Epstein-Barr Virus Encodes Three Bona Fide Ubiquitin-Specific Proteases

Citation for published version:

Digital Object Identifier (DOI):
10.1128/JVI.01113-08

Link:
Link to publication record in Edinburgh Research Explorer

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

Published in:
Journal of Virology

Publisher Rights Statement:
Copyright © 2013 by the American Society for Microbiology.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Epstein-Barr Virus Encodes Three Bona Fide Ubiquitin-Specific Proteases

Ramakrishna Sompallae,1† Stefano Gastaldello,1,2† Sebastian Hildebrand,1 Nikolay Zinin,1 Gerco Hassink,1 Kristina Lindsten,1 Juergen Haas,3 Bengt Persson,1,4 and Maria G. Masucci1*

Department of Cell and Molecular Biology, Karolinska Institutet, S-171 77 Stockholm, Sweden; Department of Biomedical Sciences, University of Padua, 35121 Padua, Italy; Division of Pathway Medicine, School of Biomedical Sciences, College of Medicine, University of Edinburgh, Edinburgh, United Kingdom, and Max von Pettenkofer-Institute, LMU-München, Munich, Germany; and IFM Bioinformatics, Linköping University, S-581 83 Linköping, Sweden

Received 27 May 2008/Accepted 8 August 2008

Manipulation of the ubiquitin proteasome system (UPS) is emerging as a common theme in viral pathogenesis. Some viruses have been shown to encode functional homologs of UPS enzymes, suggesting that a systematic identification of these products may provide new insights into virus-host cell interactions. Ubiquitin-specific proteases, collectively known as deubiquitinating enzymes (DUBs), regulate the activity of the UPS by hydrolyzing ubiquitin peptide or isopeptide bonds. The prediction of viral DUBs based on sequence similarity with known enzymes is hampered by the diversity of viral genomes. In this study sequence alignments, pattern searches, and hidden Markov models were developed for the conserved C- and H-boxes of the known DUB families and used to search the open reading frames (ORFs) of Epstein-Barr virus (EBV), a large gammaherpesvirus that has been implicated in the pathogenesis of a broad spectrum of human malignancies of lymphoid and epithelial cell origin. The searches identified a limited number of EBV ORFs that contain putative DUB catalytic domains. DUB activity was confirmed by functional assays and mutation analysis for three high scoring candidates, supporting the usefulness of this bioinformatics approach in predicting distant homologues of cellular enzymes.

The posttranslational modification of proteins by ubiquitin (Ub) and Ub-like molecules (UbLs) and the degradation of polyubiquitinated substrates by the proteasome regulate fundamental cellular processes, including cell growth and differentiation, intracellular signaling, protein trafficking, apoptosis, and the recognition of virus-infected or malignant cells by the host immune response (21, 27, 33, 44, 46). Modulation of the ubiquitin-proteasome system (UPS) is emerging as a central theme in viral pathogenesis, and several examples have been reported of viral proteins that mimic or redirect the activity of the system in order to modify the cellular environment in favor of virus persistence or replication (22, 28, 31, 41, 45). Thus, the identification of viral products that interfere with the UPS may yield new insights on important features of viral pathogenesis and could lead to the development of new means of therapeutic intervention.

Conjugation of Ub or UbLs is achieved through the activity of a cascade of enzymes, including activating enzymes (E1s), conjugating enzymes (E2s), and specific ligases (E3s) that catalyze the covalent linkage of the modifier to Lys residues in the substrate (11). In addition, deconjugating enzymes act as regulators of the system by maintaining the pool of free Ub and UbLs and by determining the rate of turnover of the conjugates (14, 50). In line with this proposed regulatory function, recent evidence points to a critical role of Ub deconjugases (deubiquitinating enzymes [DUBs]) in intracellular signaling, as exemplified by the activation of the transcription factor nuclear factor-κB (NF-κB) by cylindromatosis tumor suppressor, A20, Cezanne, and ubiquitin-specific protease 31 (25, 33) and by the initiation of DNA repair by USP1 (37).

Based on similarity with known family members, the human DUBs have been classified in five subfamilies including: ubiquitin C-terminal hydrolases (UCHs), ubiquitin-specific proteases (USPs), Machado-Joseph disease proteases (MJDs or Josephins), ovarian tumor proteases (otubains, OTUs), and the JAMM (Jab1/MPN/Mov34 metalloenzyme) motif proteases (2, 38). With the exception of JAMMs, the DUBs are cysteine proteases identified by a catalytic triad of Cys, His, and Asp/Asn residues and by the presence of conserved amino acid domains known as Cys- and His-boxes (C- and H-boxes, respectively) that are unique for each family (38). Mutational studies have shown that the Asp/Asn residue is not essential for catalytic activity, although it may contribute to the enzymatic activity by stabilizing the active-site thiolate and imidazolium ion pair (26, 35). The JAMM metalloproteases lack a C-box, while two conserved His residues in the JAMM motif were shown to be essential for activity (49).

