B-aggressive lymphoma (BAL) family proteins have unique domains which modulate transcription and exhibit PARP activity

Ricardo C.T. Aguiar, Kunihiko Takeyama, Chunyan He, Katherine Kreinbrink and Margaret A. Shipp

From the Division of Hematological Malignancies, Department of Medical Oncology, Dana-Farber Cancer Institute - Harvard Medical School
44 Binney Street Boston, MA, 02115.

Running title: BAL transcriptional modulation and PARP activity
Address correspondence to: Margaret Shipp, Dana-Farber Cancer Institute
44 Binney Street, Room M513, Boston MA, 02115. Phone: 617-6323880, Fax: 617-6324734, email: margaret_shipp@dfci.harvard.edu

Abstract:

*BAL1* (B-Aggressive Lymphoma 1) was originally identified as a risk-related gene in diffuse large B-cell lymphoma (DLBCL). BAL1 encodes a nuclear protein with N-terminal macro domains and a putative C-terminal poly(ADP-ribose) polymerase (PARP) active site. Macro domains are sequences homologous to the non-histone region of histone macroH2A. Several lines of evidence suggest that these domains may modulate transcription, including: high concentration of histone macroH2A in the inactive X chromosome, direct interference with transcription factor binding in a positioned nucleosome and structural similarity to DNA binding domains. Poly(ADP-ribosyl)ation is a critical post-translational modification that regulates chromatin configuration and transcription. In this report, we describe two additional BAL family members, *BAL2* and *BAL3*, with N-terminal macro domains and putative C-terminal PARP active sites and assess the function of these specific regions in BAL family members. Herein, we demonstrate that BAL macro domains repress transcription when tethered to a promoter. In addition, we show that BAL2 and BAL3, but not BAL1, exhibit PARP activity. In agreement with these data, BAL1 lacks several critical donor and acceptor residues that are conserved in the BAL2 and 3 PARP active sites. Of interest, BAL family members with inactive or functional PARP domains differed in their ability to repress transcription. BAL family members are the only described proteins with both PARP and macro domains, underscoring the potential functional significance of this unique combination.

Introduction:

The *BAL1* (B-Aggressive Lymphoma 1) gene was originally identified in a genome-wide search for risk-related genes in diffuse large B-cell lymphoma (DLBCL) (1). In a pilot series of primary DLBCLs, BAL expression was significantly higher in chemoresistant tumors. In our initial report, we also showed that *BAL1* encodes a nuclear protein with a duplicated N-terminal domain homologous to the non-histone region of histone macroH2A (mH2A) (2). This ≈ 135 aa non-histone region, which contains a short stretch of basic residues and a putative leucine zipper motif, has been termed a macro domain (3,4).

Several lines of evidence suggest that macro domains may modulate transcription. First, histone macroH2A is enriched in the inactive mammalian X chromosome potentially implicating the macro domain-containing protein in gene silencing (5). More specifically, the macro domain of histone macroH2A represses transcription *in vitro* (6) and directly interferes with transcription factor
binding in nucleosomes containing this variant histone (7). In addition, the crystallographic structure of a macro domain-only protein (AF1521, from *Archaeoglobus fulgidus*) includes a known DNA binding motif, further suggesting that macro domains might interact with nucleic acids (8).

In addition to containing N-terminal candidate macro domains, the BAL1 protein includes a C-terminal region with homology to the catalytic domain of PARP (poly [ADP-ribose] polymerase) proteins. PARPs catalyze the transfer of ADP-ribose onto acceptor proteins, using NAD$^+$ as a substrate (9,10). The resulting poly(ADP-ribosyl)ation of acceptor proteins has been principally associated with functions of the prototype PARP-1 enzyme and its role in DNA repair (9). Recently, other PARP domain-containing proteins have been identified and implicated in additional nuclear processes (11-13). This diverse group of proteins contains a minimal PARP active site embedded within a larger polypeptide (9). In these proteins, termed the PARP superfamily, the functional consequences of PARP activity may be dependent on additional functional domains. For example, the PARP family member, Tankyrase 1 controls telomere physiology via poly(ADP-ribosyl)ation and inhibition of TRF-1, a negative regulator of telomere length (14).

In our initial description of BAL-1, we noted homology between the BAL-1 C-terminus and the Tankyrase 1 PARP-like catalytic domain (1). However, BAL-1 did not exhibit PARP-like activity in preliminary functional assays. More recently, we identified additional BAL family members with potential critical differences in their C-terminal PARP-like regions. Given the emerging evidence that PARP family proteins represent possible cancer treatment targets (15-17) and the reported association between BAL-1 expression and chemoresistance in DLBCL, we performed detailed structure/function analyses of BAL family members.

**Materials and methods:**

**Cloning and chromosomal localization of BAL2 and BAL3.**

*BAL1* cDNA and protein sequences were used to search human nucleotide and protein databases (http://www.ncbi.nlm.nih.gov/BLAST/). Candidate homologue genes were further characterized by 5’RACE-PCR and RT-PCR. For 5’ RACE-PCR, RNAs from normal mature B-cells and DLBCL cell lines were reverse transcribed with a gene-specific primer (sequence available upon request) and submitted to successive rounds of nested amplification as previously described (18). To confirm the RACE-PCR findings, cDNAs from normal mature B-cells and DLBCL cell lines were also used for RT-PCR with primers derived from the newly identified 5’ and 3’ sequences.

To define the chromosomal location of *BAL2* and *BAL3*, we initially searched the human genome database with these genes sequences. To confirm these in silico findings, we obtained a series of YAC (CEPH-Généthon, Paris, France) and BAC clones (BACPAC Resources Center – Children’s Hospital Oakland Research Institute, Oakland, California) with previously assigned chromosomal locations (http://www.ncbi.nlm.nih.gov/mapview). DNA from these clones was isolated and PCR performed as previously described (19).

