Identification of Major Histocompatibility Complex Class II-Restricted Antigens and Epitopes of the Epstein-Barr Virus by a Novel Bacterial Expression Cloning Approach

Slavoljub Milosavic, Uta Behrends, Dinesh Adhikary, and Josef Mautner*

Clinical Cooperation Group, Technical University Munich, Children’s Hospital, Hematology-Oncology, Kölner Platz 1, 80804 Munich, Germany, and Institute for Clinical and Molecular Biology, GSF-National Research Center for Environment and Health, Marchioninistr. 25, 81377 Munich, Germany

Received 8 June 2006/Accepted 7 August 2006

Epstein-Barr virus (EBV)-specific T cells have been successfully used to treat or prevent EBV-positive lymphoproliferative disease in hematopoietic stem cell transplant recipients, but the antigens recognized by the infused CD4+ T cells have remained unknown. Here, we describe a simple procedure that permits the identification of viral T-helper (Th)-cell antigens and epitopes. This direct antigen identification method is based on the random expression of viral polypeptides fused to chloramphenicol acetyltransferase (CAT) in bacteria, which are subsequently fed to major histocompatibility complex class II+ antigen-presenting cells and probed with antigen-specific T cells. The fusion of antigenic fragments to CAT offers several advantages. First, chloramphenicol treatment allows the selection of bacteria expressing antigen-CAT fusion proteins in frame, which greatly reduces the number of colonies to be screened. Second, antigenic fragments fused to CAT are expressed at high levels, even when derived from proteins that are toxic to bacteria. Third, the uniformly high expression level of antigen-CAT fusion proteins permits the establishment of large and representative pool sizes. Finally, antigen identification does not require knowledge of the restriction element and often leads directly to the identification of the T-cell epitope. Using this approach, the BALF4 and BNRF1 proteins were identified as targets of the EBV-specific T-helper-cell response, demonstrating that lytic cycle antigens are a relevant component of the EBV-specific Th-cell response.

CD4+ T helper (Th) cells play a central role in the immune response against viruses and tumors. Th cells recognize peptide/major histocompatibility complex (MHC) class II complexes on the surface of professional antigen-presenting cells (pAPC), e.g., macrophages, B cells, and dendritic cells. The peptides presented on MHC class II molecules are derived mostly from exogenous or cell membrane proteins processed in the endosomal/lysosomal compartment (18, 23). Given the central role of Th cells in adaptive immune responses (3, 10), the identification of antigens recognized by virus-specific Th cells is critical for the development of vaccines and other forms of immunotherapy (20, 21).

The Epstein-Barr virus (EBV) is a ubiquitous gammaherpesvirus implicated in the pathogenesis of a number of human malignancies (6). Protective immunity against EBV is mediated by T cells, but the targets of the EBV-specific Th-cell response are still poorly defined. Although CD4+ T cells specific for several latent and few lytic cycle antigens have been isolated, the low frequency of EBV-specific Th cells in peripheral blood and the lack of generic methods for the identification of MHC class II antigens have hampered a detailed characterization of the EBV-specific Th-cell response (4, 19). The EBV genome codes for approximately 100 different proteins, most of which are expressed only during lytic replication of the virus (7). Which of these proteins elicit Th-cell responses and how T cells specific for these proteins contribute to EBV immunity are mostly unknown (4, 19).

For the identification of MHC class II-restricted antigens, two methods are currently used: the biochemical fractionation of proteins from cells expressing the antigen combined with mass spectrometric sequencing of highly purified fractions containing the antigen (12, 17) and genetic approaches based on the testing of T cells with pools of cDNA libraries expressed in prokaryotic (22) or eukaryotic (24) cells. Since spontaneous lytic replication of EBV occurs in only a small percentage of virus-infected cells (7), viral mRNAs and proteins account for only a small proportion of all mRNAs and proteins expressed in these cells. Thus, current methods of antigen identification are impractical to define the antigens recognized by these T cells.

