## SHORT COMMUNICATION

# Peptide matching between Epstein–Barr virus and human proteins

Giovanni Capone<sup>1</sup>, Michele Calabrò<sup>1</sup>, Guglielmo Lucchese<sup>2</sup>, Candida Fasano<sup>1</sup>, Bruna Girardi<sup>3</sup>, Lorenzo Polimeno<sup>3</sup> & Darja Kanduc<sup>1</sup>

1 Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari, Bari, Italy

2 Brain and Language Laboratory, Cluster of Excellence "Languages of Emotions", Free University of Berlin, Berlin, Germany

3 Section of Gastroenterology, Department of Emergency and Organ Transplantation (DETO), University of Bari, Bari, Italy

Abstract

EBV and the human host.

This article describes a high level of peptide similarity between human proteins and proteins encoded by the Epstein Barr virus (EBV), in particular in the glycine-alanine (GA) repeat region of the nuclear antigen 1 of EBV (EBNA1). Some of the human proteins that share similarities with EBV are implicated in brain development and function. These similarities could contribute to preventing immune recognition of EBV and form the basis for molecular mimicry in autoimmune diseases.

Epstein-Barr virus proteins were examined for amino acid sequence matching to

human proteins at the decapeptide level. We report that numerous EBV peptides

of different length (from 10- to 13-mer) are present in 28 human proteins. The viral vs. human peptide overlap mainly involves the glycine-rich region allocated in the

NH2 terminus of Epstein-Barr nuclear antigen 1 protein and host cellular

components that play crucial roles in basic biochemical pathways, such as

chromatin remodeling, RNA splicing, transmission across chemical/electrical

synapses, and neurogenesis, and that, when altered, may characterize various

pathologies such as immunodeficiency, systemic lupus erythematosus, myelina-

tion, and speech disorders. The present results might contribute to understand and

define the (physio) pathological relationships and interactions occurring between

#### Keywords

EBV immunoevasion; EBV proteins; human proteins; peptide matching; EBV EBNA1; glycine-rich region.

#### Correspondence

Darja Kanduc, Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari, Bari 70126, Italy. Tel.: +390805443321 fax: +390805443317 e-mails: dkanduc@gmail.com; darja.kanduc@uniba.it

Received 19 March 2013; revised 4 July 2013; accepted 9 July 2013. Final version published online 6 August 2013.

doi:10.1111/2049-632X.12066

Editor: Alfredo Garzino-Demo

gens

The human herpesvirus 4, also called Epstein–Barr virus (EBV), is a common virus in humans (Rickinson & Kieff, 2007; http://www.cdc.gov/ncidod/diseases/ebv.htm). EBV infection, although usually asymptomatic (Klein *et al.*, 2010; Saha & Robertson, 2011), may cause mononucleosis, neoplasms, and autoimmune diseases (Cesarman, 2002; Farrell & Jarrett, 2011; Saha & Robertson, 2011; Gourzones *et al.*, 2012). However, the pathogenic contribution of EBV to tumors as well as the etiology of autoimmune diseases associated with EBV infection such as multiple sclerosis (Farrell & Jarrett, 2011; Niller *et al.*, 2011; Tselis, 2012) and systemic lupus erythematosus (SLE) remain unclear (Draborg *et al.*, 2012).

Other poorly understood issues are EBV immunoevasion, latency, and (re)activation (Jochum *et al.*, 2012; Kalla & Hammerschmidt, 2012; Severa *et al.*, 2013) as well as the

EBV ubiquitous presence in the human population, with approximately 95% of adults worldwide infected (http://www. cdc.gov/ncidod/diseases/ebv.htm; Rickinson & Kieff, 2007). It seems that a number of EBV proteins can protect the virus from the host immune attack. Such EBV proteins, also called immunoevasins, function by targeting MHC class I and MHC class II antigen presentation pathways (Ressing *et al.*, 2008; Rowe & Zuo, 2010); for example, the EBV immunoevasin interleukin-10 homolog (IL-10H) protects infected B cells from immune recognition and elimination (Jochum *et al.*, 2012), thus contributing to EBV successful persistence and immune escape.

However, EBV immunoevasin IL-10H might contribute through other additional pathways to the EBV immunoevasion phenomenon given its sequence identity to the human IL-10. In fact, IL-10H is a striking example of the conservation of



amino acid (aa) sequence between EBV and the human host (Moore *et al.*, 1990; Moore *et al.*, 2001; Yoon *et al.*, 2005), with viral IL-10H and human IL-10 sharing four continuous identical stretches ranging from 17 to 42 aa (e.g. MLRDLRDAFSRVKTFFQ, DNLLLKESLLEDFKGYLGCQAL SEMIQFYLEEVMPQAENQDP, HVNSLGENLKTLRLRLRR CHRFLPCENKSKAVEQ, KNAFNKLQEKGIYKAMSEFDIFI-NYIEAYMT). The sequence alignment of EBV IL-10H (P0CAP9, IL10H\_EBVG) and human IL-10 (P22301, IL10\_HUMAN) shows a percent identity equal to 77.2 at the aa level, with 139 identical positions.

In this regard, we already observed that a striking level of sequence identity to the human host might act as a camouflage mechanism of infectious agents (Natale et al., 2000; Tindle, 2002; Lucchese et al., 2009; Capone et al., 2013), because when high levels of sequence/structure identity are present between microbial and human molecules, the breaking of the self-tolerance mechanisms that prevent self-reactivity is highly improbable (Silverstein, 2001). Thus, the sharing of continuous aa sequences with host molecules may represent an elective microbial mechanism to escape immune recognition attack (Kanduc et al., 2008; Trost et al., 2010). Here, to further our understanding of how can EBV establish a persistent infection in the human host (Ressing et al., 2008; Rowe & Zuo, 2010; Jochum et al., 2012), we analyzed EBV proteins for aa sequence identity to human proteins to investigate whether other identity regions are present between EBV and the human host in addition to the above-cited sequence identities between EBV IL-10H and human IL-10.

To this aim, we examined the polyprotein derived from EBV, strain GD1, GenBank: AY961628.3, NCBI taxonomic identifier: 10376 (Zeng et al., 2005), consisting of 68 proteins (total number of aa: 34 503) listed and described at http:// www.ncbi.nlm.nih.gov/nuccore/AY961628. Sequence identity analyses of EBV proteins to the human proteome were conducted using viral decapeptides as probes to scan the Homo sapiens proteome, searching for exact peptide matches. Each probe was shifted by one residue; that is, viral decapeptides sequentially overlapped by nine residues such as MVHVLERALL, VHVLERALLE, HVLERALLEQ, were used in scanning the human proteome for peptide matching using Protein International Resource (PIR) peptide match program (pir.georgetown.edu/pirwww/search/peptide.shtml) (Wu et al., 2003). The human proteins containing viral matches were analyzed using UniProt database (http:// www.uniprot.org) (UniProt Consortium, 2009). Fragments, duplicated sequences, and obsolete entries were filtered out manually.

The peptide-by-peptide comparison of EBV and *Homo* sapiens proteomes at the decapeptide level is presented in Table 1. It can be seen that: (1) viral decapeptides repeatedly occur in 28 human proteins; (2) the peptide sharing also occurs at 11-, 12-, and 13-mer levels; (3) five of 68 EBV proteins (i.e. BDLF2, EBNA1, DEN, EBNA2, and Q3KSS2) are implicated in the viral vs. human peptide sharing. In particular, Table 1 highlights a heavy involvement of the Epstein–Barr nuclear antigen 1 (EBNA1) in the peptide sharing, with a clustering of identity regions in the gly-

cine-rich region (GRR) allocated along the NH2 terminus of the 641-aa-long EBV EBNA1 sequence.

Biologically, the 28 human proteins (ARI1B, BD1L1, BMP2K, CHD5, CHTOP, DIAP3, FRM4A, FUS, FZD8, JUND, LAR1B, LARP1, MBD2, MLL4, NOVA2, NOXA1, ONEC3, PCSK6, RBM26, RBM27, RS2, SFR15, SHSA7, SKOR2, SMD1, TSN3, UNG, and ZN579) implicated in the sharing exert critical functions in crucial processes such as myelination, chromatin remodeling, RNA splicing, and proteolysis. For example:

(1) The EBV BDLF2<sub>247–256</sub>VYTLIPAVVI decapeptide is shared with human tetraspanin-3 (TSN3) that regulates the proliferation and migration of oligodendrocytes, a process essential for normal myelination and repair (Tiwari-Woodruff *et al.*, 2004).

(2) The GAGAGGGAGG decapeptide is repeated eight times in EBNA1 and is also present in the neuro-oncological ventral antigen 2 (NOVA2). NOVA2 is a neuron-specific splicing factor that regulates neuronal migration (Yano *et al.*, 2010), is necessary for physiologic motor neuron firing (Ruggiu *et al.*, 2009), and has been identified as a target in autoimmune motor disease (Yang *et al.*, 1998).

(3) The GAGGAGGAGAG 11-mer is present 12 times in EBNA1 and is shared with the human bi-orientation of chromosomes in cell division protein 1-like 1 (BD1L1) (Porter *et al.*, 2007).

(4) The EBV EBNA1<sub>314–325</sub>GGGAGAGGAGAG 12-mer is shared with the human AT-rich interactive domain–containing protein 1B (ARI1B). ARI1B is involved in gene transcriptional activation and repression by chromatin remodeling (Santen *et al.*, 2012). Of note, ARI1B is important in human brain development and function in general and in the development of corpus callosum in particular (Halgren *et al.*, 2012). Indeed, ARI1B alterations are associated with mental retardation, impairments in adaptative behavior, and speech disorders with expressive speech more severely affected than receptive function (Santen *et al.*, 2012).

(5) The EBV EBNA1<sub>40–51</sub>GRGRGRGRGRGRGG and the EBV EBNA2<sub>311–322</sub>RGRGRGRGRGRGRG are common to small nuclear ribonucleoprotein Sm D1 (SMD1). SMD1 may act as a charged protein scaffold to promote snRNP assembly or strengthen snRNP–snRNP interactions through nonspecific electrostatic contacts with RNA. As a note of special importance, antinuclear antibodies with SMD1 specificity are developed in the autoimmune SLE disease (Poole *et al.*, 2006).

(6) The EBV EBNA1<sub>202-214</sub>AGAGGAGGAGGAGGAG 13-mer is shared with proprotein convertase subtilisin/kexin type 6 (PCSK6). Of note, PCSK6 is associated with handedness in individuals with dyslexia (Scerri *et al.*, 2013).

(7) The GGRGRGGSGG decapeptide is present three times in EBNA1 and is shared with RNA-binding protein FUS (FUS). Accumulation of FUS protein as cytoplasmic inclusions in neurons and glial cells in the central nervous system is the pathological hallmark of amyotrophic lateral sclerosis as well as certain subtypes of frontotemporal lobar degeneration (Takeuchi *et al.*, 2013).

(8) The EBV EBNA2<sub>311–320</sub>RGRGRGRGRG decapeptide is shared with the histone-lysine N-methyltransferase MLL4 or

EBV Protein*	Pos⁺	10-mer <sup>‡</sup>	11-mer <sup>‡</sup>	12-mer <sup>‡</sup>	13-mer <sup>‡</sup>	Human Proteins <sup>§</sup>
BDLF2	247	VYTLIPAVVI				TSN3
EBNA1	39			HGRGRGRGRGRG		RBM26
	40		GRGRGRGRGRG			CHTOP LAR1B <sup>¶</sup> MBD2 <sup>‡‡</sup> RS2 <sup>¶</sup>
	40			GRGRGRGRGGG		LARP1 SMD1 ZN579
	42	GRGRGRGRGG				RBM27
	92	GAGAGGAGAG				ARI1B
	93	AGAGGAGAGG				NOXA1
	96	GGAGAGGAGA				SKOR2
	96		GGAGAGGAGAG			ARI1B
	98	AGAGGAGAGG				NOXA1
	108	GAGAGGGAGG				NOVA2
	112		GGGAGGAGGAG			ONEC3
	113		GGAGGAGGAGG			FRM4A PCSK6 SHSA7
	114		GAGGAGGAGGA			FZD8
	114			GAGGAGGAGGAG		PCSK6
	115	AGGAGGAGGA				DND
	115		AGGAGGAGGAG			SHSA7
	116	GGAGGAGGAG				FRM4A ONEC3
	117		GAGGAGGAGAG			BD1L1
	129	GAGAGGGAGG				NOVA2
	133		GGGAGGAGGAG			ONEC3
	134	GGAGGAGGAG				FRM4A PCSK6 SHSA7
	135		GAGGAGGAGAG			BD1L1
	147	GAGAGGGAGG				NOVA2
	150		AGGGAGGAGAG			BMP2K.
	156	GAGAGGGAGG				NOVA2
	160		GGGAGGAGGAG			ONEC3
	161	GGAGGAGGAG				FRM4A PCSK6 SHSA7
	162		GAGGAGGAGAG			BD1L1
	174	GAGAGGGAGG				NOVA2
	177		AGGGAGGAGAG			BMP2K
	183	GAGAGGGAGG				NOVA2
	187		GGGAGGAGGAG			ONEC3
	188	GGAGGAGGAG				FRM4A PCSK6 SHSA7
	189		GAGGAGGAGAG			BD1L1
	199	GGGAGAGGAG				ARI1B
	202				AGAGGAGGAGGAG	PCSK6
	203	GAGGAGGAGG				FRM4A SHSA7
	203		GAGGAGGAGGA			FZD8
	204	AGGAGGAGGA				JUND
	204		AGGAGGAGGAG			SHSA7
	205	GGAGGAGGAG				FRM4A ONEC3

Pathogens and Disease (2013), 69, 205–212, © 2013 Federation of European Microbiological Societies. Published by John Wiley & Sons Ltd. All rights reserved 207