Proteins with DUB activity are encoded in the genome of human adenovirus, herpesvirus, coronavirus, and bunyavirus (5, 6, 18, 29, 42). Furthermore, DUB activity was recently demonstrated in proteins encoded by some pathogenic bacteria that lack an intrinsic UPS (40, 52), suggesting that these enzymes play specific roles in the regulation of both viral and bacterial infection. Although these findings make the systematic identification of microbial DUBs a worthy endeavor, the task is complicated by the wide sequence variation of the known enzymes. Moreover, viral and bacterial proteins are often considerably different in sequence and domain organization compared to their mammalian counterparts, which fur-
the identification of distant functional homologues. We have used a bioinformatics strategy here to identify putative virus-encoded DUBs. To overcome the difficulty posed by sequence variation, we focused on the conserved C- and H-boxes. Family-specific sequence alignments, pattern searches, and hidden Markov models (HMMs) were generated based on the amino acid sequences of known members of the five DUB subfamilies and then used to search a custom-made database of open reading frames (ORFs) from Epstein-Barr virus (EBV), a human gamma herpesvirus that establishes latent infections and is associated with a broad spectrum of malignancies of lymphoid and epithelial cell origin.

**Materials and Methods**

Data sets. The sequences of DUBs were extracted from the UniProt database version 51.2 (4). The DUB data set consisted of both experimentally verified human enzymes and highly conserved sequence homologs (Table 1). For families with fewer than 10 human sequences, homologues from other species were included. The JAMM metallopeptases contain only H-box domains, and sequence homologs with catalytic His residues were considered for deriving the patterns.

**TABLE 1. Data sets of DUBs used in this study**

<table>
<thead>
<tr>
<th>Protein family</th>
<th>Characteristic domain</th>
<th>No. of human DUBs</th>
<th>No. of verified DUBs</th>
<th>No. of orthologs</th>
<th>No. of DUBs in pattern search set</th>
<th>No. of DUBs in HMM training set</th>
</tr>
</thead>
<tbody>
<tr>
<td>USP</td>
<td>Peptidase 19</td>
<td>45</td>
<td>37</td>
<td>15</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td>UCHL</td>
<td>Peptidase 12</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>OTU</td>
<td>OTU</td>
<td>8</td>
<td>5</td>
<td>9</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>MJD</td>
<td>Josephin</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>JAMM</td>
<td>MPN+</td>
<td>8</td>
<td>2</td>
<td>29</td>
<td>34</td>
<td>21</td>
</tr>
</tbody>
</table>

*a* That is the number of DUBs with experimentally verified enzymatic activity.

*b* Due to the small number of identified sequences in the human UCH, OTU, JAMM, and MJD subfamilies, orthologs from other species were included in the data set.

*c* The JAMM metallopeptases contain only H-box domains, and sequence homologs with catalytic His residues were considered for deriving the patterns.

*d* Only C- and H-boxes with <90% sequence identity were used for training the HMMs.

**DUB searches.** (i) Sequence alignment. Sequence comparisons between the human DUBs and the EBV ORFs were carried out by using CLUSTAL W (47) and BLAST (1). CLUSTAL W calculates multiple sequence alignments and was therefore used to investigate whether any of the EBV ORFs clusters with the specific DUB subfamilies. BLAST searches were used to identify putative C- and H-boxes. To this end, members of each DUB family were aligned by using CLUSTAL W, and conserved catalytic domains were extracted with the GoCore program (http://www.helsinki.fi/project/ritvos/GoCore/). The C- and H-box domains of each family were defined on the basis of the conserved regions (UCH, C-box 23aa and H-box 24aa; USP, C-box 16aa and H-box 21aa; OTU, C-box 19aa and H-box 14aa; MJD, C-box 21aa and H-box 14aa; JAMM domain 25aa), and the sequences were then searched against the EBV ORF database. Hits with arbitrarily chosen E-values of ≤10 and containing Cys or His residues were selected for further analysis.

(ii) Pattern search. C- and H-box specific patterns were constructed to search for putative catalytic domains. The C- and H-boxes of each DUB family were aligned with respect to the conserved Cys and His residues, and search patterns were manually derived from the sequence alignments based on the physicochemical properties.

(iii) HMM search. HMMs were constructed by using the HMMER (17) suite of programs. For each DUB subfamily the HMMs were trained with multiple sequence alignments of C- and H-boxes from a nonredundant set of sequences (Table 1, HMM training set). The HMM results were further categorized based on E-values and raw scores. In every search with subfamily-specific HMMs, matches with E-values of <2.5 were selected and manually screened for the presence of Cys or His residues in the alignment. Candidates containing putative C- or H-boxes were further categorized based on the HMM scores of the domains. The HMM score is a base 2 log of the probability of alignment to the HMM divided by the probability of random alignment. A score value of zero represents 50% likelihood that the sequence is a true match to the model. Since the EBV ORF database contains 106 entries, scores greater than 6.7 correspond to a 99% probability to detect true homologs while a score of −6.6 corresponds to 1% probability. Due to the fact that the genomes of viral and other pathogenic organisms undergo higher mutation rates (16), low thresholds of statistical parameters were used to detect catalytic domains of DUBs.

**Protein family** | **Characteristic domain** | **No. of human DUBs** | **No. of verified DUBs** | **No. of orthologs** | **No. of DUBs in pattern search set** | **No. of DUBs in HMM training set** |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>USP</td>
<td>Peptidase 19</td>
<td>45</td>
<td>37</td>
<td>15</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td>UCHL</td>
<td>Peptidase 12</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>OTU</td>
<td>OTU</td>
<td>8</td>
<td>5</td>
<td>9</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>MJD</td>
<td>Josephin</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>JAMM</td>
<td>MPN+</td>
<td>8</td>
<td>2</td>
<td>29</td>
<td>34</td>
<td>21</td>
</tr>
</tbody>
</table>

*a* That is the number of DUBs with experimentally verified enzymatic activity.

