Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural genomics of the Epstein–Barr virus

Nicolas Tarbouriech,[a,b] Marlyse Buisson,[b,c] Thibault Géoui,[a,b] Susan Daenke,[d] Stephen Cusack[a] and Wim Pascal Burmeister[b,e]*

[a]EMBL-Grenoble Outstation, BP 181, F-38042 Grenoble CEDEX 9, France, [b]Institut de Virologie Moléculaire et Structurale, FRE 2854 CNRS–UJF, BP 181, F-38042 Grenoble CEDEX 9, France, [c]Laboratoire de Virologie, CHU Michallon, BP 217, F-38043 Grenoble CEDEX 9, France, [d]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, England, and [e]Institut Universitaire de France, 103 Boulevard Saint-Michel, F-75005 Paris, France

Correspondence e-mail: wpb@embl-grenoble.fr

Received 15 November 2005
Accepted 1 August 2006

Epstein–Barr virus is a herpesvirus that causes infectious mononucleosis, carcinomas and immunoproliferative disease. Its genome encodes 86 proteins, which provided targets for a structural genomics project. After updating the annotation of the genome, 23 open reading frames were chosen for expression in *Escherichia coli*, initially selecting for those with known enzyme activity and then supplementing this set based on a series of predicted properties, in particular secondary structure. The major obstacle turned out to be poor expression and low solubility. Surprisingly, this could not be overcome by modifications of the constructs, changes of expression temperature or strain or renaturation. Of the eight soluble proteins, five were crystallized using robotic nanolitre-drop crystallization trials, which led to four solved structures. Although these results depended on individual treatment rather than standardized protocols, a high-throughput miniaturized crystallization screening protocol was a key component of success with these difficult proteins.

1. Introduction

Human herpesviruses comprise three subfamilies: (i) α-herpesviruses [herpes simplex viruses (HSV) 1 and 2 and varicella zoster virus (VZV)], (ii) β-herpesviruses [cytomegalovirus (CMV) and human herpesvirus (HHV) 6 and 7] and (iii) γ-herpesviruses, comprising the Kaposi's sarcoma-associated herpesvirus (KSHV or HHV8) and Epstein–Barr virus (EBV or HHV4). The last infects the vast majority of the world's human population, establishing and maintaining a lifelong persistence in the infected host.

Primary infection typically occurs in childhood and is frequently asymptomatic. In contrast, a delayed primary infection in adolescents or young adults results in infectious mononucleosis (IM) in approximately half of cases, with symptoms including fever, pharyngitis, lymphadenopathy and splenomegaly. IM is a self-limiting lymphoproliferative disorder characterized by an expansion of EBV-infected B-lymphocytes associated with viral lytic replication in the oropharynx, controlled by a vigorous $CD8^+$ cytotoxic T-cell immune response. The majority of cases of acute IM recover, but serious complications can occasionally lead to death. EBV is associated with a number of cancers in the immunocompetent host (Rickinson & Kieff, 1996), in particular Burkitt's lymphoma and nasopharyngeal carcinoma, which are endemic in African and Asian populations (Raab-Traub, 2005). Furthermore, EBV can lead to immunoproliferative disease in immunosuppressed patients, notably those infected with HIV (Rickinson & Kieff, 1996). Currently licensed anti-herpesvirus drugs (acyclovir and related compounds) directed against viral DNA synthesis (Coen & Schaffer, 2003) show little effect against EBV.

Table 1
Proteins of EBV.

Accession, SwissProt, TrEMBL or PIR (Protein Information Resource) accession number. NCBI, GI numbers assigned by NCBI. Function, information on name, synonyms and function of the protein. ORF EBV, name of the EBV open reading frame. F, classification based on the function into C, capsid; M, membrane (glyco)protein; N, nucleotide metabolism; L, latency; P, packaging; R, replication; S, transcription factors, transactivators, signalling; T, tegument. S, an × means translated from spliced messenger RNA. Occ., occurrence in herpesvirus subfamilies, no entry for proteins present only in EBV and very closely related viruses such as rhesus lymphocryptovirus (LCV). Homologue, name of the homologue in HSV, if existing, otherwise of human CMV. *N*, number of constructs used in the project. St, current status. Proteins with enzymatic activity are shown in bold. For space reasons, only a limited number of references to original work are given; otherwise, review articles are cited. The annotation extensively used the *BLAST* program at NCBI (Altschul *et al.*, 1997).

Accession	NCBI	Function	ORF EBV	F	S	Occ.	Homologue	Reference	*N*	St	Comment
P03229	gi:140202	Bcl-2 homologue, negative regulator of anti-apoptosis protein BHRF1	BALF1	S		γ		Marshall *et al.* (1999), Bellows *et al.* (2002), Cabras *et al.* (2005)	1	†	Toxic
P03227	gi:118744	Single-stranded DNA-binding protein, part of replication fork/machinery	BALF2	R		$\alpha\beta\gamma$	UL29	Decaussin *et al.* (1995), Robertson *et al.* (1996)			
P25939	gi:124087	**Terminase large subunit/ATPase**	BALF3	P		$\alpha\beta\gamma$	UL28	Alba (2002), Hwang & Bogner (2002), Savva *et al.* (2004)			
P03188	gi:138191	Membrane glycoprotein B (gB, gp110) precursor, fusion and co-receptor binding	BALF4	M		$\alpha\beta\gamma$	UL27	Gong *et al.* (1987), Spear & Longnecker (2003)			
P03198	gi:118858	**DNA polymerase**	BALF5	R		$\alpha\beta\gamma$	UL30	Kieff (1996)	12	‡	Insoluble
Q8AZJ4	gi:23893662	BARF0, nuclear/perinuclear in epithelial cells, start at amino acid 298 of SWISS-PROT, homologue in rhesus LCV	BARF0	U	×			Hitt *et al.* (1989), Gilligan *et al.* (1991), Fries *et al.* (1997)			
P03175	gi:132620	**Ribonucleoside reductase, small 38 kDa subunit**	BaRF1	N		$\alpha\gamma$	UL40	Kieff (1996)	1	†	No expression
P03228	gi:115128	CD80 homologue (p33), oncogene, soluble glycoprotein, 2 Ig domains, binds CSF-1	BARF1	S				Wei & Ooka (1989), Strockbine *et al.* (1998), Tarbouriech *et al.* (2006)	1	§	PDB 2ch8
P03216	gi:136813	Myristoylated phosphoprotein in tegument (MyrP)	BBLF1	T		γ		Johannsen *et al.* (2004)			
Not in DB	Not in DB	Primase-associated factor, spliced, full sequence not in DB, part of helicase–primase complex	BBLF2/3	R	×	$\alpha\beta\gamma$	UL8	Yokoyama *et al.* (1999)			
P03214	gi:122807	**Helicase, part of helicase–primase complex**	BBLF4	R		$\alpha\beta\gamma$	UL5	Yokoyama *et al.* (1999)	1	†	No expression
P03213	gi:136792	Portal protein UL6 homologue, function by homology	BBRF1	P		$\alpha\beta\gamma$	UL6	Newcomb *et al.* (2001)			
P29882	gi:267198	Unknown function	BBRF2	U		$\alpha\beta\gamma$	UL7	Johannsen *et al.* (2004)			
P03215	gi:140278	Glycoprotein M, part of gN–gM complex involved in envelope–tegument interaction	BBRF3	M		$\alpha\beta\gamma$		Lake & Hutt-Fletcher (2000), Johannsen *et al.* (2004)			
P03226	gi:137569	Major capsid protein, MCP, VP5, 155K	BcLF1	C		$\alpha\beta\gamma$		Kieff (1996)			
P03180	gi:114886	Viral interleukin-10 homologue precursor, vIL-10	BCRF1	S		$\beta\gamma$		Hsu *et al.* (1990)			PDB 1vlk, 1y6m
P25215	gi:136975	Unknown function	BcRF1	U		$\beta\gamma$	hcmvUL84	Alba (2002)			
Q8AZK7	gi:7441833	EBNA-Lp (EBNA-5) nuclear phosphoprotein, highly spliced, 12 exons, proline-rich, enhances EBNA-2 transactivation	BWRF1	L	×	γ		Kieff (1996), Waltzer *et al.* (1996)			
P25214	gi:139189	Minor capsid protein (mCP), triplex protein HSV-1 VP23 homologue	BDLF1	C		$\alpha\beta\gamma$	UL18	Trus *et al.* (2001)	1	¶	Purified
P03225	gi:136827	Tegument protein, transmembrane protein?	BDLF2	T		γ		Johannsen *et al.* (2004)	5	‡	Insoluble
P03224	gi:138369	Envelope glycoprotein gp150 (gp117)	BDLF3	M		γ		Borza & Hutt-Fletcher (1998), Johannsen *et al.* (2004)			
P03223	gi:136983	Protein (gp115)	BDLF4	U		$\beta\gamma$	hcmvUL92	Alba (2002), Boeckmann *et al.* (2003)			
Q66541	gi:1334910	Assemblin without protease domain coded by 2nd reading frame	BdRF1	C		$\alpha\beta\gamma$	UL26.5				
P03203	gi:119113	EBNA-3B nuclear protein (EBNA-4)	BERF2a/ BERF2b	L	×			Kieff (1996), Young & Rickinson (2004)			