**Analyses of BAL family sequences.**

The full length, N- or C-termini domains of BAL1-related proteins were analyzed in detail for conservation of critical residues using the MegAlign™ 5.8 software package (DNAStar, Inc Madison, WI). These alignments were also displayed as phylogenetic trees and sequence distance matrices. Phylogenetic trees depict the evolutionary relationships predicted from the multiple sequence alignment where the length of each pair of branches represents the distance between sequence pairs. The distance
matrix displays the divergence and percent identity values for each sequence pair in an alignment; divergence is calculated by comparing sequence pairs in relation to the phylogeny reconstructed by MegAlign\textsuperscript{TM}. Percent identity compares sequences directly, without accounting for phylogenetic relationships. Note that divergence is not usually the inverse of percent identity (i.e. the sum of the percent identity and divergence values for a given pair is not usually 100).

**PARP activity assays**

The putative C-terminal PARP catalytic domains of the BAL family proteins (BAL1 aa 590-854, BAL2B aa 1455-1638, BAL3 aa 448-673) were fused to GST and the recombinant proteins generated using the pGEX-4T3 expression system (Pharmacia, Piscataway, NJ). Poly(ADP-ribose) activity assays were carried out as previously described (14,20). Reactions contained \( \approx 2\mu g \) of purified BAL proteins or 0.1 \( \mu g \) of high-purity recombinant PARP protein (Alexis Biochemicals, San Diego CA) as a positive control. Reactions were incubated at 25° C for 30 min in assay buffer (0.1 ml) containing 50 mM Tris-HCl, pH 8.0, 4 mM MgCl\(_2\), 0.2 mM DTT and with or without 200 ng of activated DNA (Sigma). In radioactive reactions, 1.3\( \mu M \) of \[^{32}P\]NAD\(^{+}\) (4 \( \mu Ci\); Perkin Elmer) was used whereas .1 to 1 mM of unlabeled NAD\(^{+}\) (Roche, Indianapolis, IN) served as substrate in non-radioactive assays. In certain assays, the PARP inhibitor, 3-aminobenzamide (3ABA) (Calbiochem, San Diego CA), was included at 1 mM final concentration. Reactions were stopped by the addition of 20% TCA. Precipitated proteins were rinsed once in 5% TCA, suspended in SDS loading buffer and fractionated by SDS-PAGE. In radioactive reactions, proteins were visualized by Coomassie blue stain and exposed to autoradiography. For non-radioactive reactions, after size fractionation, the proteins were transferred to PVDF membranes (Millipore, Bedford, MA) and immunoblotted with a polyclonal antibody to poly(ADP-ribose) (LP98-10, Alexis Biochemicals).

**Transcription repression assays**

Human embryonic kidney 293 (HEK293) were cultured in Dulbecco’s modified Eagle’s medium (DMEM) (GIBCO, Rockville, MD), containing 10% fetal calf serum (FCS) (GIBCO). For these experiments, HEK293 cells were seeded on six-well plates, at \( 1 \times 10^5 \) cells/well, one day before transfection. Cells were triple transfected (Effectene Transfection Reagent, Qiagen, Valencia, CA) with combinations of plasmid DNAs including 100ng of the report plasmid (Gal4-TK-Luc), 100ng of the specific effector plasmid and 100 ng pCMVBeta (Clontech, Palo Alto, CA) as previously described (6).

The effector plasmids were generated by PCR and included the Gal4 DNA binding domain (DBD) fused to full-length BAL1 (aa 2-854), BAL1 N-terminus complete, including the macro domains (aa 2-464), BAL1 N-terminus proximal (aa 2-119), BAL3 full-length (aa 2-673), BAL3 macro domains (aa 58-454) and BAL3 PARP domain (aa 450-673). All these fragments were cloned in-frame with amino acids 1-147 of Gal4 protein in a pcDNA3 backbone. Gal4 DBD-HDAC5 (histone deacetylase domain of HDAC5) served as a positive control. The latter construct, as well as the reporter plasmid were gifts of Saadi Kochbin (Grenoble, France) and were described previously (6). All constructs were verified by sequencing.

Transfected cells were harvested after 24h. Luciferase and \( \beta \)-Galactosidase activities were measured using Luciferase assay system and \( \beta \)-Galactosidase Enzyme Assay, respectively (Promega, Madison, WI). The \( \beta \)-Galactosidase activity from pCMVBeta was used as control for transfection efficiency. All experiments were performed in triplicates. In dose-dependent experiments, the total amounts of plasmid DNA were kept constant by adding appropriate amounts of empty vectors. Western blot with an anti-Gal4 antibody indicated that all fusion proteins were expressed (data not shown).
Results:

The *BAL* gene family is comprised of three genes

We previously identified an additional highly conserved *BAL1*-like sequence, suggesting that *BAL1* is a member of a larger gene family (1). For this reason, we used *BAL1* cDNA and protein sequences to search human EST, nucleotide and protein databases (http://www.ncbi.nlm.nih.gov) and identify two related genes termed *BAL2* and *BAL3* (Fig. 1).

*BAL2* corresponds to the partially characterized KIAA1268 sequence (Entrez GeneID: 54625) (1). Using 5’ RACE-PCR, we identified additional *BAL2* 5’ sequences, generating a ≈ 4.5kb open reading frame which is 120 aa longer than the previously described N-terminus (locus NM_017554, updated on December 2004). However, the extended *BAL2* cDNA lacks an in-frame stop codon upstream to the most 5’ ATG, raising the possibility that *BAL2* encodes a protein with additional N-terminal sequences. The existence of *BAL2* rat and chicken orthologues with longer N-termini supports this hypothesis (data not shown). Of interest, *BAL2* includes three N-terminal macro domains (Fig. 1).