With the cloning of the whole EBV genome of about 170 kb on a F-factor-derived plasmid in bacteria, viral DNA has become available as a potential source of antigens (5, 14). Libraries established from viral DNA, as opposed to cDNA from virus-infected cells, are highly enriched for viral antigens, which should facilitate their molecular identification. Such a genomic expression cloning approach has been successfully applied previously to identify mycobacterial Th-cell antigens (2). In those experiments, genomic DNA of the pathogen was sheared to a size range of 1 to 4 kb, the DNA fragments were expressed in *Escherichia coli*, and the bacterial cultures were subsequently added to dendritic cells. Following phagocytosis, peptides derived from bacterially expressed proteins were presented on MHC class II molecules for recognition by pathogen-specific CD4+ T cells. Since almost all of the approxi-
ultimately 80 genes of EBV expressed during lytic replication lack introns, such an approach might also be applicable to the identification of EBV antigens. However, the level of expression of most EBV proteins is low in bacteria, and truncating the open reading frames (ORFs) by shearing may further compromise protein stability and hence expression levels.

In previous experiments, we noted that peptides fused to green fluorescent protein are expressed at high levels, even when derived from proteins that are difficult to express in bacteria (11). Due to their green appearance, bacterial colonies expressing antigen-green fluorescent protein fusion proteins in frame are easily identified. Based on this observation, we developed a simple method to identify T-helper-cell epitopes within known antigens by picking and analyzing single green bacterial colonies (11). Identification of unknown antigens with this method, however, is not feasible because of the much larger number of colonies that have to be screened. Here, we adapted this bacterial expression cloning approach for the identification of unknown viral T<sub>H</sub>-cell antigens and applied this method to identify antigens recognized by EBV-specific T<sub>H</sub>-cell clones.

**MATERIALS AND METHODS**

**Bacterial strain and culture.** *Escherichia coli* strain XL1-Blue MR<sup>F</sup> (Stratagene) was used in all experiments described here. The following concentrations of antibiotics were used: 100 μg/ml ampicillin, 15 μg/ml tetracycline, and 30 μg/ml chloramphenicol. Protein expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 1 mM.

**Construction of expression vectors.** To generate the expression vector, the open reading frame of chloramphenicol acetyltransferase (CAT) was cloned into this plasmid together with an MCS. This MCS contains unique recognition sequences for several restriction enzymes that allow the insertion of blunt-ended antigenic fragments or sticky-end DNA linkers. The Stul and the MscI sites are separated by stop codons in all three reading frames (shaded boxes) to prevent CAT expression from nonrecombinant plasmids. Plasmids F315B and F315C differ from F315A by one or two nucleotides inserted in front of the multiple cloning site, respectively. (B) Fusion proteins are highly expressed in bacteria. To assess protein expression from these plasmids, short antigenic fragments were randomly inserted into the MCS. Following transformation, 13 chloramphenicol-resistant bacterial colonies were further examined. Bacterial extracts were separated by SDS-PAGE and analyzed by Coomassie staining and Western blotting using the AntiXpress antibody.

**FIG. 1.** Bacterial expression of short antigen-CAT fusion proteins. (A) Schematic depiction of the expression vector. The expression plasmids F315A, F315B, and F315C are derivatives of the bacterial expression plasmid pTrcHisA (Invitrogen), which contains an IPTG-inducible promoter and translational start site (ATG) for the expression of N-terminally His- and antibody epitope (AntiXpress)-tagged fusion proteins. The coding sequence of CAT was inserted into this plasmid together with an MCS. This MCS contains unique recognition sequences for several restriction enzymes that allow the insertion of blunt-ended antigenic fragments or sticky-end DNA linkers. The Stul and the MscI sites are separated by stop codons in all three reading frames (shaded boxes) to prevent CAT expression from nonrecombinant plasmids. Plasmids F315B and F315C differ from F315A by one or two nucleotides inserted in front of the multiple cloning site, respectively. (B) Fusion proteins are highly expressed in bacteria. To assess protein expression from these plasmids, short antigenic fragments were randomly inserted into the MCS. Following transformation, 13 chloramphenicol-resistant bacterial colonies were further examined. Bacterial extracts were separated by SDS-PAGE and analyzed by Coomassie staining and Western blotting using the AntiXpress antibody.

