doi:10.1016/j.jmb.2006.03.056 J. Mol. Biol. (2006) 359, 667–678

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Structure of the Epstein-Barr Virus Oncogene BARF1

Nicolas Tarbouriech[1,2], Florence Ruggiero[3], Mireille de Turenne-Tessier[4] Tadamasa Ooka[4] and Wim P. Burmeister[1,2,5]*

[1]*EMBL Grenoble outstation BP181, F-38042, Grenoble cedex 9, France*

[2]*Institut de Virologie Moléculaire et Structurale FRE 2854, CNRS—UJF, BP181 F-38042, Grenoble cedex 9 France*

[3]*Institut de Biologie et Chimie des Protéines, UMR 5086 CNRS—UCB, 7 passage du Vercors, F-69367 Lyon Cedex 07 France*

[4]*Laboratoire de Virologie Moléculaire, UMR 5537 CNRS—Faculté de Médecine R.T.H. Laennec, Rue G. Paradin F-69372 Lyon cedex 08, France*

[5]*Institut Universitaire de France, 103, bd Saint-Michel F-75005 Paris, France*

**Corresponding author*

The Epstein-Barr virus is a human gamma-herpesvirus that persistently infects more than 90% of the human population. It is associated with numerous epithelial cancers, principally undifferentiated nasopharyngeal carcinoma and gastric carcinoma. The BARF1 gene is expressed in a high proportion of these cancers. An oncogenic, mitogenic and immortalizing activity of the BARF1 protein has been shown. We solved the structure of the secreted BARF1 glycoprotein expressed in a human cell line by X-ray crystallography at a resolution of 2.3 Å. The BARF1 protein consists of two immunoglobulin (Ig)-like domains. The N-terminal domain belongs to the subfamily of variable domains whereas the C-terminal one is related to a constant Ig-domain. BARF1 shows an unusual hexamerisation involving two principal contacts, one between the C-terminal domains and one between the N-terminal domains. The C-terminal contact with an uncommonly large contact surface extends the beta-sandwich of the Ig-domain through the second molecule. The N-terminal contact involves Ig-domains with an unusual relative orientation but with a more classical contact surface with a size in the range of dimer interactions of Ig-domains. The structure of BARF1 is most closely related to CD80 or B7-1, a co-stimulatory molecule present on antigen presenting cells, from which BARF1 must have been derived during evolution. Still, domain orientation and oligomerization differ between BARF1 and CD80. It had been shown that BARF1 binds to hCSF-1, the human colony-stimulating factor 1, but this interaction has to be principally different from the one between CSF-1 and CSF-1 receptor.

Keywords: Epstein-Barr virus; BARF1; immunoglobulin domain; CD80; B7-1

Introduction

The Epstein Barr virus (EBV) is a human gamma-herpesvirus that persistently infects more than 90% of the world population. If the first infection occurs early during childhood, it often goes unrecognized but EBV can lead to infectious mononucleosis if the primo-infection occurs during adolescence or adulthood. The virus persists life-long in B cells and is able to promote B cell immortalization *in vitro*. In immunocompromised and immunosuppressed patients it can lead to lymphoproliferative disease due to an absence of the control of the virus by the immune system. EBV is not only one of the causative agents of Burkitt's lymphoma, a childhood cancer common in parts of Africa where *Plasmodium falciparum* malaria is endemic,[1] but EBV is also associated with numerous human epithelial cancers including undifferentiated nasopharyngeal carcinoma (NPC) and gastric carcinoma. NPC is a major health problem in South-East Asia and North Africa, with an incidence of 5–40 per 100,000 individuals.[2–4] Its association with EBV is constant except in a few atypical highly differentiated cases.[5,6] About 5–20% of gastric carcinoma, a widespread cancer, are associated with EBV.[7]

Among the about 86 proteins encoded by the EBV genome,[8] two viral oncogenes, LMP1 and BARF1, were shown to induce malignant transformation in rodent fibroblasts.[9–11] LMP1 and BARF1 are respectively expressed in 50% and 90% cases of NPC.[12,13] BARF1 transcripts are detected in a high proportion of EBV-positive gastric carcinomas,

Abbreviations used: EBV, Epstein-Barr virus; Ig, immunoglobulin.

E-mail address of the corresponding author: wpb@embl-grenoble.fr

where LMP1 expression is consistently negative.[14] BARF1-encoded protein was also found in epithelial cells immortalised by NPC-derived EBV[15] and in NPC biopsies.[12] The oncogenic activity of the BARF1 gene was first demonstrated in rodent fibroblasts.[10] BARF1 protein can also immortalize primary monkey kidney epithelial cells[16] and induce malignant transformation in other established cell lines, such as human Louckes B cells[17] and EBV-negative AKATA[18] cells. BARF1-immortalised epithelial cells are not tumourigenic in nude mice, while BARF1-transfected immortalised cell lines induce tumours.[10,17] This suggests that BARF1 can intervene in two oncogenic processes: immortalisation and malignant transformation. Several cellular genes (Bcl2, CD21, CD23 and CD71) are transactivated in BARF1-transfected cells.[11,17,18] Following transfection of rodent fibroblast or simian epithelial cell lines, or after infection of human epithelial cells with recombinant adenovirus, or upon viral cycle activation in EBV-positive B cell lines the BARF1 protein is secreted into the medium after signal sequence (residues 1–20) cleavage.[19–21]The addition of purified BARF1 protein into serum-free culture medium of Balb/c3T3 fibroblasts, human B cells and primary monkey epithelial cells resulted in cell cycle activation,[20] suggesting that BARF1 protein can act as a cell growth factor.

Moreover, it has been shown also that BARF1 binds to the human cytokine hCSF-1 and reduces the action of hCSF-1 on the proliferation of macrophages[19] and the alpha interferon production by mononuclear cells,[22] suggesting that BARF1 protein acts also in immunomodulation.

The BARF1 gene, its protein and its function have recently been reviewed.[23] A BARF1 homologue with 75% sequence identity is found in Cercopithecine herpesvirus 15 (Rhesus lymphocryptovirus, rLCV,[24]), a lymphocryptovirus closely related to EBV present in old world primates. As LCV is strictly species-specific and as there is no BARF1 homologue present in LCV of new world primates it seems to be acquired late in evolution after the separation of old and new world primates.[25]

Here we report the crystal structure of the BARF1 protein purified from human cells at 2.3 Å resolution. The structure presents two immunoglobulin (Ig) domains in a unique hexameric assembly.

Results

Protein production, crystallization and structure determination

The BARF1 protein was expressed in HeLa cells using a recombinant adenovirus system. As it is secreted, it was purified from the culture media by concanavalin A affinity chromatography followed by size-exclusion chromatography where the protein eluted as an oligomer with an approximative molecular mass of 240 kDa. Crystals were first obtained by the vapour diffusion method using nanodrop crystallization trials using standard screens. Initial conditions were manually refined using the hanging-drop method in order to obtain crystals of suitable size for data collection. Rhombohedric crystals belonging to spacegroup $H3$ with cell parameters $a=b=179.25$ Å, $c=95.72$ Å diffracted to 2.3 Å. Despite systematic twinning of the crystals by merohedry with twinning fractions between 0.1 and 0.45 the structure was solved by the single anomalous dispersion (SAD) method using a platinum derivative of a crystal with a twinning fraction of 0.12. All steps of the structure determination used less than 0.5 mg of pure protein. The final model was refined to an R-factor of 17.5% ($R_{free}=23.4\%$). The asymmetric unit contains four protein chains with the visible residues 21 to 161 and 173 to 220 organized in two dimers. Residue 21 is the first residue of secreted BARF1 after cleavage of the signal peptide. Statistics of the data collection and refinement are shown in Table 1.