In addition, we used 3’ RACE-PCR to fully characterize the 3’ end of *BAL2* and identify 2 alternatively spliced transcripts that encode 33aa or 154aa unique C-termini (*BAL2A* and *BAL2B* respectively, Fig. 1; GenBank accession numbers, DQ063585 and DQ063584). Whereas *BAL2B* encodes a protein containing a putative PARP catalytic domain, *BAL2A* lacks these sequences, suggesting that these C-terminal differences might be functionally relevant. In northern blot analysis of multiple human tissues, we found similar patterns of *BAL1* and *BAL2* expression with highest transcript levels in lymphocyte-rich tissues (spleen, lymph node and peripheral blood leukocytes) and lowest in total brain (data not shown and (1)).

An additional group of incompletely characterized *BAL*-like cDNA clones were identified that did not correspond to *BAL2* (RefSeq accession NM_152615.1: mRNA for predicted protein FLJ40597, recently reassigned to Entrez GeneID: 165631). Using RT-PCR, these sequences were organized into an open reading frame of ≈ 2025 nucleotides, which encoded a 656 amino acid protein termed *BAL3* (Fig. 1) (GenBank accession number DQ063586). The *BAL3* N-terminus is substantially longer than that of the predicted protein encoded by GeneID: 165631. However, there are still unidentified coding sequences at the 5’ end of *BAL3* whereas cloning of 3’ *BAL3* is complete. Using electronic resources (UniGene EST ProfileViewer and Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/), the expression pattern of *BAL3* in human tissues was found to be similar to that of *BAL1* and *BAL2* (data not shown).

*BAL1* was previously mapped to chromosome 3q21 (1). By searching the human genome database with the *BAL2* and 3 cDNA sequences, we found that these genes also mapped to chromosome band 3q21. In fact, the three genes are localized in tandem within ~ 200kb (contig, NT 005612), reinforcing the notion that they are functionally (and evolutionarily) related. These in silico findings were confirmed by PCR amplification of BAC/YAC clones assigned to this chromosomal region (data not shown).

**BAL2 and BAL3, but not BAL1, exhibit PARP activity**

All three *BAL* family members have C-termini which include sequences with partial identity to the PARP-1 catalytic domain (Fig. 1 and 2A). The putative PARP active site (the minimal PARP-1 fragment retaining catalytic activity) (9) of *BAL2B* and *BAL3* is ~ 20% identical to that of the prototype enzyme, PARP-1, and ~26% identical to that of another PARP family member, Tankyrase 1 (14) (Fig. 2B). Of note, *BAL2B* and *BAL3* are ~ 72% identical at their
PARP active site; these sequences also include the requisite catalytic glutamic acid residue (aa 1631 in BAL2B and 649 in BAL3) and additional highly conserved donor (NAD$^+$) and acceptor (polymer binding) sites (Fig. 2A) (10,21-23). In contrast, the BAL1 C-terminus is only $\approx$ 14% identical with the PARP-1 catalytic domain; in addition, this BAL-1 sequence lacks the required catalytic glutamic acid (position 988 in PARP-1) and additional donor and acceptor sites conserved in BAL2B and 3 (Figs. 2A and B). Further analyses of the divergence and evolutionary relationships between these PARP active sites indicates that the BAL1 sequence is more closely related to that of BAL2 and 3 than that of PARP-1 or Tankyrase 1 (Figs. 2B and C).

To determine whether any of the BAL family members exhibit PARP activity, we generated recombinant GST-tagged BAL1, BAL2B and BAL3 C-termini for use in PARP functional assays (8,13). In these experiments, recombinant PARP-1 was included as a positive control. We first measured the ability of these proteins to add radiolabeled ADP-ribose moieties to protein acceptors using $^{32}\text{P}\text{NAD}^+$ as a substrate. Coomassie blue staining of the gel before exposure to X-ray film confirmed that comparable amounts of BAL1, 2B and 3 were used in these assays (Fig. 3A, left panel). The C-terminal domains of BAL2B and BAL3, but not BAL1, were capable of auto(ADP-ribosyl)ation (Fig. 3A, right panel). Inclusion of the PARP-specific inhibitor, 3-ABA, markedly decreased the auto(ADP-ribosyl)ation of BAL2B and 3, confirming the specificity of these reactions (Fig. 3A, right panel). In similar reactions carried out with cold NAD+, auto(ADP-ribosyl)ated PARP-1, BAL2B and 3, but not BAL1, were also identified with an antiserum directed against poly(ADP-ribose) [anti-PAR], (Fig. 3B). The observed auto(ADP-ribosyl)ation of BAL2B, BAL3 and PARP-1 was inhibited by 3-ABA, again confirming the specificity of the assay (Fig. 3B). Of interest, the PARP activity of BAL3 was DNA-independent; absence of DNA in the reaction buffer did not change the extent of BAL3 auto poly(ADP-ribosylation), whereas it markedly decreased PARP-1 activity (Fig. 3C).

The BAL proteins contain multiple N-terminal macro domains

BAL1, 2 (isoforms A and B) and 3 include N-terminal sequences with partial homology to previously described macro domains (Figs. 1 and 4A). BAL1 and BAL3 sequences include two macro domains whereas BAL2 A and B contain three similar regions (Fig. 1 and 4A). These seven macro domains are 26 – 37% identical to the consensus macro sequence (identified by reciprocal PSI-BLAST searches in multiple species and described as the A1pp domain [conserved domains database - CDD http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=cdd]) (Figs. 4A and B).