Table 1 (continued)

Accession	NCBI	Function	ORF EBV	F	S	Occ.	Homologue	Reference	*N*	St	Comment
P03204	gi:119114	EBNA-3C nuclear protein (EBNA-6, EBNA-4B) (−) effect on transactivation, EBNA-2 and cell cycle, essential for immortalization	BERF3/ BERF4	L	×			Kieff (1996), Young & Rickinson (2004)			
P03184	gi:136878	Role in DNA packaging, cytosolic zinc-binding protein, cysteine-rich	BFLF1	P		$\alpha\beta\gamma$	UL32	Chang *et al.* (1996)			
P03183	gi:136873	Nuclear membrane phosphoprotein, part of intracellular virions, egress protein, complex with BFRF1	BFLF2	M		$\alpha\beta\gamma$	UL31	Lake & Hutt-Fletcher (2004), Gonnella *et al.* (2005)	2	‡	Insoluble
P03185	gi:136886	Nuclear membrane protein p38, transmembrane with large cytoplasm domain, complex with BFLF2	BFRF1	M		$\beta\gamma$	hcmvUL50	Lake & Hutt-Fletcher (2004)			
P14347	gi:140660	Unknown function, not included in virions	BFRF2	U		$\beta\gamma$	hcmvUL49	Bellows *et al.* (2002), Johannsen *et al.* (2004)			
P14348	gi:139195	Smallest capsid protein (sCP) on outer capsid surface, HSV1-VP26 homologue, KSHV ORF 65 homologue	BFRF3	C		$\alpha\beta\gamma$	UL35	Nealon *et al.* (2001), Bowman *et al.* (2003), Johannsen *et al.* (2004)			
AAA45868	gi:330334	DNA-cleavage and packaging protein, part of the DNA-packaging machinery (BFRF0.5, HS4BAM)	BFRF4	P		$\alpha\beta\gamma$	UL33	Alba (2002), Beard & Baines (2004)			
P03222	gi:136833	Tegument protein (gp115)	BGLF1	T		γ		Johannsen *et al.* (2004)	1	‡	Insoluble
P03221	gi:114952	Tegument protein, MyrPBP	BGLF2	T		$\alpha\beta\gamma$	UL16	Boeckmann *et al.* (2003), Johannsen *et al.* (2004)			
P03220	gi:136988	Homologue to HSV-1 tegument protein, but not included in virion, gp118	BGLF3	U		$\alpha\beta\gamma$	UL14	Johannsen *et al.* (2004)			
P13288	gi:125627	**Ser/Thr kinase, phosphorylation of nucleoside analogues, tegument protein**	BGLF4	N		$\alpha\beta\gamma$	UL13	Smith & Smith (1989), Marschall *et al.* (2002), Johannsen *et al.* (2004)	4	‡	Insoluble
P03217	gi:119691	**Alkaline exonuclease, involved together with BALF2 in DNA recombination**	BGLF5	R		$\alpha\beta\gamma$	UL12	Cheng *et al.* (1980), Reuven *et al.* (2004)			
P03219	gi:23893636	DNA-packaging protein, terminase small subunit	BGRF1/ BDRF1	P	×	$\alpha\beta\gamma$	UL15	Alba (2002)			
P03181	gi:140778	Proline-rich protein LF3, unknown function, tandem repeats, *Not*I repeat (125 bp), homologue in rhesus LCV	BHLF1	U				Laux *et al.* (1985), Rivailler *et al.* (2002), Farell (2005)			
P03182	gi:119102	Anti-apoptotic factor bcl-2 homologue, early antigen protein R (EA-R), nuclear antigen	BHRF1	S		γ		Huang *et al.* (2003)	2	††	PDB 1q59 (NMR)
P03208	gi:138777	G-protein coupled receptor (G-PCR), 7 transmembrane helices, 6 glycosylation sites, 2 disulfide bridges gP64	BILF1	M		γ		Hutt-Fletcher (2005), Paulsen *et al.* (2005)			
P03218	gi:138183	Membrane glycoprotein gp55/80, Ig-like, gp78	BILF2	M				Mackett *et al.* (1990), Boeckmann *et al.* (2003), Johannsen *et al.* (2004)			
P03211	gi:119110	Latent nuclear protein EBNA-1, assures EBV episome maintenance replication, Gly-rich domain, essential for immortalization	BKRF1	L				Bochkarev *et al.* (1996)			PDB 1b3t, 1vhi
P03212	gi:140976	Glycoprotein L precursor, gp25, in gL–gH complex involved in viral fusion together with gB	BKRF2	M		$\alpha\beta\gamma$	UL1	Hutt-Fletcher (2005), Spear & Longnecker (2003)			
P12888	gi:137034	**Uracil-DNA glycosylase**	BKRF3	N		$\alpha\beta\gamma$	UL2	Winters & Williams (1993)	10	§	Publication in preparation
P30117	gi:267499	Tegument phosphoprotein	BKRF4	T		γ		Johannsen *et al.* (2004)			
P03200	gi:138368	Envelope glycoprotein gp350 (gp340) initial cell binding through complement receptor 2 (CR2, CD21)	BLLF1	M		$\alpha\gamma$		Spear & Longnecker (2003), Hutt-Fletcher (2005)			

