International Journal of Infectious Diseases 114 (2022) 65–71

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

# Identifying Epstein–Barr virus peptide sequences associated with differential IgG antibody response

Anna E. Coghill[a,b,†,*], Jianwen Fang[c,†], Zhiwei Liu[a], Chien-Jen Chen[d], Ruth F. Jarrett[e], Henrik Hjalgrim[f,g], Carla Proietti[h], Kelly J. Yu[a], Wan-Lun Hsu[d,i], Pei-Jen Lou[j], Chen-Ping Wang[j], Yingdong Zhao[c], Denise L. Doolan[h], Allan Hildesheim[a]

[a] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA
[b] Cancer Epidemiology Program, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA
[c] Biometric Research Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Rockville, Maryland, USA
[d] Genomics Research Centre, Academica Sinica, Taipei, Taiwan
[e] MRC-University of Glasgow Centre for Virus Research, Glasgow, UK
[f] Statens Serum Institut, Copenhagen, Denmark
[g] Department of Haematology, Rigshospitalet, Copenhagen, Denmark
[h] Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, Australia
[i] College of Public Health, National Taiwan University, Taipei, Taiwan
[j] Department of Otolaryngology, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan

ARTICLE INFO

*Article history:*
Received 17 September 2021
Revised 22 October 2021
Accepted 27 October 2021

*Keywords:*
Epstein–Barr virus
IgG antibody array
LMP1/2
EBNA protein

ABSTRACT

*Background:* Epstein–Barr virus (EBV) infection contributes to cancers in a fraction of seropositive individuals, but much remains to be learned about variation in EBV-directed humoral immunity in cancer-free adults.

*Methods:* A protein microarray was used to probe serum from 175 Taiwanese and 141 Northern European adults for immunoglobulin G (IgG) antibody responses to 115 different peptide sequences, representing protein segments or protein variants, from 45 EBV proteins. It was posited that this antibody-based approach could identify EBV peptide sequences representing immunodominant regions relevant for B-cell immunity.

*Results:* Analyses of 45 EBV proteins with multiple protein segments or variants printed on the array identified eight EBV peptide sequences that appear to play a role in immunogenicity. This included: (1) three proteins with segments/regions associated with IgG reactivity (BALF5, LMP1, LMP2A); and (2) five proteins with sequence variants/amino acid changes associated with IgG reactivity (BDLF4, EBNA3A, EBNA3B, EBNA-LP, LF1).

*Conclusion:* This examination of IgG antibody responses against 115 EBV peptide sequences in 316 cancer-free adults represents an important step toward identifying specific EBV protein sequences that play a role in generating B-cell immunity in humans.



## Introduction

Epstein–Barr virus (EBV) is a common pathogen that is generally acquired during early childhood, with nearly all adults harbouring evidence of prior EBV exposure in the form of systemic immunoglobulin G (IgG) antibodies against the EBV nuclear antigen 1 (EBNA1) and/or viral capsid antigen (VCA) proteins (Kangro et al., 1994; Young and Rickinson, 2004; Cohen et al., 2011; Chen et al., 2015). Although infection is primarily asymptomatic, this virus does contribute to cancers in a fraction of seropositive individuals, including nasopharyngeal carcinoma (NPC), paediatric Burkitt lymphoma, and a subset of adult B-cell and Hodgkin lymphomas. Studies have evaluated these associations between aberrant antibody responses to EBV and diseases such as NPC (Hjalgrim et al., 2007; Coghill and Hildesheim, 2014; Coghill et al., 2018, 2020; Liu et al., 2020), but much remains to be learned

* Corresponding author. Address: Cancer Epidemiology Program, H. Lee Moffitt Cancer Center and Research Institute, 12902 Usf Magnolia Dr, Tampa, FL 33612, USA.
*E-mail address:* anna.coghill@moffitt.org (A.E. Coghill).
† Authors contributed equally to this work.

https://doi.org/10.1016/j.ijid.2021.10.054

about the breadth of EBV-directed immune responses mounted by disease-free individuals.

The authors recently used virome-wide protein microarray technology to characterize antibody responses to 86 known EBV proteins in 289 cancer-free adults from Taiwan (Liu et al., 2018). This represented a notable departure from historical descriptions of EBV antibody responses, which have focused primarily on responses to EBNA1, VCA and the EBV early antigen alone. All individuals in this investigation mounted detectable IgG antibody responses to at least 21 different EBV proteins, and it was possible to divide the population into four distinct clusters according to their IgG and IgA response patterns.

While informative, this prior protein-level analysis did not fully capture the potential of the technology. The microarray included at least two unique segments and/or variants of 45 assayed EBV proteins. In the analysis reported here, we used this within-protein variability in the array design to better understand whether (1) different segments of a given EBV protein and/or (2) different variants of the protein (i.e., amino acid changes in a protein segment) elicited differential IgG antibody reactivity. We posited that this approach would identify EBV peptide sequences representing immunodominant regions relevant for B-cell immunity, vis-à-vis antibody production.