However, the percent identity among the BAL-derived macro domains varies from 18 to 61% (Fig. 4B). Analyses of the divergence and evolutionary relationships of BAL macro domains identify specific pairings and four separate subgroups (Figs. 4B and C). As shown, the first (most N-terminal) macro domains of BAL1 and BAL2 form one cluster, the second macro domain of BAL2 and the first of BAL3 define another pair and the third macro domain of BAL2 segregates with the second macro domain of BAL3 (Fig. 4C). Although it is most related to the latter pair, the second macro domain of BAL1 branches independently from the others (Fig. 4C). These patterns suggest that the individual macro domains within a given BAL protein might have subtle functional distinctions, which are conserved in all three BAL proteins.

The macro domain superfamily

Analyses of the most recent human genome assemblies revealed the existence of at least five additional human genes encoding macro domain proteins (Figs. 4A-C). Two of these genes are the well-characterized and related histone macroH2A1 and macroH2A2. Histone macroH2A1 (H2AFY), the original
member of this family (2), is expressed as two functionally distinct alternatively spliced transcripts, macroH2A1.1 and 1.2 (H2AFY isoform 1 and 2). The more recently described histone macroH2A2 (H2AFY2) (24,25) encodes a single protein that is more closely related to macroH2A1.2 (Fig. 4C).

Much less is known about the other three human genes encoding single macro domain proteins, GDAP2, LRP16 and c20orf113. The first two, which are conserved from plants to high mammals, have been implicated in ganglioside-induced neural differentiation (GDAP2) (26) and increased proliferation of MCF-7 cells (LRP16) (27). C20orf113 (GeneID: 170486) was cloned as part of a human cDNA sequencing project and does not yet have an ascribed function.

In summary, there are eight human genes encoding nine proteins with thirteen distinct macro domains. Additional structural and functional features, such as the presence of histone or PARP domains, delineate specific gene families (histone macroH2A and BAL). A third cluster, GDAP2, LRP16 and c20orf113, is defined by the evolutionary relationship of their respective macro domains (Fig. 4C). Of note, BAL family members are the only known genes with multiple macro domains in the same protein (Fig. 1).

The BAL proteins mediate transcription repression: role of the macro-domains.

Given the described association of the macro domain of histone macroH2A with transcriptional repression (6), we next assessed the possibility that the prototype BAL family member, BAL1, might have a similar function. Initially, we investigated the ability of BAL1 to repress transcription when tethered (via a Gal4 DNA binding domain) to the TK promoter (Fig. 5A). This experimental approach, which is often used to characterize the repressive activity of chromatin-associated proteins, previously defined the repressive activities of the macro domain of histone macroH2A1 (6).

The Gal4DBD alone markedly increased TK-driven luciferase activity as previously described (28) (Figs. 5B and C). For this reason, we used a more stringent approach to assess BAL-mediated repression, comparing TK-driven luciferase activity of Gal4DBD-BAL1 transfected cells with that of cells transfected with reporter alone (no effector) (21). As shown, Gal4DBD-BAL1 decreased TK-driven transcription in a dose-dependent manner (Fig. 5B); the magnitude of transcription repression was comparable to that reported for the macro domain of histone macroH2A (6).

To define the role of BAL1 macro domains in the observed transcriptional repression, an additional series of BAL1 deletion constructs were generated: N-terminus proximal (aa 2-119) and N-terminus complete (aa 2-464, including macro domains aa 119-446) (Fig. 5B). As expected, a BAL1 proximal N-terminus construct which lacks the macro domains had no effect on TK-driven luciferase activity (Fig. 5C). In contrast, the macro domain-containing BAL1 N-terminus and the full length BAL1 protein were equally effective in repressing transcription (Fig. 5C).

BAL1 mediated transcription repression was not promoter dependent as similar results were obtained using the TATA-containing Gal4-MLP-Luc reporter (data not shown). In addition, the Gal4DBD-BAL fusion proteins did not repress the activity of promoter constructs (TK-luc or MLP-luc) which lacked the five Gal4-binding sites (data not shown). These data confirmed that BAL1-mediated repression was due to physical association with the promoter area rather than an indirect (general) inhibitory effect.

PARP activity and transcription repression.

The BAL1 protein, which represses transcription via its N-terminal macro domain, lacks C-terminal PARP activity. However, BAL2 and 3 contain macro domains and active PARP sites, of interest given the known association between PARP-1 function and
transcriptionally active chromatin (9,11-13). For these reasons, we also investigated the ability of full-length BAL3 and the N-terminal BAL3 macro domain to repress transcription (Fig. 5D). Full-length BAL3 was significantly less effective than full-length BAL1 in repressing transcription (p = .004), likely due to the PARP activity of BAL3 (Fig. 5D).

Discussion:

The BAL gene family is comprised of three members mapping in tandem on chromosome band 3q21. BAL proteins have multiple N-terminal macro domains capable of mediating transcription repression; BAL2B and BAL3 also exhibit PARP activity.

The macro domain is an evolutionary conserved sequence of \( \approx 135 \) amino acids initially described as the non-histone region of the variant histone macroH2A (2). Some proteins consist almost entirely of the macro domain (notably in lower species); in other instances, this sequence is embedded within much larger polypeptides and associated with a variety of unrelated motifs such as histone domain in the macroH2A family, the PARP domain in the BAL family and the sec14 region (a putative lipid binding domain) in GDAP2. In our genome-wide search, we identified thirteen human macro domains encoded by eight unique genes including the three BAL family members. Although BAL macro domains have highly similar sequences, the divergence analysis and the evolutionary tree indicates that the two or three macro domains within a given BAL protein do not cluster with each other; instead, they cluster with corresponding macro domains in other BAL family members. This structural organization likely ensures the representation of specific macro regions in all three BAL proteins, prompting speculation regarding subtle functional differences between these macro sequences. Consistent with this possibility, recent studies suggest that the alternatively spliced macro domains in macroH2A (macroH21.1 and macro H2A1.2) differ in their ability to bind NAD metabolites (29).