Table 1 (continued)

Accession	NCBI	Function	ORF EBV	F	S	Occ.	Homologue	Reference	*N*	St	Comment
AAA45880	gi:330361	Envelope glycoprotein gp220, obtained through in-frame splicing from BLLF1	BLLF1b	M	×	$\alpha\gamma$		Beisel *et al.* (1985)			
P03199	gi:140999	Unknown function	BLLF2	U		γ					
P03195	gi:118952	**dUTP pyrophosphatase, dUTPase**	BLLF3	N		$\alpha\gamma$	UL50	Sommer *et al.* (1996), Tarbouriech *et al.* (2005)	2	§	PDB 2bsy, 2bt1
P03196	gi:141001	Membrane glycoprotein gN, part of the gM–gN complex, part of the envelope–tegument interaction	BLRF1	M		$\beta\gamma$	hcmvUL73	Lake & Hutt-Fletcher (2004)			
P03197	gi:141002	Tegument protein	BLRF2	T		γ		Johannsen *et al.* (2004)		‡‡	Soluble
P12977	gi:119112	EBNA-3A nuclear protein (EBNA-3), (−) effect on transactivator EBNA-2 and cell cycle, essential for immortalization	BLRF3/ BERF1	L	×			Kieff (1996), Waltzer *et al.* (1996), Young & Rickinson (2004)	1	†	No expression
P03191	gi:119098	Processivity factor, sliding clamp, early antigen protein D (EA-D, polymerase accessory protein)	BMRF1	R		$\alpha\beta\gamma$	UL42	Tsurumi *et al.* (1993)			
P03192	gi:141066	Receptor for cellular integrins, needed for infection of epithelial cells, 10 TM helices, RGD motif	BMRF2	M		γ		Modrow *et al.* (1992), Tugizov *et al.* (2003)			
P03230	gi:126373	Latent membrane protein 1 (LMP-1), interferes with signalling, TRAF-binding through CTAR1 and 2, essential for immortalization	BNLF1	L	×	γ		Young & Rickinson (2004)			
B28918	gi:23893667	Potential membrane protein	BNLF2a	U				Rivailler *et al.* (2002), Farell (2005)			
A28918	gi:7460898	Potential gp141	BNLF2b	U				Rivailler *et al.* (2002), Farell (2005)			
P03179	gi:139165	Major tegument protein (MTP), viral surface protein involved in B lymphocyte binding	BNRF1	T		γ		Johannsen *et al.* (2004), Lopez *et al.* (2005)			
P03189	gi:136893	Large tegument protein-binding protein (LTPBP)	BOLF1	T		$\alpha\beta\gamma$	UL37	Johannsen *et al.* (2004)			
P03187	gi:139172	Minor capsid protein-binding protein (mCP-BP), Triplex protein HSV-1 VP19C homologue	BORF1	C		γ	UL38	Johannsen *et al.* (2004)			
P03190	gi:132602	**Ribonucleoside reductase, large 140 kDa subunit**	BORF2	N		$\alpha\beta\gamma$	UL39	Kieff (1996)	2	†	No expression
P03186	gi:135574	Large tegument protein (LTP)	BPLF1	T		$\alpha\beta\gamma$	UL36	Johannsen *et al.* (2004)			
P03209	gi:115130	Transcription activator, R transactivator, Rta, dimeric (TAF50)	BRLF1	S		γ		Hardwick *et al.* (1988), Gruffat *et al.* (1990)	3	‡	Insoluble
Not in DB	Not in DB	Spliced BRLF1-BZLF1 protein, repressor of BZLF1, RAZ	BRLF1/ BZLF1	S	×			Manet *et al.* (1989), Furnari *et al.* (1994), Segouffin *et al.* (1996)			
Not in DB	Not in DB	Hypothetical BRLF1-BZLF1 splice variant	BRLF1/ BZLF1b	S	×			Farell (2005)			
P03207	gi:141395	Enhancement of the induction of the lytic cycle	BRRF1	S		γ		Segouffin-Cariou *et al.* (2000), Hong *et al.* (2004)	1	†	No expression
P03210	gi:141396	Tegument protein, unknown function	BRRF2	T		γ		Johannsen *et al.* (2004)			
P03193	gi:136937	Primase, subunit of the helicase–primase complex, part of the replication machinery	BSLF1	R		$\alpha\beta\gamma$	UL52	Yokoyama *et al.* (1999)		†	No expression
Q04360	gi:1708410	mRNA-export factor (EB2, Mta,SM), mRNA splicing, interaction with human Spen proteins, IE63 (ICP27) homologue	BSLF2/ BMLF1	S	×	$\alpha\beta\gamma$	UL54	Boeckmann *et al.* (2003), Hiriart *et al.* (2005), Swaminathan (2005)	6	¶	In crystallization
P03194	gi:141432	Palmitoylated tegument protein (PalmP)	BSRF1	T		$\alpha\beta\gamma$	UL51	Johannsen *et al.* (2004)			
P30119	gi:267575	Capsid maturation, capsid-associated	BTRF1	C		$\alpha\gamma$	UL21	Wagenaar *et al.* (2001)			
P03233	gi:136861	Portal plug (EC-RF2), capsid-associated tegument protein, seals DNA inside capsid	BVRF1	C		$\alpha\beta\gamma$	UL25	Sheaffer *et al.* (2001), Johannsen *et al.* (2004)			

Table 1 (continued)

Accession	NCBI	Function	ORF EBV	F	S	Occ.	Homologue	Reference	*N*	St	Comment
P03234	gi:139231	**Protease/assemblin**	BVRF2	C		$\alpha\beta\gamma$	UL26	Buisson *et al.* (2002)	2	†	PDB 1o6e
P03177	gi:1170666	**Thymidine kinase**	BXLF1	N		$\alpha\gamma$	UL32	Littler *et al.* (1986)	3	‡	Insoluble
P03231	gi:138312	Glycoprotein gp85, gH, part of gHgLgp42, fusion	BXLF2	M		$\alpha\beta\gamma$	UL22	Spear & Longnecker (2003), Hutt-Fletcher (2005)			
P03232	gi:136852	Nucleoprotein	BXRF1	U		$\alpha\beta\gamma$	UL24	Pearson & Coen (2002)	2	‡	Insoluble
P12978	gi:119111	EBNA-2 nuclear protein, transcription factor, interacts with RBPJ-K, essential for immortalization	BYRF1	L	×			Henkel *et al.* (1994), Waltzer *et al.* (1994), Young & Rickinson (2004)			
P03206	gi:115196	Trans-activator ZEBRA, origin binding protein (EB1, Zta), bZip similar to CCAAT/enhancer binding protein α	BZLF1	S	×	γ		Chevallier-Greco *et al.* (1986), Rooney *et al.* (1988), Giot *et al.* (1991), Petosa *et al.* (2006)			PDB 2c9n, 2c9l
QQBE27	gi:73982	Function unknown, BZLF1 splice variant	BZLF1b	U	×			Countryman & Miller (1985), Farell (2005)			
P03205	gi:141578	gp42, MHC class II binding protein, part of gHgLgp42 complex	BZLF2	M		γ		Mullen *et al.* (2002), Ressing *et al.* (2003)			PDB 1kg0
P03235	gi:140616	EC-RF4 (ECRF4) protein	ECRF4	U				Rivailler *et al.* (2002), Ressing *et al.* (2003)			
Q8AZJ5	gi:23893655	Protein LF1, contains a dUTPase like domain, γ-herpes ORF10 family	LF1	U		γ		Rivailler *et al.* (2002), Davison & Stow (2005), Farell (2005)			
Q99306	gi:23893654	Protein LF2, contains a dUTPase like domain, γ-herpes ORF11 family	LF2	U		γ		Rivailler *et al.* (2002), Davison & Stow (2005), Farell (2005)			
Not in DB	Not in DB	RK-BARF0, interaction with notch	RK-BARF0	S	×			Fries *et al.* (1997), Kusano & Raab-Traub (2001)			
P13285	gi:126379	LMP-2A, interference with protein kinase signalling, gene terminal protein, essential for immortalization		L	×			Kieff (1996), Young & Rickinson (2004)			
Q8AZK9	gi:23893578	LMP-2B, negative regulator of LMP-2A		L	×			Kieff (1996), Young & Rickinson (2004)			