*b* Due to the small number of identified sequences in the human UCH, OTU, JAMM, and MJD subfamilies, orthologs from other species were included in the data set.

*c* The JAMM metallopeptases contain only H-box domains, and sequence homologs with catalytic His residues were considered for deriving the patterns.

*d* Only C- and H-boxes with <90% sequence identity were used for training the HMMs.

**DUB functional validation of candidate DUBs: cloning of EBV ORFs into prokaryotic expression vector.** ORFs derived from the B95.8 virus were transferred from the GATEWAY expression vector pDONR207 (48) to the prokaryotic expression vector pDEST-GST (ampicillin resistance), and E. coli DH5α was used for transformation.

**UB-GFP construct.** Green fluorescent protein (GFP) was amplified by PCR from pEGFP-N1 vector (Clontech, Palo Alto, CA) with the sense primer 5'-GGTTTCAAGATCTAAAGGAGAAGACGTGGTACCCCGGCTAGCAAGGCGCAGGAGC GTTACC-3' and the antisense primer 5'-GTTTCGAGACTTGTGACTGTCGCTTCCATGGCGCAGTATG-3' cloned in the BglII and XhoI sites of the pACYC-Duet-1 vector (Novagen, Darmstadt, Germany). Ubiquitin was amplified by PCR from an Ub-βGal plasmid with the primer 5'-GTTTTGAATTCATGCAAATCTTCGTGAGAAGACTCTGACTGTTGA-3' and the antisense primer 5'-GTTAATTCGAGCTACCGCGGCGGCGGAATGAT-3' and cloned in the BamHI and XhoI sites of the pACYC-Duet-1 vector containing GFP. The underlined sequences represent restriction sites.

**DUB assays.** The GST-ORFs and Ub-GFP reporter plasmids were cotransformed by electroporation into competent BL21(DE3) bacteria that lack endogenous DUBs (9). Antibiotic-resistant colonies were selected in LB agar supplied with 100 μg of ampicillin/ml and 30 mg of chloramphenicol/ml. Exponentially growing bacteria (OD at 600 nm of 0.5) were induced for 5 h at 30°C with 0.5 mM IPTG (isopropyl-β-D-thiogalactopyranoside) and then lysed by sonication in phosphate-buffered saline supplemented with protease inhibitors (Complete Mini protease inhibitor cocktail tablets; Roche, Germany). The lysates were centrifuged for 10 min at 13,000 rpm, and cleavage of the reporter was analyzed by fractionation in acrylamide bis-Tris 4 to 12% precasted gradient gel (Invitrogen), followed by Western blotting. After transfer to polyvinylidene difluoride membrane (Millipore), the filter was blocked in phosphate-buffered saline supplemented with 5% nonfat milk and 0.1% Tween 20 and incubated for 1 h with anti-rabbit-GFP serum (Abcam, Cambridge, United Kingdom) diluted in 1% nonfat milk.
D in the supplemental material. In a first attempt, stringent search patterns were derived for each DUB family (UCH, C-box −11/+11 and H-box −7/+16; USP, C-box −8/+7 and H-box −17/+3; OTU, C-box −8/+10 and H-box −8/+5; MJD, C-box −10/+10 and H-box −6/+7; and JAMM motif −7/+17) by including at each position all residues observed in the data set (Fig. 1B, boldface residues) and then used to search the EBV ORF data set, allowing for only a single mismatch. The derived patterns were in all cases able to identify the DUBs belonging to the family but failed to identify any candidate in the EBV ORFs (results not shown), suggesting that they may be too stringent to recognize distant homologues. Less-stringent search patterns were then derived by allowing at each position additional amino acids with similar physicochemical characteristics or, when residues with different physicochemical characteristics were observed, all amino acids (indicated by an “X” in Fig. 1B and in Fig. S1A to D in the supplemental material). Regions of similarity with the C- and/or H-boxes of different DUB families were identified in 30 EBV ORFs (see Table S2 in the supplemental material). Four ORFs contained at least one putative C-box and one H-box, while three ORFs contained only one putative C-box and twenty-three ORFs contained one or more H-box-like domain.

HMM search. Family-specific HMMs were derived for the aligned C- and H-boxes of the different DUB families. In order to increase the probability of identifying distant homologues, only sequences with <90% identity were included in the training set (Table 1). Since the EBV ORF database contains only 106 sequences and the HMMs are based on short C- and H-box domains of 13 to 24 aa, high E-values and low scores were expected. HMM alignments with E-values of ≤25 were therefore considered for further analysis. All hits containing Cys or His residues were further categorized based on the HMM scores (Table 2). Scores of ≥0, corresponding to more than 50% likelihood to be a true match, were considered high. Intermediate scores were between 0 and −6.6 (down to 1% probability to be a true match), while scores less than −6.6 were considered low. Seven EBV ORFs—BALF2, BALF5, BBRF3, BcLF1, BERF3-BERF4, BRLF1, and the predicted ORF8—contained at least one C- or H-box domain with score above 0, suggesting a potential homology with the corresponding DUB domains. Two ORFs—BNRF1 and BPLF1—contained both C- and H-box-like domains with scores ≥−6.6, and eighteen additional ORFs contained either a C- or an H-box-like domain with scores of less than −6.6.

Search for conserved Cys and His residues. A further attempt to identify candidate DUBs was undertaken based on the assumption that amino acid residues that are critical for protein function may be conserved in homologues encoded by different members of the HHV family. Homologous ORFs encoded by the known HHVs were identified by sequence comparison, and the conserved Cys and His residues were located. Twenty-one EBV ORFs contain at least one Cys or His residue that is conserved in homologues encoded by all HHV members, while Cys or His residues conserved only in α- or β-HHVs were detected in two and seven ORFs, respectively; thirteen EBV ORFs shared conserved Cys or His residues only with homologues encoded by the other member of the γ-HHV subfamily, KSHV, that resembles EBV in cell tropism and oncogenic capacity (see Table S3 in the supple-

RESULTS

Four search strategies were used to identify putative DUBs encoded in the EBV genome: (i) sequence alignment with the conserved C- and H-boxes of known DUB families, (ii) pattern search of conserved catalytic domains, (iii) HMM-based searches, and (iv) identification of Cys and His residues that are conserved in homologues encoded by other members of the HHV family, followed by an HMM search.

Sequence alignment searches. A first attempt to identify putative DUBs was performed using CLUSTAL W to align the EBV ORFs with the human DUB data set. Three ORFs clustered with specific DUB subfamilies, BGLF4 and BBLF4 clustered with the OTU subfamily and BcLF1 ORF aligned with JAMMs (not shown), but the homologous regions did not contain Cys or His residues, and these viral ORFs are therefore unlikely to possess enzymatic activity. Since alignment of the entire amino acid sequence may not detect homology restricted to short domains that are critical for enzymatic activity, BLAST was used to identify EBV ORFs containing sequences of homology with the conserved C- and H-boxes of each DUB family using an arbitrary cutoff E-value of ≤10. Twenty-five candidates containing Cys or His residues in homologous regions were found in this search (see Table S1 in the supplemental material), including for example, the BBRF3 ORF that contains a region of strict homology with the USP C-box (E-value = 0.018).

Pattern search. Additional search strategies were conducted to achieve a more stringent assessment of the likelihood that the identified domains might be true C- or H-box homologues. Family-specific search patterns were derived from the aligned C- and H-box sequences of the known DUBs. To increase the stringency of the search, when fewer than 10 family members were found in the human data set, homologues from other species were also included in the alignment. The derivation of search patterns for the C- and H-boxes of the UCH family is illustrated in Fig. 1, and the search patterns for USP, OTU, MJD and JAMM families are given in Fig. S1A to
mental material). The sequences adjacent to the conserved residue were searched for putative C- and H-boxes using the DUB family specific HMMs. Nine EBV ORFs were found to contain conserved Cys or His residues located in regions homologous to the DUB C- or H-boxes, and one ORF, BGRF1/BDRF1, contained both C- and H-box homologous domains (see Table S3 in the supplemental material).

**DUB score and exclusion criteria.** A relatively large number of candidate DUBs were identified by at least one search, and several were identified by more than one. To select the most likely candidates for functional studies, scoring criteria were devised for each search, and these were then compiled in a global DUB score. The scores assigned to each search reflected the likelihood of identifying candidates that possess critical features associated with DUB activity (see Table S4 in the supplemental material).

The high-scoring candidates are structural proteins or glycoproteins associated with the virus capsid and envelope. These proteins are unlikely to possess enzymatic activity, although this cannot be excluded a priori. The remaining candidates include immediate-early, early, late, or latent proteins that are expressed in...
The type of DUB domain found in each candidate is indicated as follows: USP, ubiquitin-specific protease; JAMM, ubiquitin C-terminal hydrolase; MJD, Josephin domain protease. Boldfacing indicates high scores (>0); italics indicates intermediate scores (between 0 and −6.6); regular typeface indicates low scores (<−6.6).

### Functional validation of the candidates

In order to validate the DUB score, GST fusions of 11 EBV ORFs with scores of ≥4 and an approximately equal number of ORFs with scores of ≤3 were tested for their capacity to hydrolyze a Ub-GFP reporter plasmid coexpressed in bacteria. The results of a representative assay wherein the ORFs were tested alongside the human USP19 and a catalytic mutant that lacks enzymatic activity are shown in Fig. 3.

The reporter was efficiently cleaved by USP19, as confirmed by the detection of both free GFP (Fig. 3A) and free ubiquitin (data not shown), while cleavage was abolished by mutation of the catalytic Cys residue in the USP19Mut. In line with the reported DUB activity of this ORF, Ub-GFP was cleaved almost completely in bacteria expressing the N terminus of BPLF1 (Fig. 3A and B). The levels of cleavage significantly above background were also detected with BSLF1 and BLXLF1, while the activity of BGRF1 and BALF5 was just below the cutoff (3 × the mean percent cleavage in the presence of ORFs with scores ≤3 [Fig. 3B]). The expression of the fusion proteins was in all cases confirmed in Western blots probed with anti-GST antibodies (data not shown). In order to further characterize their DUB activity, bacterially expressed GST-BPLF1-N, GST-BSLF1, and GST-BLXLF1 were purified, and their enzyme activity was assayed by cleavage of the fluorogenic substrate Ub-AMC (Fig. 3C). The three EBV ORFs hydrolyzed Ub-AMC with different kinetics and \( K_m \) values, suggesting very different affinities for the substrate, but the specificity of the reaction was in all cases confirmed by blocking with the cysteine protease inhibitor NEM. Interestingly, the enzymatic activity of BSLF1 and BLXLF1 relative to BPLF1 was significantly improved when equal amounts of immunoprecipitated proteins expressed in mammalian cells were compared for the cleavage of Ub-AMC (compare Fig. 3C and D), suggesting that posttranslational modifications that are not achieved in bacterial cells may be required for optimal activity.

In order to further validate the DUB activity of the candidates, the predicted catalytic Cys residues in BPLF1, BSLF1, and BLXLF1 were mutated to Ala by using PCR-mediated site-directed mutagenesis, and the enzymatic activity of recombinant GST tagged proteins was tested against the Ub-AMC substrate (Fig. 4). Two putative C-boxes were predicted in the BPLF1 and BLXLF1 ORFs, while a single domain containing two Cys residues was predicted in BSLF1. Alignment of the corresponding sequences in other members of the HHV family revealed that, while BPLF1 Cys61 and BSLF1 Cys819 and Cys824 are conserved in all HHVs, other potentially catalytic Cys residues identified by the bioinformatics search are unique for EBV or shared only by some of the family members. Mutation of the conserved C61 of BPLF1 to Ala abolished the enzymatic activity, while the C65A mutation had no effect, confirming that C61 is the catalytic residue in the C-box (Fig. 4). Single mutation of the nonconserved C462 and conserved C819 of BSLF1 did not affect the enzymatic activity, while the C824A mutation was significantly reduced. Interestingly, the hydrolysis of Ub-AMC was further decreased in the double mutant C819/824A, suggesting that C824 may partly substitute for the catalytic C819 residue. Homologs of BLXLF1 were present only in alpha- and gammaherpesviruses, and the putative catalytic cysteine C491 is conserved only in gammaherpesviruses. Mutation of this residue abolished the enzymatic activity of BLXLF1, while mutation of the nonconserved C84 had no effect (Fig. 4). Thus, catalytic Cys residues were iden-
tified by mutation analysis in BPLF1, BSLF1, and BXLF1, confirming that the three EBV ORFs are bona fide viral DUBs.

To finally validate our bioinformatics approach, the scoring strategy used to identify BPLF1, BSLF1, and BXLF1 was applied to their homologues encoded by other human herpesviruses. All homologs of BPLF1 and BSLF1 received scores of ≥4, and several scored even higher than the EBV-encoded protein (Table 4). This is in line with the reported DUB activity of several BPLF1 homologues. In agreement with the very low sequence similarity of the BXLF1 homologues expressed in alpha- and gammaherpesvirus these ORFs received very low DUB scores, suggesting that this protein may have different functions in different viruses.

**DISCUSSION**

In this study we used a bioinformatics approach to identify distant functional homologues of human ubiquitin deconjugases encoded in the EBV genome. This task is complicated by the great genetic variability of viruses, which results in poor sequence conservation even between homologous proteins expressed by different members of the same virus family. Indeed, currently used bioinformatics tools, such as BLAST searches

---

**TABLE 3. ORF annotation and filtering criteria**

<table>
<thead>
<tr>
<th>EBV ORF</th>
<th>DUB score</th>
<th>EBV protein</th>
<th>Homologue</th>
<th>Expression in virus cycle</th>
<th>Location</th>
<th>Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBRF3</td>
<td>9</td>
<td>Glycoprotein M</td>
<td>HSV gM, CMV gM, KSHV gM</td>
<td>Late Envelope</td>
<td>M (−)</td>
<td></td>
</tr>
<tr>
<td>BcLF1</td>
<td>7</td>
<td>Major capsid protein</td>
<td>UL19 UL86 ORF25</td>
<td>Late</td>
<td>Capsid</td>
<td>C (−)</td>
</tr>
<tr>
<td>BALF5</td>
<td>7</td>
<td>DNA pol catalytic subunit</td>
<td>UL30 UL54 ORF9</td>
<td>Early</td>
<td>Nucleus</td>
<td>R (−)</td>
</tr>
<tr>
<td>BGRF1/BDRF1</td>
<td>7</td>
<td>DNA packaging, Terminator</td>
<td>UL15 UL89 ORF29</td>
<td>Late</td>
<td>Nucleus</td>
<td>P (−)</td>
</tr>
<tr>
<td>BNRF1</td>
<td>6</td>
<td>Major tegument protein</td>
<td>UL36 UL48 ORF64</td>
<td>Late</td>
<td>Tegument</td>
<td>T (−)</td>
</tr>
<tr>
<td>BPLF1</td>
<td>6</td>
<td>Large tegument protein</td>
<td>UL22 UL75 ORF22</td>
<td>Late</td>
<td>Envelope</td>
<td>M (−)</td>
</tr>
<tr>
<td>BXLF2</td>
<td>5</td>
<td>gp85, gH homologue</td>
<td>UL29 UL57 ORF6</td>
<td>Late</td>
<td>Nucleus</td>
<td>R (−)</td>
</tr>
<tr>
<td>ORF35</td>
<td>5</td>
<td>Hypothetical</td>
<td>UL29 UL57 ORF6</td>
<td>Early</td>
<td>Membrane</td>
<td>M (−)</td>
</tr>
<tr>
<td>ORF88</td>
<td>5</td>
<td>Hypothetical</td>
<td>UL29 UL57 ORF6</td>
<td>Early</td>
<td>Membrane</td>
<td>M (−)</td>
</tr>
<tr>
<td>BALF2</td>
<td>5</td>
<td>ssDNA binding protein</td>
<td>UL29 UL57 ORF6</td>
<td>Early</td>
<td>Nucleus</td>
<td>R (−)</td>
</tr>
<tr>
<td>BSLF1</td>
<td>5</td>
<td>Helicase/primase complex</td>
<td>UL52 UL70 ORF36</td>
<td>Early</td>
<td>Nucleus</td>
<td>R (−)</td>
</tr>
<tr>
<td>BILF1</td>
<td>4</td>
<td>gp64, constitutive GPCR</td>
<td>UL52 UL70 ORF36</td>
<td>Early</td>
<td>Membrane</td>
<td>M (−)</td>
</tr>
<tr>
<td>BERF3-BERF4</td>
<td>4</td>
<td>EBNA3C</td>
<td>UB40 ORF60</td>
<td>Early</td>
<td>Nucleus</td>
<td>L (−)</td>
</tr>
<tr>
<td>BaRF1</td>
<td>4</td>
<td>RNase reductase</td>
<td>UB40 ORF60</td>
<td>Early</td>
<td>Nucleus</td>
<td>N (−)</td>
</tr>
<tr>
<td>BRLF1</td>
<td>4</td>
<td>Rta, transactivator</td>
<td>ORF50</td>
<td>I-Early</td>
<td>Nucleus</td>
<td>S (−)</td>
</tr>
<tr>
<td>BXLF1</td>
<td>4</td>
<td>Thymidine kinase</td>
<td>UB23 ORF21</td>
<td>Early</td>
<td>Nucleus</td>
<td>N (−)</td>
</tr>
</tbody>
</table>

*For the details on the scoring, see Table S4 in the supplemental material.

*b Classification of protein based on function: C, capsid; M, membrane protein; N, nucleotide metabolism; L, latency; P, packaging; R, replication; S, signaling, transactivator, transcription factor; T, tegument.
or more sophisticated threading programs that use sequence alignment and protein fold recognition algorithms, fail to detect homologies between known viral and cellular DUBs, suggesting that the structural properties of the viral proteases are clearly distinct. This was recently confirmed by the crystal structure of a DUB encoded by the murine gammaherpesvirus M48 that, based on the arrangement of the active-site residues and the architecture of the ubiquitin interacting interface, was classified as the first member of a new class of DUBs (43).

We reasoned that, in spite of strongly divergent sequences, a higher degree of conservation might be expected in regions that are important for enzymatic activity. Thus, although the homology between the catalytically relevant regions of viral and cellular enzymes falls below the statistically significant cutoffs used in conventional searches, distant homologues may be identified using less-stringent search criteria. The human genome encodes for approximately 95 DUBs subdivided in five families (38). Although the sequences of individual DUBs vary widely, the members of each family share a high degree of homology in regions surrounding conserved Cys and His residues that form the catalytic core of the enzymes. We therefore focused on the identification of EBV ORFs containing a region of homology with the C- and H-boxes of each DUB family and scored the candidates based on the degree of homology with the consensus sequences and on the presence of only one or both types of consensus domains. Four different strategies based on sequence alignment, pattern search, HMM motifs, and detection of conserved Cys and His residues in the context of HMM motifs were used to search for putative C- and H-boxes. The use of multiple searches provides a distinct advan-