Although recent progress has been made, the role of the macro domains is still incompletely characterized. Histone macroH2A is enriched on the inactive mammalian X chromosome (5) and the macro domain from this variant histone represses transcription in vitro (6). A molecular mechanism was recently proposed for the domain-specific transcriptional repression of histone macro2A (7). The macro domain of histone macroH2A in a positioned nucleosome was found to interfere with transcription factor binding whereas the histone region disrupted SWI/SNF nucleosome remodeling. This observation was further substantiated by crystallographic studies that identified similarities between macro domains and DNA binding structures (8). Recent studies also suggest that certain macro domains may bind monomeric or polymeric ADP-ribose moieties (29,30). Although the functional consequences of these interactions remain to be defined, they are of particular interest given the demonstrated PARP activity of BAL 2B and 3.

In our studies, the BAL N-terminal macro domain repressed transcription when brought into the close proximity of a promoter. The magnitude of BAL macro domain-mediated repression was similar to that seen with the macro domains of histone macroH2A (6), highlighting the functional relationship between macro domains from distinct proteins. Like histone macroH2A, BAL proteins may sterically block the access of transcription factors and coactivators to specific chromatin regions (8). The macro domains of BAL2B and 3 might also serve as guides to direct BAL2B and 3 PARP activity to specialized compartments in the nucleosome. In our in vitro studies, BAL family members with functional PARP domains exhibited significantly less transcription repression, suggesting that BAL activity is determined by family member-specific macro domains and PARP sites. In this regard, it is of interest that enzymatically active or inactive PARP-1 can have opposite effects on transcription (9,11-13).
It is not yet clear whether BAL proteins modulate transcription independently or in association with a multi-protein co-repressor complex. However, considering the recently identified role of ubiquitination in transcription modulation (31), it is of interest that BAL1 binds to BBAP (B-lymphoma and BAL Associated Protein), a DELTEX family member with ubiquitin E3 ligase activity (32). Based on C-terminal sequence homologies, BAL proteins were predicted to be members of the PARP superfamily and termed PARP-9 (BAL1), PARP-14 (BAL2) and PARP-15 (BAL3) (9). However, several residues critical for PARP catalytic activity, elongation and branching are missing in BAL1, including glutamic residue E988 of PARP-1 (22,23,33,34). Consistent with these structural differences, BAL1 lacks PARP activity whereas BAL2B and 3 exhibit poly(ADP-ribose) polymerase activity. Although BAL2B and 3 have highly similar PARP active sites and a higher level of conservation with the prototype PARP-1 catalytic site, less than half of the identified acceptor and donor sites (21) are present in these new BAL family members. These findings suggest that the requirements for a functional PARP domain might be less stringent than initially described.

We specifically found that BAL2B and BAL3 are capable of poly(ADP-ribosyl)ating themselves, in a DNA-independent manner. In PARP-1 and Tankyrase, auto-modification serves as a regulatory mechanism that limits the activities of these enzymes toward their specific targets (9,35). It is likely that auto-modification of the BAL proteins serves a similar objective. Identification of additional BAL binding partners and putative targets for this post-translational modification will help clarify the role of BAL-mediated poly(ADP-ribosyl)ation in cellular systems.

Interestingly, recent studies also implicate PARP activity in DNA-repair independent nuclear processes that regulate chromatin configuration and transcription (11-13). In general, these studies point to an association between enzymatically active PARP-1 and high transcription activity via decondensation of chromatin (13), poly(ADP-ribosyl)ation of co-repressor complexes (11) or automodification (12). Conversely, by binding to the nucleosome (12) or co-repressor complex (11), inactive PARP-1 plays a structural role that adversely influences gene expression. These recent studies and the data presented here suggest a model whereby the transcription repression mediated by N-terminal macro-domains of the BAL proteins might be curtailed by C-terminal PARP activity, via auto-modification and/or poly(ADP-ribosyl)ation of nucleosomal histones. Alternatively, poly(ADP-ribosyl)ation might be the primary functional attribute of the BAL2B and BAL3 proteins and the macro domains would promote the critical interaction with specific compartments of the chromatin. The noted differences between BAL1 and BAL2/BAL3 will provide the ideal tools to test these models and to define the signals necessary to trigger BAL PARP activity in vivo. Given emerging evidence that PARP proteins represent possible cancer treatment targets (15-17), functional characterization of BAL family members may have therapeutic applications in lymphoma.
References:


Footnotes:

We thank Susan Smith (New York University, New York) for input regarding PARP activity assays. Ricardo Aguiar was supported by a career development award (Special Fellow) from the Leukemia and Lymphoma Society of America.

Figure Legends:

Figure 1. Diagrammatic representation of the BAL family of proteins. The macro domains and putative PARP domains of BAL1, 2 and 3 are highlighted. The numbering corresponds to the amino acid positions at the beginning and end of these specific domains. The first and last amino acid positions of each BAL family member are indicated. For BAL2 and BAL3, the first methionine in the current sequence is designated as amino acid 1 (both sequences may have longer N-termini [see text for details]). The BAL2A isoform results from premature termination; BAL2B encodes an additional 154 amino acids including a putative PARP domain.