† Cloned. ‡ Expressed. § Structure solved. ¶ Purified protein. †† Crystals. ‡‡ Soluble protein.

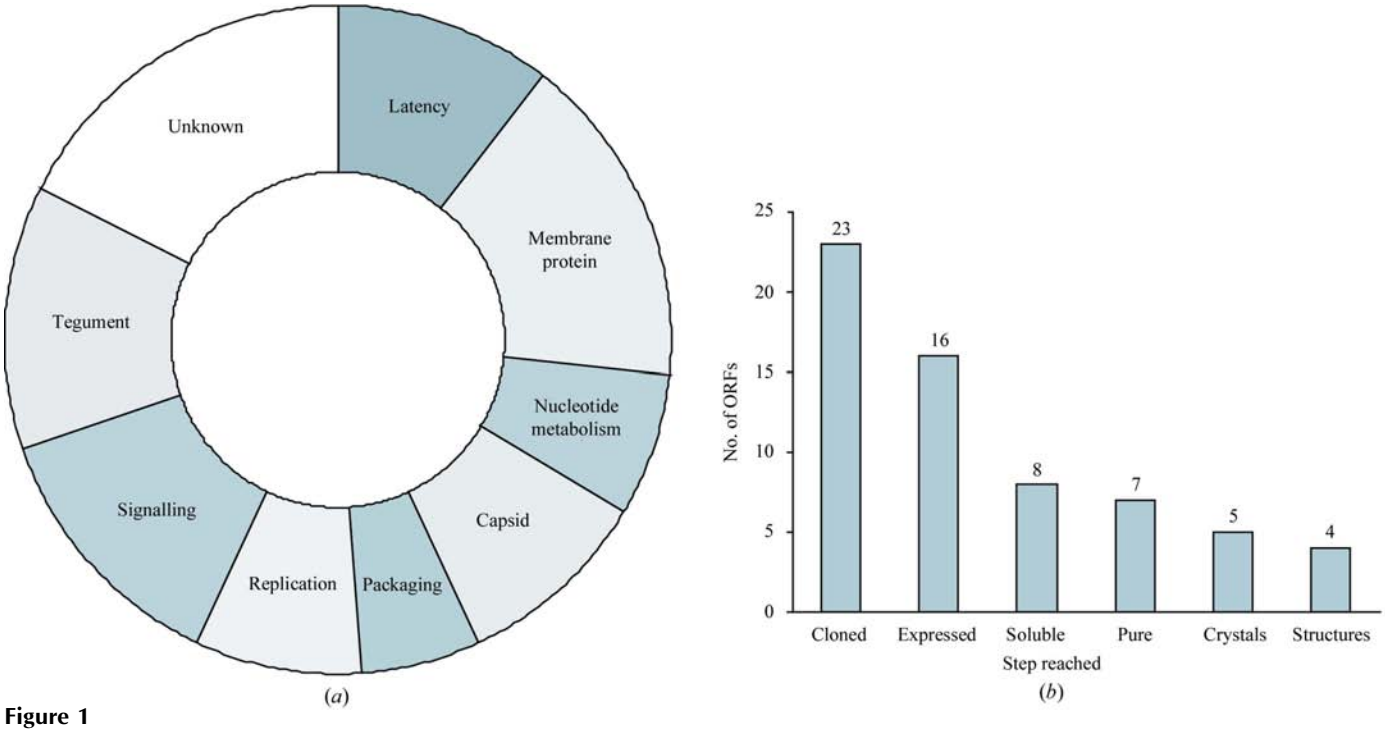

Figure 1
(*a*) Classification of the EBV proteins according to function. (*b*) Outcome for the proteins entering into the structural genomics project.

Table 2
Crystallographic results.

Protein reference	EBV ORF, PDB code	Construct†	Inhibitor	Protein buffer	Precipitant	Space group	Unit-cell parameters (Å)	Resolution (Å), method‡	Structure
Protease§	BVRF2, 1o6e	GSHMAS + (2–235) + RS	Diisopropyl-fluorophosphate	100 m*M* NaCl, 20 m*M* Tris–HCl pH 7.5, 1 m*M* EDTA, 10 m*M* β-mercaptoethanol	1.4 *M* sodium formate, 100 m*M* sodium acetate pH 4.6, 5 m*M* EDTA	$P3_121$, twinned	$a = b = 52.8$, $c = 330.5$	2.3, MR	Fig. 2(*a*)
dUTPase¶	BLLF3, 2bsy	GAMGSGIP + (1–278)	dUMP	250 m*M* NaCl, 20 m*M* imidazole, 10 m*M* $MgCl_2$, 20 m*M* HEPES pH 7.5	20% PEG 3350, 200 m*M* $LiSO_4$, 100 m*M* Tris–HCl pH 8.5	$P2_12_12_1$	$a = 56.7$, $b = 55.8$, $c = 81.1$	1.5, SAD	Fig. 2(*b*)
	2bt1	GAMGSGIP + (1–278)	α,β-Imino-dUTP	250 m*M* NaCl, 20 m*M* imidazole, 10 m*M* $MgCl_2$, 20 m*M* HEPES pH 7.5	3 *M* NaCl, 100 m*M* Bis-Tris–HCl pH 6.5	*P*622	$a =$ $b = 146.6$, $c = 77.1$	2.7, MR	
BARF1††	BARF1, 2ch8	(21–221)	—	100 m*M* NaCl, 20 m*M* HEPES–NaOH pH 7.5	1 *M* $(NH_4)_2SO_4$, 2% PEG 3350, 100 m*M* Bis-Tris–HCl pH 6.0	*H*3, twinned	$a = b = 179.3$, $c = 95.7$	2.3, SAD	Fig. 2(*c*)
Uracil DNA glycosylase‡‡	BKRF3, N/A	GAM + (25–255)	UGI of bacteriophage, PBS-2	100 m*M* NaCl, 20 m*M* Tris pH 7.5 10 m*M* DTT	20% PEG 3350, 50 m*M* NH_4Cl	$C222_1$	$a = 62.8$, $b = 83.5$, $c = 273.4$	2.3, MR	Fig. 2(*d*)

† Extra residues arising from cloning are given; the residues of the viral ORF are indicated in parentheses. ‡ Method used for structure determination. MR, molecular replacement; SAD, single-wavelength anomalous dispersion on a heavy-atom derivative. § Buisson *et al.* (2002). ¶ Tarbouriech *et al.* (2005). †† Tarbouriech *et al.* (2006). ‡‡ Unpublished.

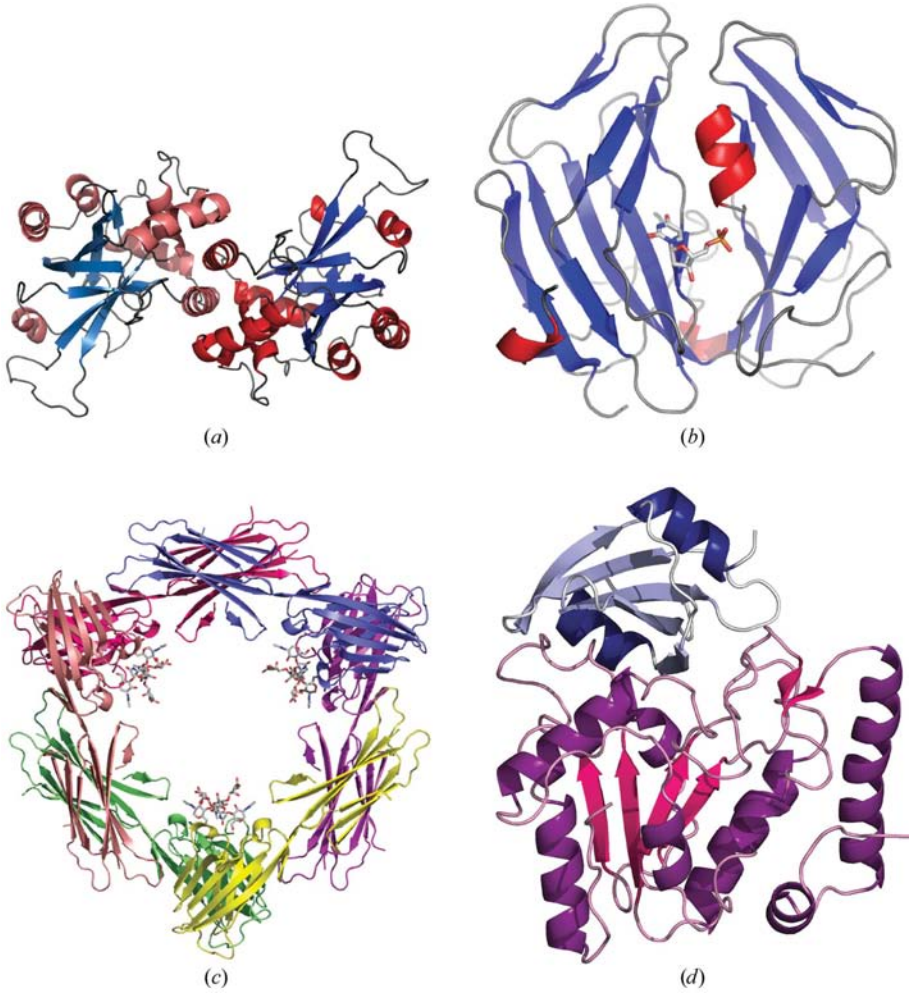

Figure 2
Protein structures (see Table 2).

EBV is composed of an inner capsid that contains the viral double-stranded DNA genome, surrounded by a membrane carrying various surface glycoproteins. Tegument fills the space between the capsid and the membrane. During the latent stage of infection in B-lymphocytes a very limited set of proteins is expressed. The viral DNA forms a circular episome which is associated with the cellular chromosomes and is replicated by the cellular machinery during cell division. After activation, the infection can switch to the lytic cycle, leading to the expression of the full set of viral proteins and production of viral particles. This complex lifestyle utilizes about 86 predicted proteins (Table 1), meaning that EBV has one of the largest genomes of human viruses. The principal viral functions are receptor binding and cell entry, maintenance of latency, nucleotide metabolism, DNA replication and packaging and capsid assembly (Fig. 1*a*, Table 1). EBV also codes for a number of immune-modulators. Some little-studied proteins shuttle viral particles from the nucleus, the site of viral replication, to the extracellular space and a number of proteins still have no assigned function. With the aim of obtaining insight into

the protein functions and in order to identify new drug targets, SPINE (Structural Genomics In Europe) included the structural proteomics of herpesviruses in workpackage 9 (human pathogen targets; see Fogg *et al.*, 2006) and here we report our contribution to this, namely the analysis of a cohort of 23 EBV proteins.

2. Project design, methods and results

2.1. Target annotation

The project included a major continued effort in protein annotation since the information available in databases [principally SWISS-PROT (Boeckmann *et al.*, 2003) and VIDA (Alba, 2002)] was rather incomplete, in particular for spliced reading frames, or no longer up to date. Our annotation is given in Table 1 with the results on the SPINE targets, together with as much bibliographic information as possible. We identified 86 proteins encoded by the EBV genome. The existence of a few of these remains questionable, owing to alternative splicing. The function of 15 proteins is unknown and could not be inferred from sequence homology or bibliographic information (Table 1, Fig. 1*a*). In general, little is known about the role of the tegument proteins, even though they have been recently localized unambiguously in the virus particle (Johannsen *et al.*, 2004).

2.2. Target selection

As one aim of the project was to obtain structures of potential new drug targets, we first targeted proteins with known enzymatic activity (11 ORFs; Table 1). Next, proteins were ranked according to several predicted properties. Firstly, they were given priority if they had a high predicted secondary structure by the NSP@ server (Deleage *et al.*, 1997), small size and a high stability index according to the ExPASy *ProtParam* tool (Gasteiger *et al.*, 2005). Known membrane proteins, surface glycoproteins and proteins involved in the packaging mechanism were omitted in order to avoid redundancy with other teams of the SPINE project. Furthermore, we selected against components of known multi-protein assemblies and eliminated proteins containing transmembrane domains using the *DAS* software (Cserzo *et al.*, 1997) and the TMHMM server (Krogh *et al.*, 2001) available from the ExPASy web site.

2.3. Cloning and protein production

We opted for a small-scale parallel approach using simple restriction-based cloning into a vector containing a tobacco etch virus protease (TEV) cleavable N-terminal His_6 tag, allowing the targets to be closely followed through purification and crystallization.