## Methods

EBV-directed IgG antibody responses were measured in the serum of 175 adults who served as controls for a community-based study of NPC in Taiwan (Hildesheim et al., 2002). Sixty-six percent were male, approximately half (49%) were aged <40 years and 18% were aged ≥60 years. Serological data from 141 disease-free adults who served as controls in prior studies of classical Hodgkin lymphoma from the UK, Denmark and Sweden were also included (Jarrett et al., 2003; Smedby et al., 2005; Johnson et al., 2015). Samples from the UK were derived from two population-based case–control studies. Samples from Denmark and Sweden were collected as part of the population-based Scandinavian Lymphoma Etiology study. Sixty-eight percent were male, 29% were aged <40 years and 17% were aged ≥60 years.

### *EBV protein array*

The microarray used for this study has been described previously (Coghill et al., 2018). Briefly, the array was designed to characterize antibody responses against 199 different peptide sequences from 86 known EBV proteins. The predicted EBV sequences for these proteins were generated from five EBV strains (AG876, Akata, B95-8, Mutu, Raji) using genomic coordinates from GenBank (http://www.ncbi.nlm.nih.gov/), as well as known splice variants (i.e. different versions) of EBV proteins identified in the literature and non-overlapping segments from EBV proteins that were too large (>1000 bp) to print on to the array as a single peptide. High coverage was achieved across the five EBV strains, with 97% of the predicted sequences for the 86 proteins from each strain represented on the microarray at ≥99% homology.

### *Analytical approach*

After probing serum from each study participant on a single array, the normalized microarray output for each EBV sequence was divided by a person-specific background value, defined as the mean plus 1.5 standard deviations of the output from four no-DNA negative controls that were included on the array. Using this standardized dataset, each IgG antibody with output ≥1.0 was classified as a positive response. After computing the IgG response to each peptide sequence, the maximum reactivity for each EBV protein (i.e. the highest IgG antibody response observed for any of the peptide sequences from a given EBV protein) was listed. This enabled the identification of proteins that were more/less immunoreactive overall.

Next, IgG serological data for 115 peptide sequences from the 45 EBV proteins for which multiple segments or variants of the protein were included on the microarray were examined. Twenty-six of these 45 EBV proteins had two different sequences on the array, seven had three sequences, and 12 had at least four sequences. IgG reactivity was compared across sequences from the same protein. For example, if three different variants of EBV protein 'X' were included on the array, three across-sequence comparisons would be made to determine whether one of those EBV protein variants elicited a different IgG antibody response (i.e. was more/less immunogenic). These across-sequence comparisons by EBV protein were made using (1) a McNemar's test to ascertain across-sequence imbalance in IgG antibody reactivity and (2) Kappa coefficients to ascertain a lack of correlation in across-sequence IgG reactivity. In total, 306 comparisons were conducted for the 115 peptide sequences examined across 45 EBV proteins.

### *Filtering criteria*

After making these initial comparisons, the following criteria were applied to identify peptide sequences whose differential IgG antibody reactivity was pronounced and consistent: (1) McNemar's test *P*-value met a strict Bonferroni-corrected threshold of 0.00016, accounting for 306 statistical comparisons; (2) Kappa coefficient <0.01, confirming low correlation of IgG responses across different sequences; (3) presence of criteria 1 and 2 in both the Taiwanese and Northern European study populations; (4) difference in IgG reactivity between sequences of magnitude >1.5-fold; and (5) consistent directionality of IgG reactivity in both study populations (i.e. if IgG antibody response to sequence 'x' of an EBV protein was higher than sequence 'y' in the Taiwanese population, the IgG response to 'x' must also be higher than 'y' in Northern Europeans). Two investigators (AEC and AH) reviewed these criteria independently and agreed on a final peptide sequence list. This list included eight EBV proteins with specific sequences that appeared more/less immunogenic across both the Taiwanese and Northern European study populations.

### *Amino acid characterization*

Sequence mapping was performed using BLASTP suite from the NCBI BLAST server, with default values for all parameters (Johnson et al., 2008). Amino acid variances were identified in the pairwise alignment of reference and non-reference sequences. Structural models of the selected EBV proteins were developed using the i-TASSER server (Yang and Zhang, 2015). Predicted secondary structures and solvent accessibility of identified amino acid changes were extracted from these models, and chemical property changes associated with the noted variances were annotated using https://www.compoundchem.com/2014/09/16/aminoacids/.

## Results

Among the 175 Taiwanese and 141 Northern European adults examined, the proportion of adults who mounted detectable IgG antibody responses to at least one peptide sequence from each of the 86 known EBV proteins on the array was listed (Figure 1). Table 1 highlights 15 proteins that elicited nearly ubiquitous humoral immunity (IgG antibody reactivity ≥95%) in either or both study populations. Notably, for 47 of the 86 known proteins, reactivity was observed in ≥50% of participants in both populations,

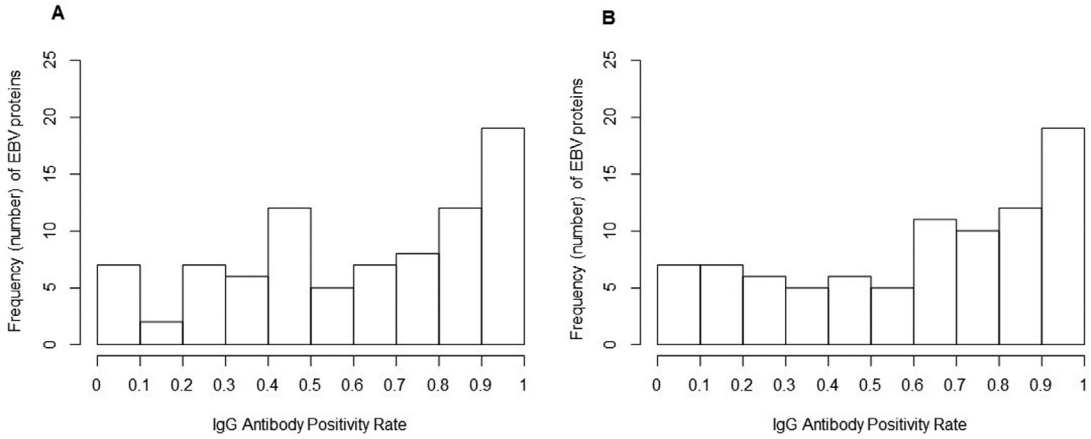


**Figure 1.** Distribution of immunoglobulin G (IgG) reactivity to Epstein–Barr virus (EBV) proteins across the full virome in cancer-free adults from Taiwan (A) and Northern European countries (B).

**Table 1**
Epstein–Barr (EBV) proteins with immunoglobulin G (IgG) antibody reactivity rates ≥95% and ≤10%, by study population.

| EBV protein | Protein description | IgG reactivity in Taiwanese | IgG reactivity in Northern Europeans |
|---|---|---|---|
| ≥95% IgG reactivity | | | |
| BRRF2 | Lytic phase (virion release) | 100% | 99% |
| BFRF3 | Viral capsid protein | 100% | 97% |
| BILF1 | Lytic-phase (immune modulation) | 100% | 87% |
| BZLF2 | gp42 (B-cell fusion) | 100% | 99% |
| BBRF3 | Tegument protein | 97% | 86% |
| BDLF3 | gp150 (immune modulation) | 100% | 92% |
| BDLF4 | Late lytic-phase expression | 99% | 97% |
| BLRF2 | Tegument protein | 99% | 94% |
| EBNA1 | Latency (episome maintenance) | 98% | 88% |
| EBNA3A | EBNA2 regulation | 99% | 96% |
| A73 | BART family of viral RNAs | 98% | 38% |
| BPLF1 | Tegument protein | 97% | 97% |
| BBLF1 | Lytic phase | 97% | 92% |
| BMRF2 | Glycoprotein | 97% | 50% |
| BZLF1 | Latent-to-lytic activator | 55% | 97% |
| <10% IgG reactivity | | | |
| BGLF4 | Early antigen kinase | 2% | 20% |
| BFLF2 | Lytic phase (virion release) | 3% | 36% |
| BORF2 | Early antigen ribonuclease | 5% | 17% |
| BBRF2 | Tegument protein | 8% | 67% |
| LF3 | Lytic phase | 37% | 5% |
| BFRF2 | Late lytic-phase expression | 38% | 7% |

indicating that the majority of EBV proteins elicit an immune response. A small fraction of proteins with evidence of low immunoreactivity was identified. For example, IgG reactivity <10% was observed in at least one study population for six proteins, and two proteins elicited IgG responses of <20% in both populations (BORF3 and BGLF4).

After applying the filtering criteria (see above) to the 45 EBV proteins with at least two different protein segments or variants included on the microarray, eight EBV proteins with specific sequences that appeared more/less immunogenic across both the Taiwanese and Northern European study populations were identified. Three had segments (i.e. protein regions) associated with higher/lower IgG responses, while five had variant sequences (i.e. amino acid changes within a protein region) that elicited differential IgG reactivity. Details are presented in Table 2 and summarized below.

*Across-segment differences*

Three EBV proteins were identified by comparing IgG antibody responses across non-overlapping protein segments.

*BALF5*

For this protein, segment 1–800 (i.e. amino acid positions 1–800) was less reactive than segment 801–1015. This pattern was consistent across all peptide sequences examined. For example, IgG responses to both the 1–800 reference sequence and the 1–800 sequence with a D776E amino acid change were both significantly lower than IgG responses to the 801–1015 reference sequence and the 801–1015 sequence with a T888A amino acid change.

*Latent membrane proteins (LMP1 and LMP2A)*

For LMP1, the examined sequence for segment 1–89, which included two amino acid changes compared with the reference sequence (M61I and I85L), was less reactive compared with three examined sequences for segment 120–387. For LMP2A, segment 1–118 appeared to be less reactive than segment 225–284.

*Within-segment differences*

Five EBV proteins were identified by comparing IgG antibody responses between different variants of the same protein segment.

**Table 2**
Twenty-nine peptide sequences from eight Epstein–Barr proteins observed to have a segment or variant that elicited a uniquely different immunoglobulin G (IgG) antibody response.

| Protein | Sequence details | IgG% in Taiwanese | IgG% in Northern Europeans | Amino acid segment | Specific amino acid changes by sequence[c] |
|---|---|---|---|---|---|
| BALF5 | Reference: CAA24805.1 | | | | |
| | CAA24805.1-156746-153699-1 | 5.1% (2.5–9.2) | 24.1% (17.6–31.7) | 1–800 | Reference |
| | YP_001129507.1-157772-154725-1 | 8.6% (5.1–13.4) | 4.3% (1.7–8.6) | 1–800 | D776E |
| | YP_001129507.1-157772-154725-2 | 21.7% (16.1–28.2) | 51.8% (43.5–59.9) | 801–1015 | Reference |
| | AFY97980.1-156149-153102-2 | 72.0% (65.0–78.3) | 69.5% (61.5–76.7) | 801–1015 | T888A |
| LMP1 | Reference: YP_401722.1 | | | | |
| | YP_001129515.1-170457-170190 | 14.9% (10.3–20.9) | 3.5% (1.3–7.7) | 1–89 | M61I, I85L |
| | YP_401722.1-168507-167702 | 82.3% (76.0–87.2) | 19.9% (13.9–27.1) | 120–386 | Reference |
| | YP_001129515.1-169948-169188 | 66.3% (59.0–72.9) | 19.9% (13.9–27.1) | 120–386 | [a] |
| | AFY97987.1-168367-167562 | 79.4% (72.8–84.8) | 34.8% (27.2–42.9) | 120–386 | M129I, L151I, G212S, S229T, S309N, G318K, Q322E, Q334R, L338S, H352R, S366T |
| LMP2A | Reference: YP_001129436.1 | | | | |
| | YP_001129436.1-167587-167942 | 26.3% (20.3–33.3) | 28.4% (21.4–36.2) | 1–118 | Reference |
| | AFY97909.1-165963-166318 | 45.1% (38.0–52.5) | 12.1% (7.4–18.2) | 1–118 | S36Y, Y64D, T79N, Q82P |
| | YP_001129436.1-540-788 | 79.4% (72.8–84.8) | 48.2% (40.1–56.5) | 225–284 | Reference |
| | AFY97909.1-540-788 | 88.0% (82.4–92.0) | 68.1% (60.1–75.4) | 225–284 | V244F |
| BDLF4 | Reference: BDLF4 | | | | |
| | YP_001129488.1-117560-116883 | 99.4% (97.2–100.0) | 96.5% (92.3–98.7) | 1–225 | Reference |
| | AFY97877.1-116284-115607 | 32.0% (25.4–39.2) | 7.8% (4.2–13.2) | 1–225 | Y78D |
| | AFY97959.1-116599-115922 | 43.4% (36.2–50.9) | 11.3% (6.9–17.4) | 1–225 | K114Q |
| EBNA3A | Reference: YP_401669.1 | | | | |
| | YP_401669.1-80382-82877 | 91.4% (86.6–94.9) | 92.2% (86.6–95.8) | 114–944 | Reference |
| | YP_001129463.1-80447-82888 | 64.0% (56.7–70.9) | 95.7% (91.4–98.3) | 116–944 | [b] |
| | AFY97830.1-80050-82545 | 9.7% (6.0–14.8) | 19.9% (13.9–27.1) | 114–944 | L219P, V267I, S293N, I333L, **T357A**, P459T, F492S, I561F, P620T, Q733R, **T811A** |
| | AFY97915.1-80252-82747 | 93.7% (46.9–61.6) | 93.6% (88.6–96.8) | 114–944 | L219P, V267I, S293N, I333L, P459T, F492S, R496Q, I561F, P620T, I675T, V681A, I763M, T811P, G814A, E877G, G908S, G914D, P915H, T917P, V925I |
| EBNA3B | Reference: YP_001129464.1 | | | | |
| | YP_001129464.1-83074-83430 | 78.3% (71.7–83.9) | 82.3% (75.3–88.0) | 1–119 | Reference |
| | AFY97829.1-82733-83089 | 42.9% (35.7–50.3) | 13.5% (8.6–19.9) | 1–119 | Q33T, **G36E**, D38E, A43E, D77H, E89D, **V90F** |
| | CAA24858.1-95353-95709 | 78.3% (71.7–83.9) | 83.0% (76.1–88.5) | 1–118 | Q33T, D38E, A43E, D77H, E89D |
| EBNA-LP | Reference: YP_001129440.1 | | | | |
| | YP_001129440.1-35558-35662 | 9.1% (5.5–14.1) | 4.3% (1.7–8.6) | 473–506 | Reference |
| | AFY97832.1-35494-35598 | 52.6% (45.2–59.9) | 60.3% (52.0–68.1) | 473–506 | S482R, E483R, D486E, H488P, A491T, R493Q |
| | AFY97917.1-35572-35676 | 6.3% (3.3–10.7) | 8.5% (4.7–14.0) | 473–506 | S482R, E483R, D486E, H488P, P489T, V490E, A491T, R493Q, **V497I** |
| | YP_401636.1-35590-35694 | 46.9% (39.5–54.3) | 73.1% (65.3–79.9) | 473–506 | S482R, E483R, D486E, H488P, P489T, V490E, A491T, R493Q |
| LF1 | Reference: YP_001129505.1 | | | | |
| | YP_001129505.1-153178-151769 | 77.1% (70.5–82.9) | 62.4% (54.2–70.1) | 1–469 | Reference |
| | AFY97897.1-151239-149830 | 11.4% (7.3–16.8) | <1% | 1–469 | Q147K, I163T, S183Y, G224E, A241T, L324M |
| | AFY97978.1-151556-150147 | 73.1% (66.2–79.3) | 51.8% (43.5–59.9) | 1–469 | D338N |

[a] Shared with reference sequence: identity: 243/267 (91%); number of differences too high to present in tabular format but sequence details provided in online supplementary material.
[b] Shared with reference sequence: identity: 706/829 (85%); number of differences too high to present in tabular format but sequence details provided in online supplementary material.
[c] Specific amino acid changes are shown in the following format: (a) amino acid identity on the reference sequence; (b) position of amino acid change; (c) amino acid identify on variant sequence. For example, the Y78D change in BDLF4 represents tyrosine (Y) to aspartic acid (D) at position 78. Conventional, single-letter amino acid nomenclature was used.

*BDLF4*

Three sequences from the 1–225 segment of BDLF4 were examined. Both non-reference sequences, one containing a Y78D amino acid change and one containing a K114Q change, elicited lower IgG antibody reactivity compared with the ubiquitous IgG response observed for the reference sequence. The location and chemical characteristics of these amino acid changes are outlined in Table 3 and Figure 2A. Each change resulted in altered chemical properties; for Y78D, a hydroxytic and aromatic amino acid residue (Y) was replaced by a negatively charged residue (D); and for K114Q, a positively charged residue (K) was replaced by a polar amidic residue (Q). Such changes could feasibly impact antibody–antigen binding affinity and therefore B-cell immunogenicity. Notably, the K114Q change is located closer to the surface of BDLF4, leading the authors to posit that this amino acid change is more likely to be associated with the observed change in IgG response.

*EBNA3A*

Four sequences from the 115–925 segment of EBNA3A were examined, and two amino acid changes specific to the less reactive AFY98730.1-80050-82545 sequence – T357A and T811A (in bold in Table 2) – were noted. These two differences resulted in hydroxytic (T) to non-polar (A) amino acid changes near the surface of the EBNA3A protein (Figure 2B). Such changes may cause the loss of a hydrogen bond reaction that could impact antibody–epitope binding and contribute to low IgG reactivity.

*EBNA3B*

Three sequences from the 1–119 segment of EBNA3B were examined. Both non-reference sequences shared five amino acid changes, but only sequence AFY97829.1-82733-83089 was associated with a lower IgG antibody response. This sequence contained two unique amino acid changes – G36E and V90F. The V90F

**Table 3**
Characteristics of amino acid changes identified as eliciting a uniquely different immunoglobulin G antibody response.

| Protein | Amino acid change | Location[a] | Secondary structure | Chemical properties |
|---|---|---|---|---|
| BDLF4 | | | | |
| | Y78D | 1 | Coil | Polar, hydroxytic, aromatic » negative |
| | K114Q | 4 | Coil | Positive » polar, amidic |
| EBNA3A | | | | |
| | T357A | 4 | Coil | Polar, hydroxytic » non-polar |
| | T811A | 4 | Coil | Polar, hydroxytic » non-polar |
| EBNA3B | | | | |
| | G36E | 4 | Coil | Non-polar » negative |
| | V90F | 1 | Helix | Non-polar » non-polar, aromatic |
| EBNA-LP | | | | |
| | V497I | 2 | Strand | Non-polar » non-polar |
| LF1 | | | | |
| | Q147K | 6 | Coil | Polar, amidic » positive |
| | I163T | 2 | Coil | Non-polar » polar, hydroxytic |
| | S183Y | 0 | Strand | Polar, hydroxytic » polar, aromatic, hydroxytic |
| | G234E | 3 | Coil | Non-polar » negative |
| | A241T | 4 | Coil | Non-polar » polar, hydroxytic |
| | L324M | 2 | Strand | Non-polar » non-polar |

[a] Values range from 0 (buried residue) to 9 (highly exposed residue).

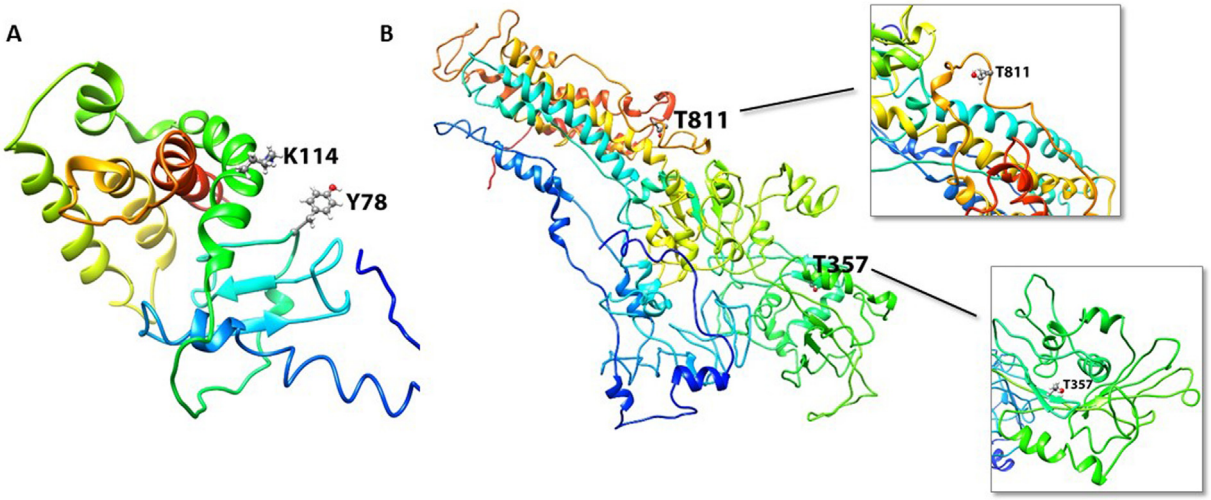


**Figure 2.** Amino acid changes present in sequence variants associated with differential immunoglobulin G antibody responses for (A) BDLF4 and (B) EBNA3A. Structural models displayed are for the reference sequence, with rainbow colouring representing amino acids moving from the N-terminus to the C-terminus. Insets of amino acid changes are provided for the larger EBNA3A protein.

change occurred far from the protein surface and did not change amino acid polarity significantly (Figure 3A); in contrast, the G36E change occurred closer to the protein surface and added a negatively charged residue (E), suggesting that this change may be responsible for the differential B-cell immunogenicity.

*EBNA-LP*

Four sequences from the 473–506 segment of the EBNA-latency protein (EBNA-LP) were examined. The three non-reference sequences each contained between six and nine amino acid changes, many of which were shared. The presence of such changes was associated with higher IgG antibody responses for two of the three non-reference sequences, although it was not possible to narrow down specific amino acids associated with this reactivity. In contrast, the AFY97917.1-35572-35676 sequence did not elicit higher IgG reactivity. This sequence contained one unique amino acid change (V497I), but this change did not appear to alter the chemical properties of this protein (Figure S1, see online supplementary material).

*LF1*

Three sequences from the 1–469 segment of LF1 were examined. The AFY97978.1-151556-150147 sequence included a D338N amino acid change, but this change did not appear to affect B-cell immunogenicity relative to the reference sequence. The AFY97897.1-151239-149830 sequence, which varied from the reference sequence by six amino acids (Q147K, I163T, S183Y, G224E, A241T, L324M), was associated with a lower IgG antibody response (Figure 3B). The Q147K and A241T amino acid changes occurred closest to the protein surface, and both resulted in chemical property changes. The S183Y and L324M changes were located furthest from the protein surface and did not elicit similar chemical property changes.

## Discussion

A virome-wide protein microarray was used to understand the degree to which select EBV peptide sequences play a role in eliciting B-cell immunity. This study identified eight EBV proteins with at least one sequence associated with differential IgG antibody response, including three proteins with segments that were more or less reactive, as well as five proteins with specific amino acid changes that appeared to alter IgG antibody response rates in both Taiwanese and Northern European cancer-free adults.

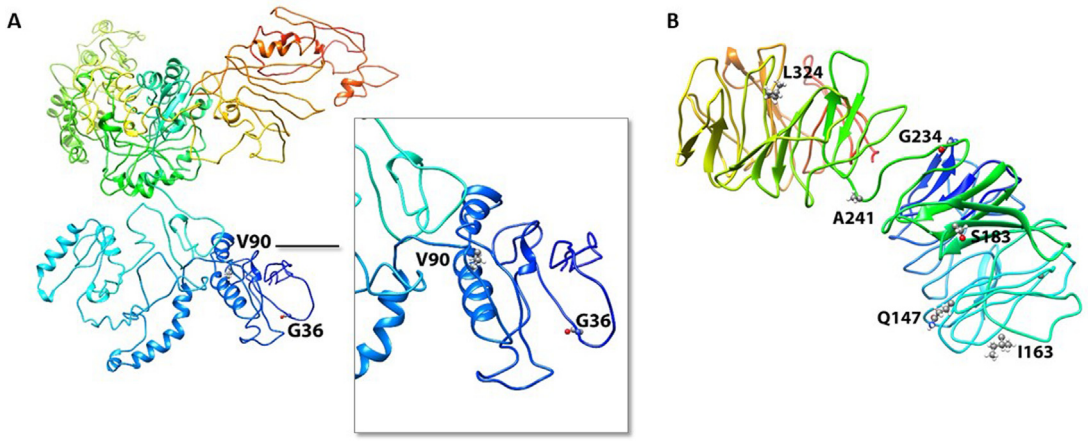


**Figure 3.** Amino acid changes present in sequence variants associated with differential immunoglobulin G antibody responses for (A) EBNA3B and (B) LF1. Structural models displayed are for the reference sequence, with rainbow colouring representing amino acids moving from the N-terminus to the C-terminus. Insets of amino acid changes are provided for the larger EBNA3B protein.

Research conducted in East Asian populations suggests that healthy adults display variability in EBV-directed B-cell immunity (i.e. antibody reactivity) at the protein level, and that behaviours such as smoking can impact these anti-EBV serological profiles (Liu et al., 2018; Hsu et al., 2020). However, that research did not address whether specific EBV protein segments or amino acid variations can impact the likelihood of an infected adult mounting a detectable IgG antibody response, and few EBV-peptide-specific serological response data are available from other human studies for comparison. Serological evaluations of the constitutively expressed EBNA1 protein in human sera have identified an immunodominant epitope in the 390–450 amino acid region; interestingly, a recent comparison of different EBV-based antibody assays used in cancer studies observed high agreement across the EBNA1 IgA tests, and attributed this to the common inclusion of this important region (Fachiroh et al., 2006; Middeldorp 2015; Liu et al., 2019). It is notable that preclinical laboratory studies describing immunogenic epitopes have focused primarily on defining sequences that elicit EBV-directed CD4+/CD8+ T-cell immunity rather than B-cell immunity (Taylor et al., 2015; Brooks et al., 2016).

After strict control for statistical chance and the requirement of consistency across the study population, many of the sequences identified as potentially important for B-cell immunity originated in viral proteins from crucial stages of the EBV life cycle. This includes sequences from three EBV nuclear antigen proteins expressed during the original B-cell transformation process (EBNA3A, -3B and -LP), as well as two latent membrane proteins defined as EBV oncogenes (LMP1 and -2A). This report also identified a viral DNA polymerase unit (BALF5) and two lytic-cycle proteins (BDLF4, LF1), one of which (LF1) has been identified in recent serological evaluations of cancer patients as being associated with the risk of future NPC diagnosis (Coghill et al., 2018), as well as being actively expressed by children with Burkitt lymphoma (Tierney et al., 2015).

This study is the largest serological survey of EBV-peptide-specific IgG antibody responses to date, and the inclusion of two geographically distinct populations increases the relevance of the study findings for EBV-seropositive adults globally. Despite these strengths, this study represents only a first step. The examination of peptide-specific IgG responses was necessarily limited to the sequences printed on the array (i.e. 115 sequences for 45 EBV proteins); the possibilities that: (1) a proportion of study participants were exposed to, and thus mounted EBV-directed antibodies against, EBV strains with limited homology to the five used to create the microarray; or that (2) protein segments and/or variants not selected for inclusion on the array could elicit even more pronounced differential IgG responses than those observed cannot be excluded. Partial sequences from five EBV genomes originating from saliva collected in Taiwan were recently made publicly available. This sequence data included one of the eight proteins identified in the present study as a key result (BALF5). However, limited variation was noted across those five sequences, with identical sequence length and 99.5% homology across the published sequences. Finally, future laboratory research is required to identify the functional B-cell epitopes within the amino acid sequences highlighted in this report, and to probe additional EBV proteins not included here for immunogenic B-cell epitopes.

In summary, peptide sequences from eight EBV proteins that appear to be important for eliciting anti-EBV serological responses in cancer-free adults were identified. Amino acid variation in these sequences may play an important role in mounting B-cell immunity against this ubiquitous herpesvirus. A more complete understanding of the EBV-directed serological repertoire in seropositive individuals will be crucial for the design of EBV-based diagnostic or therapeutic tools.

## Declaration of Competing Interest

None declared.

## Funding

This work was funded through the National Cancer Institute intramural research program.

## Ethical approval

This study was approved as exempt research by the institutional review board at the National Cancer Institute.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijid.2021.10.054.

## References

Brooks JM, Long HM, Tierney RJ, Shannon-Lowe C, Leese AM, Fitzpatrick M, et al. Early T cell recognition of B cells following Epstein–Barr virus infection: identifying potential targets for prophylactic vaccination. PLoS Pathog 2016;12.