Figure 2. PARP domains in the BAL protein family. A. Alignment of the PARP active sites of PARP-1, Tankyrase 1, BAL1, BAL2B and BAL3. Numbers correspond to amino acid positions. Identical and similar residues are highlighted in black and gray, respectively. Conserved polymer-binding residues (acceptor sites) are indicated by a triangle and donor sites (NAD+) by a circle; those conserved in all five proteins are represented as solid boxes and those missing in BAL1 are shaded. The critical catalytic glutamic acid residue (aa 988 in PARP-1), which is not conserved in BAL1, is boxed. Previously described donor and acceptor residues that are not conserved in the BAL proteins include: S864, I876, L881, R882, M993, K906, N909, Y910 and L985 (residue numbers from PARP-1 sequence). B. Matrix alignment display of PARP-1, Tankyrase and BAL1, 2B and 3 PARP active sites generated using the MegAlign™ 5.8 software package (DNAStar, Inc Madison, WI). The distance matrix displays the divergence and percent identity values of each sequence pair in the alignment. Divergence is calculated by comparing sequence pairs in relation to the phylogeny reconstructed by MegAlign™. Percent identity compares sequences directly, without accounting for phylogenetic relationships. Note that divergence is not usually the inverse of percent identity. BAL2B and BAL3 are ~ 20% identical to PARP-1 whereas BAL1 is only
14% identical to PARP-1; BAL1, 2B and 3 percent identities with Tankyrase 1 are similar (25-26%)
. BAL2B and BAL3 have the highest percent identity, ≈72%. C. The evolutionary relationship
of these sequences is shown in a phylogenetic tree which reveals higher conservation
of BAL PARP active sites with each other than with Tankyrase 1 or PARP-1.

Figure 3. PARP activity of BAL proteins. A. Radioactive assays. C-terminal sequences encompassing the putative PARP active site domains of the BAL proteins (BAL1 aa 590-854, BAL2 aa 1455-1638, BAL3 aa 433-656 [Fig. 1]) were fused to GST and the recombinant proteins generated using the pGEX expression system. Reactions were performed with \([^{32}P]NAD^+ \), in the presence or absence of the PARP inhibitor 3-aminobenzamide (3-ABA) (designated + vs. – in the figure). Samples contained ≈ 2µg of purified GST (alone) or GST-BAL proteins or 0.1 µg of high-purity recombinant PARP protein. The products were analyzed by Coomassie blue staining (left), which confirmed equal loading, and autoradiography (right) of SDS-PAGE gels. In these assays, BAL2B, BAL3 and PARP-1 ADP-ribosylate themselves, whereas no activity was detected for BAL1 or GST alone. These enzymatic activities were clearly inhibited by 3-ABA, confirming their specificity. B. Anti-PAR [poly(ADP-ribose)] immunoblots. Reactions were performed as above but supplemented with 1 mM unlabeled NAD\(^+ \) instead of \([^{32}P]NAD^+ \). After size fractionation, the proteins were transferred to PVDF membranes (Millipore, Bedford, MA) and immunoblotted with a polyclonal antibody to poly(ADP-ribose). As in the radioactive reactions, BAL2B and BAL3 PARP activities are readily detectable and inhibited by 3-ABA. The experiments shown are representative examples of multiple similarly performed assays. C. DNA-independent BAL3 PARP activity. Increasing amounts of BAL3 recombinant protein or 0.1 µg of PARP-1 protein were analyzed for PARP activity (as above) in the presence or absence of activated DNA. Samples were immunoblotted with an antibody to poly(ADP-ribose). The poly(ADP-ribosyl)ation of BAL3 was unchanged by the presence or absence of DNA whereas the PARP-1 activity was DNA dependent.

Figure 4. Macro domain superfamily. A. Alignment of the thirteen macro domains encoded by eight unique human genes. Numbers correspond to amino acid positions. Identical and similar residues are highlighted in black and gray, respectively. The consensus macro domain sequence was identified by reciprocal PSI-BLAST searches in various species and corresponds to the SMART (Simple Modular Architecture Research Tool) sequence #00506.10 (http://smart.embl-heidelberg.de/). This domain is also referred to as the A1pp domain (29). B. Matrix alignment display of the macro domains and consensus sequence generated using the MegAlign™ 5.8 software package. The distance matrix displays the divergence and percent identity values of each sequence pair in the alignment. C. Evolutionary relationship of the multiple macro domains is shown in a phylogenetic tree. The evolutionary tree indicates that the two or three macro domains within a given BAL protein do not cluster with each other; instead, they cluster with corresponding macro domains in other BAL family members (highlighted in yellow). This structural organization likely ensures the representation of specific macro regions in all three BAL proteins. Two additional clusters are recognized; one is composed of the macro domains of the histone macroH2A proteins and the other, the macro domains encoded by the largely uncharacterized genes, GDAP2, LRP16 and c20orf113.

Figure 5. Transcriptional repression properties of BAL1. A. Schematic representation of the BAL1 and BAL3 proteins, effector plasmids and luciferase reporter Gal4-TK-Luc. The indicated constructs were fused in-frame to the DNA-binding domain (DBD) of the transcription activator, Gal4. Gal4-DBD fused to the histone deacetylase domain of HDAC5 served as a positive control. The luciferase reporter plasmid contains five copies of the Gal4-binding site (5XGal4) upstream from the thymidine kinase (TK) core promoter (nucleotides 2105 to 152). B. Dose-dependent transcription repression mediated by BAL1. As previously reported (28), Gal4DBD alone (100
ng) increased the activity of the TK promoter by ≈ 50%. Conversely, Gal4DBD-BAL1 decreased TK-driven transcription in a dose-dependent manner, and at 100 ng, reduced luciferase to ~ 40% of no effector baseline or ~ 30% of Gal4DBD alone. Results are shown as the luciferase activity normalized by β-galactosidase (± SD). C. Domain mapping. The BAL1 N-terminus proximal construct, which lacks the macro domains, had no effect on TK-driven luciferase activity. In contrast, the N-terminal construct including macro domains and the full length BAL1 protein were equally effective in repressing transcription. D. Analysis of BAL3 transcription repression. Gal4DBD-BAL3 full-length was significantly less effective than Gal4DBD-BAL1 full-length in repressing TK-driven transcription (p = .004), likely due to the PARP activity of BAL3. All the experiments were repeated at least three times and performed in triplicate. Representative assays are shown.
Figure 1
Figure 2A
### Figure 2B