Table 1. Data collection and model refinement statistics

	BARF1 + Pt
Data collection statistics	
Wavelength (Å)	0.9393
Resolution (Å)	30–2.3 (2.42–2.3)
Unique reflections	50,910
Completeness (%)	100 (100)
R_{sym} (%)	5.9 (22.8)
$I/\sigma I$	10.7 (3.1)
Multiplicity	5.3 (4.7)
Refinement statistics	
R_{cryst} (%)	17.5 (17.9)
R_{free} (%)	23.4 (27.0)
Mean atomic B-factor (Å^2)	28.1
RMS on bond length (Å)	0.022
RMS on bond angles (deg.)	2.15
Model composition (in the asymmetric unit)	
Residues	748
Water molecules	533
Platinum atoms	12
N-acetyl glucosamine	(Nag)
Mannose (Man)	4

Values for the highest resolution bin are given in parentheses.

Overall fold

The BARF1 protein structure is composed of two domains belonging to the immunoglobulin fold superfamily, the first one ranging from residues 21 to 123 and the second one from residues 125 to 220 (Figure 1). The N-terminal domain is clearly related to the variable subfamily (V-set) with an ABED/C″C′CFG organization where the A strand is hydrogen bonding to the G strand rather than the B strand (Figure 1(b)). The C-terminal domain is more difficult to classify due to the disordered region 162–172, potentially containing the strand C′ or D. Nevertheless, an ABED/CFG organization characteristic of the constant subfamily (C-set) is most probable as the few residues visible in electron

density before the E strand are hydrogen bonding to it and form the beginning of strand D. With a length of only two residues, the inter-domain region is very short leading to many contacts between the two domains (salt-bridge between Lys123 and Asp180, hydrophobic interactions around residues Phe151 and Ile96). The link between the two domains is further rigidified by inter-molecular contacts within the BARF1 hexamer (salt-bridge between Glu27 of one chain and Arg133 of the other chain, see below). The long axes of the N-terminal and the C-terminal Ig-domains are almost aligned, so that the difference in orientation of 123° between the two domains corresponds roughly to a rotation around the long axis.

Glycosylation

BARF1 has been shown to be glycosylated by O-glycosylation and by N-glycosylation (M. de T.-T., unpublished results), where the N-glycosylation is of the high-mannose type as indicated by the endoglycosidase-H sensitiveness of BARF1, and its binding to concanavalin A, a lectin specific for this type of glycosylation (M. de T.-T., unpublished results). The O-glycosylation site is predicted on Thr169 using NetOGlyc, a server for the prediction of mucin-like O-glycosylation sites.[26] As Thr169 is part of a disordered loop region it agrees with the absence of an O-linked glycosylation site in the electron density.

In contrast, attached to Asn95, three sugar residues (Nag-Nag-Man) of the core structure of the high-mannose glycosylation site are clearly visible (Figure 2(b) and (c)), whereas electron density for even more poorly ordered sugar residues is present. The carbohydrate is located on the inner side of the hexameric BARF1 ring. Matrix assisted laser desorption/ionization (MALDI) mass spectrometric analysis of the purified BARF1 protein used for crystallization gives a mass of 24,890 Da, in agreement with a theoretical mass in the range of 24,814 to 24,960 Da for the protein including one high-mannose glycosylation ($GlcNac_2$-Man_9) and a trisaccharide sugar structure at the O-glycosylation site.[27]

Oligomerization

BARF1 forms hexameric rings (Figure 2(a)–(c)) of molecules related by 2-fold and 3-fold axes. In the crystal the rings are oriented along the *c*-axis and stack on top of each other. The two BARF1 dimers present in the asymmetric unit of the crystal correspond to two subsequent layers of these rings, which are generated by the crystallographic 3-fold symmetry.

On gel filtrations experiments, BARF1 protein showed a hydrodynamic radius of about 5 nm. In solution, a bimodal size distribution was observed using dynamic light-scattering. A small fraction of aggregates corresponded to hydrodynamic radii

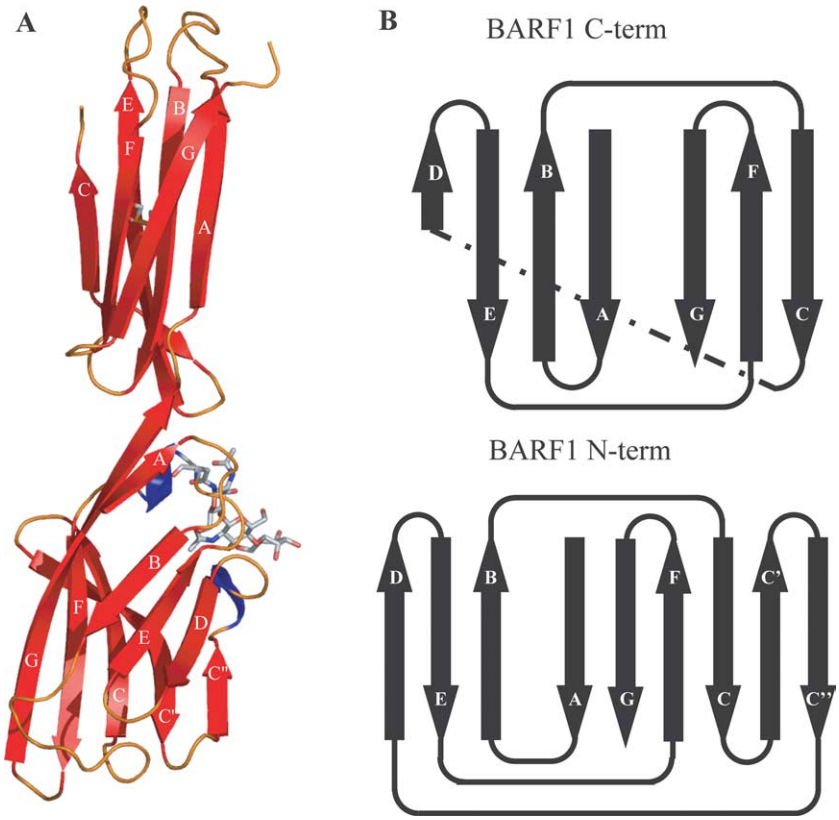


Figure 1 (*legend next page*)

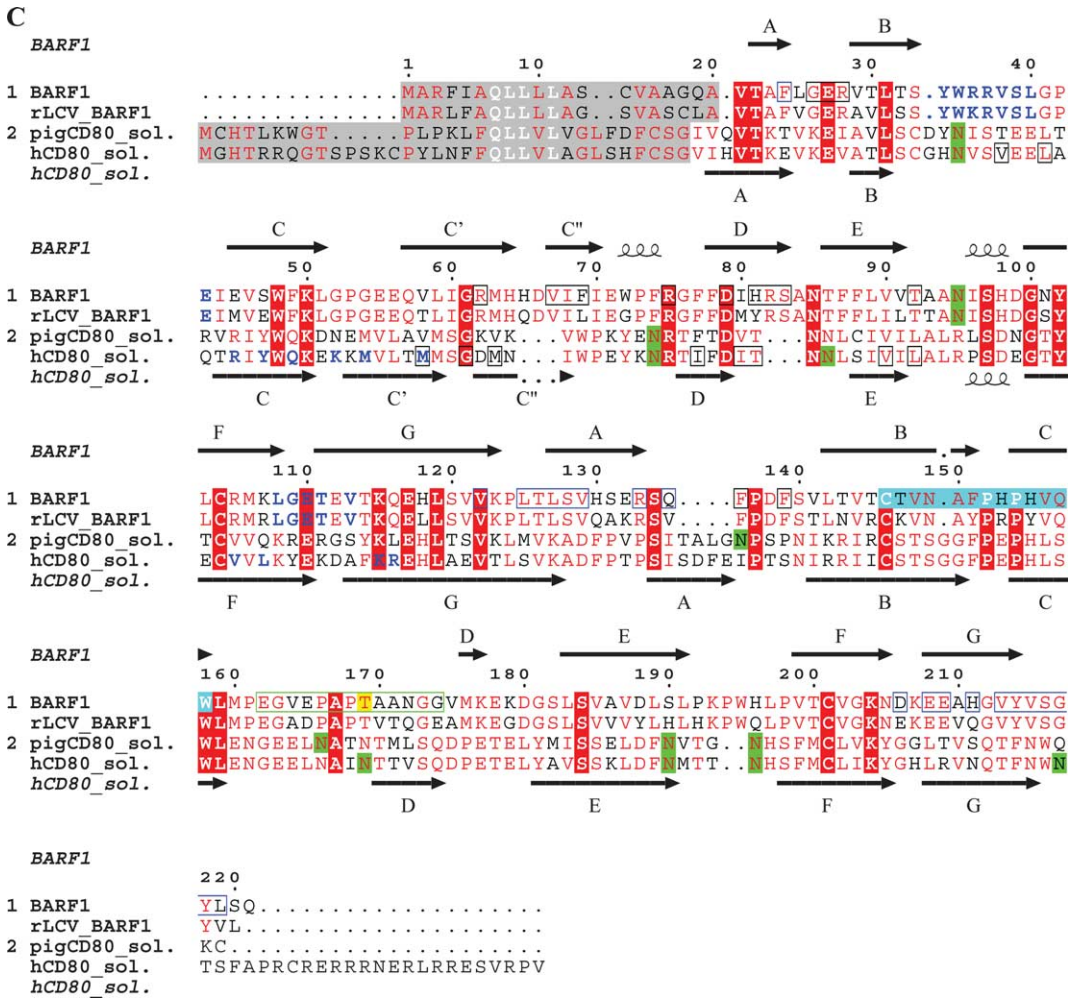
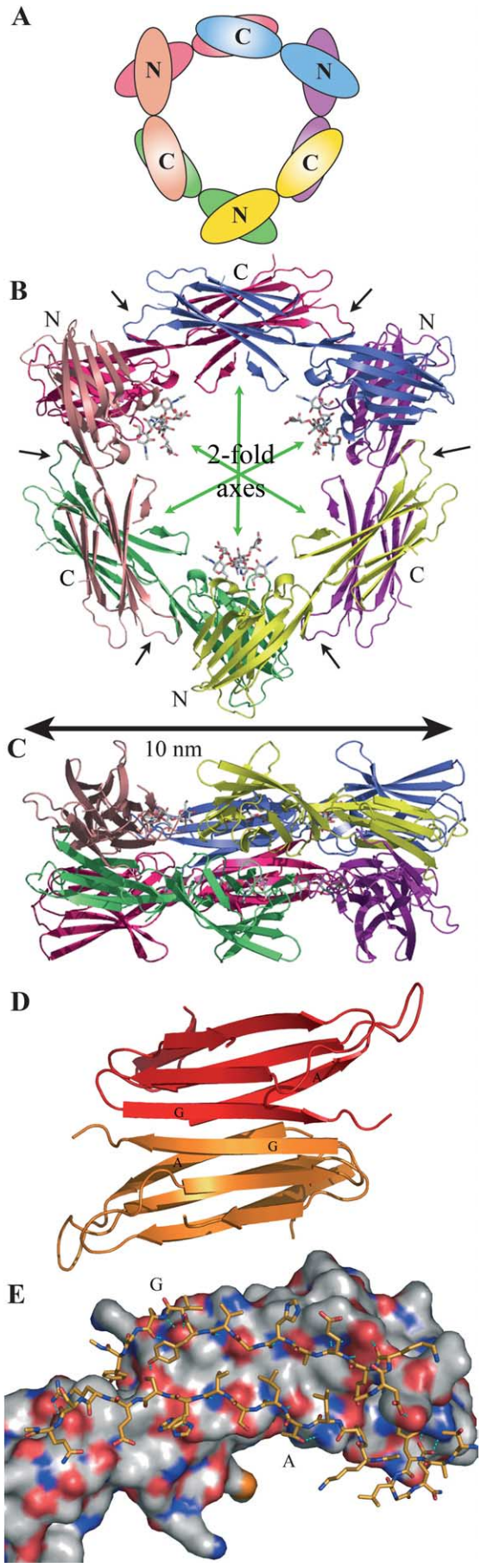


Figure 1. Fold and sequence of BARF1. (a) View of the BARF1 monomer, α-helices are shown in blue. (b) Topology diagram of BARF1. The diagram for CD80 is identical apart from the absence of the disordered region. (c) Alignment of BARF1 (UniProt entry number EBV: Q777A5, rLCV: Q8UZD0) and CD80 (human: Q5DTB0, porcine: Q9TT71) protein precursor sequences. Human and porcine CD80 sequences share about 70% identity, similar to the 75% identity between human and rLCV BARF1. Gray shade, signal peptide; red letters, conservation within the two BARF1 sequences or the two CD80 sequences; red background, strict conservation; green box, disordered residues in the BARF1 structure; yellow background, likely O-glycosylation site; green background, (putative) N-glycosylation sites; cyan background, growth factor receptor homology region;[19] blue characters, experimental CD152-binding site for CD80,[42] suggested ligand binding site for BARF1. Black boxes, dimerization site of the N-terminal domain of BARF1 or CD80;[29] blue box, dimerization site of the C-terminal domain of BARF1.

above 100 nm and a large majority of the protein with a hydrodynamic radius of about 6 nm. Using the crystal structure of the BARF1 hexamer, a radius of 5.1 nm and a sedimentation coefficient of 6.3 S (Svedberg) could be calculated using HYDRO-PRO.[28] The calculated values do not take into account the disordered part of the N-linked carbohydrate structure or the invisible residues (162–172) together with their O-linked glycosylation. Furthermore, the hydration of the protein is not considered. Altogether this leads to an underestimation of the hydrodynamic radius.

Rotary shadowing of purified BARF1 revealed a homogenous population of very small globular structures; most of them appeared as ring-shaped particles (Figure 3(a)). At higher magnification, the globules were unambiguously identified as ring-shaped particles with an external diameter of 11($\pm$2) nm (n=60) and a central empty region of about 3–4 nm in diameter (Figure 3(b)–(d) arrows, and inset). Sometimes two facing elongated structures (Figure 3(c) and (d), arrowheads) were observed, which could correspond to lateral views of the ring-shaped particles. The dimensions of these objects correspond to the ones of BARF1 hexamers obtained from the X-ray structure (Figure 2(b) and (c)).