2.3.1. Cloning and expression tests. The selected genes were cloned by PCR amplification of EBV DNA extracted from the B95-8 cell line using primers introducing restriction sites at the 5′ and 3′ ends of the gene and ligated into the pPROEX-HTb plasmid (Invitrogen) using standard methods. The PCR products were cloned between *Nco*I or *Bam*HI sites as a first choice, *Eco*RI as a second choice and *Hin*dIII or *Xho*I sites. The ligated products were directly transformed into *Escherichia coli* BL21(DE3) GOLD cells (Invitrogen), which were used for both DNA preparation for sequencing and small-scale expression tests. DNA preparation was performed either manually or automatically on the RoBioMol platform at the IBS (Grenoble). Small-scale expression tests used 1 ml LB media inoculated with single colonies. Protein production was induced with 0.5 m*M* isopropyl β-D-thiogalactoside and continued for 3–5 h at 310 and 303 K and overnight at 296 and 289 K. Cells were lysed with BugBuster (Novagen). Protein solubility was checked on SDS–PAGE by loading both the cell extract and the soluble fraction after centrifugation at 18 000*g* for 20 min. If soluble protein was not detected, the *E. coli* strains Rosetta, Origami, BL21 (DE3) STAR (Invitrogen), C41 and C43 (Avidis) were tested with overnight induction at 289 K.

2.3.2. Protein expression and purification. Proteins were produced using either classical LB or an auto-inducible medium (Studier, 2005). Cells were lysed by sonication and cell debris was removed by centrifugation at 30 000*g* for 30 min. The supernatant was loaded onto an Ni–NTA (Qiagen) column equilibrated with 20 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl and 20 m*M* imidazole, washed using the same buffer containing 50 m*M* imidazole and eluted at an imidazole concentration of 500 m*M*. After buffer exchange back to the loading buffer, the protein was incubated overnight at room temperature with a ratio of 1/100 of recombinant His-tagged TEV protease. This was loaded again on an Ni–NTA column and the eluate of this column was concentrated by ultrafiltration and loaded onto a Superdex S75 or S200 gel-filtration column (GE/Amersham), depending on the protein size.

2.3.3. Refolding. When good expression levels of insoluble protein were obtained, refolding was attempted. Following large-scale production with induction at 310 K for 4–5 h, the protein was purified from inclusion bodies using buffers supplemented with 8 *M* urea. After purification and concentration to 5 mg ml^{-1}, a 20-fold dilution in refolding buffers was followed by 24 h incubation at 277 K. Refolding buffers varied in salt concentration (0 or 500 m*M* NaCl), pH (Bis-Tris–HCl pH 5, Tris–HCl pH 7 or Tris–HCl pH 9) or divalent cation contents (10 m*M* EDTA or 5 m*M* $CaCl_2$/5 m*M* $MgCl_2$), leading to 12 different basic conditions. Samples were centrifuged for 15 min at 16 000*g* and supernatants were assayed for soluble protein either by ammonium sulfate precipitation and SDS–PAGE or by concentration followed by gel filtration.

2.4. Crystallization

Proteins were analyzed by dynamic light scattering (Protein Solutions) prior to crystallization. Crystallization screening was carried out at the High Throughput Crystallization Laboratory of the EMBL Grenoble Outstation (HTX Lab). Typically, 576 conditions were tested per sample using a PixSys4200 robot (Cartesian) and the vapour-diffusion method in CrystalQuick (Greiner Bio-One) 96-well sitting-drop crystallization plates with square wells. Drops contained 100 nl protein solution and 100 nl buffer solution. Crystal

Screen, Crystal Screen II, PEG/Ion Screen, Crystal Screen Lite, Natrix, Membfac, Grid Screens and Index Screen (Hampton Research) were used as well as Clear Strategy Screens (Molecular Dimensions). Crystallization plates were stored and automatically imaged by a CrystalMation robot (RoboDesign) including a RoboIncubator and a Minstrel III module. Successful crystallizations were reproduced and refined manually using 1 + 1 µl hanging drops.

3. Discussion

A significant bottleneck in the structure-determination pipeline for EBV proteins was obtaining levels of protein expression (16/23) and soluble protein sufficient for crystallization (7/23; Fig. 1*b*, Table 1), although the success rate at crystallization was unexpectedly high (5/7). Surprisingly, changing the bacterial strain or expression temperature did not increase soluble expression levels compared with our standard protocol using BL21 cells at 303 K. A bioinformatics analysis using secondary-structure prediction (Deleage *et al.*, 1997) and *ClustalW*-based alignments (Thompson *et al.*, 1994) only rarely suggested obvious truncations. Perhaps as a consequence of this, modification of the constructs by N-terminal and C-terminal truncations, although attempted for the majority of the studied reading frames (Table 1), was successful in only one case, uracil-DNA glycosylase (UNG), where deletion of the N-terminal 24 residues increased expression levels and led to diffraction-grade crystals. The deleted residues may contain a nuclear localization signal based on sequence identity with human UNG2 (Otterlei *et al.*, 1998). Seven soluble proteins were expressed in *E. coli*: the EBV protease domain, dUTPase, uracil-DNA glycosidase, BHRF1, BLRF2, BDLF1 and a fragment of BMLF1 (EB2), but the last three proteins were unstable after purification. In the case of the dUTPase, the low solubility of the protein necessitated intensive optimization of purification and crystallization conditions (Tarbouriech *et al.*, 2005). Work on the EBV protease domain predated the SPINE project (Buisson *et al.*, 2002). Structural determination of BHRF1 was abandoned despite the existence of small crystals when an NMR structure was reported (Huang *et al.*, 2003). BARF1 was obtained through an external collaboration and expressed in eukaryotic cells (de Turenne-Tessier *et al.*, 2005) before entering our structure-determination pipeline. Protein purification using an N-terminal His_6 tag together with a TEV protease cleavage site, sometimes including size-exclusion chromatography, reliably produced pure protein for crystallization. In line with other unpublished results in SPINE, refolding from inclusion bodies failed to produce soluble protein from any of the 12 cases. However, we subsequently tested expression in insect cells using baculovirus and obtained three soluble proteins from six ORFs. Overall in SPINE the experience has been that viral proteins tend to be more difficult to express in bacterial systems than prokaryotic proteins (*e.g.* 27% of viral proteins were expressed in *E. coli* compared with 33–77% of some bacterial proteins; Fogg *et al.*, 2006). It is clear that eukaryotic expression is a real alternative for difficult viral proteins. Crystallization screening used 200 nl sitting drops dispensed robotically and achieved a very high success rate; however, for proteins except BARF1 this required the addition of enzyme inhibitors (Table 2). Crystallographic details for each EBV structure are given in Table 2 and further details on the structure determinations and refinement have been or will be published elsewhere.

The study described here highlights the particular problems associated with the application of pipeline technologies to difficult proteins. In this case, EBV proteins were poorly suited to bacterial expression systems and success was dependent on a much more individual approach to protein production. Although a simple pipeline approach with standard protocols is unlikely to be universally applicable for structural determination, pipeline components can be extremely effective, exemplified here by the high-throughput nanolitre crystallization platform. This major breakthrough in crystallization screening undoubtedly contributed to the high crystallization rates observed with the soluble EBV proteins.

This work was undertaken as part of the European Union Framework Programme 'Quality of Life and Management of Living Resources', Integrated Project SPINE (Structural Proteomics In Europe), contract No. QLG2-CT-2002-00988. We thank Jean-Marie Seigneurin for providing DNA from the B95.8 cell line and support for the project, Florine Dupeux, Benoit Gallet, José-Antonio Marquez, Martin Rower and Thierry Vernet for the operation of high-throughput facilities at the Grenoble Partnership for Structural Biology (PSB), and Lucy Freeman and Lucie Rivail for work on individual proteins. We are grateful to Henri Gruffat and Tadamasa Ooka for help with the genome annotation.

References

Alba, M. M. (2002). *VIDA Database.* http://www.biochem.ucl.ac.uk/bsm/virus_database/.
Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Beard, P. M. & Baines, J. D. (2004). *Virology*, **324**, 475–482.
Beisel, C., Tanner, J., Matsuo, T., Thorley-Lawson, D., Kezdy, F. & Kieff, E. (1985). *J. Virol.* **54**, 665–674.
Bellows, D. S., Howell, M., Pearson, C., Hazlewood, S. A. & Hardwick, J. M. (2002). *J. Virol.* **76**, 2469–2479.
Bochkarev, A., Barwell, J. A., Pfuetzner, R. A., Bochkareva, E., Frappier, L. & Edwards, A. M. (1996). *Cell*, **84**, 791–800.
Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. & Schneider, M. (2003). *Nucleic Acids Res.* **31**, 365–370.
Borza, C. M. & Hutt-Fletcher, L. M. (1998). *J. Virol.* **72**, 7577–7582.
Bowman, B. R., Baker, M. L., Rixon, F. J., Chiu, W. & Quiocho, F. A. (2003). *EMBO J.* **22**, 757–765.
Buisson, M., Hernandez, J. F., Lascoux, D., Schoehn, G., Forest, E., Arlaud, G., Seigneurin, J. M., Ruigrok, R. W. & Burmeister, W. P. (2002). *J. Mol. Biol.* **324**, 89–103.
Cabras, G., Decaussin, G., Zeng, Y., Djennaoui, D., Melouli, H., Broully, P., Bouguermouh, A. M. & Ooka, T. (2005). *J. Clin. Virol.* **34**, 26–34.
Chang, Y. E., Poon, A. P. & Roizman, B. (1996). *J. Virol.* **70**, 3938–3946.

Cheng, Y. C., Chen, J. Y., Hoffmann, P. J. & Glaser, R. (1980). *Virology*, **100**, 334–338.
Chevallier-Greco, A., Manet, E., Chavrier, P., Mosnier, C., Daillie, J. & Sergeant, A. (1986). *EMBO J.* **5**, 3243–3249.
Coen, D. M. & Schaffer, P. A. (2003). *Nature Rev. Drug Discov.* **2**, 278–288.
Countryman, J. & Miller, G. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 4085–4089.
Cserzo, M., Wallin, E., Simon, I., von Heijne, G. & Elofsson, A. (1997). *Protein Eng.* **10**, 673–676.
Davison, A. J. & Stow, N. D. (2005). *J. Virol.* **79**, 12880–12892.
Decaussin, G., Leclerc, V. & Ooka, T. (1995). *J. Virol.* **69**, 7309–7314.
Deleage, G., Blanchet, C. & Geourjon, C. (1997). *Biochimie*, **79**, 681–686.
Farell, P. J. (2005). *Epstein–Barr Virus*, edited by E. S. Robertson, pp. 263–287. Norwich: Caister Academic Press.
Fogg, M. J. *et al.* (2006). *Acta Cryst.* D**62**, 1196–1207.
Fries, K. L., Sculley, T. B., Webster-Cyriaque, J., Rajadurai, P., Sadler, R. H. & Raab-Traub, N. (1997). *J. Virol.* **71**, 2765–2771.
Furnari, F. B., Zacny, V., Quinlivan, E. B., Kenney, S. & Pagano, J. S. (1994). *J. Virol.* **68**, 1827–1836.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa, NJ, USA: Humana Press.
Gilligan, K. J., Rajadurai, P., Lin, J. C., Busson, P., Abdel-Hamid, M., Prasad, U., Tursz, T. & Raab-Traub, N. (1991). *J. Virol.* **65**, 6252–6259.
Giot, J. F., Mikaelian, I., Buisson, M., Manet, E., Joab, I., Nicolas, J. C. & Sergeant, A. (1991). *Nucleic Acids Res.* **19**, 1251–1258.
Gong, M., Ooka, T., Matsuo, T. & Kieff, E. (1987). *J. Virol.* **61**, 499–508.
Gonnella, R., Farina, A., Santarelli, R., Raffa, S., Feederle, R., Bei, R., Granato, M., Modesti, A., Frati, L., Delecluse, H. J., Torrisi, M. R., Angeloni, A. & Faggioni, A. (2005). *J. Virol.* **79**, 3713–3727.
Gruffat, H., Manet, E., Rigolet, A. & Sergeant, A. (1990). *Nucleic Acids Res.* **18**, 6835–6843.
Hardwick, J. M., Lieberman, P. M. & Hayward, S. D. (1988). *J. Virol.* **62**, 2274–2284.
Henkel, T., Ling, P. D., Hayward, S. D. & Peterson, M. G. (1994). *Science*, **265**, 92–95.
Hiriart, E., Gruffat, H., Buisson, M., Mikaelian, I., Keppler, S., Meresse, P., Mercher, T., Bernard, O. A., Sergeant, A. & Manet, E. (2005). *J. Biol. Chem.* **280**, 36935–36945.
Hitt, M. M., Allday, M. J., Hara, T., Karran, L., Jones, M. D., Busson, P., Tursz, T., Ernberg, I. & Griffin, B. E. (1989). *EMBO J.* **8**, 2639–2651.
Hong, G. K., Delecluse, H. J., Gruffat, H., Morrison, T. E., Feng, W. H., Sergeant, A. & Kenney, S. C. (2004). *J. Virol.* **78**, 4983–4992.
Hsu, D. H., de Waal Malefyt, R., Fiorentino, D. F., Dang, M. N., Vieira, P., de Vries, J., Spits, H., Mosmann, T. R. & Moore, K. W. (1990). *Science*, **250**, 830–832.
Huang, Q., Petros, A. M., Virgin, H. W., Fesik, S. W. & Olejniczak, E. T. (2003). *J. Mol. Biol.* **332**, 1123–1130.
Hutt-Fletcher, L. M. (2005). *Epstein–Barr Virus*, edited by E. S. Robertson, pp. 359–378. Norwich: Caister Academic Press.
Hwang, J. S. & Bogner, E. (2002). *J. Biol. Chem.* **277**, 6943–6948.
Johannsen, E., Luftig, M., Chase, M. R., Weicksel, S., Cahir-McFarland, E., Illanes, D., Sarracino, D. & Kieff, E. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 16286–16291.
Kieff, E. (1996). *Fields Virology*, edited by B. N. Fields, pp. 2343–2396. Philadelphia: Lipincott–Raven.
Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). *J. Mol. Biol.* **305**, 567–580.
Kusano, S. & Raab-Traub, N. (2001). *J. Virol.* **75**, 384–395.
Lake, C. M. & Hutt-Fletcher, L. M. (2000). *J. Virol.* **74**, 11162–11172.
Lake, C. M. & Hutt-Fletcher, L. M. (2004). *Virology*, **320**, 99–106.
Laux, G., Freese, U. K. & Bornkamm, G. W. (1985). *J. Virol.* **56**, 987–995.
Littler, E., Zeuthen, J., McBride, A. A., Trost Sorensen, E., Powell, K. L., Walsh-Arrand, J. E. & Arrand, J. R. (1986). *EMBO J.* **5**, 1959–1966.
Lopez, R., Urquiza, M., Patino, H., Suarez, J., Reyes, C., Patarroyo, M. A. & Patarroyo, M. E. (2005). *Biochimie*, **87**, 985–992.
Mackett, M., Conway, M. J., Arrand, J. R., Haddad, R. S. & Hutt-Fletcher, L. M. (1990). *J. Virol.* **64**, 2545–2552.
Manet, E., Gruffat, H., Trescol-Biemont, M. C., Moreno, N., Chambard, P., Giot, J. F. & Sergeant, A. (1989). *EMBO J.* **8**, 1819–1826.
Marschall, M., Stein-Gerlach, M., Freitag, M., Kupfer, R., van den Bogaard, M. & Stamminger, T. (2002). *J. Gen. Virol.* **83**, 1013–1023.
Marshall, W. L., Yim, C., Gustafson, E., Graf, T., Sage, D. R., Hanify, K., Williams, L., Fingeroth, J. & Finberg, R. W. (1999). *J. Virol.* **73**, 5181–5185.
Modrow, S., Hoflacher, B. & Wolf, H. (1992). *Arch. Virol.* **127**, 379–386.
Mullen, M. M., Haan, K. M., Longnecker, R. & Jardetzky, T. S. (2002). *Mol. Cell*, **9**, 375–385.
Nealon, K., Newcomb, W. W., Pray, T. R., Craik, C. S., Brown, J. C. & Kedes, D. H. (2001). *J. Virol.* **75**, 2866–2878.
Newcomb, W. W., Juhas, R. M., Thomsen, D. R., Homa, F. L., Burch, A. D., Weller, S. K. & Brown, J. C. (2001). *J. Virol.* **75**, 10923–10932.
Otterlei, M., Haug, T., Nagelhus, T. A., Slupphaug, G., Lindmo, T. & Krokan, H. E. (1998). *Nucleic Acids Res.* **26**, 4611–4617.
Paulsen, S. J., Rosenkilde, M. M., Eugen-Olsen, J. & Kledal, T. N. (2005). *J. Virol.* **79**, 536–546.
Pearson, A. & Coen, D. M. (2002). *J. Virol.* **76**, 10821–10828.
Petosa, C., Morand, P., Baudin, F., Moulin, M., Artero, J. B. & Muller, C. W. (2006). *Mol. Cell*, **21**, 565–572.
Raab-Traub, N. (2005). *Epstein–Barr Virus*, edited by E. S. Robertson, pp. 71–92. Norwich: Caister Academic Press.
Ressing, M. E., van Leeuwen, D., Verreck, F. A., Gomez, R., Heemskerk, B., Toebes, M., Mullen, M. M., Jardetzky, T. S., Longnecker, R., Schilham, M. W., Ottenhoff, T. H., Neefjes, J., Schumacher, T. N., Hutt-Fletcher, L. M. & Wiertz, E. J. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 11583–11588.
Reuven, N. B., Willcox, S., Griffith, J. D. & Weller, S. K. (2004). *J. Mol. Biol.* **342**, 57–71.
Rickinson, A. & Kieff, E. (1996). *Fields Virology*, edited by B. N. Fields, Vol. 2, pp. 2397–2446. Philadelphia: Lippincott–Raven.
Rivailler, P., Jiang, H., Cho, Y. G., Quink, C. & Wang, F. (2002). *J. Virol.* **76**, 421–426.
Robertson, E. S., Ooka, T. & Kieff, E. D. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 11334–11340.
Rooney, C., Taylor, N., Countryman, J., Jenson, H., Kolman, J. & Miller, G. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 9801–9805.
Savva, C. G., Holzenburg, A. & Bogner, E. (2004). *FEBS Lett.* **563**, 135–140.
Segouffin-Cariou, C., Farjot, G., Sergeant, A. & Gruffat, H. (2000). *J. Gen. Virol.* **81**, 1791–1799.
Segouffin, C., Gruffat, H. & Sergeant, A. (1996). *J. Gen. Virol.* **77**, 1529–1536.
Sheaffer, A. K., Newcomb, W. W., Gao, M., Yu, D., Weller, S. K., Brown, J. C. & Tenney, D. J. (2001). *J. Virol.* **75**, 687–698.
Smith, R. F. & Smith, T. F. (1989). *J. Virol.* **63**, 450–455.
Sommer, P., Kremmer, E., Bier, S., Konig, S., Zalud, P., Zeppezauer, M., Jones, J. F., Mueller-Lantzsch, N. & Grasser, F. A. (1996). *J. Gen. Virol.* **77**, 2795–2805.
Spear, P. G. & Longnecker, R. (2003). *J. Virol.* **77**, 10179–10185.
Strockbine, L. D., Cohen, J. I., Farrah, T., Lyman, S. D., Wagener, F., DuBose, R. F., Armitage, R. J. & Spriggs, M. K. (1998). *J. Virol.* **72**, 4015–4021.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Swaminathan, S. (2005). *Epstein–Barr Virus*, edited by E. S. Robertson, pp. 631–649. Norwich: Caister Academic Press.
Tarbouriech, N., Buisson, M., Seigneurin, J. M., Cusack, S. & Burmeister, W. P. (2005). *Structure*, **13**, 1299–1310.
Tarbouriech, N., Ruggiero, F., de Turenne-Tessier, M., Ooka, T. & Burmeister, W. P. (2006). *J. Mol. Biol.* **359**, 667–678.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Trus, B. L., Heymann, J. B., Nealon, K., Cheng, N., Newcomb, W. W., Brown, J. C., Kedes, D. H. & Steven, A. C. (2001). *J. Virol.* **75**, 2879–2890.
Tsurumi, T., Kobayashi, A., Tamai, K., Daikoku, T., Kurachi, R. & Nishiyama, Y. (1993). *J. Virol.* **67**, 4651–4658.
Tugizov, S. M., Berline, J. W. & Palefsky, J. M. (2003). *Nature Med.* **9**, 307–314.
Turenne-Tessier, M. de, Jolicoeur, P., Middeldorp, J. M. & Ooka, T. (2005). *Virus Res.* **109**, 9–18.
Wagenaar, F., Pol, J. M., de Wind, N. & Kimman, T. G. (2001). *Vet. Res.* **32**, 47–54.
Waltzer, L., Logeat, F., Brou, C., Israel, A., Sergeant, A. & Manet, E. (1994). *EMBO J.* **13**, 5633–5638.
Waltzer, L., Perricaudet, M., Sergeant, A. & Manet, E. (1996). *J. Virol.* **70**, 5909–5915.
Wei, M. X. & Ooka, T. (1989). *EMBO J.* **8**, 2897–2903.
Winters, T. A. & Williams, M. V. (1993). *Virology*, **195**, 315–326.
Yokoyama, N., Fujii, K., Hirata, M., Tamai, K., Kiyono, T., Kuzushima, K., Nishiyama, Y., Fujita, M. & Tsurumi, T. (1999). *J. Gen. Virol.* **80**, 2879–2887.
Young, L. S. & Rickinson, A. B. (2004). *Nature Rev. Cancer*, **4**, 757–768.