<table>
<thead>
<tr>
<th>Divergence</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>20.7</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>192.9</td>
<td></td>
<td>25.6</td>
<td>24.8</td>
<td>26.3</td>
</tr>
<tr>
<td>3</td>
<td>317.0</td>
<td>171.1</td>
<td></td>
<td>35.0</td>
<td>30.9</td>
</tr>
<tr>
<td>4</td>
<td>213.0</td>
<td>175.5</td>
<td>131.5</td>
<td></td>
<td>71.5</td>
</tr>
<tr>
<td>5</td>
<td>208.0</td>
<td>165.4</td>
<td>154.4</td>
<td>35.8</td>
<td></td>
</tr>
</tbody>
</table>

**Percent Identity**

- PARP-1
- Tankyrase
- BAL1
- BAL2
- BAL3

### Figure 2C

- PARP-1
- Tankyrase
- BAL1
- BAL2
- BAL3
Figure 3A

Coomassie-blue

NAD$^+$ $^{32}$P

Figure 3B

anti-PAR
Figure 3C
<table>
<thead>
<tr>
<th>Protein</th>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>consensus</td>
<td>1</td>
<td>75</td>
<td>LKVVKGDITKPRADAIVNAANSDGAHGGG-VAGAIARAAAGGEESKEREKRKG--CPVGIAVTPCGNLKAKYVI</td>
</tr>
<tr>
<td>BAL1-1</td>
<td>119</td>
<td>594</td>
<td>ELSVWKDDITTHAHDVAIVNAANEDILHCGG-LAADVKAGGEQGEBSKQKVVRKVSAGELVTCAGRIPCGQI</td>
</tr>
<tr>
<td>BAL1-2</td>
<td>319</td>
<td>794</td>
<td>LQIVCGELEWQATDADTVSNVPDDITVGP-VAKSILQCOAGVMKSEPFLATKQKO-FQRSQVLTVKGNFLGCKYI</td>
</tr>
<tr>
<td>BAL2-1</td>
<td>646</td>
<td>1121</td>
<td>LIVVQGQANLRPVDVNVNASNLKIRDHG-GCRAALSKAGGELQADCQ-VKRRHELPLGFATKLYPHII</td>
</tr>
<tr>
<td>BAL2-2</td>
<td>860</td>
<td>1235</td>
<td>LVKEGQANAKDVTWVGVSPLVLSRQPLSKSLEKAPQEL-QEELDVTQVGAVSMQTVLFSSWNLDCGVL</td>
</tr>
<tr>
<td>BAL2-3</td>
<td>1070</td>
<td>1545</td>
<td>IFQVAASDITKEEADVIVNSNSFNKLGA-VSRAALLEAG---NVERECSCOQQQRKNDYITTCGFLRCKNII</td>
</tr>
<tr>
<td>BAL3-1</td>
<td>85</td>
<td>360</td>
<td>NKLISGDVLITYADTVNVSVPNILQGCGLPLQAFLOMQAPMT-QKELDTRRETEEKGVNIFGSTCSNLDCGAVL</td>
</tr>
<tr>
<td>BAL3-2</td>
<td>301</td>
<td>776</td>
<td>FOVATDIAETEOVIVNSNNSFNKLGA-VSRAALLEAG---QAVESECALVLAOQPRHDFITTPGCLLCKENDII</td>
</tr>
<tr>
<td>macroH2A1-1</td>
<td>196</td>
<td>571</td>
<td>KLOVQCDIASDSDAVVIVNTANSLNNGG-VGNTLEKKGGKEFVEAVLREKKNQPLEVGAASPACHLPAKFI</td>
</tr>
<tr>
<td>macroH2A1-2</td>
<td>205</td>
<td>680</td>
<td>VLKWHSEFPKPA--------------QILKHAMKECLEKIQNITSFRAILGTNMEIERTAAPP</td>
</tr>
<tr>
<td>macroH2A2</td>
<td>196</td>
<td>675</td>
<td>KLSLQSDISEGISRIVERGIVHPATTABDLKDEDI-GKAEvkGKEEFLSETKYERKSCQPLEVAABAASQSSGLAAKFI</td>
</tr>
<tr>
<td>GDAP2</td>
<td>55</td>
<td>230</td>
<td>KVLWKGDPALLNCTAIVNTSLESRTDKNP-VSEGTAIGPLPLDLQKLC-CERTGPAKTCKFLNIRAP</td>
</tr>
<tr>
<td>LRP16</td>
<td>153</td>
<td>538</td>
<td>KISSLRSITKLEAVNAANASSLGGG-VDGCHRAAGPPLLTDCECRQOS------KTGGRKTRCCYRLPAKFI</td>
</tr>
<tr>
<td>C20orf113</td>
<td>71</td>
<td>256</td>
<td>KVSLYRGDITLEVDAIVNANASLGGGG-VDGCHRAAGPCLAEVNHNC------CTGDEAKYTCGDLPKAKFI</td>
</tr>
</tbody>
</table>