The hexameric rings are formed by two kinds of contacts (Table 2), both relating domains by a 2-fold axis: One contact involves the C-terminal domain of BARF1 molecules which interact through an extension of the β-sheets of the immunoglobulin fold through strands A and G in order to form continuous sheets (Figure 2(d) and (e)). This



interaction has a contact surface of 1400 Å^2 per monomer. This interface is a very striking feature of the BARF1 protein structure. The interaction is mediated by hydrophobic contacts and 24 hydrogen bonds in total, of which 15 involve main-chain atoms through the β-sheet structure. The hydrophobic contacts are mediated by Leu127 and Val129 inside the β-sandwich (Figure 2(e)), furthermore by Val122 of strand G of the N-terminal Ig-domain, Tyr218 of the C terminus and Phe24 of the AB loop of the N-terminal domain.

The second interface involves the N-terminal Ig-domains, which interact around a 2-fold axis through their face involving strands C″ and D as well as the loops: AB, EF and C″D, leading to a contact surface of 580 Å^2 (Figure 4(b)). The two domains cross under a 65° angle. The interaction includes hydrophobic interactions involving principally Arg28, His81 and Ile68, together with hydrogen bonds between arginine residues and carbonyl oxygen atoms (Arg75–Thr92, Arg75–Gly26, Arg82–Ser83) and salt-bridges (salt-bridge Arg28–Asp79). Furthermore an asymmetric interaction of the AB loop of the C-terminal domain with strands C′ and C″ contributes with a contact surface of 250 Å^2 to the interaction (hydrophobic contacts involving in particular Arg62, Val67 and Phe69 on one side and Gln135, Phe139 and Phe136).

Discussion

Comparison with CD80

Human CD80 (or B7-1) is, with an identity of 18% (Figure 1(c)), not only the most closely related protein, but has as well the same topology shown in Figure 1(b) and its two domains can be superposed onto the corresponding domains of BARF1 (90 residues aligned with an rms of 1.94 Å for the N-terminal domain and 78 residues aligned with an rms of 1.91 Å for the C-terminal domain, Figure 4(a) and (b)) although their relative orientations are completely different. BARF1 has an inter-domain

Figure 2. BARF1 hexamerisation. (a) Schematic arrangement of BARF1 molecules and domains in the hexamer. (b) Top view of a cartoon of the BARF1 hexamer coloured according to peptide chains as in (a). The N-linked glycosylation is shown in a stick representation. The directions of the 2-fold axes relating N-terminal and C-terminal domains are indicated by green arrows. Black arrows show the AB loops of the C-terminal domains involved in oligomerization through a contact with the N-terminal domains. (c) Side view of the BARF1 hexamer. (d) Dimer of C-terminal domains. (e) Dimerization interface of the C-terminal domains. One subunit is shown as an accessible surface coloured according to atom type, the two stretches of residues forming the main part of the interface and containing β-strands A and G are shown on top of the surface in a stick representation. His130 has an alternate conformation. Hydrogen bonds are shown as cyan dotted lines.

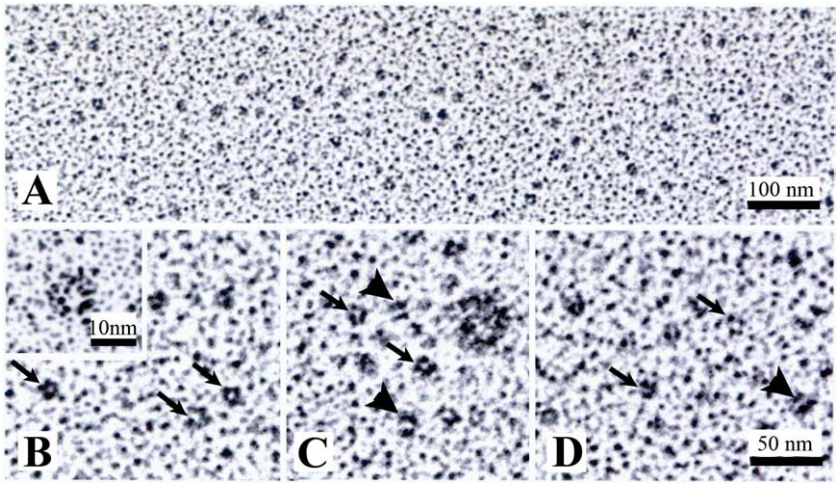


Figure 3. Electron micrographs. (a) Representative micrograph of the ring-shaped structure of BARF1 oligomers after rotary shadowing. (b)–(d) and inset: gallery of BARF1 images at higher magnification. Arrows show ring-shaped structures, arrowheads indicate potential lateral views.

angle of about 125° and CD80 of about 70°. In contrast to BARF1, CD80 is a dimer formed by a contact at the level of the N-terminal domains[29] (Figure 4(c)).

Structural differences between the N-terminal domains of CD80 and BARF1 concern mainly the loops BC, CC′ and FG. Additionally a difference in the twist of strands C′ and C″ in BARF1 leads to a repositioning of the C′C″ loop (Figure 4(b)). In contrast, the loops contacting the C-terminal domain are much more conserved. When the C-terminal domains are compared, especially the AB and EF loops differ, whereas the loops facing the N-terminal domain are again more conserved (Figure 4(a)).

CD80 is a membrane glycoprotein present on antigen presenting cells, such as B lymphocytes, closely related to CD86 (B7-2) but CD80 and CD86 share only 26% sequence identity. Upon interaction with CD152 (CTLA-4) or CD28, CD80 transmits inhibitory or activating signals to T lymphocytes. CD80 and CD86 are primarily expressed on activated dendritic cells and B lymphocytes.[30] Soluble forms of human (GenBank entry AY197778) and porcine[31] CD80 generated through alternate splicing have been observed. The soluble form of human CD80 lost its C-terminal transmembrane domain but still carries its cytoplasmic part, whereas the porcine protein is generated through a C-terminal truncation, which yields a protein with an extension similar to the one of BARF1 (Figure 1(c)). As judged from the structural similarity in agreement with the late acquisition of the BARF1 gene during viral evolution, BARF1 appears to be derived from a primate CD80 molecule, possibly from its soluble form.

Solution structure and oligomerization

The observation of a hexameric molecule in the crystal structure agrees with our results from gel filtration, electron microscopy and diffuse light-scattering, which showed a predominant presence of a large species in solution. Furthermore the theoretical sedimentation coefficient of 6.3 S is similar to previous estimations of the size of the assembly obtained from the speed of sedimentation on a sucrose gradient (around 8 S).[20]

Both kinds of interactions between BARF1 monomers, the ones forming a continuous β-sheet between the C-terminal Ig-domains and the less extensive dimerization of the N-terminal domains do not have equivalents in the CD80 structure. The area involved in the contact between two N-terminal domains of BARF1 is similar to the one that mediates dimerization for CD80, but there are no similarities concerning the residues forming the interaction (Figure 1(c)) and the relative orientations of the immunoglobulin domains of the second molecule differ by about 100° between CD80 and BARF1 (Figure 4(b)).

The interaction between C-terminal domains of the BARF1 molecule mediated by the formation of continuous β-sheets by two Ig-domains through

Table 2. Comparison of oligomerization interfaces

	C-term	N-term	N–C interface	CD80	IgG-V 1FMA[a]	IgG-V 8FAB[a]
Interface accessible surface area (Å^2)	1400	580	250	570	840	530
Length and breadth (Å)	50 and 22	26 and 26	17 and 10	30 and 17	30 and 26	26 and 18
% Non-polar atoms in interface	62	55	66	78	63	61
Hydrogen bonds	24	9	0	0	2	4
Salt-bridges	2	2	0	0	0	0

[a] These pdb entries have been selected because of the similarity of their V-domain to the BARF1 N-terminal domain.

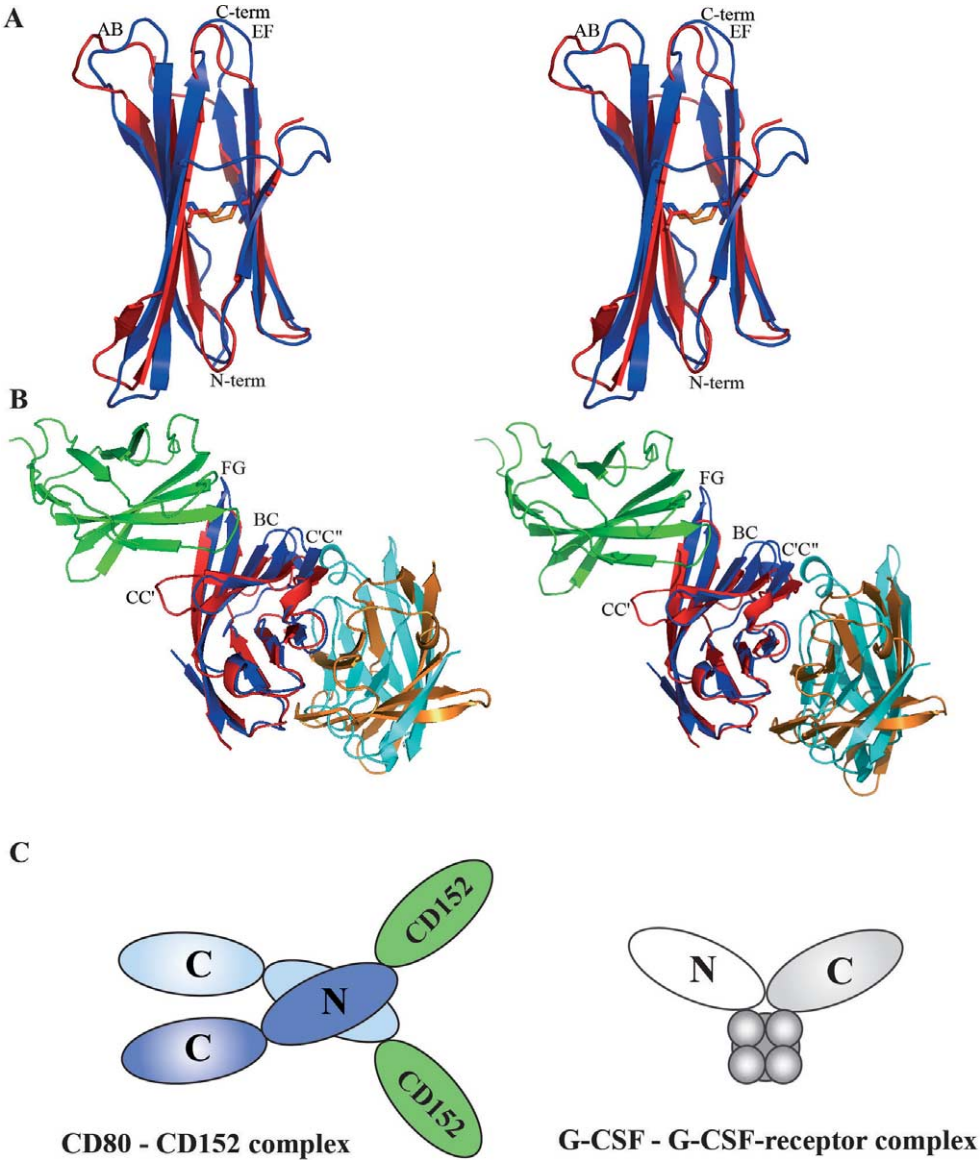


Figure 4. Comparisons and interactions of BARF1 and CD80. (a) Stereoview of a cartoon of the C-terminal domains of BARF1 (red) superposed with the C-terminal domain of CD80 (blue) seen from the side of the stretch of disordered residues in BARF1. The disulfide bridge is shown. (b) Two interacting N-terminal domains of BARF1 (red and orange) superposed onto the N-terminal domains (blue and cyan) of the CD80 dimer are shown. The CD80 N-terminal domain shown in red and the one from BARF1 shown in blue have been used in a least-squares alignment. The position of CD152 (green) in the CD80–CD152 complex (pdb 1I8L)[42] is also shown. (c) Schemes of the CD80–CD152 interaction and the typical interaction of four-helix-bundle cytokines with their receptors.

hydrogen bonding involving strand A which interacts with another strand A and strand G which interacts with strand G is unique. There are no other examples of such an interaction as judged from the results of a SSM[32] search with the dimer of the two domains. At a first glimpse, the most similar interaction is the dimerization of filamin through its sixth rod domain (pdb 1WLH).[33] But the individual domains align only poorly (37 matched C^{α} atoms). In addition for filamin only one of the β-sheets composed of strands B, E, D (there is no equivalent of strand A) extends across the dimer interface, where strand B is the strand making the contact. The interaction surface formed by the C-terminal domain is much larger than the one observed for other dimeric Ig-domains such as the dimerization of CD80 or of variable domains of IgG (Table 2).

The dimerization of the N-terminal domains is more similar to classical interactions between Ig-domains, which interact most of the time *via* side-chain contacts between the faces of their β-sheets. Examples are the interactions between the different Ig-domains of antibodies,[34] the interaction between β2-microglobulin and the heavy chain in MHC class I molecules[35] and the dimerization of CD80,[29] CD152[36] or coxsackie and adenovirus receptor.[37]

The size of the interface of BARF1 is for example in the typical range observed for the dimerization of Ig-V domains (Table 2). But the relative orientations of the domains of the N-terminal dimer do not have equivalents as judged by the results of a search with SSM.

BARF1–hCSF-1 interactions

Binding of BARF1 to activated peripheral blood T-cells was observed by Strockbine *et al.*[19] The authors identified the cytokine hCSF-1 expressed on the activated T-cells as the molecule mediating this interaction. hCSF-1 functions either as a membrane-bound molecule present in three isoforms or as a soluble cytokine generated by proteolysis.[38] The soluble C-terminal part belongs to the four-helix bundle family of cytokines. Immunoprecipitation showed that all isoforms interact with BARF1 and hence situate the BARF1 binding site on the four-helix bundle domain. Furthermore, the authors observed a neutralization of the hCSF-1 activity in a mouse bone marrow macrophage non-adherent proliferation assay, suggesting that the anti-proliferative action of BARF1 may be due to a sequestration of hCSF-1.

In contrast to the other monomeric growth factors of the same family, hCSF-1 forms a disulfide-linked dimer in solution. The dimerization still leaves the side of the four-helix bundle accessible for interactions as for the related molecules. As only the C^{α} backbone structure of hCSF-1 is available (pdb 1HMC),[39] the absence of full coordinates precludes a further analysis of the surface of hCSF-1. The interaction of the CSF-1 with its receptor is likely to be similar to the interaction of granulocyte colony stimulating factor (G-CSF) with its receptor[40] or of other four-helix bundle cytokines with their receptors, like the one between erythropoietin (EPO) and EPO-receptor.[41] In the known structures this interaction involves the hinge region between the Ig-domains (Figure 4(c)), which interacts with a side of the four-helix-bundle. Such a hinge does not exist for BARF1 due to the different orientation of its domains. The surprising orientation of the two immunoglobulin-like domains of BARF1 in respect to each other together with the unusual hexamerisation of the molecules does not agree well with the idea that BARF1 is a soluble form of a hCSF-1 receptor.[19] These observations can only be reconciled with the structure of BARF1 if we assume that during evolution BARF1 has acquired the function of hCSF-1 binding using a different interaction site than the one usually present within the family of receptors for four-helix bundle cytokines.

The sequence homology detected between BARF1 and the CSF-1 receptor tyrosine kinase *c-fms*[19] actually corresponds only to a general sequence homology within the Ig-constant domains as the conserved residues have a structural role. They are either part of the disulphide bridge, hydrophobic residues or proline residues stabilizing a loop structure (Figure 1(c)). The homology does not allow us to infer a common ligand-binding site.

Further functional studies using BARF1 mutants are necessary in order to establish the role of the BARF1–hCSF-1 interaction, in particular, whether the mitogenic effect of BARF1 is mediated through the interaction with hCSF-1 or whether there are still other, unknown partners in interactions with BARF1. This may lead to an explanation of the apparently contradictory actions of BARF1, on one hand as a mitogen and on the other hand a reduction of macrophage proliferation due to the sequestration of hCSF-1 from its receptor.

Putative ligand-binding site

For CD80, the N-terminal immunoglobulin domain is involved in the dimerization through one face and in the interaction with CD152 through the opposite face[42] (Figure 4(b) and (c)). Experimental evidence suggests that the interaction of CD80 with CD28[43,44] as well as the one between CD86 and CD28[45] use a similar surface. The equivalent of the CD152 binding surface of CD80 differs completely for BARF1 concerning sequence conservation and at the level of the structure of loop regions (FG, CC′) involved in the contact so that any interaction of BARF1 similar to the CD80–CD152 interaction can rather be excluded. Still, this area is accessible on the BARF1 hexamer and is thus a possible site of interaction.

Several arguments are in favour of a location of the ligand-binding site on the N-terminal domain. (i) In order to create the intricate dimerization site on the C-terminal domain, this domain has certainly been subject to more evolutionary pressure potentially incompatible with the preservation of a ligand-binding site on this domain. (ii) Part of the outer surface of the C-terminal domain is occupied by the disordered region 162–172. A direct involvement of this region is unlikely due to its O-glycosylation site and poor conservation with rLCV BARF1.

As it is unlikely that the binding site would be located inside the ring and as a large surface is already involved in oligomerization, the possible sites for ligand binding on the N-terminal domain are considerably narrowed down. A visual inspection of the outer surface shows a relatively hydrophobic area formed by residues Trp35, Val38, Leu40, Leu108 and the aliphatic part of Arg36 and Arg37 surrounded by several polar residues (Figure 5(a)). Such an arrangement of hydrophobic and hydrophilic residues is typical of a binding "hotspot".[46] Furthermore, these residues, besides a mutation Arg36Lys, are conserved between the rLCV and EBV BARF1 (Figure 5(b)). This makes this zone a good candidate for a putative binding site. This hypothesis may help to guide an identification of the interaction site by mutagenesis.

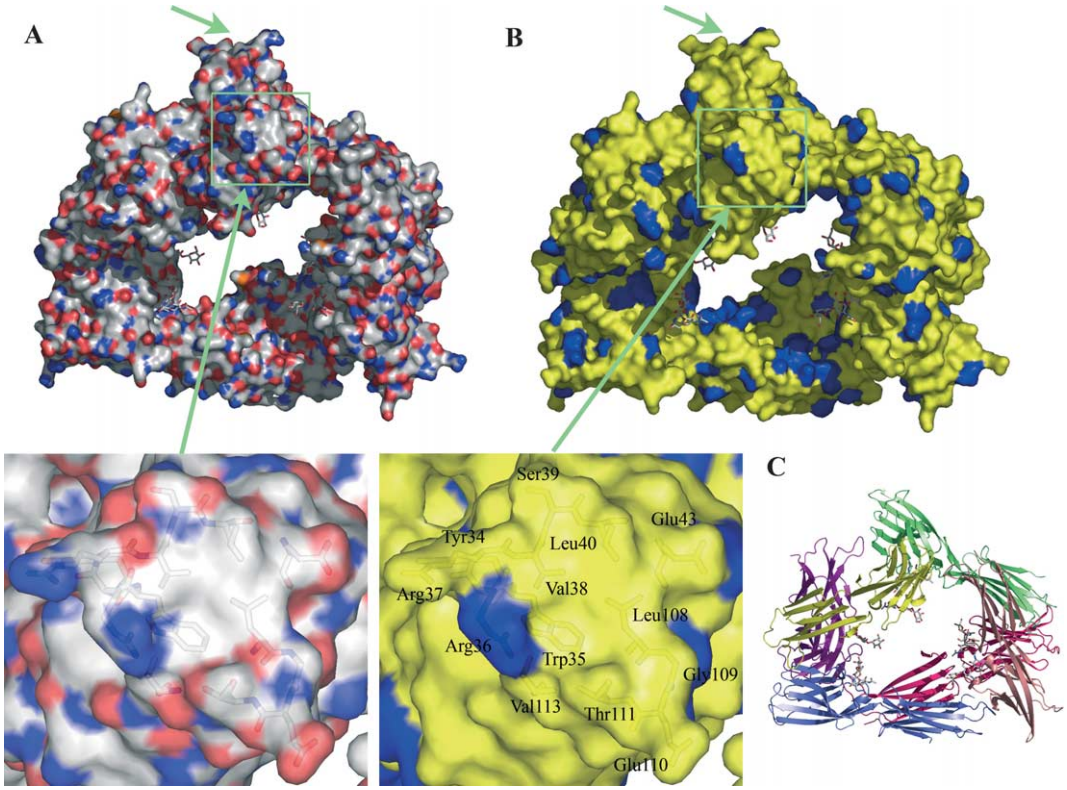


Figure 5. Surface and possible interactions of the BARF1 hexamer. (a) Surface coloured according to atom type (C, white; O, red; N, blue; S, orange). The green arrows show the visible putative ligand-binding sites of which the framed one is shown in detail in an inset. (b) Surface coloured in yellow for conserved residues and in blue for residues differing between EBV and rLCV BARF1. (c) View of a cartoon of the hexamer from the same viewpoint as the surface representations in (a) and (b) coloured as in Figure 2.

Materials and Methods

Protein production and purification

Protein was produced in Hela cells by a recombinant adenovirus expression system as described[21] and purified by concanavalin A affinity chromatography (M. de T.-T., unpublished results). The fractions containing the purified protein were concentrated and loaded on a Superdex 200 10/300 FPLC column (GE-Healthcare) equilibrated in 20 mM Hepes/NaOH (pH 7.5), 100 mM NaCl. The protein eluted at the position of a 240 kDa globular protein from a comparison with catalase (232 kDa) (GE-Healthcare). It was concentrated to 7 mg/ml using Ultrafree 0.5 concentrators with a 10 kDa cutoff (Millipore).

Electron microscopy

Purified BARF1 was dialysed against 0.2 M ammonium bicarbonate. Samples were diluted in the same buffer and mixed with glycerol (1:1) to obtain final concentrations of BARF1 ranging from 5 μg/ml to 20 μg/ml. A drop of the solution was placed onto freshly cleaved mica sheets, using the "mica sandwich" technique and immediately transferred to the holder of a MED 010 evaporator (Balzers). Rotary shadowing was carried out by evaporating platinum at an angle of 8°, followed by evaporation of carbon at 90°. Replicas were floated onto distilled water, picked up on copper grids and examined with a Philips CM120 microscope at the Centre Technique des Microstructures (Université Claude Bernard, Lyon I).

Mass spectrometry

MALDI mass spectrometry was performed with a Perseptive Biosystems (Framingham, MA) Voyager Elite XL time-of-flight mass spectrometer at the IBS, Grenoble.

Crystallization, data collection and reduction

Protein size and homogeneity were checked using a DynaPro dynamic light-scattering apparatus (Wyatt Technology). Extensive crystallization trials were performed with a PixSys4200 robot (Cartesian) using crystal screens I, II and Index (Hampton research) and the vapour diffusion method in drops of 0.1 μl protein solution and 0.1 μl reservoir solution. Successful crystallization conditions were refined by hand using 2 μl hanging drops leading finally to a crystallization condition with 1 M ammonium sulphate, 1–2% (w/v) PEG 3350, 100 mM BisTris–HCl (pH 6.0) in the reservoir.

The platinum derivative was prepared by soaking crystals overnight in the reservoir solution containing additional 1 mM tetrachloroplatinate (II) (Fluka). Single crystals were harvested, dipped into paraffin oil from the panjelly kit (Jena Biosciences) and frozen directly at 100 K in a nitrogen gas stream (Oxford Cryosystems). Data collection (Table 1) was performed at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) on the beamline ID14-4. Data were integrated using

MOSFLM[47] and further processed using the CCP4 package.[48] The crystals belonging to space group *H*3 were systematically twinned by merohedry described by the operator (*k*,*h*,−*l*) with twinning fractions varying from 0.1 to 0.45 according to DETWIN.[49] The data used for the structure determination and refinement showed a twinning fraction of 0.12 and were used without detwinning for phasing as this improved neither the maps nor the statistics, probably due to a degradation of the anomalous signal.

Phase determination and refinement

Phases were determined by the SAD method on the platinum derivative. The heavy-atom sites were found by SOLVE[50] at 3.2 Å resolution. A first model was built with RESOLVE using the iterative model-building script.[51] Further model building was performed using COOT[52] and model refinement was done with REFMAC5 using TLS refinement.[53] Refinement statistics and model composition are shown in Table 1. Refinement of the structure in CNS[54] using the refined twinning fraction of 0.12 improved neither statistics nor maps.

Structure analysis and comparisons

Structure alignments were performed with CCP4MG[55] or LSQMAN.[56] PyMol[57] was used for the generation of Figures. Interfaces were analysed with the protein–protein interaction server†[58] and visual inspection with PyMol in order to exclude marginal contacts.

Protein Data Bank accession code

Atomic coordinates have been deposited with the RCSB Protein Data Bank and are available under accession code 2CH8.

Acknowledgements

We are very grateful to Dr Joanne McCarthy (ESRF, Grenoble) for access to synchrotron beamtime, to the mass spectrometry team at the IBS (Grenoble) and to Dr Guy Schoehn (IVMS, Grenoble), Marilyne Malbouyres (Centre Technique des Microstructures, UCB, Lyon) and Simone Peyrol (CECIL, Faculté de Médecine, R.T.H. Laennec, Lyon) for electron microscopy. This project is funded by the European Commission as SPINE, contract No. QLG2_CT_2002_00988 under the integrated programme "quality of life and management of living resources", by the "Association pour la Recherche contre le Cancer" and by the "Ligue nationale contre le cancer".

References

1. Rochford, R., Cannon, M. J. & Moormann, A. M. (2005). Endemic Burkitt's lymphoma: a polymicrobial disease? *Nature Rev. Microbiol.* **3**, 182–187.
2. zur Hausen, H., Schulte-Holthausen, H., Klein, G., Henle, W., Henle, G., Clifford, P. & Santesson, L. (1970). EBV DNA in biopsies of Burkitt tumours and anaplastic carcinomas of the nasopharynx. *Nature*, **228**, 1056–1058.
3. De Thé, G. (1982). Epidemiology of Epstein Barr virus and associated diseases in man. In *The Herpesviruses* (Roizman, B., ed.), vol. 1, pp. 25–87, Plenum, New York.
4. Ooka, T., de Turenne-Tessier, M. & Stolzenberg, M. C. (1991). Relationship between antibody production to Epstein-Barr virus (EBV) early antigens and various EBV-related diseases. *Springer Semin. Immunopathol.* **13**, 233–247.
5. Bouzid, M., Djennaoui, D., Dubreuil, J., Bouguermouh, A., Ellouz, D., Abdelwahab, J. *et al*. (1994). Epstein-Barr virus genotypes in NPC biopsies from north Africa. *Int. J. Cancer*, **56**, 468–473.
6. Raab-Traub, N. (2005). Epstein barr virus in the pathogenesis of NPC. In *Epstein-Barr Virus* (Robertson, E. S., ed.), pp. 71–92, Caister Academic Press, Norfolk, UK.
7. Iwakiri, D. & Takada, K. (2005). Epstein-Barr virus and gastric cancers. In *Epstein-Barr Virus* (Robertson, E. S., ed.), pp. 157–169, Caister Academic Press, Norfolk, England.
8. Tarbouriech, N., Buisson, M., Geoui, T., Cusack, S. & Burmeister, W. P. (2006). Structural genomics of the Epstein-Barr virus. *Acta Crystallog. sect. D* in the press.
9. Wang, D., Liebowitz, D. & Kieff, E. (1985). An EBV membrane protein expressed in immortalized lymphocytes transforms established rodent cells. *Cell*, **43**, 831–840.
10. Wei, M. X. & Ooka, T. (1989). A transforming function of the BARF1 gene encoded by Epstein-Barr virus. *EMBO J.* **8**, 2897–2903.
11. Sheng, W., Decaussin, G., Sumner, S. & Ooka, T. (2001). N-terminal domain of BARF1 gene encoded by Epstein-Barr virus is essential for malignant transformation of rodent fibroblasts and activation of BCL-2. *Oncogene*, **20**, 1176–1185.
12. Decaussin, G., Sbih-Lammali, F., de Turenne-Tessier, M., Bouguermouh, A. & Ooka, T. (2000). Expression of BARF1 gene encoded by Epstein-Barr virus in nasopharyngeal carcinoma biopsies. *Cancer Res.* **60**, 5584–5588.
13. Seto, E., Yang, L., Middeldorp, J., Sheen, T. S., Chen, J. Y., Fukayama, M. *et al*. (2005). Epstein-Barr virus (EBV)-encoded BARF1 gene is expressed in nasopharyngeal carcinoma and EBV-associated gastric carcinoma tissues in the absence of lytic gene expression. *J. Med. Virol.* **76**, 82–88.
14. zur Hausen, A., Brink, A. A., Craanen, M. E., Middeldorp, J. M., Meijer, C. J. & van den Brule, A. J. (2000). Unique transcription pattern of Epstein-Barr virus (EBV) in EBV-carrying gastric adenocarcinomas: expression of the transforming BARF1 gene. *Cancer Res.* **60**, 2745–2748.
15. Danve, C., Decaussin, G., Busson, P. & Ooka, T. (2001). Growth transformation of primary epithelial cells with a NPC-derived Epstein-Barr virus strain. *Virology*, **288**, 223–235.
16. Wei, M. X., de Turenne-Tessier, M., Decaussin, G., Benet, G. & Ooka, T. (1997). Establishment of a monkey kidney epithelial cell line with the BARF1 open reading frame from Epstein-Barr virus. *Oncogene*, **14**, 3073–3081.