![FIG. 3. Functional validation of the candidate DUBs. (A) Representative Western blot illustrating cleavage of the Ub-GFP reporter by a restricted panel of EBV ORFs. BL21 cells were cotransfected with Ub-GFP, and the indicated ORF and protein expression was induced by treatment with IPTG. Cell lysates were fractionated on 4 to 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis gradient gels, and GFP was detected by Western blotting. Cleavage of the reporter in lysates expressing active DUBs is visualized by appearance of a free GFP band. (B) Enzymatic activity of candidate EBV DUBs (score ≥ 4) and controls (score ≤ 3). The intensity of the bands corresponding to the uncleaved reporter and free GFP was determined by densitometry, and the percent specific cleavage was calculated as follows: intensity of the free GFP band – background/total GFP (uncleaved + cleaved) – background × 100. A cutoff activity value was calculated as 3 × the mean percent specific cleavage in cells expressing ORFs with a score of ≥3. The mean of three experiments in which all of the ORFs were tested in parallel is shown in the figure. (C) Time course of Ub-AMC hydrolysis. Ub-AMC (1.2 μM) was mixed with 200 nM concentrations of purified GST-USP19wt, GST-USP19mut, GST-BPLF1-N, GST-BSLF1, and GST-BXLF1 in the absence (empty symbols) or presence (filled symbols) of NEM, and the reactions were monitored for 180 min. The release of fluorescent AMC, expressed in arbitrary units (AU), shows the activity of these enzymes to cleave Ub-AMC. The Km of the reactions was calculated in separate experiments and is indicated in the Figure. (D) Time course of Ub-AMC hydrolysis of BPLF1, BSLF1, and BXLF1 expressed in mammalian cells. The ORFs were cloned in frame in the 3XFlag tag vector and expressed in HEK293 cells. The proteins were immunoprecipitated from the transfected cells using antibodies to the Flag tag, and the cleavage of Ub-AMC was assayed as described for panel C. Equal amounts of proteins were used in all enzymatic assays. The faster kinetics of Ub-AMC cleavage (compare panels C and D) indicates that the enzymatic activity of mammalian expressed BSLF1 and BXLF1 is significantly improved compared to the proteins expressed in bacteria.]
tage since each search is biased by different criteria. Thus, while sequence alignment scores for the presence of a continuous sequence of homologous residues, the pattern search restricts the hits to sequences that share similar residues at critical positions in the motif and HMM gives a global score based on the likelihood that a given residue will be found in each position of the motif. A further refinement of the HMM search was based on the assumption that functionally relevant proteins will be conserved among members of a virus family, and therefore conserved Cys and His residues will be found in the putative catalytic domains. By assigning individual scores to each of these searches and then combining the scores into a global DUB score, we have identified 16 candidates with DUB scores of \( \frac{1}{10^4} \) out of 106 annotated or predicted ORF in the EBV genome. Five of the candidates were excluded from further analysis because the putative catalytic domains are located within or at the opposite sides for transmembrane domains or because they are either known structural components of the virus capsid or envelope glycoproteins and are therefore unlikely to possess DUB activity.

A previously identified viral DUB encoded by the BPLF1 ORF was found among the high score candidates. BPLF1 is the EBV homologue of UL36, a presumably multifunctional protein of the alphaherpesvirus HSV-1. DUB activity was previously mapped to the N-terminal portion of UL36 by using a functional proteomic approach based on labeling of total cell extracts with and epitope-tagged, Ub-based suicide substrates followed by immunoprecipitation, and identification of the labeled enzyme by mass spectrometry (23). Using a reporter substrate that carries ubiquitin fused to GFP in a conformation that mimics the ubiquitin precursors and, to some extent, a true ubiquitinated substrate, we have now confirmed that the N terminus of BPLF1 is a very potent ubiquitin deconjugase. Interestingly, two additional ORFs in the high-score group, BSLF1 and BXLF1, were also capable of cleaving Ub-GFP at levels significantly above background. The enzymatic activity of

FIG. 4. Identification of the catalytic Cys residues by site-directed mutagenesis. Sequence alignments of the putative catalytic C-boxes in BPLF1, BSLF1, and BXLF1 and their homologs encoded by other HHV members shows conservation of Cys residues that are involved in DUB activity. The conserved residues are shown in white text with a black background, while similar residues in the alignment are shaded in gray. The asterisk above the alignment indicates the Cys residues identified in the bioinformatics search. Mutations were introduced into the predicted catalytic Cys residues, and the ability of the mutants to hydrolyze Ub-AMC was assayed as described in the legend to Fig. 3.
the new putative DUBs was in both cases confirmed by the capacity of the purified GST fusion proteins to cleave the fluorogenic substrate Ub-AMC. Furthermore, the catalytic Cys residues were identified in both proteins by mutation analysis, confirming that these EBV ORFs are bona fide DUBs. It is noteworthy that bacterially expressed full-length BSLF1 and BXLFL1 exhibited a poor enzymatic activity compared to both the N-terminal portion of BPLF1 and the human DUB USP19 that was included as control. Comparison of the enzymatic activity of purified proteins expressed in mammalian cells suggests that this may be due to the requirement for posttranslational modifications that are not achieved in bacteria. Indeed, only the N terminus of UL36 was identified in the functional screen performed in HSV-1-infected cells, suggesting that the protein may be processed during infection.

The new viral DUBs identified in the present study are expressed in the nucleus of EBV-infected cells during the early phases of the productive virus cycle and, based on their annotated functions, are involved in DNA replication and nucleotide metabolism. It is noteworthy that the efficiency and fidelity of DNA replication is regulated by posttranslational modification of several components of the replication complex, including for example, monoubiquitination, polyubiquitination, and sumoylation of PCNA (36). The BSLF1 gene product shares 23% sequence identity with HSV UL52, the primase component of the trimolecular helicase-primase complex was shown to be essential to HSV replication (12, 13). In EBV the complex is made up of BSLF1, BBLF4 (UL5 homolog), and BBLF2/3 (UL8 homolog) (31), but its activity in viral replication has not been directly confirmed. Similar to UL52, BSLF1 contains several conserved amino acid domains, such as the DxD motif (aa 481 to 498) and a putative zinc finger (Znf; aa 782 to 829) motif at the C-terminal end of protein that are also found in prokaryotic and eukaryotic primases (7, 15, 24). Mutational analysis of the Znf of HSV UL52 showed that this region is required for the virus replication (7, 8, 10). The Znf of BSLF1 partially overlaps with the predicted catalytic C-box.

