Theory Biosci. (2014) 133:79–89
DOI 10.1007/s12064-013-0196-1

ORIGINAL PAPER

Short peptide sequence identity between human viruses and HLA-B27-binding human 'self' peptides

Shipeng Sun · Tao Wang · Bo Pang · Huamin Wei · Guijian Liu

Received: 20 May 2013 / Accepted: 2 December 2013 / Published online: 22 December 2013

Abstract Molecular mimicry and arthritogenic peptides form the basis of hypotheses that attempt to explain the pathogenesis of HLA-B27-positive ankylosing spondylitis (AS). We propose, therefore, that certain human viruses may possess peptide sequences that mimic HLA-B27-binding human 'self' peptides which might induce or play a significant role in AS. In the present study, we performed bioinformatic analysis, using BLASTP, of the human virus proteome and HLA-B27-binding human 'self' peptides including peptides derived from arthritogenic sequences. We identified that some HLA-B27-binding peptides, particularly those present in proteins of the cartilage and bone, are highly similar to those present in viruses known to cause chronic infection. We suggest that the identical short amino acid sequences shared between human viruses and HLA-B27 peptides may play a role in the pathogenesis of AS.

Keywords Short peptide sequences · Ankylosing spondylitis · Virus · HLA-B27

Electronic supplementary material The online version of this article (doi:10.1007/s12064-013-0196-1) contains supplementary material, which is available to authorized users.

S. Sun (✉) · T. Wang · B. Pang · G. Liu (✉)
Clinical Laboratories, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, No. 5 Bei Xian Ge St., Xi Cheng District, Beijing 100053, People's Republic of China
e-mail: shipengsun@gmail.com

G. Liu
e-mail: liuguijian@hotmail.com

H. Wei
Cardiology Department, Beijing Royal Integrative Medicine Hospital, Beijing, People's Republic of China

Introduction

Human leukocyte antigen B27 (HLA-B27) is the best studied HLA-class I molecule because of its strong association with ankylosing spondylitis (AS) and other spondylarthritides (Brewerton et al. 1973; Reveille 2012; Gaston 2006). Approximately 85–90 % of patients with AS are HLA-B27 positive (Reveille 1998; Thomas and Brown 2010) and are at a 5.6–16 times greater recurrent risk of the disease than HLA-B27-negative individuals (Calin et al. 1983). Despite extensive investigation, the understanding of the role of HLA-B27 in disease is limited. Molecular mimicry, arthritogenic peptides, free heavy chains, and the unfolded protein response form the basis of the most plausible hypotheses (Allen et al. 1999).

The molecular mimicry hypothesis proposes that a cross-reactive peptide derived from an infecting bacterial pathogen stimulates T cells that can respond to an HLA-B27-associated "self-peptide" or to peptides derived from HLA-B27 (Benjamin and Parham 1990). Sequence similarities between short peptides of HLA-B*2705 and those of *Klebsiella pneumoniae*, *Yersinia*, and *Salmonella* have been identified (Tsuchiya et al. 1989; Fielder et al. 1995; Lahesmaa et al. 1991). Serologic studies indicate that patients with AS frequently produce antibodies against microbial pathogens (Ewing et al. 1990). In the arthritogenic peptide theory, HLA-B27 binds specific arthritogenic peptide(s) derived from microbial proteins or those of self-origin and presents them to CD8+ cytotoxic T cells. The immune system mistakenly reacts to these self-antigens and triggers AS (Benjamin and Parham 1990). PRGLLAWISR is derived from chondroitin sulfate *N*-acetyl galactosamine transferase 1, which is the homologue of *Yersinia enterocolitica* FhuB

and the intracellular attenuator protein A from *Salmonella typhimurium* (Ben Dror et al. 2010).

Viruses are obligate intracellular parasites, which hijack the host cell's biosynthetic machinery to enable the translation of viral proteins and the replication of the viral genome. The intracellular lifestyle of human viruses particularly viruses that cause chronic infections, which is a type of persistent infection that is eventually cleared, or last the life of the host, could result in the MHC class I presentation of virus peptides. For a long time it has been suspected that, relative to other class I molecules, HLA-B27 is a better restriction element that may dominate anti-viral responses (Gomard et al. 1984). HLA-B27 presents peptides to cytotoxic T lymphocytes and binds peptides derived from influenza virus, Epstein–Barr virus, human immunodeficiency virus-1 (HIV-1), hepatitis C virus (HCV) which have already been identified (Tanasescu et al. 1999). The binding peptides, that determine both folding stability and immune recognition, are probably central to the pathogenetic role of this molecule. Two studies have reported an extremely high prevalence of hepatitis B virus (HBV) surface antigen (HBsAg) in patients with AS when compared with that of the general population and patients with other spondyloarthritis, rheumatoid arthritis, or osteoarthritis (Zheng et al. 2012; Tanasescu et al. 1999). The high prevalence of HBsAg in patients with AS might be associated with a high frequency of the HLA-B27 gene. The prevalence of HBsAg or infection by HBV in patients with AS patients may be involved in the pathogenesis of autoimmune disease. Therefore, we propose that certain human viruses, particularly viruses that cause chronic infections, display peptides that mimic HLA-B27-binding peptides and that they may trigger AS or play a significant role in its progression. However, there is a scarcity of data regarding sequence identity between the proteomes of human viruses and the sequence of HLA-B27. The aim of this study was to perform a bioinformatics analysis of human viral proteomes and HLA-B27-binding peptides.

Materials and methods

Search for sequence identity between the peptides of human virus and HLA-B*2705 peptides derived from cartilage/bone-related proteins

To identify candidate viral sequences that mimic HLA-B27-binding peptide, all available amino acid sequences of human viruses were acquired from the viral genome database of the National Center for Biotechnology (NCBI) and filtered using the query "human" (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?opt=virus&taxid=10239&host=human). The NCBI taxid number of each virus present in the non-redundant protein sequences (nr) database was then accessed using the BLASTP algorithm with the virus name as query (Supplementary 1).

26 peptides derived from cartilage/bone-related proteins and two peptides from the HLA-B27 itself were identified in the HLA-B27 peptidome (Ben Dror et al. 2010) and the other three natural HLA-B27 GlnP2 peptides (IQRTPKIQ, IQRTPKIQVY, DQLQEQLQR) derived from cartilage/bone-related proteins (Infantes et al. 2013) are considered as candidate arthritogenic peptides. BLASTP was used to analyze these 31 peptides to all query viral sequences (Supplementary 1) available in the non-redundant GenBank protein (nr) database. The criterion for a significant match was the identity between at most three mismatch amino acid (aa) residues according to the results provided in the previous study (Ben Dror et al. 2010).

Search for sequence identity between the peptides of human virus that cause chronic infections and all identified HLA-B*2705 peptides

As long-term stimulation of the viruses that cause chronic infections may induce strong cytotoxic T cell reactivity, we suppose peptides of these viruses shared similarities with HLA-B27-binding peptides that derived from non-cartilage/bone-related proteins may also take effects in AS progression. We blasted the peptidomes of viruses that cause chronic infections, such as HBV, HCV, HIV, human T-lymphotropic virus 1 (HTLV-1), human herpesvirus 4 (HHV-4), human herpesvirus 5 (HHV-5), and human herpesvirus 8 (HHV-8) with 569 HLA-B27-binding peptides (Ben Dror et al. 2010) and peptide RQPQVSI (Infantes et al. 2013) to determine whether their sequences shared similarities. The criterion for a significant match was the identity between at most three mismatch amino acid residues according to the results provided in previous study (Ben Dror et al. 2010).

HLA-B*2705-binding prediction of identified viruses peptides

All identified human viruses peptides were further analyzed using SYFPEITHI (Rammensee et al. 1999) and NetMHC 3.4 Server (Nielsen et al. 2003; Lundegaard et al. 2008a, b) to discriminate HLA-B*2705 candidate epitopes versus non-binding sequences. NetMHC 3.4 server predicts binding of peptides to a number of different HLA alleles using artificial neural networks (ANNs). For ANNs prediction high-binding peptides have an IC50 value below 50 nM, and weak-binding peptides an IC50 values below 500 nM.

Table 1 HLA-B27-binding peptides related to cartilage highly similar to those of all available human viruses

Alignment	Human protein (top peptide)	Viral protein (lower peptide)	GenBank accession number of viral protein
362 QRSKSPL 368 QRSKSP- 166 QRSKSPR 172	ADAM metallopeptidase with thrombospondin type 1 motif, 15, isoform CRA_b	Core protein of HBV	ACP20752.1
362 QRSKSPL 368 QRSK-P- 157 QRSKRPR 163	ADAM metallopeptidase with thrombospondin type 1 motif, 15, isoform CRA_b	Protease of HIV-1	ABG45992.1
1514 GRHANTKVGL 1523 GR-AN-VGL 21 GRAANSLVGL 30	Collagen, type VI, alpha 3, isoform CRA_a	Polyprotein of HCV	AAK63490.1
948 VRNPSVVVK 956 VRN-SV+ V- 255 VRNESVIVR 262	Collagen, type VI, alpha 3, isoform CRA_a	Polyprotein of hepacivirus AK-2012	AFJ20708.1
948 VRNPSVVVK 956 -RNP-VV-K 155 ERNPAVVEK 163	Collagen, type VI, alpha 3, isoform CRA_a	E1B gene product of human adenovirus A	ABY47590.1
948 VRNPSVVVK 956 VRNP-V-VK 6 VRNPQVLVK 14	Collagen, type VI, alpha 3, isoform CRA_a	Rev protein of HIV-1	ACA09218.1
948 VRNPSVVVK 956 VR-PS-+ V- 191 VRTPSLVIN 198	Collagen, type VI, alpha 3, isoform CRA_a	Hemagglutinin–neuraminidase of human parainfluenza virus 3	ADX01306.1
1597 FRVGNVQEL 1605 FRV-VQ-L 289 FRVVAVQNL 297	Collagen, type VI, alpha 3, isoform CRA_a	Large S protein of HBV	ACF39609.1
1597 FRVGNVQEL 1605 FRVG-Q-L 20 FRVGASQKL 28	Collagen, type VI, alpha 3, isoform CRA_a	Envelope protein 2 of HCV	ACZ60972.1
1597 FRVGNVQEL 1605 -RV-NV+ EL 103 RRVMNVHEA 111	Collagen, type VI, alpha 3, isoform CRA_a	Envelope glycoprotein K of HHV-2	AEV91393.1
1597 FRVGNVQEL 1605 -RVGN-+ E- 764 ERVGNMDEM 772	Collagen, type VI, alpha 3, isoform CRA_a	BcLF1 of HHV4	YP_401697.1
1597 FRVGNVQEL 1605 -RVGN-Q-L 69 QRVGNTQIL 77	Collagen, type VI, alpha 3, isoform CRA_a	Rev protein of HIV1	AEO81800.1
1597 FRVGNVQEL 1605 -RVGN+ -E- 28 GRVGNITEE 36	Collagen, type VI, alpha 3, isoform CRA_a	Envelope glycoprotein of HIV1	ACX43882.1
294 SRSEVDMLK 302 -RS-VDML- 67 LRSHVDMLV 75	Annexin A2	E1 of HCV	AAV49494.1
294 SRSEVDMLK 302 -RS-VDML- 1407 ARSTVDMLV 1415	Annexin A2	Polyprotein of human enterovirus 68	AAR98503.1

Table 1 continued

Alignment	Human protein (top peptide)	Viral protein (lower peptide)	GenBank accession number of viral protein
295 SRSEVDMLK 302 -RSEVML- 721 IRSEVTMLL 729	Annexin A2	Tegument protein of HHV-3	ACL67875.1
117 GRNMYLTGL 125 -RN-YLTG- 32 ARNTYLTGE 40	Integrin alpha subunit precursor	E2 of HCV	AAX36001.1
117 GRNMYLTGL 125 GRN-YLT– 261 GRNQYLTAD 269	Integrin alpha subunit precursor	Polyprotein of human enterovirus C	BAH78284.1
13 GRSPVVLSL 21 -RSPV+ LS- 529 RRSPVILSS 537	Interleukin 17 receptor C, isoform CRA_c	Gag-Pro-Pol of HTLV-2	NP_041003.2
13 GRSPVVLSL 21 -R-PVVLS- 126 LRRPVVLST 134	Interleukin 17 receptor C, isoform CRA_c	Phosphotransferase of HHV-5	AEF33987.1
13 GRSPVVLSL 21 GR-PVVLS- 4 GRRPVVLTQ 12	Interleukin 17 receptor C, isoform CRA_c	Envelope glycoprotein of HHV-1	AAT41237.1
1 MRMSVGLSL 9 -RM-VGLS- 515 IRMGVGLSP 523	Dystroglycan 1 (dystrophin-associated glycoprotein 1	Polymerase of HBV	BAF63221.1
21 IQRTPKIQ 28 IQ–P-IQ 322 IQKNPTIQ 329	β2-Microglobulin	Attachment glycoprotein of Avian metapneumovirus	AFG33177.1
21 IQRTPKIQ 28 IQ-TP-I – 64 IQVTPNIA 71	β2-Microglobulin	Polyprotein of HCV	ABI23180.1
545 DQLQEQLQR 553 -Q-QEQL+ R 186 LQAQEQLER 194	CTCL antigen HD-CL-01/L14–2	RNA-dependent RNA polymerase of HCV	ABV24726.1
545 DQLQEQLQR 553 DQLQ-LQ- 176 DQLQRTLQD 184	CTCL antigen HD-CL-01/L14–2	L4 33K of Human adenovirus 61	AEK79926.1
545 DQLQEQLQR 553 +QL+ EQL-R 89 EQLHEQLDR 97	CTCL antigen HD-CL-01/L14–2	UL47 of Human herpesvirus 5	ABV71577.1
545 DQLQEQLQR 553 +QLQE-L– 1200 NQLQELLNH 1208	CTCL antigen HD-CL-01/L14–2	U31, large tegument protein of Human herpesvirus 6	CAA58411.1
545 DQLQEQLQR 553 +QL-EQLQ- 24 Q QLXEQLQS 32	CTCL antigen HD-CL-01/L14–2	gag polyprotein, p17 region, partial of HIV-1	CCN26679.1
545 DQLQEQLQR 553 +Q -+ EQLQ- 68 EQSEEQLQK 76	CTCL antigen HD-CL-01/L14–2	replication protein of Human papillomavirus type 5	NP_041367.1
545 DQLQEQLQR 553 - QL-EQL-R 50 RQLREQLAR 58	CTCL antigen HD-CL-01/L14–2	TA5L of Vaccinia virus Tian Tan	AAF33994.1
545 DQLQEQLQR 553 - QL-EQL-R 50 RQLREQLAR 58	CTCL antigen HD-CL-01/L14–2	membrane associated core protein of Vaccinia virus	AAB96458.1

Table 1 continued

Alignment	Human protein (top peptide)	Viral protein (lower peptide)	GenBank accession number of viral protein
545 DQLQEQLQR 553 - QL-EQL-R 50 RQLREQLAR 58	CTCL antigen HD-CL-01/L14–2	39 kDa core protein of Variola virus	ABG43681.1
545 DQLQEQLQR 553 D-LQ-Q-QR 2026 DALQAQVQR 2034	CTCL antigen HD-CL-01/L14–2	RNA-dependent RNA polymerase of Reston ebolavirus	AGC02896.1

The HLA-B27-binding peptide is on top in all the alignments. The virus peptide is on bottom in all the alignments. Amino acids (aa) marked − are non-matching, and those marked + indicate a non-identical amino acid residue with similar characteristics. The rank of human protein indicates the name of human protein encoding the alignment peptide (top peptide). The rank of viral protein indicates the name of viral protein encoding the alignment peptide (lower peptide)

Results and discussion

Sequence identity between the peptides of human virus and HLA-B27-binding peptides derived from cartilage/bone-related proteins

HLA-B27-binding peptides, particularly cartilage/bone-related peptides, may serve as candidate arthritogenic peptides, which may represent mimetic peptides involved either in initiating disease initiation or in the subsequent spread of the epitope. After BLASTP analysis of these 31 peptides to all query viral sequences (Supplementary 1), we identified some peptides (Table 1) highly similar to cartilage-related peptides (QRSKSPL, GRHANTKVGL, VRNPSVVVK, FRVGNVQEL, SRSEVDMLK, GRN MYLTGL, GRSPVVLSL, MRMSVGLSL, IQRTPKIQ and DQLQEQLQR) in this study. The mismatching amino acids of highly homologous peptides are <3. Then, we used SYFPEITHI and NetMHC 3.4 server to analyze the binding affinity of these viral peptides with HLA-B*2705. However, peptide RQPQVSI is 7-mer, it cannot be analyzed by these two bioinformatics tools. Results are listed in Table 2. We suppose that those peptides identified by both SYFPEITHI and NetMHC 3.4 server were more plausible HLA-B*2705 epitopes.

Peptide GRAANSLVGL present in polyprotein of HCV is highly similar to peptide GRHANTKVGL derived from human collagen, type VI, alpha 3, isoform CRA_a with only 3 aa mismatching. Further epitope prediction by SYFPEITHI found that the score of GRHANTKVGL is 25. It is a weak-binding peptide of HLA-B*2705 identified by NetMHC 3.4 server, as IC50 value of GRAANSLVGL is 76 nM that below 500 nM. Two peptides RAANSLVGL and GRAANSLVG derived from GRAANSLVGL were identified by SYFPEITHI. However, IC50 values of these two peptides were both above 500 nm when analyzed by NetMHC 3.4 server, which means they were unlikely to be bind by HLA-B*2705.

We also show here that peptide FRVGASQKL derived from envelope protein 2 of HCV is highly similar to FRVGNVQEL derived from collagen, type VI, alpha 3, isoform CRA_a with only 3 aa mismatching. SYFPEITHI score of FRVGASQKL is 26 and IC50 value is 110. VRNPQVLVK present in rev protein of HIV-1 is highly similar to peptide VRNPSVVVK derived from human collagen, type VI, alpha 3, isoform CRA_a with only 2 aa mismatching. These HLA-B*2705 potential epitopes of HCV and HIV may be presented to cytotoxic T lymphocytes and trigger AS or play a significant role in its progression. However, some studies (Neumann-Haefelin 2011; Mathieu et al. 2009) found that HLA-B27 has a positive effect in HIV and HCV infection. HLA-B27-positive patients have low HIV viral loads, CD4+ T cell counts decline slowly and AIDS progresses slowly (Neumann-Haefelin 2011). In acute HCV infection, HLA-B27 is associated with a very high rate of spontaneous viral clearance (Mathieu et al. 2009). This is likely to be on the other side of the double-edged sword of HLA-B27, as the existing immunodominant 'self' peptide epitopes are highly similar to viral immunodominant epitopes to CD8+ T cells, although the mechanism might be more complex and might involve innate immunity.

The high prevalence of HBsAg in patients with AS might be associated with a high frequency of the HLA-B27 gene (Zheng et al. 2012; Tanasescu et al. 1999). In this study, we found FRVVAVQNL present in large S protein of HBV is highly similar to peptide FRVGNVQEL derived from human collagen, type VI, alpha 3, isoform CRA_a with 3 aa mismatching. SYFPEITHI score of FRVVAVQNL is 24 and IC50 value is 164. It might be a HLA-B*2705 epitope that associated with the high prevalence of HBsAg in patients with AS. The immunogenicity and prevalence of FRVVAVQNL in AS patients are suggested to be evaluated in the future.

RRVMNVHEA present in envelope glycoprotein K of HHV-2 is highly similar to peptide FRVGNVQEL derived

Table 2 HLA-B27-binding prediction of identified viruses' peptides highly similar to those of HLA-B27-binding peptides related to cartilage

Virus peptides	mer	SYFPEITHI score	NetMHC affinity IC50 values (nM)
GRAANSLVGL	10	25	76
RAANSLVGL	8	17	–
GRAANSLVG	8	16	–
VRNESVIVR	9	25	–
ERNPAVVEK	9	24	–
VRNPQVLVK	9	24	451
VRTPSLVIN	9	14	–
FRVVAVQNL	9	24	164
FRVGASQKL	9	26	110
RVGASQKL	8	–	247
RRVMNVHEA	9	17	159
RVMNVHEA	8	–	449
ERVGNMDEM	9	22	–
QRVGNTQIL	9	23	–
GRVGNITEE	9	22	–
LRSHVDMLV	9	13	–
ARSTVDMLV	9	14	–
IRSEVTMLL	9	24	–
ARNTYLTGE	9	14	–
GRNQYLTAD	9	18	–
RRSPVILSS	9	21	358
LRRPVVLST	9	15	–
GRRPVVLTQ	9	18	–
IRMGVGLSP	9	16	–
IQKNPTIQ	8	–	–
IQVTPNIA	8	–	–
LQAQEQLER	9	17	–
DQLQRTLQD	9	6	–
EQLHEQLDR	9	16	–
NQLQELLNH	9	17	–
QQLXEQLQS	9	7	–
EQSEEQLQK	9	16	–
RQLREQLAR	9	20	153
DALQAQVQR	9	18	–

SYFPEITHI score marked "−" indicate no results, and NetMHC affinity marked "−" indicate IC50 values over 500 nM. For ANNs prediction high-binding peptides have an IC50 value below 50 nM, and weak-binding peptides an IC50 values below 500 nM

from human collagen, type VI, alpha 3, isoform CRA_a with 2 aa mismatching and 1 similar characteristic aa. SYFPEITHI score of FRVVAVQNL is 17 and IC50 value is 159. RRSPVILSS present in Gag-Pro-Pol of HTLV-2 is highly similar to peptide GRSPVVLS derived from interleukin 17 receptor C, isoform CRA_c with 2 aa mismatching and 1 similar characteristic aa. SYFPEITHI score of FRVVAVQNL is 21 and IC50 value is 358. HHV-2 and HTLV-2 might be not associated with AS, as no related report can be found so far. Nevertheless, these two potential epitopes also require consideration when patients with AS are HLA-B27 positive infected with these virus.