ERV Date     Deri     Tendi     <	Table 1 (continued)						
200 CAGGAGAGA CAGGAGAGAG CAGGAGAGAG CAGGAGAGAG CAGGAGAGAG CAGGAGAGAGA CAGGAGGAGAG CAGGAGAGAGA CAGGAGAGAGA CAGGAGGAGAG CAGGA	EBV Protein*	Pos <sup>†</sup>	10-mer <sup>‡</sup>	11-mer <sup>‡</sup>	12-mer <sup>‡</sup>	13-mer <sup>‡</sup>	Human Proteins <sup>§</sup>
11     00AM0GAAG     9000       211     00AM0GAAG     00A605AAG     9000       211     00A605AAG     00A605AAG     9000       211     00A605AAG     00A605AAG     9000       212     00A605AAG     00A605AAG     9000       213     00A605AAG     00A605AAG     9000       214     00A605AAG     00A605AAG     9000       215     00A605AAG     00A605AAG     9000       216     00A605AAG     00CAA605A     9000       216     00A605AAG     00CAA605A     9000       216     00A605AAG     00CAA605A     9000       216     00A605AAG     00CAA605A     9000       216     00A605AGAG     00CA605AGAG     9		206		GAGGAGGAGAG			BD1L1
211     Condecision     Condecision     Condecision     Antile       211     Galadecision     Galadecision     Condecision		211	GGAGAGGAGA				SKOR2
131     0.46466A0.466     NOX       271     0.46466A0.466     0.0000.400     0.0000       272     0.46466A0.466     0.46466A0.46     0.0000       273     0.644666A0.46     0.46666A0.46     0.0000       283     0.644666A0.46     0.66666A0.46     0.0000       283     0.64666A0.46     0.66666A0.46     0.0000       284     0.64666A0.46     0.66666A0.46     0.0000       284     0.64666A0.46     0.66666A0.46     0.0000       284     0.66666A0.46     0.66666A0.46     0.0000       284     0.66666A0.46     0.66666A0.46     0.0000       284     0.66666A0.46     0.66666A0.46     0.0000       284     0.66666A0.46     0.66666A0.46     0.0000       284     0.666660A0.46     0.666660A0.46     0.0000       284     0.666660A0.46     0.666660A0.46     0.0000       284     0.666660A0.46     0.666660A0.46     0.0000       284     0.666660A0.46     0.666660A0.46     0.0000       284     0.6666660A0.46     0.666666		211		GGAGAGGAGAG			ARI1B
217     0.04040GA404     0.002       228     664064054     0.4640GA404     0.4640GA404       228     664064054     0.46640GA40     0.46640GA40       228     6640640540     0.46640GA40     0.46640GA40       229     6640640540     0.46640GA40     0.46640GA40       229     6640640540     0.46640GA40     0.46640GA40       230     640640GA40     0.46640GA40     0.46640GA40       230     640640GA40     0.46640GA40     0.46640GA40       230     640640GA40     0.46640GA40     0.46640GA40       231     400640GA40     0.46640GA40     0.46640GA40       232     400640GA40     0.46640GA40     0.46640GA40       233     400640GA40     0.46640GA40     0.46640GA40       234     400640GA40     0.46640GA40     0.46640GA40       231     4040340GA40     0.46640GA40     0.46640GA40       232     4040340GA40     0.46640GA40     0.46640GA40       233     4040340GA40     0.46640GA40     0.46640GA40       234     4040340GA40		213	AGAGGAGAGG				NOXA1
21     Conditional     Condion     Condional     Condion<		217	GAGAGGGAGG				NOVA2
222     CGAGGAGGAS     FIMA PCNG SHOT       223     GGAGAGAS     GGAGAGAS     GGAGAGAS     BTII       224     GGAGAGAS     GGAGGAGAS     GGAGGAGAS     BTII       224     GGAGAGAS     GGAGGAGAS     GGAGGAGAS     COL       224     GGAGGAGAS     GGAGGAGAS     COL     COL       224     AGAGGAGAS     GGAGGAGAS     COL     COL       225     AGAGGAGAS     GGAGGAGAS     COL     COL       226     AGAGGAGAS     GGAGGAGAS     COL     COL       227     CAGGGAGAS     GGAGGAGAS     COL     COL       228     AGAGGAGAS     GGAGGAGAS     GGAGGAGAS     COL <t< td=""><td></td><td>221</td><td></td><td>GGGAGGAGGAG</td><td></td><td></td><td>ONEC3</td></t<>		221		GGGAGGAGGAG			ONEC3
223     GAGGAGAGA     GAGGAGAGA     BDU1       224     GAGGAGAGA     GAGGAGAGA     SCAGGAGAGA     SCAG       226     AdAGGAGAGA     GGAGGAGAGA     SCAGGAGAGA     SCAG       236     AdAGGAGAGA     GGAGGAGAGA     SCAGGAGAGA     SCAG       236     AdAGGAGAGA     GGAGGAGAGA     SCAG     SCAG       236     AdAGGAGAGA     GGAGGAGAGA     SCAG     SCAG       236     AdAGGAGAGA     GGAGGAGAGA     GGAGGAGAGA     SCAG       237     AdAGGAGAGA     GGGGAGGAGA     GGGGAGGAGA     SCAG       237     AdAGGAGAGA     GGGGAGGAGA     GGGGGAGAGA     SCAG       237     AdAGGAGAGA     GGGGAGGAGA     GGGGGAGAGA     SCAG       236     AdAGGAGAGA     GGGGAGGAGA     GGGGGAGGAGA     SCAGGAGGAGA       237		222	GGAGGAGGAG				FRM4A PCSK6 SHSA7
280     Conductands     SOOR       281     Condicional     SOOR     SOOR       282     Condicional     SOOR     SOOR       283     Condicional     Condicional     SOOR       284     Condicional     Condicional     Condicional		223		GAGGAGGAGAG			BD1L1
200     Anticachadad     Galacachada     Mitile       201     Cadacachada     Cadacachada     Codacachada     Codacachada       201     Cadacachada     Cadacachada     Cadacachada     Codacachada       202     Cadacachada     Cadacachada     Cadacachada     Codacachada       203     Cadacac		228	GGAGGGGGAGA				SKOR2
200     6006040405     NOX1     NOX1       236     6006040405     600604040     NOX1       236     600604040     600604040     NOX1       236     600604040     600604040     NOX1       236     600604040     600604040     NOX1       246     600604040     600604040     NOX1       246     600604040     600604040     NOX1       247     7606040640     600604040     NOX1       248     7606040640     6006060404     NOX1       277     600604040     6006060404     9011       277     6006060404     6006060404     9011       277     6006060404     6006060404     9011       277     6006060404     6006060404     9011       277     6006060404     6006060404     9011       278     60060604040     6006060404     9011       278     60060604040     6006060404     9011       278     600606060404     60060606040     9011       278     6006060604		228		GGAGAGGAGAG			ARI1B
281     GARAGRAGG     ANIB     NAIIB       282     GARAGRAGG     GARAGRAGG     NAIB       283     GARAGRAGG     GARAGRAGG     NAIB       284     AAGGAGAG     GARAGRAGG     NAIB       285     AAGGAGAGAG     GAGAGGAGAG     NAIB       286     AAGGAGAGAG     GAGAGGAGAG     NAIB       286     AAGGAGAGAG     GAGAGGAGAG     NAIB       286     AGGAGGAGAG     GAGAGGAGAG     NAIB       287     AGGAGGAGAG     GAGAGGAGAG     NAIB       288     AGGAGGAGAG     GAGAGGAGAG     NAIB       287     AGGAGGAGAG     GAGAGGAGAG     NAIB       288     AGGAGGAGAG     GAGAGGAGAG     NAIB       289     AGGAGGAGAG     GAGAGGAGAG     NAIB       281     AGGAGGAGAG     AGGAGGAGAG     NAIB       281     AGGAGGAGAG     AGGAGGAGAG     NAIB       281     AGGAGGAGAG     AGGAGGAGAG     NAIB       281     AGGAGGAGAG     AGGAGGAGAG     NAIB       281     AGGAGGAGA		230	AGAGGAGAGG				NOXA1
28     GAAGGAAGA     NOXAI       28     GAAGGAAGA     GAAGGAAGA     SCORE       28     KAAGGAAGAG     GAAGGAAGAG     SCORE       26     KAAGGAGGAG     GAGGAGGAGA     CAGGAGGAGA       28     KAAGGAGGAG     GAGGAGGAGA     CAGGAGGAGA       28     KAAGGAGGAG     GAGGAGGAGA     CAGGAGGAGA       28     KAAGGAGGAG     GAGGAGGAGA     CAGGAGGAGA       28     AGAGGAGGAG     GAGGAGGAGAG     GAGGAGGAGA       27     GACGAGGAGAG     GAGGAGGAGAG     CAGGAGGAGAG       28     AGAGGAGGAG     AGGGAGGAGAG     GAGGAGGAGAG       29     AGAGGAGGAGA     GAGGAGGAGAG     GAGGAGGAGAG       29     AGAGGAGGAG     GAGGAGGAGAG     GAGGAGGAGAG       29     AGAGGAGGAGA     GAGGAGGAGAG     GAGGAGGAGAG       29     AGAGGAGGAGA     GAGGAGGAGAG     GAGGAGGAGAG       29     AGAGGAGGAGA     GAGGAGGAGAG     GAGGAGGAGAG       29     GAGGAGGAGAG     GAGGAGGAGAG     GAGGAGGAGAG       29     GAGGAGGAGAG     GAGGAGGAGAG		234	GAGAGGAGAG				ARI1B
28     GGAGGAGAG     SCORE       20     AGAGGAGAG     GGAGGAGAG     SCORE       26     AGAGGAGAG     GGAGGAGAG     SCORE       26     AGAGGAGAG     GGAGGAGAG     SCORE       26     AGAGGAGAG     GGAGGAGAG     SCORE       26     AGAGGAGAG     GAGGAGGAG     SCORE       27     GAGGAGGAG     GAGGAGGAG     SCORE       28     AGAGGAGAG     GAGGAGGAG     GAGGAGGAG       27     GAGGAGGAG     GAGGAGGAG     GAGGAGGAG       28     GAGGAGGAG     GAGGAGGAG     GAGGAGGAG       28     GAGGAGGAG     GGGGAGGAG     GAGGAGGAG       29     GAGGAGGAG     GGGGAGGAG     GGGGAGGAG       29     GAGGAGGAG     GGGGAGGAG     GGGGAGGAG       29     GAGGAGGAGA     GGGGAGGAGA     GGGGAGGAGA       29     GAGGAGGAGAG     GGGGAGGAGAG     GGGGAGGAGA       29     GAGGAGGAGAG     GGGGAGGAGA     GGGGAGGAGA       29     GGGGAGGAGAG     GGGGAGGAGAG     GGGGGAGGAGA       29		235	AGAGGAGAGG				NOXA1
28     Cidadedadd     Mills       24     Adadadad     Cidadedad     Noxi       24     Adadadad     Cidadedad     Noxi       24     Adadadad     Cidadedad     Noxi       24     Adadedad     Cidadedad     Noxi       25     Adadedad     Cidadedad     Noxi       261     Adadedad     Cidadedad     Novi       271     Gadadedad     Gadeadada     Novi       272     Gadadedad     Gadeadada     Novi       273     Gadadedad     Gadadedada     Gadeadada       280     Gadadadada     Gadadadada     Novi<		238	GGAGAGGAGA				SKOR2
240     6.406.64.64.64     NOX41       245     4.64.66.46.64.64     6.46.66.46.64.64     PCSK6     PCSK6       246     A.64.66.46.64.64     6.46.66.46.64.64     PCSK6     PCSK6     PCSK6       257     A.64.66.46.64.64     6.46.66.46.64     PCSK6     PCSK6     PCSK6       261     A.64.66.46.64     6.46.64.64.64     PCSK6     PCSK6     PCSK6       277     G.46.66.66.64     G.46.66.66.64     PCSK6     PCSK6     PCSK6       287     A.64.66.66.64     G.46.66.66.64     BD11     PCSK6     PCSK6       288     A.64.66.66.64     G.46.66.66.64     BD11     PCSK6     PCSK6       288     A.64.66.66.64     G.46.66.66.64     BD11     PCSK6     PCSK6       288     A.64.66.66.64     G.46.66.66.64     G.46.66.66.64     PCSK6     PCSK6       288     A.66.66.66.64     G.46.66.66.64     PCSK6     PCSK6     PCSK6       211     G.66.66.66.64     G.66.66.66.64     G.66.66.66.64     PCSK6     PCSK6       212     G.66.66.66.64		238		GGAGAGGAGAG			ARI1B
245     0.64.0G.66.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G.G		240	AGAGGAGAGG				NOXA1
246     CadGaGGAGAG     BDIL       247     CaGGAGGAGAG     GAGGAGGAGA     BDIL       247     CaGGAGGAGAG     GAGGAGGAGAG     BDIL       267     AGAGGAGAG     GAGGAGGAGA     BDIL       277     GAGGAGGAGA     GAGGAGGAGA     BDIL       287     GAGGAGGAG     GAGGAGGAGA     BDIL       287     GAGGAGGAG     GAGGAGGAGA     BDIL       287     GAGGAGGAG     GAGGAGGAGA     BDIL       288     AGAGGAGGAG     GAGGAGGAGA     BDIL       289     AGAGGAGGAG     GAGGAGGAGA     BDIL       289     AGAGGAGGAG     GAGGAGGAGAG     BDIL       280     AGAGGAGGAG     GAGGAGGAGAG     BDIL       281     AGAGGAGGAGA     GAGGAGGAGAG     BDIL       281     AGAGGAGGAGA     GAGGAGGAGAG     BDIL       281     AGAGGAGGAGAG     GAGGAGGAGAG     BDIL       313     AGAGGAGGAGAG     GAGGAGGAGAG     BDIL       314     GAGGAGGAGAG     GAGGAGGAGAG     BDIL       315     GGAGGAGGAG		245	AGAGGAGGAG				PCSK6