BXLFL1 gene encodes a protein of 607 aa that was annotated as the EBV thymidine kinase (TK) gene due to its capacity to complement TK-negative strains of E. coli (30) and partial sequence homology with regions of HSV-1 TK (19). We found that the regions that identify BXLFL1 as a putative DUB and the catalytic Cys residue are not conserved in other herpesvirus homologues. Moreover, in contrast to other family members, BXLFL1 was shown to localize to the centrosomes, where it encircles the tubulin-rich centrioles in a microtubule-independent manner (20). It remains to be seen whether this atypical localization may reflect a dual enzymatic activity that is not present in the homologues.

In conclusion, we have shown that it is possible to use bioinformatics tools to identify distant homologues of DUBs encoded by viruses. Our strategy of choosing candidates based on a global DUB score obtained by combining the results of multiple low stringency searches identified 16 high scoring ORFs in a database of 106 putative or confirmed EBV ORFs. Enzymatic activity was confirmed for 3 of 11 ORFs tested, whereas ORFs with low score and ORFs lacking putative C- or H-boxes were uniformly negative. Thanks to its relatively good predictive capacity, our approach could provide a valuable complement to the functional screening described by Kattenhorn et al. (23), which, due to the requirement for validation by mass spectrometry, is strongly biased toward enzymes that are over-expressed in the infected cells. Further studies will be required to assess how the DUB activity of BPLF1, BSLF1, and BXL1 contributes to their role in virus replication and to the remodeling of the cellular environment during productive EBV infection.

**ACKNOWLEDGMENTS**

We thank Joel Hedlund for help with the sequence clustering procedure.

This study was supported by grants from the Swedish Cancer Society, the Swedish Medical Research Council, and the Karolinska Institutet, Stockholm, Sweden (to M.G.M. and K.L.); by the Carl Trygger Foundation and Linköping University (to B.P.); by Bayerisches Genomforschungszentrum and MRC G0501453 (to J.H.); and by the European Community Integrated Project on Infection and Cancer, project LSIC-CT-2005-018704 (to M.G.M.).

**REFERENCES**

merase to herpes simplex virus type 1 replication foci in cells expressing 

biquitinating protease expressed by the K3 family of viral and cellular ubiquitin E3 

tations in the putative zinc-binding motif of UL52 demonstrate a complex 
interdependence between the UL5 and UL52 subunits of the human herpes 


engineering study of the role of aspartate 158 in the catalytic mechanism of 


18. Frias-Staheli, N., N. V. Giannakopoulou, M. Kikkt, S. L. Taylor, A. Brid- 
gen, J. Paragas, J. A. Rict, R. R. Rowland, C. S. Schmaljohn, D. J. Len- 
schow, E. J. Snijder, A. Garcia-Sastre, and H. W. ten Virgim. 2007. Oavian 
tumor domain-containing viral proteases evade ubiquitin- and ISG15-depen- 

Structural basis for the dual thymidine and thymydilate kinase activity of herpes thymidine kinases. Structure 11:1265–1277.

thymidine kinase is a centrosomal resident precisely localized to the periph- 


Lehner. 2002. Ubiquitylation of MHC class I by the K3 viral protein signals 

Ploegh. 2005. A deubiquitinating enzyme encoded by HSV-1 belongs to a 
family of cysteine proteases that is conserved across the family Herpesviridae. 

herpes simplex virus type 1: a mutation in the UL52 subunit abolishes 

25. Kovaleva, A., C. Chable-Bessia, G. Cantarella, A. Israel, D. Wallach, and 
G. Courtous. 2003. The tumour suppressor CYLD negatively regulates 


27. Rice, P., I. Longden, and A. Bealess. 2000. EMBOSS: the European molecu- 


29. Rajewsky, A., J. Poh, J. Garmananda, C. Boyle, A. Thompson, M. Liu, P. 
Freeman, J. C. Hinton, and D. W. Holden. 2007. SceI, a Salmonella deu- 
Sci. USA 104:3502–3507.

Howley. 1990. The E6 oncoprotein encoded by human papillomavirus types 


32. Schiek, C., W. A. Weihofen, E. Frijns, L. M. Kattenhorn, R. Gaudet, and 
H. L. Ploegh. 2007. Structure of a herpesvirus-encoded cysteine protease 

33. Schwartz, A. L., and A. Ciechanover. 1999. The ubiquitin-proteasome path- 

5093.

35. Sun, L., and Z. J. Chen. 2004. The novel functions of ubiquitination in 

36. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: 
improving the sensitivity of progressive multiple sequence alignment through 
sequence weighting, position-specific gap penalties and weight matrix choice. 

37. Ueda, Y., A. Dong, C. Zeretzke, C. Atzler, B. Berger, S. V. 
Rajagopala, M. Roupelieva, D. Rose, E. Fossum, and J. Haas. 2006. Her- 
pesviral protein networks and their interaction with the human proteome. 

Koonin, and R. J. Deshaies. 2002. Role of Rpn11 metalloprotease in 
deubiquitination and degradation by the 26s proteasome. Science 298: 
611–615.

35:590–593.

40. Yokoyama, N., K. Fujii, M. Hirata, K. Tamai, T. Kiyono, K. Kuzushima, Y. 
BBLF4, BBLF1 and BBLF2/3 proteins and their interactive properties. J. Gen. 

Hixen. 2005. Yeast virulence factor Vpl1 acts as a deubiquitinate to inhibit 