**Protein expression and purification.** To isolate proteins, bacteria were incubated in Superbroth medium containing ampicillin and tetracycline. After cultivation overnight, 100 μl of bacterial suspension from each well was transferred into 1.5 ml Superbroth medium/well containing 1 mM IPTG and 30 μg/ml chloramphenicol in deep-well blocks (Peqlab). One hundred microliters of Superbroth medium containing 40% glycerol was added to the rest of the bacterial culture in microtiter plates, and the cultures reached an OD<sub>600</sub> of 5, the bacterial pools in deep-well blocks were incubated under vigorous agitation at 37°C in a bacterial shaker. When the optical density at 600 nm (OD<sub>600</sub>) of the wells reached 1, the bacteria were harvested by centrifugation (3,000 × g for 15 min), and the supernatant was discarded. To obtain bacterial suspensions with an OD<sub>600</sub> of 5, the bacterial pellets in the deep-well blocks were resuspended in one-fifth of the original volume (300 μl/well).

**Protein expression and purification.** To isolate proteins, bacteria were incubated in Superbroth medium containing 1 mM IPTG and 30 μg/ml chloramphenicol and harvested by centrifugation (3,000 × g for 15 min) when the cultures reached an OD<sub>600</sub> of 0.8. The bacterial pellet was resuspended in 50 ml of lysis buffer (100 mM Na<sub>2</sub>HPO<sub>4</sub>, 10 mM Tris-HCl, 8 M urea, 10 mM imidazole, 0.05% Tween 20, pH 8.0). Following centrifugation (5,000 × g for 15 min) to pellet insoluble bacterial debris, histidine (His)-tagged proteins were purified by restriction enzyme digestion and sequence analysis of the modified regions.

For the insertion of antigenic DNA fragments into the expression vector, equal amounts of plasmids F315A, F315B, and F315C were mixed and cut with Stul, respectively. The vector DNA was purified using a Qiagen kit (Qiagen). To define T-cell epitopes, 20 complementary oligonucleotides were synthesized, which, upon annealing, generated cohesive ends for directed ligation into the expression vector (Fig. 1A).
using nickel-nitrotriacetic acid agarose beads according to the guidelines of the manufacturer (QIAGEN). The protein eluate was dialyzed against phosphate-buffered saline for 2 days, the concentration was determined using Bradford reagent (Bio-Rad), and the solutions were brought to a concentration of 500 μg/ml. The proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and identity and purity were analyzed by Coomassie staining and by Western blotting using the AntiXpress antibody (Invitrogen) and the ECL Plus detection system (Amersham Biosciences).

**Cell culture conditions.** Lymphoblastoid cell lines (LCL) and mini-LCL are EBV-immortalized B-cell lines efficient in MHC class I and class II antigen presentation (13). LCL and mini-LCL were generated by infecting primary B cells with wild-type EBV derived from the B95.8 marmoset producer cell line and a genetically engineered EBV mutant (mini-EBV), respectively. In contrast to LCL-mini-LCL fail to express lytic cycle proteins of EBV (1, 13). LCL and mini-LCL were grown as suspension cultures in LCL medium (RPMI 1640 supplemented with 2 mM L-glutamine, 1% nonessential amino acids, 1 mM sodium pyruvate, 50 μg/ml gentamicin, and 10% fetal calf serum). The cells were incubated for 24 h with either purified proteins or whole bacteria. At the end of the incubation period, 100 μl of supernatant was removed, and 1 × 10^5 T cells in 100 μl LCL medium were added. Cytokine secretion by the T cells was measured 20 h later by enzyme-linked immunosorbent assay (ELISA) (R&D Systems) as described previously (9).

The following T<sub>H</sub>-cell clones were used in this study: 20-4/A4, recognizing neomycin phosphotransferase II (Neo<sup>r</sup>) amino acids (aa) 216 to 229 (DRYQD IALATRDIA) (16); M1-E5, recognizing influenza virus M1 aa 234 to 248 (LE NLQAYQKRMGVQL) (15); and 3A-3D5, recognizing EBNA3A aa 142 to 156 (IALATRDIA) (16); M1-E5, recognizing influenza virus M1 aa 234 to 248 (LE NLQAYQKRMGVQL) (15); and 3A-3D5, recognizing EBNA3A aa 142 to 156 (IALATRDIA) (16); M1-E5 (relevant [rel.]) or a plasmid coding for an irrelevant (irrel.) CAT fusion protein were mixed at the indicated ratios, and 10 μl of each mix was added to LCL derived from the same donor as the M1-E5 T cells in medium supplemented with 50 μg/ml gentamicin to terminate bacterial growth. In parallel, purified M1-CAT fusion protein (500 μg/ml) and control protein (500 μg/ml) were mixed at the indicated ratios and added to LCL. After 24 h of incubation, M1-specific CD4<sup>+</sup> T cells were added, and granulocyte-macrophage colony-stimulating factor (GM-CSF) secretion by the T cells was determined 20 h later by ELISA. The T cells specifically recognized target cells incubated with whole bacteria or purified protein with similar efficiencies over a broad concentration range.

**RESULTS**

**Design of expression vector.** For the identification of T<sub>H</sub> antigens by expression cloning of genomic libraries in bacteria, the viral DNA fragments must be provided with translation initiation signals and must be expressed as fusion proteins, because short peptides are unstable in bacteria. The random insertion of short viral sequences between the translational initiation site and the ORF of a fusion partner results in the expression of in-frame fusion proteins in only one out of nine cases. To eliminate frameshifted bacterial recombinants, various selection markers were tested as fusion partners. Of these, the CAT gene was selected, first, because CAT fusion proteins were expressed at high levels and, second, because bacterial recombinants expressing CAT fusion proteins in frame were selectable with chloramphenicol (data not shown).

To create a bacterial expression vector that allows the random insertion of short DNA fragments, an MCS and the CAT gene were inserted downstream of the IPTG-inducible promoter in the pTrcHisA plasmid (Invitrogen), fusing a His tag and the AntiXpress antibody epitope to the N terminus of CAT. The resulting fusion proteins could be purified over nickel columns, and expression was monitored by Western blotting (Fig. 1A). To test whether bacterial colonies expressing different antigen-CAT fusion proteins were expressed at similar levels, we randomly ligated blunt-ended short DNA fragments into the vector mix. Following bacterial transformation and chloramphenicol selection, 13 randomly chosen bacterial colonies were further analyzed. Sequence analysis showed that the size of the cDNA inserts ranged from 27 to 159 bp. All colonies except colony 6 expressed fusion proteins consisting of an N-terminal His tag, the inserted peptide fragment, and CAT. These results indicated that bacterial recombinants expressing frameshifted fusion proteins are efficiently eliminated by chloramphenicol treatment, thereby reducing the number of colonies to be screened by almost 10-fold. In parallel, bacterial lysates of the colonies were separated by SDS-PAGE, and the proteins were stained with Coomassie dye and analyzed by Western blotting. All fusion proteins were expressed at similar and very high levels, comprising up to 50% of the bacterial protein lysate (Fig. 1B). Translation of the fusion protein expressed by colony 6 was initiated at an ATG within the inserted DNA fragment, resulting in an antigen-CAT fusion protein that remained undetected in the Western blot experiment (Fig. 1B).

**Efficiency of antigen presentation.** To assess the efficiency of antigen presentation following incubation of pAPC with whole bacteria expressing antigen-CAT fusion proteins, we incubated LCL with increasing amounts of bacteria expressing the epitope recognized by the influenza virus M1-specific T-cell clone M1-E5 (relevant [rel.]) or a plasmid coding for an irrelevant (irrel.) CAT fusion protein were mixed at the indicated ratios, and 10 μl of each mix was added to LCL derived from the same donor as the M1-E5 T cells in medium supplemented with 50 μg/ml gentamicin to terminate bacterial growth. In parallel, purified M1-CAT fusion protein (500 μg/ml) and control protein (500 μg/ml) were mixed at the indicated ratios and added to LCL. After 24 h of incubation, M1-specific CD4<sup>+</sup> T cells were added, and granulocyte-macrophage colony-stimulating factor (GM-CSF) secretion by the T cells was determined 20 h later by ELISA. The T cells specifically recognized target cells incubated with whole bacteria or purified protein with similar efficiencies over a broad concentration range.

**Direct epitope mapping in known antigens.** To assess the feasibility of this expression cloning approach, we first sought to define T<sub>H</sub>-cell epitopes within known antigens. The coding
sequences of the EBV proteins BRLF1 and BALF4 were digested with single or combinations of frequently cutting restriction enzymes, and the resulting short DNA fragments were ligated into the expression vector. Following transformation, bacteria were cultured in 96-well plates at 14 CFU/well in the case of BALF4 and at 60 CFU/well for BRLF1. From each well, 10 μl of bacterial cultures (OD_{600} of 5) was added to pAPC and incubated for 24 h before the addition of T cells. After 20 h, granulocyte-macrophage colony-stimulating factor (GM-CSF) secretion by the T cells was measured by ELISA. (A) The BALF4-specific B5 T cells strongly recognized five pools (left), of which pool D1 was analyzed further (right). All single positive bacterial colonies derived from this pool expressed the same peptide of 27 aa (AA) derived from BALF4 and expressed in frame with CAT. The T-cell epitope, marked as a boxed sequence, was verified using synthetic oligonucleotides partially spanning the identified region. (B) The BRLF1-specific T-cell clone 1H3 recognized 11 pools (left), of which pool H5 was analyzed further (right). All single colonies of this pool tested contained an insert coding for a peptide of 22 aa derived from BRLF1. The epitope recognized by the T cells was determined using DNA linkers (depicted as a boxed sequence).

FIG. 3. Identification of T-cell epitopes within known antigens. The coding sequences of EBV BALF4 and BRLF1 genes were cleaved with frequently cutting restriction enzymes, and the fragments were ligated into the expression vector. Following transformation, bacteria were cultured in 96-well plates at 14 CFU/well in the case of BALF4 and at 60 CFU/well for BRLF1. From each well, 10 μl of bacterial cultures (OD_{600} of 5) was added to pAPC and incubated for 24 h before the addition of T cells. After 20 h, granulocyte-macrophage colony-stimulating factor (GM-CSF) secretion by the T cells was measured by ELISA. (A) The BALF4-specific B5 T cells strongly recognized five pools (left), of which pool D1 was analyzed further (right). All single positive bacterial colonies derived from this pool expressed the same peptide of 27 aa (AA) derived from BALF4 and expressed in frame with CAT. The T-cell epitope, marked as a boxed sequence, was verified using synthetic oligonucleotides partially spanning the identified region. (B) The BRLF1-specific T-cell clone 1H3 recognized 11 pools (left), of which pool H5 was analyzed further (right). All single colonies of this pool tested contained an insert coding for a peptide of 22 aa derived from BRLF1. The epitope recognized by the T cells was determined using DNA linkers (depicted as a boxed sequence).

Identification of unknown EBV antigens with the direct antigen identification (DANI) method. In a previous attempt to define the EBV-specific T-helper-cell response, virus-specific T-cell lines were generated by repeated stimulation of peripheral blood CD4 T cells with autologous LCL in vitro, and LCL-reactive T-cell lines were cloned by limiting dilution. Single-
cell outgrowths were tested for recognition of autologous LCL and mini-LCL that had been established by the infection of B cells with a genetically engineered EBV mutant. Mini-LCL are identical to LCL in terms of latent cycle protein expression, antigen presentation, and T-cell costimulation but do not express lytic cycle proteins of EBV (1, 13). The exclusive recognition of LCL but not mini-LCL by some of the T-cell clones therefore implied that these T cells were specific for antigens expressed during lytic replication of the virus. The EBV genome contains approximately 100 putative open reading frames, but neither the precise number nor the identities of all genes expressed during lytic replication are known.

In an attempt to identify the antigens recognized by these EBV-specific CD4<sup>+</sup>-T-cell clones with the DANI method, EBV DNA was isolated and digested with frequently cutting restriction enzymes. The resulting DNA fragments were ligated into the expression vector mix as described above, and the resulting EBV library was plated at 60 CFU/well. With an average insert size of 82 bp and an EBV plasmid size of 182 kb, a single 96-well plate would cover the whole EBV genome approximately 2.5 times. Ten microliters of IPTG-induced bacterial suspensions from each well was added to autologous mini-LCL, which were subsequently probed with the EBV-specific T-cell clones F7 and G11-3. (A) The T-cell clone G11-3 recognized pool B12. Of 72 single colonies from this pool tested, only colony E1 was recognized (right panel). This colony carried a plasmid with a 147-bp insert derived from the BNRF1 gene coding for a tegument protein of EBV. Using synthetic linkers, the epitope recognized by the T cells was determined (boxed sequence). (B) The T-cell clone F7 also recognized a single positive pool, A11, from which six positive bacterial colonies were derived. Sequence analysis revealed that all carried the same 66-bp insert derived from the BALF4 gene. GM-CSF, granulocyte-macrophage colony-stimulating factor; AA, amino acids.
the His tag and CAT (Fig. 4B). These results demonstrated that unknown viral antigens are efficiently identified by using this method.

**Efficient recognition of bacteria expressing short antigen-CAT fusion proteins but not full-length proteins.** Bacterial expression of whole cDNAs has been successfully applied to identify an MHC class II-restricted murine minor histocompatibility antigen (22). To compare the efficiency and sensitivity of the two approaches, we assessed the T-cell recognition of various antigens expressed as full-length proteins or as short antigenic CAT fusion proteins. The ORF of the EBV proteins BALF4, BNRF1, BRLF1, and EBNA3A; the influenza virus matrix protein M1; and bacterial neomycin phosphotransferase II (Neo') were expressed in frame with the upstream His tag in the pTrcHis vector. Transformed bacteria were added at various dilutions to pAPC and subsequently probed with antigen-specific T cells. Protein was purified from the rest of the cultures and analyzed by Western blotting using the AntiXpress antibody. As shown in Fig. 5, all bacteria expressing antigen-CAT fusion proteins were efficiently recognized over a broad concentration range. Of the bacteria transformed with plasmids expressing whole cDNAs, only bacteria expressing influenza virus M1, Neo', and BRLF1 full-length proteins were recognized by the T cells but to a lesser extent than the CAT fusion constructs. Bacteria transformed with plasmids encoding the EBV proteins BALF4, BNRF1, and EBNA3A were not recognized. The results of T-cell recognition assays correlated with a low or undetectable level of expression of the full-length proteins by Western blotting (data not shown). These experiments demonstrated that antigens derived from proteins that are difficult to express in bacteria are efficiently detected by this method.

**DISCUSSION**

The molecular identification of antigens recognized by CD4+ T cells is important for understanding and enhancing antigen-specific immune responses. Here, we describe a simple and fast procedure to identify viral T-cell epitopes and antigens (Fig. 6). This method is based on the expression of small DNA fragments fused to CAT in bacteria. Bacteria expressing the antigenic fragment in frame with CAT are added directly to pAPC, which efficiently present peptides derived from bacterially expressed proteins on MHC class II molecules.

The fusion of short antigenic fragments to CAT offers two advantages. First, chloramphenicol treatment allows selection for bacterial recombinants expressing the inserted DNA fragments in frame with the upstream ORF coding for the His tag as well as the downstream ORF coding for CAT, thereby reducing the number of colonies to be screened by almost 10-fold. Second, polypeptides that are <6 kDa, which are unstable in bacteria (8), are expressed at uniformly high levels when fused to CAT.

Bacteria expressing antigen-CAT fusion proteins are fed directly to pAPC. In our experiments, we used LCL and mini-LCL as pAPC because these cells express high levels of MHC class II molecules and can be easily obtained from small blood samples. LCL efficiently ingested whole bacteria and presented
bacterially expressed antigens on MHC class II molecules, resulting in specific T-cell recognition even when diluted 1:500 in irrelevant bacteria. The use of autologous LCL as pAPC also offers the advantage that knowledge of the restriction molecule is not required.

The small DNA fragments were generated from genomic DNA or cDNA by restriction enzyme digestion. Because recognition sequences for restriction enzymes are rarely evenly distributed over a given DNA sequence, and because cleavage of the DNA into small fragments inherently bears the risk of epitope destruction, different frequently cutting restriction enzymes were used, alone and in combinations, to cleave the DNA partially and to completion. As expected, the size of the resulting DNA fragments varied depending on the reaction conditions applied. Of note, no bacterial colonies carrying inserts shorter than 20 bp or longer than 250 bp were detected, although such DNA fragments must have been generated by restriction enzyme digestion. DNA fragments shorter than 20 bp, which would code for incomplete MHC class II epitopes, were probably lost during the DNA preparation process involving phenol-chloroform extractions and ethanol precipitations. Fragments longer than 250 bp either ligated inefficiently into the expression vector or impaired CAT function, e.g., by interfering with protein folding. The latter possibility is in line with control experiments showing that the level of protein expression decreases when DNA fragments longer than 250 bp are inserted (data not shown).

The DNA fragments created by restriction enzyme digestion differ not only in length but also in the reading frame. To ensure expression of all fragments within the given size range, a vector system that accommodates DNA inserts in all possible reading frames was designed. As a consequence, the DNA fragments may be expressed in any of the six (sense and antisense) possible reading frames, resulting in the expression of nonsense proteins in five of six cases. However, 3 of the 64 codons of the genetic code cause termination of translation. Thus, except for natural ORFs, translation of DNA fragments expressed in irrelevant reading frames will terminate statistically after 64 bp. In the sequence analysis described here, we did not find colonies expressing fusion proteins in which the insert was expressed in an irrelevant reading frame. In more recent experiments, such colonies were occasionally detected, especially in libraries with relatively short inserts. These results define the optimal insert size range as >64 bp and <250 bp.

Compared to previously published methods of MHC II antigen identification entailing the expression of cDNAs (22, 24), the fragmentation of antigens into short peptides inevitably increases the number of colonies to be screened. However, libraries established from viral DNA contain exclusively viral inserts, whereas cDNA libraries generated from virus-infected cells contain a high percentage of inserts derived from cellular transcripts, which increases the number of colonies to be screened proportionally. Furthermore, many pathogens are able to adopt different life cycles associated with the expression of distinct subsets of viral genes. Therefore, not all potential antigens may be identified with cDNA libraries established from infected cells.

Most importantly, the expression levels of viral as well as eukaryotic proteins in bacteria or eukaryotic cells may vary over several orders of magnitude, thereby rendering representative pool sizes difficult to establish. While highly expressed antigens may be successfully identified within large pools, the identification of most antigens will require much smaller pool sizes. In our experiments, several full-length proteins remained undetected even when undiluted bacteria were used. Therefore, some antigens may remain undiscovered following bacterial expression of whole cDNAs or large genomic fragments.

In the DANI method described here, all antigenic fragments are expressed at similarly high levels in bacteria, irrespective of the protein from which they are derived. Thus, large represen-
tative pool sizes may be established for all antigens. Moreover, the small size of the antigenic peptides greatly facilitates the definition of the T-cell epitopes. Using this method, the lytic cycle proteins BALF4 and BNRF1 were identified as targets of EBV-specific T cells that had been generated by repeated stimulation of peripheral blood CD4+ T cells with autologous LCL. Given the low percentage of cells in an LCL culture that spontaneously become permissive for lytic replication (7), the identification of lytic cycle proteins as targets of EBV-specific T cells was unexpected. Recent experiments, however, have suggested that T cells specific for virion proteins are present in the peripheral blood of healthy virus carriers and may play an important role in the control of EBV infection (1). With the DANI method described here, breadth and immunodominance patterns of the EBV lytic cycle antigen-specific T cells can be assessed. These analyses may help to delineate the role of T cells in the control of EBV infection and in addition may provide novel targets for immunotherapy. Moreover, the identification of MHC class II antigens by the DANI method is not limited to EBV. The high sensitivity of this method may also allow high-throughput analysis of expression libraries derived from other pathogens.

ACKNOWLEDGMENT

This study was supported by the Deutsche Forschungsgemeinschaft (SFB455).

REFERENCES