253     0.40.60.60.60     PCSK6     PCSK6       264     3.64.60.60.60     6.66.60.60.60     BD1.1       277     2.77     6.46.60.60.60     BD1.1       287     3.64.60.60.60     6.66.60.60.60     BD1.1       287     3.64.60.60.60     6.66.60.60.60     BD1.1       287     3.64.60.60.60     3.66.60.60.60     BD1.1       288     3.64.60.60.60     6.66.60.60.60     BD1.1       289     3.66.60.60.60     6.66.60.60.60     BD1.1       281     3.66.60.60.60     6.66.60.60.60     BD1.1       313     3.66.60.60.60     6.66.60.60.60     BD1.1       314     3.66.60.60.60     6.66.60.60.60     BD1.1       315     6.66.66.60.60     6.66.66.60.60     BD1.1       316     6.66.66.60.60     6.66.66.60.60     BD1.1       317     6.66.66.60.60     6.66.66.60.60     BD1.1       318     6.66.66.60.60     6.66.66.60.60     BD1.1       317     6.66.66.60.60     6.66.66.60.60     BD1.1       316     6.66.66.60.60		246		GAGGAGGAGAG			BD1L1
254     GAGGAGAG     GAGGAGAG     DILI       201     AGAGAGAG     GAGGAGAG     BDILI       201     GAGGAGAG     GAGGAGAG     BDILI       201     GAGGAGAG     GAGGAGAG     BDILI       201     GAGGAGAGA     GAGGAGAGA     BDILI       201     GAGGAGAGA     GAGGAGAGA     BDILI       201     GAGGAGAGA     GAGGAGAGA     BDILI       202     AGAGAGCAG     GAGGAGAGA     BDILI       203     AGAGAGCAG     GAGGAGAGAG     BDILI       204     GAGGAGGAGAG     GAGGAGAGAG     BDILI       203     AGAGGAGAGA     GAGGAGAGAG     BDILI       204     GAGGAGGAGAG     GAGGAGAGAG     BDILI       214     GAGGAGAGAG     GAGGAGAGAG     GAGGAGGAGAG       215     GGAGAGAGAG     GAGGAGAGAG     GAGGAGAGAG       216     GAGGAGAGAGA     GAGGAGAGAGA     BDILI       217     GGAGAGAGAG     GAGGAGAGAG     GAGGAGAGAG       218     GAGGAGAGAG     GAGGAGAGAGA     GAGGAGAGAG		253	AGAGGAGGAG				PCSK6
281     AGAGGAGAG     PCSK6     PCSK6       277     GAGGAGAG     GAGGAGAG     PCSK6       287     AGAGGAGAG     AGGGAGAGAG     NOVA2       287     AGAGGAGAG     AGGGAGAGAG     NOVA2       287     AGAGGAGAG     AGGGAGAGAG     PCSK6       288     AGAGGAGAG     GAGGAGAGAG     PCSK6       289     AGAGGAGAG     GAGGAGAGAG     PCSK6       286     AGAGGAGAG     GAGGAGAGAG     PCSK6       288     AGAGGAGAG     GAGGAGAGAG     PCSK6       303     AGAGGAGAGA     GAGGAGAGAG     PCSK6       314     GAGGAGAGAG     GAGGAGAGAG     PCSK6       314     GAGGAGAGAG     GAGGAGAGAG     PCSK6       315     GGAGGAGAGA     GAGGAGAGAG     PCSK6       316     GGAGGAGAGA     GAGGAGAGAG     PCSK6       317     AGAGGAGAGAG     GAGGAGAGAGA     PCSK6       318     GGAGGAGAGA     GAGGAGAGAGA     PCSK6       319     GGAGGAGAGA     GAGGAGAGAGA     PCSK6       326		254		GAGGAGGAGAG			BD1L1
262     GGGGGGGGG     GGGGGGGGG     DIL1       277     GGGGGGGG     GGGGGGGGG     NOV2       287     AGGGGGGG     GGGGGGGGG     NOV2       287     AGGGGGGG     GGGGGGGGG     NOV2       288     AGGGGGGGG     GGGGGGGGG     NOV2       289     AGGGAGGAG     GGGGGGGGG     NOV2       289     AGGGAGGAG     GGGGGGGGGG     NOV2       289     AGGGAGGAG     GGGGGGGGGG     NOV2       280     AGGGAGGAG     GGGGGGGGGG     NOV2       303     AGGGAGGAG     GGGGGGGGGG     NOV2       314     GGGGGGGGGG     GGGGGGGGGG     NOV2       314     GGGGGGGGGG     GGGGGGGGGG     NOV2       315     GGGGGGGGGG     GGGGGGGGGG     NOV3       316     GGGGGGGGGG     GGGGGGGGGGGGG     NOV3       317     AGGGGGGGGG     GGGGGGGGGGG     NOV3       318     GGGGGGGGGG     GGGGGGGGGG     NOV3       311     GGGGGGGGG     NOV3     NOV3       311     GGGGGGGGGG		261	AGAGGAGGAG				PCSK6
277     GaGaGadad     Nova2       280     AgaGaGada     Nova2       287     AgaGaGada     Nova2       287     AgaGaGada     Pcsk6       286     GaGaGaGada     Pcsk6       286     GaGaGaGada     BD1L1       296     GaGaGaGada     BD1L1       296     GaGaGaGada     BD1L1       203     AgaGaGada     GaGaGaGada       304     GaGaGaGada     GaGaGaGada       317     AgaGaGada     BD111       318     GaGaGaGada     BD111       317     AgaGaGaGada     BD111       318     GaGaGaGada     BD111       317     AgaGaGaGada     BD111       328     GaGaGaGada     BD111       327     GaGaGaGada     BD111       328     GaGaGaGada     BD111       327     GaGaGaGada     BCB       328     GaGaGaGada     BCB       331     AgaGaGaGada     BCB       331     GaGaGaGada     BCB       331     G		262		GAGGAGGAGAG			BD1L1
280     AGGGAGGAG     BMP2K       287     AGAGGAGGAG     PCSK6       288     GAGGAGGAG     PCSK6       286     GAGGAGGAG     GAGGAGGAG       286     GAGGAGGAG     PCSK6       295     AGAGGAGGAG     GAGGAGGAG       296     GAGGAGGAG     GAGGAGGAG       297     GAGGAGGAG     GAGGAGGAG       308     AGAGGAGGAG     GAGGAGGAG       314     GAGGAGGAG     GAGGAGGAG       315     GGAGGAGGA     GAGGAGGAGG       316     GGAGGAGGA     GAGGAGGAGG       317     AGAGGAGGAG     ARIB       318     GGAGGAGGG     GGGAGGAGGG       319     GGAGGAGGG     NOXAI       210     GGAGGAGGG     NOXAI       221     GGAGGAGGG     GGAGGAGGG       323     GGAGGAGGG     NOXAI       333     GGAGGAGGG     NOXAI       343     GGAGGAGGG     NOXAI       344     GAGGAGGGGG     NOXAI       351     GACGAGGGGG     NOXAI		277	GAGAGGGAGG				NOVA2
287     AGAGGAGGAG     PCSK6       288     CAGGAGGAG     GAGGAGGAG       286     CAGGAGGAG     BDIL1       295     AGAGGAGGAG     GAGGAGGAG       296     CAGGAGGAG     GAGGAGGAG       203     AGAGGAGGAG     GAGGAGGAG       204     CAGGAGGAG     GAGGAGGAG       304     CACACAGGAG     CAGGAGGAG       315     CGAGAGGAGG     CAGGAGGAGG       316     CGAGAGGAGGA     GGAGGAGAG       317     AGAGGAGGG     CACACAGGAGAG       318     CGAGGGGGGG     CACACAGGAGAG       311     CGAGGGGGG     CHDS       311     CAPPPPPP     CHDP MLL4 RBM26       311     CHDP MLL4 RBM26     CHDP MLL4 RBM26		280		AGGGAGGAGAG			BMP2K
28     GAGGAGAGA     BD1L1       296     AGAGGAGAG     EAGGAGGAG       298     AGAGGAGAG     EAGGAGGAGA       298     AGAGGAGAG     EAGGAGGAGA       298     AGAGGAGAG     EAGGAGGAGA       299     AGAGGAGAG     EAGGAGGAGA       303     AGAGGAGAG     GAGGAGGAGA       314     AGAGGAGAGA     EAGGAGGAGA       315     GGAGAGGAGA     GAGGAGGAGA       317     AGAGGAGGAG     ARIIB       317     AGAGGAGGAG     ARIIB       317     AGAGGAGGAGA     GAGGAGGAGA       318     GGAGGAGGAG     ARIIB       317     AGAGGAGGAG     ARIIB       318     GGAGGAGGAG     ARIIB       317     AGAGGAGGAG     ARAIB       318     GGAGGAGGAG     ARIB       329     GGAGGAGGGG     AGAGGAGGAG       321     SAAAAAAV     AAAAAAV       51     GAPPPPPPP     AAAAAV       51     GAPPPAPA     AAAAAV       51     BCACPAGGAG     AAAAAAV		287	AGAGGAGGAG				PCSK6
295     AGAGGAGGAG     PCSK6       296     GAGGAGGAG     BDIL1       291     GAGGAGGAG     PCSK6       303     AGAGGAGGAG     BDIL1       314     AGAGGAGGAG     BDIL1       315     GGAGGAGAGA     GAGGAGGAG       317     AGAGGAGGAG     AR11B       317     AGAGGAGGAG     AR11B       317     AGAGGAGAGG     AR11B       317     AGAGGAGAGG     AR11B       317     AGAGGAGGAG     AR11B       318     GGAGGAGGAG     AR11B       317     AGAGGAGAGG     AR11B       318     AGAGGAGAGG     AR11B       311     BGAGAGGAGG     BAAAAAVA       57     GVPPPPPPP     FUS       51     GVPPPPPPP     FUS       51     BGAGAGGAG     CHD5       51     BCNP     FUS		288		GAGGAGGAGAG			BD1L1
296     GaGaGaGaG     BD1L1       303     303     GaGaGaGaG     PCSK6       304     314     PCSK6     BD1L1       315     GaGaGaGaG     GaGaGaGaG     BD1L1       316     GaGaGaGaG     GaGaGaGaG     BD1L1       317     AGAGGAGAG     GaGaGaGaG     BD1L1       317     AGAGGAGAG     GaGaGaGaG     NoxA1       317     AGAGGAGAG     GaGaGaGaG     NoXA1       317     AGAGGAGAG     AGAGGAGAG     SKOR2       317     AGAGGAGAG     NOXA1     NOXA1       327     GGAGAGGAGG     SKOR2     NOXA1       328     GGAGAGGAGG     SKOR2     NOXA1       329     GGAGAGGAGG     SKOR2     NOXA1       321     IS01     SAAAAAAVA     SAAAAAVA     CHD5       57     GVPPPPPPP     SVDP     CHD5     CHD5       51     GVPPOPPAP     AGAGGAGGAG     CHD5     CHD5		295	AGAGGAGGAG				PCSK6
303     AGAGGAGGA     PCSK6       304     CSK6     BD1L1       305     GGAGGAGAG     GAGGAGGAG       316     GGAGGAGAG     GGGGAGGAG       317     AGAGGAGGAG     GGGGGGGAGG       317     AGAGGAGGAG     GGGGGGGGGGG       317     AGAGGAGGG     SKOR2       317     AGAGGAGGG     NOXA1       327     GGGAGGGGGG     NOXA1       328     GGGAGGGGG     NOXA1       335     GGGAGGGGG     NOXA1       336     GGAGGGGGG     NOXA1       337     GGAGGGGGG     NOXA1       338     GGAGGGGGG     NOXA1       333     GGAGG		296		GAGGAGGAGAG			BD1L1
304     BD1L1       315     GGGAGGAGG     BD1L1       316     GGAGGAGGG     AR1B       317     AGAGGAGGG     SKOR2       317     AGAGGAGGG     SKOR2       317     AGAGGAGGG     SKOR2       317     AGAGGAGGG     NOXA1       327     GGRGGGGGG     SKOR2       318     GGRGGGGGG     NOXA1       327     GGRGGGGGG     SKOR2       328     GGRGGGGGG     NOXA1       335     GGRGGGGGG     SKOR2       336     GGRGGGGGG     SKOR2       337     GGRGGGGGG     SKOR2       338     GGRGGGGGG     SKOR2       333     GGRGGGGGG     SKOR2       333     GGRGGGGGG     SKOR2       333     GGRGGGGGGG     SKOR2       333     GGRGGGGGG     GGRGGGGG       311     RGRGRGGG     STOR2       311     RGRGRGGG     STOR2       311     RGRGRGGG     STOR2		303	AGAGGAGGAG				PCSK6
314   GGGAGGAGAG   ARI1B     315   GGAGGAGAG   SKOR2     317   AGAGGAGAG   NOXA1     327   GGRGRGGGGG   NOXA1     328   GGRGRGGGGG   NOXA1     329   GGRGRGGGGG   SKOR2     335   GGRGRGGGG   NOXA1     335   GGRGRGGGG   SKOR2     343   GGRGRGGGG   SKOR2     343   GGRGRGGGG   SKOR2     343   GGRGRGGGG   SKOR2     343   GGRGRGGGG   SKOR2     344   GGRGRGGGG   SKOR2     345   GGRGRGGGG   SKOR2     346   GGRGRGGGG   SKOR2     343   GGRGRGGGG   SKOR2     344   GGRGGGGG   SKOR2     345   GGRGRGGGG   SKOR2     346   GGRGGGGGG   SKOR2     347   GVPPPPPP   SKOR2     348   GGRGGGGGG   SKOR2     349   GGRGGGGGG   SKOR2     341   GGRGGGGG   SKOR2     341   RGRGRGGG   SKOR2 <td< td=""><td></td><td>304</td><td></td><td>GAGGAGGAGAG</td><td></td><td></td><td>BD1L1</td></td<>		304		GAGGAGGAGAG			BD1L1
315     GGAGAGGAG     SKOR2       317     AGAGAGAGG     NOXA1       317     AGAGGAGGG     NOXA1       327     GGRGRGGGGG     NOXA1       328     GGRGRGGGGG     NOXA1       335     GGRGRGGGG     NOXA1       335     GGRGRGGGG     NOXA1       336     GGRGRGGGG     NOXA1       337     GGRGRGGGG     NOXA1       338     GGRGRGGGG     NOXA1       310     FU     SAAAAAVA       57     GVPPPPPPP     CHD5       311     RGRGRGGG     CHD6       311     RGRGRGGG     CHD6		314			GGGAGAGGAGAG		ARI1B
317     AGAGGAGG     NOXA1       327     GGRGRGGGGG     EUS       325     GGRGRGGGGG     EUS       335     GGRGRGGGGG     EUS       343     GGRGRGGGG     EUS       271     SAAAAAVA     ENA2       311     RGRGRGGG     CHDB       311     RGRGRGGG     CHDB       311     RGRGRGGG     CHDB		315	GGAGAGGAGA				SKOR2
327     GGRGRGGGG     FUS.       335     GGRGRGGGG     FUS.       335     GGRGRGGGG     FUS.       343     GGRGRGGGG     FUS.       DEN     1501     SAAAAAAVA       57     GVPPPPPPP     CHD5       11     RGRGRGGG     CHTOP MLL4 RBM26		317	AGAGGAGAGG				NOXA1
335     GGRGRGGGG     FUS       343     GGRGRGGGGG     FUS       343     GGRGRGGGGG     FUS       343     GGRGRGGGGG     CHD5       25     GVPPPPPPP     CHD5       311     RGRGRGGG     CHTOP MLL4 RBM26		327	GGRGRGGSGG				FUS.
343     GGRGRGGGG     FUS       DEN     1501     SAAAAAAVA     CHD5       EBNA2     57     GVPPPPPPP     DIAP3 SFR15       311     RGRGRGRGG     CHTOP MLL4 RBM26		335	GGRGRGGSGG				FUS
DEN     1501     SAAAAAAVA     CHD5       EBNA2     57     GVPPPPPPP     DIAP3 SFR15       311     RGRGRGRGG     CHTOP MLL4 RBM26		343	GGRGRGGSGG				FUS
EBNA2 57 GVPPPPPPP DIAP3 SFR15 311 RGRGRGRGG	DEN	1501	SAAAAAAVA				CHD5
311 RGRGRGRG	EBNA2	57	GVPPPPPPP				DIAP3 SFR15
		311	RGRGRGRGRG				CHTOP MLL4 RBM26

208 Pathogens and Disease (2013), 69, 205–212, © 2013 Federation of European Microbiological Societies. Published by John Wiley & Sons Ltd. All rights reserved

Table 1 (continued)						
EBV Protein*	Pos⁺	10-mer <sup>‡</sup>	11-mer <sup>‡</sup>	12-mer <sup>‡</sup>	13-mer <sup>‡</sup>	Human Proteins <sup>§</sup>
	311 312 313	RGRGRGRG	GRGRGRGRG	RGRGRGRGRGRG		LAR1B <sup>1</sup> LARP1 <sup>1</sup> MBD2 <sup>1+</sup> RS2 <sup>1</sup> SMD1** ZN579 <sup>++</sup> CHTOP RBM26 MLL4
Q3KSS2	84			KVVILGQDPYHG		UNG
The entire EBV pro matching analysis t	teome, excepter to the human pro	d EBV IL-10H, for a total of oteome by using PIR pepti	f 68 proteins was dissected into ide match program (pir.georgetc	overlapping decapeptides shifte wn.edu/pirwww/search/peptide.s	d by one residue. The decapeptic shtml) (Wu <i>et al.</i> , 2003).	les were used as probes in peptide
TEBV proteins give.	n as UniProtNb/ EBV protein seç	'owiss-Prot entry name; juence;				
‡aa peptide sequer	nces given in 1-l	letter code;				
§human proteins in	volved in the pel	ptide overlap given as UniF	<sup>o</sup> rotKB/Swiss-Prot entry name. E	3iological function of the 28 hum	an proteins (in alphabetical order)	): ARI1B. Involved in transcriptional
activation and repre	ession of select o	genes by chromatin remod∈	eling. Belongs to the neural prog	enitors-specific chromatin remoc	leling complex. BD1L1. Biorientati	on of chromosomes in cell division.
BMP2K. May be inv	volved in osteobl	last differentiation. CHD5. [	Development of the nervous sys	tem. CHTOP. A role in the active	tion of estrogen receptor target g	enes and in silencing of fetal globin
genes. DIAP3. Bind	Is to GTP-bound	form of Rho and to profilin	n. Promotes actin polymerization	. Required for cytokinesis, stress	s fiber formation, and transcription	al activation of the serum response
iactor. FHIM4A. He intercellular transmi	gulates epimella ission of polarity	ti polarity. FUS. KINA-bindir information during tissue n	ng protein mat regulates transcr mombodenesis, JUND, Transcrit	Iption, splicing and minima trans ation factor binding AP-1 sites. I	port. FZD8. Receptor for writ pro AR1B. RNA-binding protein. MBF	otentis. Involved in transduction and 22. Binds CpG islands in promoters
where the DNA is n	nethylated at pos	sition 5 of cytosine. LARP1.	. Facilitates the synthesis of prot	eins required for cellular remode	elling and migration. MLL4. Methyl	ates 'Lys-4' of histone H3. NOVA2.
Regulates RNA sp	viicing or metable	olism in a specific subset	of developing neurons. NOXA	1. Activator of superoxide-prod	ucing NADPH oxidase. ONEC3.	Transcriptional activator. PCSK6.

Represents an endoprotease activity within the constitutive secretory pathway. RBM26. RNA-binding protein. RBM27. RNA-binding protein. RS2. Participates in aminoacyl-tRNA binding to the ribosome. SFR15. Links transcription and pre-mRNA processing. SHSA7. Transmembrane adaptor. SKOR2. Has transcriptional repressor activity. Acts as a TGF-beta antagonist in the nervous system. SMD1. Acts as a charged protein scaffold to promote or strengthen snRNP-snRNP interactions. TSN3. Proliferation and migration of oligodendrocytes. UNG. Excises uracil residues from the DNA ZN579. May be involved in transcriptional regulation.

 $[],^{**}, \dagger \uparrow, \ddagger$  refer to peptide occurrences repeated 3, 4, 5, and 6 times, respectively.

myeloid/lymphoid or mixed-lineage leukemia protein 4. MLL4 is a crucial player in cell viability and cell-cycle progression and is critical for tumor growth *in vivo* (Ansari *et al.*, 2012).

(9) The EBV Q3KSS2<sub>84–95</sub>KVVILGQDPYHG 12-mer is also present in human uracil–DNA glycosylase (UNG). UNG excises uracil residues from the DNA that can arise as a result of misincorporation of dUMP residues by DNA polymerase or due to deamination of cytosine. Defects in UNG are a cause of immunodeficiency with hyper-IgM type 5 (Kavli *et al.*, 2005).

In summary, we find a significant peptide sharing between EBV and the human proteome that involves even peptides 13-mer long and is predominant at level of EBV EBNA1 protein. In fact, 42 of the total 47 decapeptides shared between EBV and human proteins are derived from EBNA1 (Table 1). Such a peptide commonality might explain the immunoevasive properties of EBNA1 protein, a viral antigen that has been reported to go undetected by the cell-mediated immune system (Münz, 2004). Indeed, it seems logical to hypothesize that human immunotolerance mechanism(s) should prevent attacks against a protein, EBNA1, endowed with such a high level of sequence identity to human proteins. However, in parallel, it has to be observed that EBNA1 antigen is also one of the most frequently recognized EBV antigens for CD4+ helper T cells (Long et al., 2005), thus representing a prime target for T-cell-based immunotherapy (Tsang et al., 2006). Interestingly, mapping of CD4+ epitopes within the primary sequences of EBNA1 reveals a specific epitope location in the EBNA1 COOH region (Long et al., 2005; Tsang et al., 2006). In molecular terms, data from Table 1 favor the view of EBNA1 protein sequence as hosting antigenicity and immunotolerance in two spatially distinct domains, with the antigenic portion allocated along the COOH terminus and the potentially immunotolerogenic GRR confined in the NH2 region. Accordingly, a thorough search through IEDB database (Peters et al., 2005) highlights that the EBNA1 COOH terminus allocates 74 T-cell epitopes, while EBNA1 NH2 terminus GRR presents only one epitope (AGAGGGAGGAGAGA, IEDB ID: 1432) (Petersen et al., 1989) comprehending a 11-mer sequence reported in Table 1 (AGGGAGGAGAG).

This study suggests a role for EBNA1 GRR in immunoevasion and, in addition, may offer hints to explore the molecular basis underlying the delayed development of CD4+ T cell and humoral immune responses against EBV EBNA1 (Hislop et al., 2007; Long et al., 2013). Indeed, Levitskaya et al. (1995) demonstrated that the GRR allocated in the NH2 terminus of EBNA1 interferes with antigen processing and MHC class I-restricted presentation, possibly by inhibiting ubiquitin-/proteasome-dependent protein degradation (Levitskaya et al., 1997). In parallel, Yin et al. (2003) showed that the GRR inhibits EBNA1 mRNA translation and proposed that minimizing translation of the EBNA1 transcript, cells expressing EBNA1 avoid cytotoxic T-cell recognition. However, the molecular mechanism by which the GRR repeat inhibits ubiquitin-/proteasome-dependent degradation remained unexplained (Levitskaya et al., 1997). As a matter of fact, GRRs are critical domains in proteins such as TAR DNA-binding protein 43 (TDP-43), heterogeneous nuclear ribonucleoprotein A1 (ROA1), and the above-mentioned RNA-binding protein FUS (Rogelj *et al.*, 2011). TDP-43, ROA1, and FUS are proteins crucially involved in the regulation of different steps of gene expression, including transcription, splicing, mRNA transport, and translation (Strong *et al.*, 2007; Bekenstein & Soreq, 2012; Dormann and Haass, 2013). Then, it can be hypothesized that the EBNA1 GRR might interfere with GRRs present in host proteins, thus subverting crucial cellular functions, such as transcription and translation.

In the present context, it is also noteworthy that the extent and significance of the peptide overlap between EBNA1 and human proteins are even more relevant following sequence identity analyses using short peptide modules as scanning probes, being thousand the EBV matches in the human proteome (Kanduc et al., 2008). As an example, the EBV EBNA1 heptapeptide GAGAGGG occurs 15 times along the NH2-terminal domain of EBV EBNA1 (http://www.uniprot. org/uniprot/Q3KSS4). Of interest, the same heptapeptide GAGAGGG occurs in the human nuclear factor NF-kappa-B p105 (NFkB1) protein, where the Gly-rich heptapeptide GAGAGGG functions as a processing signal for the generation of the p50 subunit (Lin & Ghosh, 1996; Orian et al., 1999). Hence, it might be postulated that the multiple EBNA1 GRRs could compete for the proteolytic reaction of NFkB1, a transcriptional factor crucial in the activation and function of mature B cell (Pohl et al., 2002; Ruland & Mak, 2003).

More in general, the fact that short aa modules such as pentapeptides can represent minimal functional determinants in biological interactions and immune recognition (Kanduc, 2012a, b; Kanduc, 2013) implies a wide array of physio(patho)logical viral-host relationships, being massive the level of peptide overlap between EBV and human proteins at the 5-mer level (Kanduc et al., 2008). Hence, data reported in Table 1 not only might help understand EBV escape from immunosurveillance, but could also contribute to unfold the still obscure links between EBV infection and associated cancer diseases and autoimmune disorders (Farrell & Jarrett, 2011; Niller et al., 2011; Draborg et al., 2012; Tselis, 2012). Finally, the present phenetic analyses might open the way to innovative therapeutic approaches to fight/eradicate EBV infection and related pathologic sequelae (Kanduc et al., 2007; Lucchese et al., 2011; Capone et al., 2013).

## **Authors' Contributions**

All authors contributed to the computational analysis. D.K. proposed the original idea, interpreted the data, developed the research project, and wrote the manuscript. All authors discussed the results, and commented and revised the manuscript.

## Acknowledgement

The authors declare that there are no conflict of interests.

### References

- Ansari KI, Kasiri S, Mishra BP & Mandal SS (2012) Mixed lineage leukaemia-4 regulates cell-cycle progression and cell viability and its depletion suppresses growth of xenografted tumour *in vivo*. *Br J Cancer* 107: 315–324.
- Bekenstein U & Soreq H (2012) Heterogeneous nuclear ribonucleoprotein A1 in health and neurodegenerative disease: from structural insights to post-transcriptional regulatory roles. *Mol Cell Neurosci* DOI: pii: S1044-7431(12)00212-6.
- Capone G, Lucchese G, Calabrò M & Kanduc D (2013) West Nile virus diagnosis and vaccination: using unique viral peptide sequences to evoke specific immune responses. *Immunopharmacol Immunotoxicol* 35: 64–70.
- Cesarman E (2002) Epstein–Barr virus (EBV) and lymphomagenesis. *Front Biosci* 7: e58–e65.
- Dormann D & Haass C (2013) Fused in sarcoma (FUS): an oncogene goes awry in neurodegeneration. *Mol Cell Neurosci* DOI:pii: S1044-7431(13)00048-1.
- Draborg AH, Duus K & Houen G (2012) Epstein–Barr virus and systemic lupus erythematosus. *Clin Dev Immunol* 2012: 370516.
- Farrell K & Jarrett RF (2011) The molecular pathogenesis of Hodgkin lymphoma. *Histopathology* 58: 15–25.
- Gourzones C, Barjon C & Busson P (2012) Host-tumor interactions in nasopharyngeal carcinomas. *Semin Cancer Biol* 22: 127–136.
- Halgren C, Kjaergaard S, Bak M *et al.* (2012) Corpus callosum abnormalities, intellectual disability, speech impairment, and autism in patients with haploinsufficiency of ARID1B. *Clin Genet* 82: 248–255.
- Hislop AD, Taylor GS, Sauce D & Rickinson AB (2007) Cellular responses to viral infection in humans: lessons from Epstein–Barr virus. *Annu Rev Immunol* 25: 587–617.
- Jochum S, Moosmann A, Lang S, Hammerschmidt W & Zeidler R (2012) The EBV immunoevasins vIL-10 and BNLF2a protect newly infected B cells from immune recognition and elimination. *PLoS Pathog* 8: e1002704.
- Kalla M & Hammerschmidt W (2012) Human B cells on their route to latent infection–early but transient expression of lytic genes of Epstein–Barr virus. *Eur J Cell Biol* 91: 65–69.
- Kanduc D (2012a) Homology, similarity, and identity in peptide epitope immunodefinition. *J Pept Sci* 18: 487–494.
- Kanduc D (2012b) Peptide cross-reactivity: the original sin of vaccines. *Front Biosci* 4: 1393–1401.
- Kanduc D (2013) Pentapeptides as minimal functional units in cell biology and immunology. *Curr Protein Pept Sci* 14: 111–120.
- Kanduc D, Lucchese A & Mittelman A (2007) Non-redundant peptidomes from DAPs: towards "the vaccine"? *Autoimmun Rev* 6: 290–294.
- Kanduc D, Stufano A, Lucchese G & Kusalik A (2008) Massive peptide sharing between viral and human proteomes. *Peptides* 29: 1755–1766.
- Kavli B, Andersen S, Otterlei M, Liabakk NB, Imai K, Fischer A, Durandy A, Krokan HE & Slupphaug G (2005) B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil. J Exp Med 201: 2011–2021.
- Klein G, Klein E & Kashuba E (2010) Interaction of Epstein–Barr virus (EBV) with human B-lymphocytes. *Biochem Biophys Res Commun* 396: 67–73.
- Levitskaya J, Coram M, Levitsky V, Imreh S, Steigerwald-Mullen PM, Klein G, Kurilla MG & Masucci MG (1995) Inhibition of antigen processing by the internal repeat region of the Epstein–Barr virus nuclear antigen-1. *Nature* 375: 685–688.