**Figure 4A**
Figure 4B

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36.8</td>
<td>31.6</td>
<td>34.2</td>
<td>27.4</td>
<td>29.1</td>
<td>29.9</td>
<td>25.6</td>
<td>35.0</td>
<td>28.2</td>
<td>27.4</td>
<td>31.6</td>
<td>45.3</td>
<td>34.2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>124.6</td>
<td>18.2</td>
<td>45.3</td>
<td>26.3</td>
<td>19.0</td>
<td>24.1</td>
<td>19.7</td>
<td>27.0</td>
<td>24.1</td>
<td>24.8</td>
<td>20.4</td>
<td>29.2</td>
<td>21.2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>140.9</td>
<td>246.0</td>
<td>19.5</td>
<td>32.8</td>
<td>32.8</td>
<td>30.5</td>
<td>33.6</td>
<td>25.8</td>
<td>21.1</td>
<td>21.1</td>
<td>22.7</td>
<td>24.2</td>
<td>14.8</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>133.7</td>
<td>90.4</td>
<td>244.0</td>
<td>27.6</td>
<td>23.9</td>
<td>27.6</td>
<td>24.6</td>
<td>35.1</td>
<td>32.1</td>
<td>25.4</td>
<td>25.4</td>
<td>34.3</td>
<td>22.4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>173.0</td>
<td>174.2</td>
<td>139.5</td>
<td>167.1</td>
<td>25.2</td>
<td>50.4</td>
<td>24.4</td>
<td>26.0</td>
<td>25.2</td>
<td>22.1</td>
<td>22.1</td>
<td>24.4</td>
<td>15.3</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>154.4</td>
<td>227.0</td>
<td>129.8</td>
<td>180.2</td>
<td>172.1</td>
<td>30.8</td>
<td>60.8</td>
<td>21.7</td>
<td>19.2</td>
<td>17.5</td>
<td>21.7</td>
<td>30.8</td>
<td>20.8</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>160.5</td>
<td>195.0</td>
<td>155.6</td>
<td>170.2</td>
<td>78.8</td>
<td>154.8</td>
<td>24.6</td>
<td>25.4</td>
<td>23.1</td>
<td>20.9</td>
<td>21.6</td>
<td>26.1</td>
<td>18.7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>177.1</td>
<td>216.0</td>
<td>126.0</td>
<td>173.8</td>
<td>178.4</td>
<td>53.9</td>
<td>175.5</td>
<td>26.1</td>
<td>22.7</td>
<td>20.2</td>
<td>21.8</td>
<td>27.7</td>
<td>19.3</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>132.4</td>
<td>173.3</td>
<td>186.0</td>
<td>127.3</td>
<td>184.7</td>
<td>227.0</td>
<td>189.8</td>
<td>189.0</td>
<td>81.1</td>
<td>60.6</td>
<td>23.5</td>
<td>27.3</td>
<td>16.7</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>159.9</td>
<td>186.0</td>
<td>216.0</td>
<td>134.2</td>
<td>184.3</td>
<td>241.0</td>
<td>195.0</td>
<td>205.0</td>
<td>16.0</td>
<td>65.6</td>
<td>24.0</td>
<td>25.6</td>
<td>15.2</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>176.6</td>
<td>189.8</td>
<td>226.0</td>
<td>186.4</td>
<td>214.0</td>
<td>279.0</td>
<td>228.0</td>
<td>241.0</td>
<td>54.4</td>
<td>44.9</td>
<td>20.9</td>
<td>23.9</td>
<td>14.9</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>145.5</td>
<td>224.0</td>
<td>200.0</td>
<td>181.3</td>
<td>208.0</td>
<td>214.0</td>
<td>212.0</td>
<td>212.0</td>
<td>197.0</td>
<td>194.5</td>
<td>219.0</td>
<td>38.3</td>
<td>28.9</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>92.2</td>
<td>155.2</td>
<td>192.7</td>
<td>128.3</td>
<td>192.9</td>
<td>148.6</td>
<td>178.5</td>
<td>168.6</td>
<td>171.0</td>
<td>183.8</td>
<td>196.6</td>
<td>118.2</td>
<td>49.2</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>74.8</td>
<td>126.9</td>
<td>192.2</td>
<td>119.5</td>
<td>189.6</td>
<td>136.3</td>
<td>152.3</td>
<td>148.4</td>
<td>177.3</td>
<td>189.2</td>
<td>195.0</td>
<td>90.5</td>
<td>23.3</td>
<td>14</td>
</tr>
</tbody>
</table>

Divergence

Percent Identity

macro consensus
BAL1 macro1
BAL1 macro2
BAL2 macro1
BAL2 macro2
BAL2 macro3
BAL3 macro1
BAL3 macro2
macroH2A1.1
macroH2A1.2
macroH2A2
GDAP2
LRP16
C20orf113
Figure 4C
**Figure 5A**

- **BAL1 Protein**
  - Macro domains
  - 
  - 100 200 300 400 500 600 700 800 aa

- **BAL3 Protein**
  - Macro domains
  - PARP domain
  - 100 200 300 400 500 600 aa

- **GAL4-DBD**

- **GAL4-DBD only**
  - BAL1 full-length
  - BAL1 N-term proximal
  - BAL1 N-term complete
  - BAL3 full-length
  - BAL3 macro domains
  - BAL3 PARP domain
  - HDAC5

- **GAL4-TK-Luc**
  - 5x GAL4
  - TK
  - Luciferase
Figure 5B

Luciferase/B-galactosidase activity

None

GAL4

GAL4-BAL1

GAL4-HDAC5

0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6
Figure 5C
Figure 5D
B-aggressive lymphoma (BAL) family proteins have unique domains which modulate transcription and exhibit PARP activity
Ricardo C. T. Aguiar, Kunihiro Takeyama, Chunyan He, Katherine Kreinbrink and Margaret Shipp

J. Biol. Chem. published online August 1, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M505408200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts