Like other signal transduction pathways, ADPr-dependent signaling requires both recognition and removal of the signal. Therefore, it may not be surprising that macrodomains, in addition to their binding ability, have evolved to reverse cellular ADP-ribosylation. The catalyzed

signal termination reactions included hydrolysis of PAR and MAR as well as degradation of NAD⁺-derived second messengers, such as OAADPr (8, 11, 19, 90).

MacroD1 & 2

Orthologues of MacroD-type proteins can be found in all kingdoms of life. In vertebrates, a duplication of the ancestral MacroD-type gene gave rise to MacroD1 and MacroD2 proteins (Figure 1) (20). MacroD1 and MacroD2 act as mono(ADP-ribosyl) hydrolases that reverse protein MARYlation and catalyze the cleavage of the terminal ADPr moiety e.g from proteins after PARG-mediated polymer degradation (19, 34, 35). In addition, both enzymes can hydrolyze OAADPr (8). While their catalytic activities were established *in vitro*, their exact protein targets and biological role remain largely unknown. There is a high degree of sequence similarity between the catalytic domains of MacroD1 and MacroD2, however their primary subcellular localizations are different (MacroD1 in mitochondria and MacroD2 in the cytoplasm), which implies distinct functions (42). MacroD1 (also called LRP16) was reported to act as a cofactor, modulating estrogen and androgen receptor signaling, and its overexpression was associated with proliferation and invasive growth of cancer cells (91, 92). A recent study showed that MacroD2 reverses the PARP10-mediated mono(ADP-ribosylation) and inhibition of GSK3 β , thus restoring its kinase activity (38). MacroD2 is found amplified in a subset of breast cancers. Interestingly, transgenic overexpression of MacroD2 in breast cancer cell lines results in tamoxifen resistance and estrogen independent growth (93). On the other hand, it has been suggested that MacroD2 may be a cancer-specific fragile site and function as a tumor suppressor (94, 95).

TARG1

TARG1 (terminal ADPr protein glycohydrolase 1; also known as OARD1 or C6orf130) is present in different Metazoans while scattered examples are present in bacteria (Figure 1) (20). Evolutionary analysis suggests that the *TARG1* gene may have arisen from *ALC1* as a

consequence of a partial duplication event. As of yet only limited data are available on the function of TARG1. Biochemical characterization showed that TARG1 can hydrolyze protein ADP-ribosylation (38, 39). In addition, it can also hydrolyze the sirtuin reaction product *OAADPr*, including its acetyl, propionyl and butyryl derivatives, thus yielding ADPr and corresponding carboxylic acid (37). Homozygous mutation of the *TARG1* gene in patients with severe neurodegeneration was observed, while depletion of TARG1 protein in cells leads to proliferation and DNA repair defects and cellular senescence (39).

PARG

While a simple, single-domain type of PARG (called bacterial-type PARG) is found in some bacteria, viruses and sporadically in representatives from all eukaryotic supergroups (except *Plantae*; Figure 1), the vertebrate PARGs contain a more complex structure with regulatory and accessory domains that precede the PARG catalytic macrodomain (Figure 2) (49, 52). In mammals, a single gene encodes PARG and further diversity is achieved through alternative splicing. As a consequence PARG isoforms can be identified in several subcellular compartments: PARG₁₁₁ in the nucleus, PARG₁₀₂ and PARG₉₉ are predominantly cytoplasmic and PAR-degradation deficient PARG₅₅ variant localizes to the mitochondrial matrix (numbers indicate the molecular weight of the corresponding human isoform) (96, 97). During DNA damage PARG hydrolyses PAR generated by DNA damage inducible PARPs, thereby allowing regulation of the later repair events and recycling of NAD⁺ (98, 99). Since PARG is unable to remove the terminal ADPr moiety from proteins (40), TARG1, MacroD1 and/or MacroD2 are suggested to catalyze this reaction (34, 38, 39). The knockout of all PARG isoforms leads to increased apoptosis and embryonic lethality in mouse and fruit fly (100, 101). However, sole knockout of the longest isoform, PARG₁₁₀ in mice, is viable, most likely due to nucleocytoplasmic shuttling of the cytoplasmic isoform during DNA damage (102, 103). Notwithstanding, PARG₁₁₀^{-/-} mice show hypersensitivity to genotoxic stress

induced both by ionizing radiation or alkylating agents (101, 104). In addition, PARG was implicated as modulator of the inflammatory response and PARG inactivation (PARG^{110-/-} mice and inhibition) had protective effects in models of dinitrobenzene sulfonic acid-induced colitis, splanchnic artery occlusion shock and ischemia-reperfusion as well as the associated injuries (105-109).

Generation of ADPr: A second messenger in calcium-signaling

In the context of ‘erasing’ cellular ADP-ribosylation it is worth noting that over-activation of PARP1 has long been associated with cell death (110-112). Though the underlying mechanism is still elusive, an enticing hypothesis can be raised from several observations regarding macrodomain- and PARG-derived ADPr. Under condition of oxidative stress the activation of the TRMP2 (transient receptor potential melastatin 2), a calcium-permeable cation channel, appears to be PARP1-dependent (113), but the high PAR-degrading activity of PARG localized to nucleus and cytoplasm suggests an indirect mechanism via ADPr (114). Indeed, it has been shown that ADPr is specific activator of TRPM2 via binding to its cytosolic NUDIX9-homology domain (41, 115, 116). As a result the concentration of free intracellular calcium is increased and cell death induced (41). A second, potentially synergetic, pathway requires a further degradation of ADPr to AMP and ribose-5-phosphate, a reaction efficiently catalyzed by the pyrophosphatases NUDT5 and NUDT9 (117, 118). The resulting accumulation of AMP acts as an inhibitor of the mitochondrial ADP/ATP translocase, thus preventing the nucleotide exchange with the cytoplasm and resulting in mitochondrial energy failure (117).

The sirtuin reaction product *OAADPr* has also been suggested to function as a second messenger, among other by direct interaction with macrodomain-containing proteins and both MacroD1/2 and TARG1 have been implicated in the signal clearance through *OAADPr* hydrolysis (8, 37, 68, 90).

Macrodomains in microorganisms and pathogens

Bacterial macrodomains

Macrodomain-containing proteins are found in most of the sequenced bacterial species and are most commonly of the MacroD-type (Figure 1). Overall, very little is known about intracellular ADP-ribosylation in bacteria. The first described example came from the bacterial and archaeal nitrogen fixation pathway. This highly energy demanding reaction sequence is regulated via reversible ADP-ribosylation of an arginine residue in the key enzyme nitrogenase (119). ADP-ribosylation has also been detected in species of the bacterial genera *Streptomyces* and *Myxococcus*, however the corresponding protein ADP-ribosyltransferases are still unknown (120, 121). More recently an analysis of the sequences, structures and genomic contexts of NAD-utilizing enzymes predicated novel classes of ADP-ribosyltransferases (59). Some of these exhibit genomic association with macrodomain or PARG homologs, and thus implying the wider presence of pathways regulated by ADP-ribosylation in bacteria. If present in pathogens such macrodomain-containing operons could be of additional interest, as they may be linked to virulence.

Recently we characterized a new system for reversible protein ADP-ribosylation present in pathogens such as *Staphylococcus aureus* and *Streptococcus pyogenes* (25). Central to this system is a macrodomain/sirtuin pair in which the sirtuin component (due to its macrodomain-association termed SirTM) can ADP-ribosylate a target protein derived from the same operon, whereas the MacroD-type macrodomain reverses the modification. Interestingly, the ADP-ribosylation activity of the SirTM is dependent on another protein modification, lipoylation. The latter is known to be involved in the regulation of energy metabolism and detoxification of reactive oxygen species, thereby contributing to the bacterial pathogenesis (reviewed in 122). Current data indicate that the SirTM/macrodomain

pair modulates the response to oxidative stress both in bacteria and in fungi possessing a homologous system such as *Candida albicans* (Figure 2) (25, 123, 124).

Genomic evidence suggests that the association of macrodomains and pathogenesis may be quite common. For example, a putative macrodomain-involving toxin/antitoxin system has been recently discovered in *Pseudomonas mendocina* by a large-scale cloning approach (125). The system is composed of a toxin of unknown function and a bipartite antitoxin containing an N-terminal macrodomain. The macrodomain is a diverged representative within the ALC1-like class and its orthologues can be identified in other pathogens such as *Mycobacterium tuberculosis* as well as *Vibrio* and *Xanthomonas* species (Figure 1).

These examples suggest that the bacterial macrodomains may influence processes, which have been shown to be crucial for the survival and virulence of bacteria in the host environment, thus promoting the study of these macrodomains to further our understanding of host-pathogen interactions and explore novel therapeutic routes.

Viral macrodomains

Similarly, evidence is emerging that targeting viral macrodomains could be a promising therapeutic approach to promote recovery and reduce infection severity (18, 126, 127). The evolution of viral macrodomains is peculiar and biochemical, structural as well as phylogenetic evidences indicate that viral and cellular macrodomains are strongly related (Figure 1) (42). Several different types of macrodomains are found in different virus genomes and it has been suggested that this is due to repeated, independent host-acquisition of macrodomains (42, 128, 129). Together this argues for specific and different functions in the virus-host coevolution. Macrodomain proteins were identified in more than 150 viruses, most belonging to family *Myoviridae* (which includes various phages) and are also present in other virus families: *Coronaviridae*, *Togaviridae*, *Iridoviridae*, *Poxviridae* and *Hepeviridae* (present in only one species, hepatitis E virus) (45, 130, 131). Heretofore, no studies have

been conducted on phage macrodomains, while the ones from RNA viruses are being investigated for their potential as druggable targets.

Both in corona- and alphaviruses the macrodomain (MacroD-type) is encoded as part of the multidomain non-structural protein 3 (nsP3), which is implicated to participate in host protein recruitment and viral replication (132, 133). However, beyond this broad classification the nsP3 proteins of *Togaviridae* and *Coronaviridae* are distinct in their domain makeup and consequently, have only limited overlapping functions (134, 135).

In vitro studies showed that macrodomains from both origins can hydrolyze ADPr-1''-phosphate and bind to PAR (45, 129, 136, 137) but whether this reflects the endogenous function remains elusive. On the other hand, sequential and structural similarity to human MacroD1 and MacroD2 strongly indicate that MacroD-type viral proteins might exhibit de-MARylation activity (34). This is particularly exciting as the expression of the so called antiviral PARPs (PARP7, 10 and 12) is triggered via the type I interferon response (138, 139). The activity of the antiviral PARPs decreases the rate of cellular translation and viral replication, thus counteracting the production of viral particles. Since both PARP7, 10 and 12 are mono(ADP-ribosyl) transferases (28) and their antiviral function is dependent on their catalytic activity (139), it is intriguing to speculate that the viral macrodomains have evolved to counteract this host defense.

The macrodomains appear, however, to fulfill further functions within the viruses. In a study of the Sindbis virus nsP3 macrodomain, mutagenesis had a complex phenotype. While the replication in cultured, mature neurons was macrodomain-dependent, titer levels in mice were not affected 24 hr post-infection (136). This indicates that the macrodomain, while possibly influencing replication, is not crucial for its success. Similarly, the coronaviral macrodomain appears to be dispensable for replication in IFN-deficient cells (137). However, mouse MHV infection models, using a virus strain that carried a catalytic null mutation of the

macrodomain, led to reduction of viral virulence (18, 126). It was suggested that the milder course of infection was caused by a reduction in the proinflammatory immune response, which is associated with server tissue damage in certain viral infections (140, 141).

SARS-unique domain: A RNA binding macrodomain

In addition to MacroD-type macrodomains, the nsP3 proteins of SARS-CoV and other β -coronaviruses contain a highly diverged macrodomain found in the SARS-unique domains (SUD) (Figure 1). Analysis of the structure of SUD from SARS-CoV revealed that this domain is comprised of two macrodomains (SUD-N and –M, respectively) and a C-terminal frataxin-fold domain (SUD-C) (9, 10). However it is worth noting that the sequence comparison data indicate that most SUD-containing viruses carry only one of the macrodomains, a SUD-M homologue. *In vitro* studies have demonstrated that the SARS-CoV macrodomains bind nucleic acids with preference for purine-rich RNA sequences, such as RNA G-quadruplexes, and that SUD-C increases the target specificity (9, 10). The importance of this interaction was recently highlighted by demonstrating that the binding of RNA by SUD-M is required for viral replication (142).

The therapeutic potential of macrodomain inhibitors

The study of macrodomains in their cellular context gave first insights into their pathobiological importance. Consequently, new strategies are explored to target macrodomains in order to treat or relief the burden of diseases.

Viral infections

In several viral systems intricate connections between nsP3 macrodomain function and the host immune system were described. MHV (an α -coronavirus) infection models showed that a catalytic mutation within the macrodomain attenuated the infection severity by preventing over-activation of the immune response (18, 126). In contrast, mutation of the same catalytic residue in β -coronaviruses prevented the suppression of the type I interferon response and

increased the viral susceptibility to the innate immune response (127). Despite the difference in the underlying mechanisms, these results encourage further investigation into viral macrodomains as their inhibition may help to reduce disease burden and assist recovery.

Macrodomains in cancer

Next to viral infection, the involvement of macrodomains in tumorigenesis is best studied. It has been shown that several macrodomain-containing proteins are amplified in different cancer types, which correlates with a poor prognosis and/or drug resistance. Examples for a direct link between macrodomain amplification and tumor biology include: ALC1, which can inhibit apoptosis (5, 79), MacroD2, which can induce tamoxifen resistance in estrogen receptor-positive breast cancer cells (93), and PARP9, which can increase tumor cell migration (83, 84). Altogether these observations promote the idea that macrodomain targeting could slow tumor progression and support therapy.

A second strategy for a cancer therapy, based on ‘synthetic lethality’, involves inhibition of PARG. This approach exploits the reduced DNA repair capability of BRCA1- or BRCA2-deficient tumors similarly to PARP inhibitors (143-145).

Outlook

Notwithstanding ongoing efforts (52, 146, 147), so far no soluble, efficient, bioavailable inhibitor against a macrodomain has been developed. However, the increasing numbers of available macrodomain structures, together with advancements in structure-guided drug design, have the potential to accelerate the processes. The emergence of rhodamine and phenolic hydrazide hydrazone based compounds, which exhibit PARG inhibition *in vitro*, is a further reason for optimism. However, one of the greatest remaining challenges is the limited availability of data on the biological function of macrodomains. This is true for virtually all macrodomains studied so far, as most of the data regarding binding or catalysis have been

derived from *in vitro* observations. In this context potent inhibitors would become a valuable tool to characterize macrodomains *in vivo*.

The specificity of macrodomains has been extensively studied, however, only a limited number of ADP-ribosylated metabolites have been tested as ligands/substrates for most of the macrodomains. One example for an untested compound is diadenosine 5',5'''-P¹,P⁴-tetrphosphate (Ap₄A) (an ubiquitous second messenger synthesized from ATP under condition of cellular stress), which was shown to be ADP-ribsoylated over three decades ago (148). While Ap₄A ADP-ribosylation was believed to occur only under non-physiological conditions (149), a recent study demonstrated that the concentration of both Ap₄A and its ADP-ribosylated form increase under conditions of sub-lethal stress and may act in the DNA damage response (150). Investigation of ADP-ribosylated Ap₄A and other metabolites as potential substrates for macrodomains will further our understanding of their diverse function and physiological importance.

Acknowledgments

We would like to apologize to all colleagues whose work could not be included due to space restrictions. We are grateful to Gytis Jankevicius and Kerryanne Crawford for critical comments on the manuscript. The work in I.A. laboratory is supported by the Wellcome Trust and the European Research Council.

References

1. Till S, Ladurner AG. 2009. *Front Biosci (Landmark Ed)* 14: 3246-58
2. Lee HJ, Shieh CK, Gorbalenya AE, Koonin EV, La Monica N, et al. 1991. *Virology* 180: 567-82
3. Pehrson JR, Fried VA. 1992. *Science* 257: 1398-400
4. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. 2007. *Nat Struct Mol Biol* 14: 1025-40
5. Ahel D, Horejsi Z, Wiechens N, Polo SE, Garcia-Wilson E, et al. 2009. *Science* 325: 1240-3
6. Karras GI, Kustatscher G, Buhecha HR, Allen MD, Pugieux C, et al. 2005. *EMBO J* 24: 1911-20
7. Kraus WL. 2009. *Nat Struct Mol Biol* 16: 904-7
8. Chen D, Vollmar M, Rossi MN, Phillips C, Kraehenbuehl R, et al. 2011. *J Biol Chem* 286: 13261-71
9. Johnson MA, Chatterjee A, Neuman BW, Wuthrich K. 2010. *J Mol Biol* 400: 724-42
10. Tan J, Vonnrhein C, Smart OS, Bricogne G, Bollati M, et al. 2009. *PLoS Pathog* 5: e1000428
11. Feijs KL, Forst AH, Verheugd P, Luscher B. 2013. *Nat Rev Mol Cell Biol* 14: 443-51
12. Han W, Li X, Fu X. 2011. *Mutat Res* 727: 86-103
13. Maas NM, Van de Putte T, Melotte C, Francis A, Schrandt-Stumpel CT, et al. 2007. *J Med Genet* 44: 562-9
14. Buschbeck M, Uribealago I, Wibowo I, Rue P, Martin D, et al. 2009. *Nat Struct Mol Biol* 16: 1074-9
15. Timinszky G, Till S, Hassa PO, Hothorn M, Kustatscher G, et al. 2009. *Nat Struct Mol Biol* 16: 923-9
16. Kim W, Chakraborty G, Kim S, Shin J, Park CH, et al. 2012. *J Biol Chem* 287: 5278-89
17. Chen L, Hu L, Chan TH, Tsao GS, Xie D, et al. 2009. *Hepatology* 50: 122-9
18. Eriksson KK, Cervantes-Barragan L, Ludewig B, Thiel V. 2008. *J Virol* 82: 12325-34
19. Barkauskaite E, Jankevicius G, Ladurner AG, Ahel I, Timinszky G. 2013. *FEBS J* 280: 3491-507
20. Perina D, Mikoc A, Ahel J, Cetkovic H, Zaja R, Ahel I. 2014. *DNA Repair (Amst)* 23: 4-16
21. Feijs KL, Verheugd P, Luscher B. 2013. *FEBS J* 280: 3519-29
22. Gibson BA, Kraus WL. 2012. *Nat Rev Mol Cell Biol* 13: 411-24
23. Denu JM. 2005. *Curr Opin Chem Biol* 9: 431-40
24. Sauve AA, Youn DY. 2012. *Curr Opin Chem Biol* 16: 535-43
25. Rack JG, Morra R, Barkauskaite E, Kraehenbuehl R, Ariza A, et al. 2015. *Mol Cell* 59: 309-20
26. Kowieski TM, Lee S, Denu JM. 2008. *J Biol Chem* 283: 5317-26
27. Barkauskaite E, Jankevicius G, Ahel I. 2015. *Mol Cell* 58: 935-46
28. Vyas S, Matic I, Uchima L, Rood J, Zaja R, et al. 2014. *Nat Commun* 5: 4426
29. Kleine H, Poreba E, Lesniewicz K, Hassa PO, Hottiger MO, et al. 2008. *Mol Cell* 32: 57-69
30. Langelier MF, Pascal JM. 2013. *Curr Opin Struct Biol* 23: 134-43
31. Miwa M, Ishihara M, Takishima S, Takasuka N, Maeda M, et al. 1981. *J Biol Chem* 256: 2916-21

32. Martzen MR, McCraith SM, Spinelli SL, Torres FM, Fields S, et al. 1999. *Science* 286: 1153-5
33. Shull NP, Spinelli SL, Phizicky EM. 2005. *Nucleic Acids Res* 33: 650-60
34. Jankevicius G, Hassler M, Golia B, Rybin V, Zacharias M, et al. 2013. *Nat Struct Mol Biol* 20: 508-14
35. Barkauskaite E, Brassington A, Tan ES, Warwicker J, Dunstan MS, et al. 2013. *Nat Commun* 4: 2164
36. Miwa M, Sugimura T. 1971. *J Biol Chem* 246: 6362-4
37. Peterson FC, Chen D, Lytle BL, Rossi MN, Ahel I, et al. 2011. *J Biol Chem* 286: 35955-65
38. Rosenthal F, Feijs KL, Frugier E, Bonalli M, Forst AH, et al. 2013. *Nat Struct Mol Biol* 20: 502-7
39. Sharifi R, Morra R, Appel CD, Tallis M, Chioza B, et al. 2013. *EMBO J* 32: 1225-37
40. Slade D, Dunstan MS, Barkauskaite E, Weston R, Lafite P, et al. 2011. *Nature* 477: 616-20
41. Fliegert R, Gasser A, Guse AH. 2007. *Biochem Soc Trans* 35: 109-14
42. Neuvonen M, Ahola T. 2009. *J Mol Biol* 385: 212-25
43. Chatterjee A, Johnson MA, Serrano P, Pedrini B, Joseph JS, et al. 2009. *J Virol* 83: 1823-36
44. Allen MD, Buckle AM, Cordell SC, Lowe J, Bycroft M. 2003. *J Mol Biol* 330: 503-11
45. Egloff MP, Malet H, Putics A, Heinonen M, Dutartre H, et al. 2006. *J Virol* 80: 8493-502
46. Kustatscher G, Hothorn M, Pugieux C, Scheffzek K, Ladurner AG. 2005. *Nat Struct Mol Biol* 12: 624-5
47. Gottschalk AJ, Timinszky G, Kong SE, Jin J, Cai Y, et al. 2009. *Proc Natl Acad Sci U S A* 106: 13770-4
48. Forst AH, Karlberg T, Herzog N, Thorsell AG, Gross A, et al. 2013. *Structure* 21: 462-75
49. Zaja R, Mikoc A, Barkauskaite E, Ahel I. 2012. *Biomolecules* 3: 1-17
50. Tallis M, Morra R, Barkauskaite E, Ahel I. 2014. *Chromosoma* 123: 79-90
51. Hirsch BM, Burgos ES, Schramm VL. 2014. *ACS Chem Biol* 9: 2255-62
52. Dunstan MS, Barkauskaite E, Lafite P, Knezevic CE, Brassington A, et al. 2012. *Nat Commun* 3: 878
53. Patel CN, Koh DW, Jacobson MK, Oliveira MA. 2005. *Biochem J* 388: 493-500
54. Tucker JA, Bennett N, Brassington C, Durant ST, Hassall G, et al. 2012. *PLoS One* 7: e50889
55. Kim IK, Kiefer JR, Ho CM, Stegeman RA, Classen S, et al. 2012. *Nat Struct Mol Biol* 19: 653-6
56. Lambrecht MJ, Brichacek M, Barkauskaite E, Ariza A, Ahel I, Hergenrother PJ. 2015. *J Am Chem Soc* 137: 3558-64
57. Andrabi SA, Kim NS, Yu SW, Wang H, Koh DW, et al. 2006. *Proc Natl Acad Sci U S A* 103: 18308-13
58. Daugherty MD, Young JM, Kerns JA, Malik HS. 2014. *PLoS Genet* 10: e1004403
59. de Souza RF, Aravind L. 2012. *Mol Biosyst* 8: 1661-77
60. Citarelli M, Teotia S, Lamb RS. 2010. *BMC Evol Biol* 10: 308
61. Gossmann TI, Ziegler M. 2014. *DNA Repair (Amst)* 23: 39-48
62. Rass U, Ahel I, West SC. 2008. *J Biol Chem* 283: 33994-4001
63. McLennan AG. 2006. *Cell Mol Life Sci* 63: 123-43
64. Tong L, Lee S, Denu JM. 2009. *J Biol Chem* 284: 11256-66

65. Palazzo L, Thomas B, Jemth AS, Colby T, Leidecker O, et al. 2015. *Biochem J* 468: 293-301
66. Buschbeck M, Di Croce L. 2010. *Epigenetics* 5: 118-23
67. Rasmussen TP, Huang T, Mastrangelo MA, Loring J, Panning B, Jaenisch R. 1999. *Nucleic Acids Res* 27: 3685-9
68. Posavec M, Timinszky G, Buschbeck M. 2013. *Cell Mol Life Sci* 70: 1509-24
69. Turinetto V, Giachino C. 2015. *Epigenetics* 10: 563-73
70. Creppe C, Janich P, Cantarino N, Noguera M, Valero V, et al. 2012. *Mol Cell Biol* 32: 1442-52
71. Yildirim O, Hung JH, Cedeno RJ, Weng Z, Lengner CJ, Rando OJ. 2014. *PLoS Genet* 10: e1004515
72. Doyen CM, An W, Angelov D, Bondarenko V, Miettton F, et al. 2006. *Mol Cell Biol* 26: 1156-64
73. Cantarino N, Douet J, Buschbeck M. 2013. *Cancer Lett* 336: 247-52
74. Ryan DP, Owen-Hughes T. 2011. *Curr Opin Chem Biol* 15: 649-56
75. Chen L, Chan TH, Yuan YF, Hu L, Huang J, et al. 2010. *J Clin Invest* 120: 1178-91
76. Ji X, Li J, Zhu L, Cai J, Zhang J, et al. 2013. *J Surg Res* 185: 84-91
77. Snider AC, Leong D, Wang QT, Wysocka J, Yao MW, Scott MP. 2013. *Biol Open* 2: 121-31
78. Gottschalk AJ, Trivedi RD, Conaway JW, Conaway RC. 2012. *J Biol Chem* 287: 43527-32
79. Cheng W, Su Y, Xu F. 2013. *Mol Cancer* 12: 170
80. Chen M, Huang JD, Hu L, Zheng BJ, Chen L, et al. 2009. *PLoS One* 4: e6727
81. Brockschmidt A, Chung B, Weber S, Fischer DC, Kolatsi-Joannou M, et al. 2012. *Nephrol Dial Transplant* 27: 2355-64
82. Aguiar RC, Takeyama K, He C, Kreinbrink K, Shipp MA. 2005. *J Biol Chem* 280: 33756-65
83. Aguiar RC, Yakushijin Y, Kharbanda S, Salgia R, Fletcher JA, Shipp MA. 2000. *Blood* 96: 4328-34
84. Juszczynski P, Kutok JL, Li C, Mitra J, Aguiar RC, Shipp MA. 2006. *Mol Cell Biol* 26: 5348-59
85. Hakme A, Huber A, Dolle P, Schreiber V. 2008. *Dev Dyn* 237: 209-15
86. Yan Q, Xu R, Zhu L, Cheng X, Wang Z, et al. 2013. *Mol Cell Biol* 33: 845-57
87. Camicia R, Bachmann SB, Winkler HC, Beer M, Tinguely M, et al. 2013. *J Cell Sci* 126: 1969-80
88. Guerrero-Preston R, Michailidi C, Marchionni L, Pickering CR, Frederick MJ, et al. 2014. *Epigenetics* 9: 1031-46
89. Saito K, Tautz L, Mustelin T. 2007. *Biochim Biophys Acta* 1771: 719-26
90. Tong L, Denu JM. 2010. *Biochim Biophys Acta* 1804: 1617-25
91. Han WD, Zhao YL, Meng YG, Zang L, Wu ZQ, et al. 2007. *Endocr Relat Cancer* 14: 741-53
92. Yang J, Zhao YL, Wu ZQ, Si YL, Meng YG, et al. 2009. *Endocr Relat Cancer* 16: 139-53
93. Mohseni M, Cidado J, Croessmann S, Cravero K, Cimino-Mathews A, et al. 2014. *Proc Natl Acad Sci U S A* 111: 17606-11
94. Rajaram M, Zhang J, Wang T, Li J, Kuscus C, et al. 2013. *PLoS One* 8: e66264
95. Yao F, Kausalya JP, Sia YY, Teo AS, Lee WH, et al. 2015. *Cell Rep* 12: 272-85
96. Meyer-Ficca ML, Meyer RG, Coyle DL, Jacobson EL, Jacobson MK. 2004. *Exp Cell Res* 297: 521-32

97. Niere M, Mashimo M, Agledal L, Dolle C, Kasamatsu A, et al. 2012. *J Biol Chem* 287: 16088-102
98. Feng X, Koh DW. 2013. *Int Rev Cell Mol Biol* 304: 227-81
99. Nikiforov A, Kulikova V, Ziegler M. 2015. *Crit Rev Biochem Mol Biol*: 1-14
100. Hanai S, Kanai M, Ohashi S, Okamoto K, Yamada M, et al. 2004. *Proc Natl Acad Sci U S A* 101: 82-6
101. Koh DW, Lawler AM, Poitras MF, Sasaki M, Wattler S, et al. 2004. *Proc Natl Acad Sci U S A* 101: 17699-704
102. Haince JF, Ouellet ME, McDonald D, Hendzel MJ, Poirier GG. 2006. *Biochim Biophys Acta* 1763: 226-37
103. Mortusewicz O, Fouquerel E, Ame JC, Leonhardt H, Schreiber V. 2011. *Nucleic Acids Res* 39: 5045-56
104. Cortes U, Tong WM, Coyle DL, Meyer-Ficca ML, Meyer RG, et al. 2004. *Mol Cell Biol* 24: 7163-78
105. Burkle A, Virag L. 2013. *Mol Aspects Med* 34: 1046-65
106. Cuzzocrea S, Di Paola R, Mazzon E, Cortes U, Genovese T, et al. 2005. *FASEB J* 19: 558-66
107. Cuzzocrea S, Wang ZQ. 2005. *Pharmacol Res* 52: 100-8
108. Genovese T, Di Paola R, Catalano P, Li JH, Xu W, et al. 2004. *Crit Care Med* 32: 1365-74
109. Patel NS, Cortes U, Di Paola R, Mazzon E, Mota-Filipe H, et al. 2005. *J Am Soc Nephrol* 16: 712-9
110. Burkle A. 2001. *Bioessays* 23: 795-806
111. Chiarugi A. 2002. *Trends Pharmacol Sci* 23: 122-9
112. Ha HC, Snyder SH. 1999. *Proc Natl Acad Sci U S A* 96: 13978-82
113. Fonfria E, Marshall IC, Benham CD, Boyfield I, Brown JD, et al. 2004. *Br J Pharmacol* 143: 186-92
114. Dolle C, Niere M, Lohndal E, Ziegler M. 2010. *Cell Mol Life Sci* 67: 433-43
115. Perraud AL, Fleig A, Dunn CA, Bagley LA, Launay P, et al. 2001. *Nature* 411: 595-9
116. Toth B, Iordanov I, Csanady L. 2015. *J Gen Physiol* 145: 419-30
117. Formentini L, Macchiarulo A, Cipriani G, Camaioni E, Rapizzi E, et al. 2009. *J Biol Chem* 284: 17668-76
118. Lin S, Gasmi L, Xie Y, Ying K, Gu S, et al. 2002. *Biochim Biophys Acta* 1594: 127-35
119. Nordlund S, Hogbom M. 2013. *FEBS J* 280: 3484-90
120. Eastman D, Dworkin M. 1994. *Microbiology* 140 (Pt 11): 3167-76
121. Penyige A, Keseru J, Fazakas F, Schmelczler I, Szirak K, et al. 2009. *J Microbiol* 47: 549-56
122. Spalding MD, Prigge ST. 2010. *Microbiol Mol Biol Rev* 74: 200-28
123. Enjalbert B, Rachini A, Vedyappan G, Pietrella D, Spaccapelo R, et al. 2009. *Infect Immun* 77: 4847-58
124. Surmann K, Michalik S, Hildebrandt P, Gierok P, Depke M, et al. 2014. *Front Microbiol* 5: 392
125. Sberro H, Leavitt A, Kiro R, Koh E, Peleg Y, et al. 2013. *Mol Cell* 50: 136-48
126. Fehr AR, Athmer J, Channappanavar R, Phillips JM, Meyerholz DK, Perlman S. 2015. *J Virol* 89: 1523-36
127. Kuri T, Eriksson KK, Putics A, Zust R, Snijder EJ, et al. 2011. *J Gen Virol* 92: 1899-905
128. Basta HA, Cleveland SB, Clinton RA, Dimitrov AG, McClure MA. 2009. *J Virol* 83: 10152-62

129. Malet H, Coutard B, Jamal S, Dutartre H, Papageorgiou N, et al. 2009. *J Virol* 83: 6534-45
130. Malet H, Dalle K, Bremond N, Tocque F, Blangy S, et al. 2006. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 62: 405-8
131. Putics A, Slaby J, Filipowicz W, Gorbalenya AE, Ziebuhr J. 2006. *Adv Exp Med Biol* 581: 93-6
132. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. 2006. *Virus Res* 117: 17-37
133. LaStarza MW, Lemm JA, Rice CM. 1994. *J Virol* 68: 5781-91
134. Neuman BW, Joseph JS, Saikatendu KS, Serrano P, Chatterjee A, et al. 2008. *J Virol* 82: 5279-94
135. Neuvonen M, Kazlauskas A, Martikainen M, Hinkkanen A, Ahola T, Saksela K. 2011. *PLoS Pathog* 7: e1002383
136. Park E, Griffin DE. 2009. *Virology* 388: 305-14
137. Putics A, Filipowicz W, Hall J, Gorbalenya AE, Ziebuhr J. 2005. *J Virol* 79: 12721-31
138. Atasheva S, Akhrymuk M, Frolova EI, Frolov I. 2012. *J Virol* 86: 8147-60
139. Atasheva S, Frolova EI, Frolov I. 2014. *J Virol* 88: 2116-30
140. Bertoletti A, Maini MK. 2000. *Curr Opin Microbiol* 3: 387-92
141. Gu J, Korteweg C. 2007. *Am J Pathol* 170: 1136-47
142. Kusov Y, Tan J, Alvarez E, Enjuanes L, Hilgenfeld R. 2015. *Virology* 484: 313-22
143. Fathers C, Drayton RM, Solovieva S, Bryant HE. 2012. *Cell Cycle* 11: 990-7
144. Feng FY, de Bono JS, Rubin MA, Knudsen KE. 2015. *Mol Cell* 58: 925-34
145. Min W, Wang ZQ. 2009. *Front Biosci (Landmark Ed)* 14: 1619-26
146. Finch KE, Knezevic CE, Nottbohm AC, Partlow KC, Hergenrother PJ. 2012. *ACS Chem Biol* 7: 563-70
147. Islam R, Koizumi F, Kodera Y, Inoue K, Okawara T, Masutani M. 2014. *Bioorg Med Chem Lett* 24: 3802-6
148. Tanaka Y, Matsunami N, Itaya A, Yoshihara K. 1981. *J Biochem* 90: 1131-9
149. Baltzinger M, Ebel JP, Remy P. 1986. *Biochimie* 68: 1231-6
150. Marriott AS, Copeland NA, Cunningham R, Wilkinson MC, McLennan AG, Jones NJ. 2015. *DNA Repair (Amst)* 33: 90-100

Acronyms and Definitions

Reading

The recognition of a posttranslational modification by protein domain.

Erasing

The removal of a posttranslational modification from a protein.

Interpreting

The spatial/temporal recognition of a posttranslational modification, which trigger a series of downstream events by effector proteins.

Synthetic lethality

Two mutations, which occurring alone create a viable, but synergistically a lethal phenotype.

Table 1 Distribution of macrodomains amongst different taxa.

			Macrodomain classes ^{a,b}					
			MacroD	macroH2A	ALC1	Macro2	PARG	SUD-M
Eukaryotes								
Ophisthokonta	human	<i>Homo sapiens</i>	3	9(5)	2	-	1	-
	mouse	<i>Mus musculus</i>	3	8(4)	2	-	1	-
	chicken	<i>Gallus gallus</i>	2	10(4)	2	-	1	-
	lizard	<i>Anolis carolinensis</i>	3	7(4)	2	-	1	-
	frog	<i>Xenopus tropicalis</i>	3	22(10)	1	-	1	-
	fish	<i>Danio rerio</i>	3	23(10)	2	-	2	-
	lancelet	<i>Branchiostoma floridae</i>	2	16(10)	1	-	2	-
	sea urchin	<i>Strongylocentrotus purpuratus</i>	2	1	2	1	2	-
	sea slug	<i>Aplysia californica</i>	2	5(3)	1	2	2	-
	fruit fly	<i>Drosophila melanogaster</i>	1	-	3	-	1	-
	roundworm	<i>Caenorhabditis elegans</i>	1	-	-	-	2	-
	cnidarian	<i>Hydra vulgaris</i>	3	1	2	-	1	-
	placozoan	<i>Trichoplax adhaerens</i>	2	10(9)	1	-	1	-
	sponge	<i>Amphimedon queenslandica</i>	2	20(9)	3	1	3	-
	choanoflagellate	<i>Monosiga brevicollis</i>	1	-	-	-	2	-
	filose amoeboid	<i>Capsaspora owczarzaki</i>	6	1	2	-	3	-
	yeast	<i>Saccharomyces cerevisiae</i>	-	-	1	1	-	-
Amoebozoa	slime mold	<i>Dictyostelium discoideum</i>	2	1	1	-	1	-
Excavata	african trypanosome	<i>Trypanosoma brucei</i>	1	-	-	-	1	-
Chromalveolata		<i>Phytophthora nicotianae</i>	3	-	1	2	1	-
Plantae	mouse-ear cress	<i>Arabidopsis thaliana</i>	2	-	1	-	2	-
Rhizaria	chlorarachniophyte alga	<i>Bigeloviella natans</i>	2	-	-	-	5	-
Prokaryotes								
Archaea		<i>Archaeoglobus fulgidus</i>	1	-	-	-	-	-
Bacteria		<i>Escherichia coli</i>	1	-	-	-	-	-
		<i>Streptomyces coelicolor</i>	1	-	1	-	1	-
Viruses								
Coronaviridae		SARS	1	-	-	-	-	1
Hepeviridae		HEV	1	-	-	-	-	-
Togaviridae		SFV	1	-	-	-	-	-

^a Due to a the limitations of available annotation data the table does not consider protein isoforms derived from alternative splicing

^b Number of proteins is given in parentheses

Figure legends

Figure 1 Phylogenetic representation of evolutionary relationships between different macrodomain sequences.

The phylogram shows the six major macrodomain branches and includes eukaryotic (black), archaeal (blue), bacterial (green) and viral (red) representatives. The scale bar indicates the genetic distance of the branch lengths and accession numbers of sequences used are given after species and protein names.

Figure 2 Schematic architecture of domains present in macrodomain representatives.

Schematic representation depicts the domain organization of macrodomain-containing proteins from human as well as several other interesting domain architectures from other organisms. The protein length (in amino acids) is given on the right and domain boundaries on top of the corresponding scheme and correspond to the SMART/Pfam database. Domain abbreviations: AAA, ATPase associated with diverse cellular Activities; Deltex_C, Deltex C-terminal domain; Duf3729, domain of unknown function 3729; H2A, core histone H2A-like domain; HELICc, Helicase conserved C-terminal domain; HIT, histidine tirade domain; HLH, Helix-Loop-Helix DNA binding domain; Macro, macrodomain; NUDIX, nucleoside diphosphate-linked moiety X; PARP cat., PARP catalytic domain; RdRP_2, RNA dependent RNA polymerase; RING, (Really Interesting New Gene) finger domain; Peptidase_C41, Hepatitis E cysteine protease; Pkinase, protein kinase domain; RRM, RNA recognition motif; SEC14, named after yeast Sec14p; Sir2, sirtuin domain; SNF2_N, SNF2 helicase family N-terminal domain; Vmethyltransf, Viral methyltransferase; WWE, named after the conserved tryptophan and glutamate residues; ZnF, zinc finger

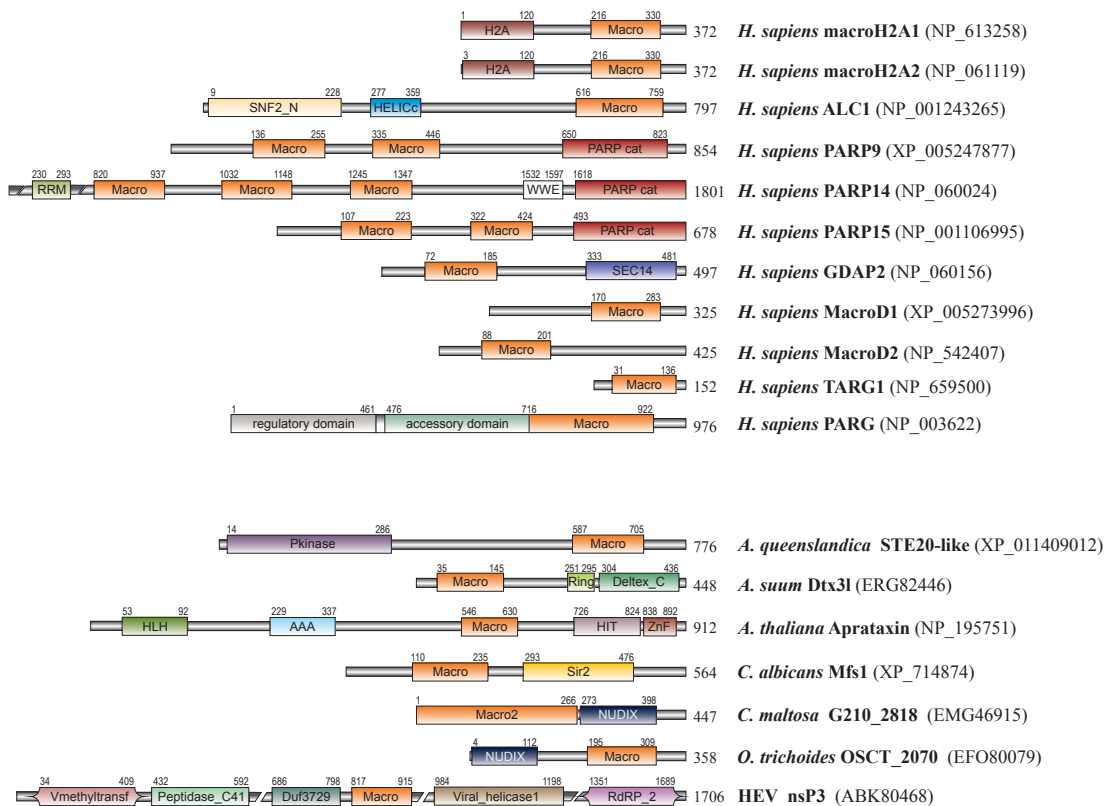


Figure 2

Figure 3 Structure and reaction mechanism of the human hydrolytic macrodomain classes.

(a) Topological representation of the macrodomain shows the organization of the central six-stranded β -sheet (red) flanked on both sides by 5 α -helices (green). Note that several macrodomains can contain additional α -helices and/or β -sheets. (b-e) The ribbon and surface representations of four human macrodomains illustrate the binding of ADPr (b+c), lysyl-ADPr intermediate (d) or dimeric ADPr (e) (yellow). The coordination of the adenosine moiety by a conserved phenylalanine and/or asparagine residue as well as the substrate binding loops 1 and 2 (red) are depicted in the magnification. (b) MacroH2A1.1 (blue; pdb: 3IID) as a reader macrodomain coordinates ADPr tightly in a relaxed conformation. (c) MacroD2 (green; pdb: 4IQY) coordinates ADPr in a strained conformation due to the presence of a structural water molecule (dark blue) and Tyr190 (loop 2). The catalytic residues Asn92 and Asp102 (loop 1) interact electrostatically with the distal ribose. (d) The catalytic Asp125 residue of TARG1 (orange; pdb: 4J5S) resides, in contrast to MacroD1/2 and PARG, in substrate loop 2. The structure contains the lysyl-ADPr intermediate formed with the catalytic Lys84. (e) Domain organization of PARG (pdb: 5A7R) showing the macrodomain in blue, the accessory domain in orange and the short, unassigned C-terminus in green. In the catalytic residue is mutated (E756N) to co-crystallize with the dimeric ADPr substrate. All structures were generate with PyMol (Molecular Graphics System, Version 1.3 Schrödinger, LLC)

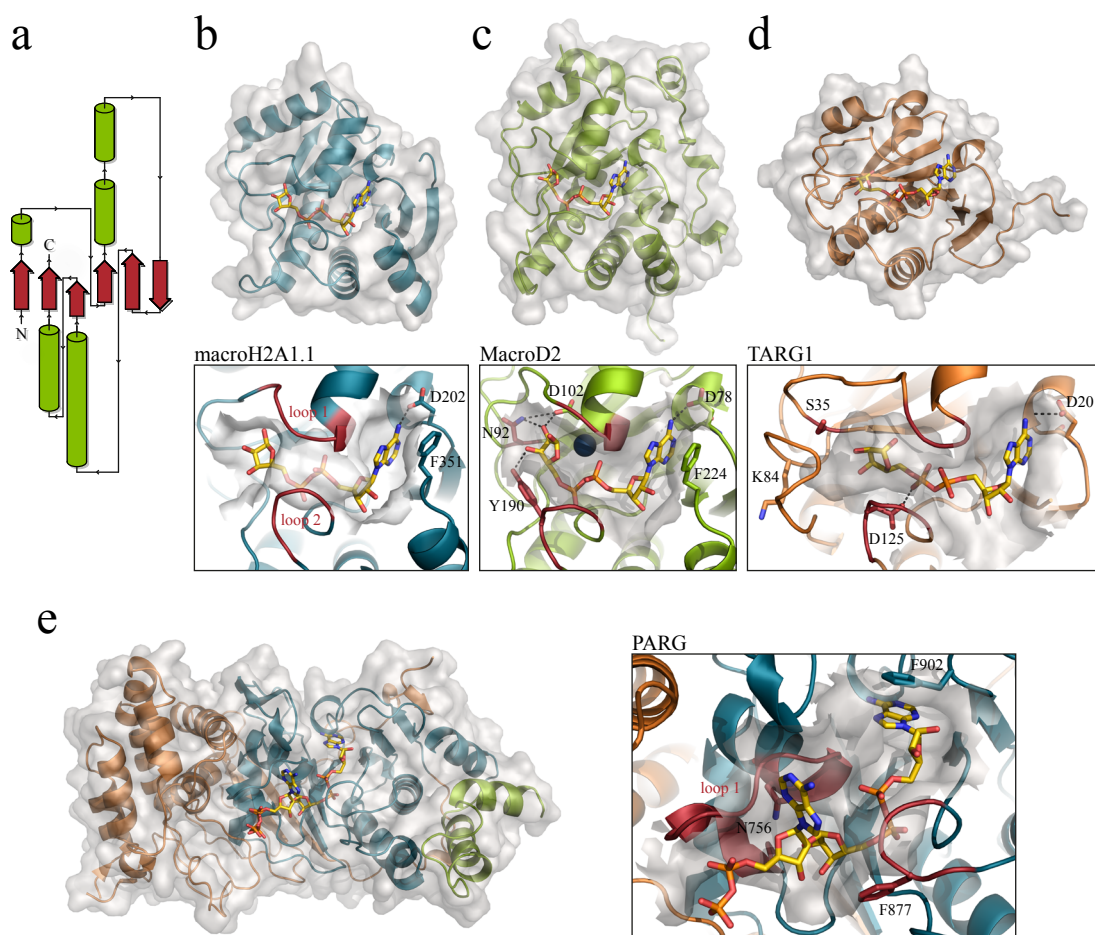


Figure 3