- Levitskaya J, Sharipo A, Leonchiks A, Ciechanover A & Masucci MG (1997) Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein–Barr virus nuclear antigen 1. *P Natl Acad Sci USA* 94: 12616–12621.
- Lin L & Ghosh S (1996) A glycine-rich region in NF-kappaB p105 functions as a processing signal for the generation of the p50 subunit. *Mol Cell Biol* 16: 2248–2254.
- Long HM, Haigh TA, Gudgeon NH *et al.* (2005) CD4+ T-cell responses to Epstein–Barr virus (EBV) latent-cycle antigens and the recognition of EBV-transformed lymphoblastoid cell lines. *J Virol* 79: 4896–4907.
- Long HM, Chagoury OL, Leese AM, Ryan GB, James E, Morton LT, Abbott RJ, Sabbah S, Kwok W & Rickinson AB (2013) MHC II tetramers visualize human CD4+ T cell responses to Epstein– Barr virus infection and demonstrate atypical kinetics of the nuclear antigen EBNA1 response. *J Exp Med* 210: 933–949.
- Lucchese G, Stufano A & Kanduc D (2009) Proteome-guided search for influenza A B-cell epitopes. *FEMS Immunol Med Microbiol* 57: 88–92.
- Lucchese G, Stufano A & Kanduc D (2011) Searching for an effective, safe and universal anti-HIV vaccine: finding the answer in just one short peptide. *Self Nonself* 2: 49–54.
- Moore KW, Vieira P, Fiorentino DF, Trounstine ML, Khan TA & Mosmann TR (1990) Homology of cytokine synthesis inhibitory factor (IL-10) to the Epstein–Barr virus gene BCRFI. *Science* 248: 1230–1234.
- Moore KW, de Waal Malefyt R, Coffman RL & O'Garra A (2001) Interleukin-10 and the interleukin-10 receptor. *Annu Rev Immunol* 19: 683–765.
- Münz C (2004) Epstein-barr virus nuclear antigen 1: from immunologically invisible to a promising T cell target. *J Exp Med* 199: 1301–1304.
- Natale C, Giannini T, Lucchese A & Kanduc D (2000) Computer-assisted analysis of molecular mimicry between human papillomavirus 16 E7 oncoprotein and human protein sequences. *Immunol Cell Biol* 78: 580–585.
- Niller HH, Wolf H, Ay E & Minarovits J (2011) Epigenetic dysregulation of Epstein–Barr virus latency and development of autoimmune disease. *Adv Exp Med Biol* 711: 82–102.
- Orian A, Schwartz AL, Israël A, Whiteside S, Kahana C & Ciechanover A (1999) Structural motifs involved in ubiquitin-mediated processing of the NF-kappaB precursor p105: roles of the glycine-rich region and a downstream ubiquitination domain. *Mol Cell Biol* 19: 3664–3673.
- Peters B, Sidney J, Bourne P *et al.* (2005) The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol* 3: e91.
- Petersen J, Rhodes G, Patrick K, Roudier J & Vaughan JH (1989) Human T cell responses to the Epstein–Barr nuclear antigen-1 (EBNA-1) as evaluated by synthetic peptides. *Cell Immunol* 123: 325–333.
- Pohl T, Gugasyan R, Grumont RJ, Strasser A, Metcalf D, Tarlinton D, Sha W, Baltimore D & Gerondakis S (2002) The combined absence of NF-kappa B1 and c-Rel reveals that overlapping roles for these transcription factors in the B cell lineage are restricted to the activation and function of mature cells. *P Natl Acad Sci USA* 99: 4514–4519.
- Poole BD, Scofield RH, Harley JB & James JA (2006) Epstein–Barr virus and molecular mimicry in systemic lupus erythematosus. *Autoimmunity* 39: 63–70.
- Porter IM, McClelland SE, Khoudoli GA, Hunter CJ, Andersen JS, McAinsh AD, Blow JJ & Swedlow JR (2007) Bod1, a novel

kinetochore protein required for chromosome biorientation. *J Cell Biol* 179: 187–197.

- Ressing ME, Horst D, Griffin BD, Tellam J, Zuo J, Khanna R, Rowe M & Wiertz EJ (2008) Epstein–Barr virus evasion of CD8(+) and CD4(+) T cell immunity via concerted actions of multiple gene products. *Semin Cancer Biol* 18: 397–408.
- Rickinson AB & Kieff E (2007) Epstein–Barr virus. Fields Virology, Vol. 2. (Knipe DM & Howley PM, eds), pp. 2655–2700. Lippincott, Williams & Wilkins, Philadelphia, PA.
- Rogelj B, Godin KS, Shaw CE & Ule J (2011) The functions of glycine-rich regions in TDP-43, FUS and related RNA-binding proteins. RNA Binding Proteins (Lorković ZJ, ed.), pp. 1–17. Landes Bioscience and Springer Science+Business Media, Austin, TX.
- Rowe M & Zuo J (2010) Immune responses to Epstein–Barr virus: molecular interactions in the virus evasion of CD8+ T cell immunity. *Microbes Infect* 12: 173–181.
- Ruggiu M, Herbst R, Kim N, Jevsek M, Fak JJ, Mann MA, Fischbach G, Burden SJ & Darnell RB (2009) Rescuing Z+ agrin splicing in Nova null mice restores synapse formation and unmasks a physiologic defect in motor neuron firing. *P Natl Acad Sci USA* 106: 3513–3518.
- Ruland J & Mak TW (2003) Transducing signals from antigen receptors to nuclear factor kappaB. *Immunol Rev* 193: 93–100.
- Saha A & Robertson ES (2011) Epstein–Barr virus-associated B-cell lymphomas: pathogenesis and clinical outcomes. *Clin Cancer Res* 17: 3056–3063.
- Santen GW, Aten E, Sun Y *et al.* (2012) Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat Genet* 44: 379–380.
- Scerri TS, Brandler WM, Paracchini S, Morris AP, Ring SM, Richardson AJ, Talcott JB, Stein J & Monaco AP (2013) PCSK6 is associated with handedness in individuals with dyslexia. *Hum Mol Genet* 20: 608–614.
- Severa M, Giacomini E, Gafa V, Anastasiadou E, Rizzo F, Corazzari M, Romagnoli A, Trivedi P, Fimia GM & Coccia EM (2013) EBV stimulates TLR- and autophagy-dependent pathways and impairs maturation in plasmacytoid dendritic cells: implications for viral immune escape. *Eur J Immunol* 43: 147–158.
- Silverstein AM (2001) Autoimmunity versus horror autotoxicus: the struggle for recognition. Nat Immunol 2: 279–281.
- Strong MJ, Volkening K, Hammond R, Yang W, Strong W, Leystra-Lantz C & Shoesmith C (2007) TDP43 is a human low

molecular weight neurofilament (hNFL) mRNA-binding protein. *Mol Cell Neurosci* 35: 320–327.

- Takeuchi R, Toyoshima Y, Tada M *et al.* (2013) Transportin 1 accumulates in FUS inclusions in adult-onset ALS without FUS mutation. *Neuropathol Appl Neurobiol* 39: 580–584.
- Tindle RW (2002) Immune evasion in human papillomavirus-associated cervical cancer. *Nat Rev Cancer* 2: 59–65.
- Tiwari-Woodruff SK, Kaplan R, Kornblum HI & Bronstein JM (2004) Developmental expression of OAP-1/Tspan-3, a member of the tetraspanin superfamily. J Neurosci 77: 166–173.
- Trost B, Kusalik A, Lucchese G & Kanduc D (2010) Bacterial peptides are intensively present throughout the human proteome. *Self Nonself* 1: 71–74.
- Tsang CW, Lin X, Gudgeon NH, Taylor GS, Jia H, Hui EP, Chan AT, Lin CK & Rickinson AB (2006) CD4+ T-cell responses to Epstein– Barr virus nuclear antigen EBNA1 in Chinese populations are highly focused on novel C-terminal domain-derived epitopes. *J Virol* 80: 8263–8266.
- Tselis A (2012) Epstein–Barr virus cause of multiple sclerosis. *Curr Opin Rheumatol* 24: 424–428.
- UniProt Consortium (2009) The Universal Protein Resource (UniProt) 2009. Nucleic Acids Res 37: D169–D174.
- Wu CH, Yeh LS, Huang H et al. (2003) The protein information resource. *Nucleic Acids Res* 31: 345–347.
- Yang YY, Yin GL & Darnell RB (1998) The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. P Natl Acad Sci USA 95: 13254–13259.
- Yano M, Hayakawa-Yano Y, Mele A & Darnell RB (2010) Nova2 regulates neuronal migration through an RNA switch in disabled-1 signaling. *Neuron* 66: 848–858.
- Yin Y, Manoury B & Fahraeus R (2003) Self-inhibition of synthesis and antigen presentation by Epstein–Barr virus-encoded EBNA1. *Science* 301: 1371–1374.
- Yoon SI, Jones BC, Logsdon NJ & Walter MR (2005) Same structure, different function crystal structure of the Epstein–Barr virus IL-10 bound to the soluble IL-10R1 chain. *Structure* 13: 551– 564.
- Zeng MS, Li DJ, Liu QL, Song LB, Li MZ, Zhang RH, Yu XJ, Wang HM, Ernberg I & Zeng YX (2005) Genomic sequence analysis of Epstein–Barr virus strain GD1 from a nasopharyngeal carcinoma patient. J Virol 79: 15323–15330.