Vaccinia virus is a large, complex, enveloped virus belonging to the poxvirus family and is well known for its role as a vaccine (its namesake) that eradicated the smallpox disease caused by the infection variola virus. With the eradication of smallpox, routine vaccination with vaccinia virus has ceased. Vaccinia virus including Tian Tan was modified as a vector for immunization against other viruses in recently (Mazzon et al. 2013; Liu et al. 2013). Interestingly, RQLREQLAR present in core protein of Variola virus, TA5L of Vaccinia virus Tian Tan and membrane-associated core protein of Vaccinia virus is highly similar to peptide DQLQEQLQR present in CTCL antigen HD-CL-01/L14–2 with 3 aa mismatching. SYFPEITHI score of FRVVAVQNL is 20 and IC50 value is 153. Experiments to confirm the immunogenicity and affinity of RQLREQLAR with HLA-B27 are suggested for the use of vaccinia virus in future research.

Sequence identity between the peptides of human virus that cause chronic infections and all identified HLA-B27 peptides

We blasted the peptides of human virus that cause chronic infections with all HLA-B27-binding peptides. Results are shown in Table 3. The identity peptides identified using SYFPEITHI and NetMHC 3.4 Server are listed in Table 4.

Here, we also found four peptides of HCV highly similar to HLA-B27-binding human 'self' peptides. They are potential HLA-B27-binding peptides analyzed by SYFPEITHI and NetMHC 3.4 Server (Table 4). All identified similarity peptides are within 2 aa mismatching (Table 3).

HHV-4 also called Epstein–Barr virus (EBV) is a virus of the herpes family, and is one of the most common viruses in humans. There is evidence that infection with the virus is associated with a higher risk of certain autoimmune diseases, especially systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and primary Sjögren's syndrome (pSS) (Toussirot and Roudier 2008). SLE patients have a dysregulated immune response against EBV. EBV antigens exhibit structural molecular mimicry with common SLE antigens and functional molecular mimicry with critical immune-regulatory components (James and Robertson 2012). Fiorillo et al. (2005) found that the viral peptide pLMP2 (RRRWRRLTV, derived from latent membrane protein 2 (residues 236–244) of Epstein–Barr virus, is presented by the HLA-B27 molecules. In this study, we show here that the viral peptide ARTLLAALF derived from BRRF2 of EB virus is highly similar to peptide ARTLLAILRL present in p85 Mcm protein with 3 aa

Table 3 HLA-B27-binding peptides highly similar to those of human viruses that cause chronic infections

Alignment	Human protein (top peptide)	Viral protein (lower peptide)	GenBank accession number of viral protein
20 IRAAPPPLF 28 +RAAPPP– 88 VRAAPPPAS 96	Cathepsin A isoform a precursor	large S protein of HBV	ADJ51372.1
608 FRGPLVINR 616 -RGPLVI-R 793 SRGPLVISR 801	KIAA0738 protein	polymerase of HBV	ABW03087.1
776 GRHSTPLHL 784 GRHSTPL- 80 GRHSTPLSP 88	Tankyrase-related protein	preS protein of HBV	AAN10078.1
100 GRLGSTVFV 108 GRLGSTV- 32 GRLGSTVPS 40	Heterogeneous nuclear ribonucleoprotein M isoform a variant	X protein of HBV	AAB47856.2
12 ARVTKVLGR 20 ARVTK+ L- 1942 ARVTKILSS 1950	Ribosomal protein S28	JK1-full length of HCV	CAA43793.1
47 GRASYGVSK 55 GRASYG+ - 31 GRASYGITS 39	Heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A), isoform CRA_b	polyprotein of HCV	ABM89846.1
21 GRSSVILLTY 30 GR-SVILLT- 60 GRDSVILLTC 69	Tumor suppressing STF cDNA 5	polyprotein of HCV	ACX44149.1
1267 KREPGTKTK 1275 -R+ PGT-TK 248 IRQPGTLTK 256	DNA topoisomerase II, alpha isozyme	polyprotein of HCV	AAC42183.1
106 KRGEPAIYR 114 -RGEP-IYR 550 GRGEPGIYR 558	Double-stranded RNA-binding protein Staufen homolog 2	polyprotein of HCV subtype 1a	ADC54619.1
154 GRLLVATTF 163 GRL-ATT+ 784 GRLVPATTY 792	Sorting and assembly machinery component 50 homolog	polyprotein of HCV	ABQ85561.1
3 RRALLLLLL 11 -R-LLLLLL 777 MRPLLLLLL 785	Prolylcarboxypeptidase isoform 1 preproprotein	polyprotein of HCV	AFQ52335.1
1960 GRFAGLTSV 1968 -RFAGL-SV 13 ARFAGLFSV 21	Cadherin-related tumor suppressor homolog precursor (protein fat homolog)	Envelope glycoprotein 2 of HCV	ABP93541.1
18 RRLLLLPLLL 27 -RLLLL-LLL 343 PRLLLLSLLL 352	Thioredoxin peroxidase	NS5 of HCV	BAA06741.1
233 HRTELETKL 241 HR-ELET- 69 HRDELETRV 77	Keratin 80	BORF2 of HHV-4	AFY97927.1

Table 3 continued

Alignment	Human protein (top peptide)	Viral protein (lower peptide)	GenBank accession number of viral protein
501 ARTLLAILRL 510 ARTLLA-L- 209 ARTLLAALF 216	p85 Mcm protein	BRRF2 of HHV-4	AFY97943.1
106 QRVSIFFDY 114 -R-SIFFDY 245 PRYSIFFDY 253	Vacuolar proton pump subunit SFD alpha isoform	EBNA-3A of HHV-4	AFY97915.1
18 RRLLLLPLL 26 RR-LLLPL- 343 PRLLLLSLLL 352	Thioredoxin peroxidase	UL150 of HHV-5	AAX63472.1
190 GRIGVITN 197 GR+ GV+ T+ 156 GRVGVVTD 163	Ribosomal protein S4, X-linked, isoform CRA_a	Membrane glycoprotein RL11 of HHV-5	AAR31236.1
320 KRLPDGLTR 328 -RLPDG+ T+ 19 IRLPDGITK 27	Mini-chromosome maintenance deficient protein 5 variant	Membrane glycoprotein US2 of HHV-5	ADV04458.1
19 ARFKERVGY 27 -RF-KERVGY 3 LRFPERVGY 11	KIAA1143	Protein US33A of HHV-5	AFP95714.1
213 ARDEEQRLR 221 -RDE+ -RLR 818 RRDEQTRLR 826	tripartite motif-containing 65	large tegument protein of HHV-5	ACZ80299.1
328 ARIGQTGTK 336 -RIG+ TGT- 334 SRIGETGTG 342	Cell division cycle 27, isoform CRA_b	L2 protein of HPV	AEX31169.1
54 GRTETTVTR 64 -RTE-TVTR 499 PRTEQTVTR 507	Heterogeneous nuclear ribonucleoprotein H3 isoform a	L1 protein of HPV	AEX31122.1
322 VRWVGGPEI 330 VR-VGGP++ 32 VRRVGGPDV 40	Chaperonin containing TCP1, subunit 5 (epsilon)	E6*I of HPV	CAA45428.1
65 SRVNVSVGY 73 SR-NVSV-Y 143 SRDNVSVDY 151	Bardet-Biedl syndrome 5	L1 protein of HPV	AAF14010.1
63 ARYSLRDEF 71 ARYSLRD-F 87 ARYSLRDHF 95	Thymidylate synthase	thymidylate synthase of HHV-8	P90463.1
16 GRGLLALLL 24 -RGLL-LLL 13 FRGLLSLLL 21	Discoidin, CUB and LCCL domain containing 1	K1 glycoprotein of HHV-8	AAT44997.1
638 RRKQEEEAR 446 RRKQEE+ -R 126 RRKQEEQER 134	Inner centromere protein antigens 135/155 kDa isoform 1	bZIP factor of human T-lymphotropic virus 1	ABB89742.1
185 RQPQVSI 191 RQ-+ VS- 22 RQLHVSL 28	Hypothetical protein	polymerase of Hepatitis B virus	AAF80682.1

Table 3 continued

Alignment	Human protein (top peptide)	Viral protein (lower peptide)	GenBank accession number of viral protein
185 RQPQVSI 191 -QP+ VS- 2343 GQPEVSS 2359	Hypothetical protein	Polyprotein of HCV subtype 4L	AFN53794.1
185 RQPQVSI 191 -+ PQVS- 331 HEPQVSG 337	Hypothetical protein	Trans-activating transcriptional regulatory protein of HTLV-1	P0C213.1 TAX_HTL1F
185 RQPQVSI 191 RQPQ+- 96 RQPQIA A 101	Hypothetical protein	tegument protein UL25 of HHV-5	AGL96626.1
185 RQPQVSI 191 -QPQVS- 245 TQPQVSG 251	Hypothetical protein	HCRF2 of HHV-6	BAA01907.1
185 RQPQVSI 191 -QP+ VS- 203 AQPHVSH 209	Hypothetical protein	ORF17.5 of HHV-8	YP_001129370.1
185 RQPQVSI 191 RQP-VSI 4 RQPLVSI 10	Hypothetical protein	protease of HIV1	ACT33801.1

The HLA-B27-binding peptide is on top in all the alignments. The virus peptide is on bottom in all the alignments. Amino acids (aa) marked − are non-matching, and those marked + indicate a non-identical amino acid residue with similar characteristics. The rank of human protein indicates the name of human protein encoding the alignment peptide (top peptide). The rank of viral protein indicates the name of viral protein encoding the alignment peptide (lower peptide)

mismatching. SYFPEITHI score of FRVGASQKL is 25 and IC50 value is 72.

We found in our earlier study that infection with HHV-8, also known as Kaposi's sarcoma (KS)-associated herpes virus, was more prevalent in patients with SLE than in healthy controls (Sun et al. 2011), suggesting that the prevalence of HHV-8 infection might be associated with the development of SLE. The prevalence of HHV-8 infection is higher in patients with SLE than in healthy controls, and HHV-8 uses a mimicry strategy that involves the expression of a Bcl-2 homolog to protect virus-infected cells from undergoing apoptosis (Mesri et al. 2010). In the current study, we identified two peptides sequence (ARYSLRDHF derived from thymidylate synthase and FRGLLSLLL derived from K1 glycoprotein) of HHV-8 that are nearly identical to two human 'self' peptides (ARYSLRDEF, derived from thymidylate synthase and GRGLLALLL, derived from discoidin) respectively. There is only 1 aa mismatching between ARYSLRDHF and ARYSLRDEF. Moreover, ARYSLRDHF is the only potential strong-binding peptide of HLA-B27 analyzed by NetMHC 3.4 in this study, as IC50 value is 45 nM. We hypothesize, therefore, that infection with HHV-8 may be associated with AS and that the identities between peptide sequences of HLA-B27 and HHV-8 may be involved in the pathogenesis of AS.

HHV-5, also known as Human cytomegalovirus (HCMV), the largest human herpesvirus, infects a majority of the world's population (Slobedman et al. 2010). After infection, HCMV can remain latent within the body for long periods. Eventually, HCMV may cause mucoepidermoid carcinoma and possibly other malignancies (Melnick et al. 2011). Three peptide sequences (PRLLLLSLLL, IRLPDGITK, and LRFPERVGY) of HCMV that matched those of HLA-B27-binding peptides were identified. PRLLLLSLLL is derived from the thymidylate synthase; IRLPDGITK is derived from membrane glycoprotein US2 which can bind newly synthesized MHC class I heavy chains (HCs) and support their dislocation into the cytosol for subsequent degradation by proteasomes (Barel et al. 2006; Oresic and Tortorella 2008); and LRFPERVGY is derived from protein US33A. HCMV encode multiple proteins that share sequence similarities with cytokines, chemokines, and their receptors. These proteins provide the virus with an arsenal of functions to counteract the host immune response, given the central role that such molecules play in controlling immune function (McSharry et al. 2012). These findings suggest the possibility that the highly similar peptides between HCMV and human play roles in AS during HCMV infection. Future research is necessary to clarify whether the identified peptides play a role in AS during HCMV infection.

Table 4 HLA-B27-binding prediction of identified viruses' peptides highly similar to those of human viruses that cause chronic infections

peptides	mer	SYFPEITHI score	NetMHC affinity IC50 values (nM)
VRAAPPPAS	9	12	–
SRGPLVISR	9	25	–
GRHSTPLSP	9	15	–
GRLGSTVPS	9	19	–
ARVTKILSS	9	18	–
GRASYGITS	9	15	–
GRDSVILLTC	10	16	–
GRDSVILLT	9	18	–
RDSVILLTC	9	8	–
IRQPGTLTK	9	26	–
GRGEPGIYR	9	26	–
GRLVPATTY	9	28	129
MRPLLLLLL	9	24	235
ARFAGLFSV	9	18	74
PRLLLLSLLL	10	25	166
PRLLLLSLL	9	26	386
RLLLLSLLL	9	20	–
HRDELETRV	9	16	–
ARTLLAALF	9	25	72
ARTLLAAL	8	–	126
PRYSIFFDY	9	21	–
PRLLLLSLLL	10	25	166
PRLLLLSLL	9	26	386
RLLLLSLLL	9	20	–
GRVGVVTD	8	–	–
IRLPDGITK	9	26	170
LRFPERVGY	9	24	146
RRDEQTRLR	9	25	–
RRDEQTRL	8	–	296
SRIGETGTG	9	16	–
PRTEQTVTR	9	23	–
VRRVGGPDV	9	13	–
RRVGGPDV	8	–	328
SRDNVSVDY	9	21	–
ARYSLRDHF	9	24	45
FRGLLSLLL	9	24	59
FRGLLSLL	8	–	120
RRKQEEQER	9	28	–

SYFPEITHI score marked "−" indicate no results, and NetMHC affinity marked "−" indicate IC50 values over 500 nM. For ANNs prediction high-binding peptides have an IC50 value below 50 nM, and weak-binding peptides an IC50 values below 500 nM

MHCl molecules preferably bind peptides comprising 8–9 amino acid residues (Bjorkman et al. 1987). However, there is evidence that other types of peptides bind to MHC1 molecules, including glycopeptides, short peptides (4–5 mers), longer peptides (up to 18 mers) and retro-inverso peptides with increased stability, all of which can induce strong responses by cytotoxic T lymphocytes (Apostolopoulos and Lazoura 2004). How these short peptides bind to MHC and induce CTL is not clear. In the present study, we identified many short (4–5 mers), HLA-B27-binding peptides related to those of human viruses (data not shown). The extent of our study is too limited to explain whether human virus infection and the presence of these peptides play roles in the inflammatory responses and pathogenesis in HLA-B27-positive AS patients. This is an exciting area for future study. Further, other factors and mechanisms must be considered along with the arthritogenic peptide model of disease.

Conflict of interest The authors declare that they have no conflict of interest.

References

Allen RL, Bowness P, McMichael A (1999) The role of HLA-B27 in spondyloarthritis. Immunogenetics 50:220–227

Apostolopoulos V, Lazoura E (2004) Non-canonical peptides in complex with MHC class I. Expert Rev Vaccines 3:151–162

Barel MT, Pizzato N, Le Bouteiller P, Wiertz EJ, Lenfant F (2006) Subtle sequence variation among MHC class I locus products greatly influences sensitivity to HCMV US2- and US11-mediated degradation. Int Immunol 18:173–182

Ben Dror L, Barnea E, Beer I, Mann M, Admon A (2010) The HLA-B*2705 proteome. Arthr Rheum 62:420–429

Benjamin R, Parham P (1990) Guilt by association: HLA-B27 and ankylosing spondylitis. Immunol Today 11:137–142

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329:512–518

Brewerton DA, Hart FD, Nicholls A, Caffrey M, James DC, Sturrock RD (1973) Ankylosing spondylitis and HL-A27. Lancet 1:904–907

Calin A, Marder A, Becks E, Burns T (1983) Genetic differences between B27 positive patients with ankylosing spondylitis and B27 positive healthy controls. Arthr Rheum 26:1460–1464

Ewing C, Ebringer R, Tribbick G, Geysen HM (1990) Antibody activity in ankylosing spondylitis sera to two sites on HLA B27.1 at the MHC groove region (within sequence 65–85), and to a *Klebsiella pneumoniae* nitrogenase reductase peptide (within sequence 181–199). J Exp Med 171:1635–1647

Fielder M, Pirt SJ, Tarpey I, Wilson C, Cunningham P, Ettelaie C, Binder A, Bansal S, Ebringer A (1995) Molecular mimicry and ankylosing spondylitis: possible role of a novel sequence in pullulanase of *Klebsiella pneumoniae*. FEBS Lett 369:243–248

Fiorillo MT, Rückert C, Hülsmeyer M, Sorrentino R, Saenger W, Ziegler A, Uchanska-Ziegler B (2005) Allele-dependent similarity between viral and self-peptide presentation by HLA-B27 subtypes. J Biol Chem 280:2962–2971

Gaston H (2006) Mechanisms of disease: the immunopathogenesis of spondyloarthropathies. Nat Clin Pract Rheumatol 2:383–392

Gomard E, Sitbon M, Toubert A, Begue B, Levy JP (1984) HLA-B27, a dominant restricting element in antiviral responses? Immunogenetics 20:197–204

Infantes S, Lorente E, Barnea E, Beer I, Barriga A, Lasala F, Jiménez M, Admon A, López D (2013) Natural HLA-B*2705 protein ligands with glutamine as anchor motif: implications for HLA-B27 association with spondyloarthropathy. J Biol Chem 288:10882–10889

James JA, Robertson JM (2012) Lupus and Epstein–Barr. Curr Opin Rheumatol 24:383–388

Lahesmaa R, Skurnik M, Vaara M, Leirisalo-Repo M, Nissila M, Granfors K, Toivanen P (1991) Molecular mimicry between HLA B27 and Yersinia, Salmonella, Shigella and Klebsiella within the same region of HLA alpha 1-helix. Clin Exp Immunol 86:399–404

Liu Z, Wang S, Zhang Q, Tian M, Hou J, Wang R, Liu C, Ji X, Liu Y, Shao Y (2013) Deletion of C7L and K1L genes leads to significantly decreased virulence of recombinant vaccinia virus Tian Tan. PLoS One 8:e68115

Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M (2008a) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res 1:W509–W512

Lundegaard C, Lund O, Nielsen M (2008b) Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9 mers. Bioinformatics 24:1397–1398

Mathieu A, Paladini F, Vacca A, Cauli A, Fiorillo MT, Sorrentino R (2009) The interplay between the geographic distribution of HLA-B27 alleles and their role in infectious and autoimmune diseases: a unifying hypothesis. Autoimmun Rev 8:420–425

Mazzon M, Peters NE, Loenarz C, Krysztofinska EM, Ember SW, Ferguson BJ, Smith GL (2013) A mechanism for induction of a hypoxic response by vaccinia virus. Proc Natl Acad Sci USA 110:12444–12449

McSharry BP, Avdic S, Slobedman B (2012) Human cytomegalovirus encoded homologs of cytokines, chemokines and their receptors: roles in immunomodulation. Viruses 4:2448–2470

Melnick M, Sedghizadeh PP, Allen CM, Jaskoll T (2011) Human cytomegalovirus and mucoepidermoid carcinoma of salivary glands: cell-specific localization of active viral and oncogenic signalling proteins is confirmatory of a causal relationship. Exp Mol Pathol 92:118–125

Mesri EA, Cesarman E, Boshoff C (2010) Kaposi's sarcoma and its associated herpesvirus. Nat Rev Cancer 10:707–719

Neumann-Haefelin C (2011) Protective role of HLA-B27 in HIV and hepatitis C virus infection. Dtsch Med Wochenschr 136:320–324

Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S, Lund O (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci 12:1007–1017

Oresic K, Tortorella D (2008) Endoplasmic reticulum chaperones participate in human cytomegalovirus US2-mediated degradation of class I major histocompatibility complex molecules. J Gen Virol 89:1122–1130

Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanović S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50:213–219

Reveille JD (1998) HLA-B27 and the seronegative spondyloarthropathies. Am J Med Sci 316:239–249

Reveille JD (2012) Genetics of spondyloarthritis—beyond the MHC. Nat Rev Rheumatol 8:296–304

Slobedman B, Cao JZ, Avdic S, Webster B, McAllery S, Cheung AK, Tan JC, Abendroth A (2010) Human cytomegalovirus latent infection and associated viral gene expression. Future Microbiol 5:883–900

Sun Y, Sun S, Li W, Li B, Li J (2011) Prevalence of human herpesvirus 8 infection in systemic lupus erythematosus. Virol J 8:210

Tanasescu C, Parvu M, Antohi I, Lazar S (1999) The significance of chronic hepatitis B and C virus infections in some connective tissue diseases: the association with chronic liver disease. Rom J Intern Med 37:53–64

Thomas GP, Brown MA (2010) Genetics and genomics of ankylosing spondylitis. Immunol Rev 233:162–180

Toussirot E, Roudier J (2008) Epstein–Barr virus in autoimmune diseases. Best Pract Res Clin Rheumatol 22:883–896

Tsuchiya N, Husby G, Williams RC (1989) Studies of humoral and cell-mediated immunity to peptides shared by HLA-27.1 and *Klebsiella pneumoniae* nitrogenase in ankylosing spondylitis. Clin Exp Immunol 76:354–360

Zheng B, Li T, Lin Q, Huang Z, Wang M, Deng W, Liao Z, Gu J (2012) Prevalence of hepatitis B surface antigen in patients with ankylosing spondylitis and its association with HLA-B27: a retrospective study from south China. Rheumatol Int 32:2011–2016