Chen CY, Huang KY, Shen JH, Tsao KC, Huang YC. A large-scale seroprevalence of Epstein–Barr virus in Taiwan. PLoS One 2015;10.

Coghill AE, Hildesheim A. Epstein–Barr virus antibodies and the risk of associated malignancies: review of the literature. Am J Epidemiol 2014;180:687–95.
Coghill AE, Pfeiffer RM, Proietti C, Hsu WL, Chien YC, Lekieffre L, et al. Identification of a novel, EBV-based antibody risk stratification signature for early detection of nasopharyngeal carcinoma in Taiwan. Clin Cancer Res 2018;24:1305–14 Erratum in Clin Cancer Res 2020;26:3496.
Coghill AE, Proietti C, Liu Z, Krause L, Bethony J, Prokunina-Olsson L, et al. The association between the comprehensive Epstein–Barr virus serologic profile and endemic Burkitt lymphoma. Cancer Epidemiol Biomarkers Prev 2020;29:57–62.
Cohen JI, Fauci AS, Varmus H, Nabel GJ. Epstein–Barr virus: an important vaccine target for cancer prevention. Sci Transl Med 2011;3:107fs7.
Fachiroh J, Paramita DK, Hariwiyanto B, Harijadi A, Dahlia HL, Indrasari SR, et al. Single-assay combination of Epstein–Barr Virus (EBV) EBNA1- and viral capsid antigen-p18-derived synthetic peptides for measuring anti-EBV immunoglobulin G (IgG) and IgA antibody levels in sera from nasopharyngeal carcinoma patients: options for field screening. J Clin Microbiol 2006;44:1459–67.
Hildesheim A, Apple RJ, Chen CJ, Wang SS, Cheng YJ, Klitz W, et al. Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. J Natl Cancer Inst 2002;94:1780–9.
Hjalgrim H, Friborg J, Melbye M, et al. The epidemiology of EBV and its association with malignant disease. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al, editors. Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge: Cambridge University Press; 2007 Chapter 53.
Hsu WL, Chien YC, Huang YT, Yu KJ, Ko JY, Lin CY, et al. Cigarette smoking increases the risk of nasopharyngeal carcinoma through the elevated level of IgA antibody against Epstein–Barr virus capsid antigen: a mediation analysis. Cancer Med 2020;9:1867–76.
Jarrett RF, Krajewski AS, Angus B, Freeland J, Taylor PR, Taylor GM, et al. The Scotland and Newcastle epidemiological study of Hodgkin's disease: impact of histopathological review and EBV status on incidence estimates. J Clin Pathol 2003;56:811–16 Erratum in J Clin Pathol 2004;57:112.
Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res 2008;36:W5–9.
Johnson PC, McAulay KA, Montgomery D, Lake A, Shield L, Gallagher A, et al. Modeling HLA associations with EBV-positive and -negative Hodgkin lymphoma suggests distinct mechanisms in disease pathogenesis. Int J Cancer 2015;137:1066–75.
Kangro HO, Osman HK, Lau YL, Heath RB, Yeung CY, Ng MH. Seroprevalence of antibodies to human herpesviruses in England and Hong Kong. J Med Virol 1994;43:91–6.
Liu Z, Coghill AE, Pfeiffer RM, Proietti C, Hsu WL, Chien YC, et al. Patterns of interindividual variability in the antibody repertoire targeting proteins across the Epstein–Barr virus proteome. J Infect Dis 2018;217:1923–31.
Liu Z, Yu KJ, Coghill AE, Brenner N, Cao SM, Chen CJ, et al. Multilaboratory assessment of Epstein–Barr virus serologic assays: the case for standardization. J Clin Microbiol 2019;57:e01107–19.
Liu Z, Jarrett RF, Hjalgrim H, Proietti C, Chang ET, Smedby KE, et al. Evaluation of the antibody response to the EBV proteome in EBV-associated classical Hodgkin lymphoma. Int J Cancer 2020;147:608–18.
Middeldorp JM. Epstein–Barr virus-specific humoral immune responses in health and disease. Curr Top Microbiol Immunol 2015;391:289–323.
Smedby KE, Hjalgrim H, Melbye M, Torrång A, Rostgaard K, Munksgaard L, et al. Ultraviolet radiation exposure and risk of malignant lymphomas. J Natl Cancer Inst 2005;97:199–209.
Taylor GS, Long HM, Brooks JM, Rickinson AB, Hislop AD. The immunology of Epstein–Barr virus-induced disease. Annu Rev Immunol 2015;33:787–821.
Tierney RJ, Shannon-Lowe CD, Fitzsimmons L, Bell AI, Rowe M. Unexpected patterns of Epstein–Barr virus transcription revealed by a high throughput PCR array for absolute quantification of viral mRNA. Virology 2015;474:117–30.
Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res 2015;43:W174–81.
Young LS, Rickinson AB. Epstein–Barr virus: 40 years on. Nat Rev Cancer 2004;4:757–68.