† http://www.biochem.ucl.ac.uk/bsm/PP/server

17. Wei, M. X., Moulin, J. C., Decaussin, G., Berger, F. & Ooka, T. (1994). Expression and tumorigenicity of the Epstein-Barr virus BARF1 gene in human Louckes B-lymphocyte cell line. *Cancer Res.* **54**, 1843–1848.
18. Sheng, W., Decaussin, G., Ligout, A., Takada, K. & Ooka, T. (2003). Malignant transformation of Epstein-Barr virus-negative Akata cells by introduction of the BARF1 gene carried by Epstein-Barr virus. *J. Virol.* **77**, 3859–3865.
19. Strockbine, L. D., Cohen, J. I., Farrah, T., Lyman, S. D., Wagener, F., DuBose, R. F. *et al*. (1998). The Epstein-Barr virus BARF1 gene encodes a novel, soluble colony-stimulating factor-1 receptor. *J. Virol.* **72**, 4015–4021.
20. Sall, A., Caserta, S., Jolicoeur, P., Franqueville, L., de Turenne-Tessier, M. & Ooka, T. (2004). Mitogenic activity of Epstein-Barr virus-encoded BARF1 protein. *Oncogene*, **23**, 4938–4944.
21. de Turenne-Tessier, M., Jolicoeur, P., Middeldorp, J. M. & Ooka, T. (2005). Expression and analysis of the Epstein-Barr virus BARF1-encoded protein from a tetracycline-regulatable adenovirus system. *Virus Res.* **109**, 9–18.
22. Cohen, J. I. & Lekstrom, K. (1999). Epstein-Barr virus BARF1 protein is dispensable for B-cell transformation and inhibits alpha interferon secretion from mononuclear cells. *J. Virol.* **73**, 7627–7632.
23. Ooka, T. (2005). Biological role of BARF1 gene encoded by Epstein-Barr virus. In *Epstein-Barr Virus* (Robertson, E. S., ed.), pp. 613–651, Caister Academic Press, Norfolk, England.
24. Rivailler, P., Jiang, H., Cho, Y. G., Quink, C. & Wang, F. (2002). Complete nucleotide sequence of the rhesus lymphocryptovirus: genetic validation for an Epstein-Barr virus animal model. *J. Virol.* **76**, 421–426.
25. Wang, F. (2005). Epstein-Barr virus related lymphocryptoviruses of old and new world nonhuman primates. In *Epstein-Barr Virus* (Robertson, E. S., ed.), pp. 691–709, Caister Academic Press, Norfolk, England.
26. Julenius, K., Molgaard, A., Gupta, R. & Brunak, S. (2005). Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology*, **15**, 153–164.
27. Van den Steen, P., Rudd, P. M., Dwek, R. A. & Opdenakker, G. (1998). Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **33**, 151–208.
28. Garcia De La Torre, J., Huertas, M. L. & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**, 719–730.
29. Ikemizu, S., Gilbert, R. J., Fennelly, J. A., Collins, A. V., Harlos, K., Jones, E. Y. *et al*. (2000). Structure and dimerization of a soluble form of B7-1. *Immunity*, **12**, 51–60.
30. Carreno, B. M. & Collins, M. (2002). The B7 family of ligands and its receptors: new pathways for costimulation and inhibition of immune responses. *Annu. Rev. Immunol.* **20**, 29–53.
31. Faas, S. J., Giannoni, M. A., Mickle, A. P., Kiesecker, C. L., Reed, D. J., Wu, D. *et al*. (2000). Primary structure and functional characterization of a soluble, alternatively spliced form of B7-1. *J. Immunol.* **164**, 6340–6348.
32. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallog. sect. D*, **60**, 2256–2268.
33. Popowicz, G. M., Muller, R., Noegel, A. A., Schleicher, M., Huber, R. & Holak, T. A. (2004). Molecular structure of the rod domain of dictyostelium filamin. *J. Mol. Biol.* **342**, 1637–1646.
34. Harris, L. J., Larson, S. B., Hasel, K. W., Day, J., Greenwood, A. & McPherson, A. (1992). The three-dimensional structure of an intact monoclonal antibody for canine lymphoma. *Nature*, **360**, 369–372.
35. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, **329**, 512–518.
36. Ostrov, D. A., Shi, W., Schwartz, J. C., Almo, S. C. & Nathenson, S. G. (2000). Structure of murine CTLA-4 and its role in modulating T cell responsiveness. *Science*, **290**, 816–819.
37. van Raaij, M. J., Chouin, E., van der Zandt, H., Bergelson, J. M. & Cusack, S. (2000). Dimeric structure of the coxsackievirus and adenovirus receptor D1 domain at 1.7 A resolution. *Structure*, **8**, 1147–1155.
38. Cerretti, D. P., Wignall, J., Anderson, D., Tushinski, R. J., Gallis, B. M., Stya, M. *et al*. (1988). Human macrophage-colony stimulating factor: alternative RNA and protein processing from a single gene. *Mol. Immunol.* **25**, 761–770.
39. Pandit, J., Bohm, A., Jancarik, J., Halenbeck, R., Koths, K. & Kim, S. H. (1992). Three-dimensional structure of dimeric human recombinant macrophage colony-stimulating factor. *Science*, **258**, 1358–1362.
40. Aritomi, M., Kunishima, N., Okamoto, T., Kuroki, R., Ota, Y. & Morikawa, K. (1999). Atomic structure of the GCSF–receptor complex showing a new cytokine-receptor recognition scheme. *Nature*, **401**, 713–717.
41. Syed, R. S., Reid, S. W., Li, C., Cheetham, J. C., Aoki, K. H., Liu, B. *et al*. (1998). Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature*, **395**, 511–516.
42. Stamper, C. C., Zhang, Y., Tobin, J. F., Erbe, D. V., Ikemizu, S., Davis, S. J. *et al*. (2001). Crystal structure of the B7-1/CTLA-4 complex that inhibits human immune responses. *Nature*, **410**, 608–611.
43. Peach, R. J., Bajorath, J., Naemura, J., Leytze, G., Greene, J., Aruffo, A. & Linsley, P. S. (1995). Both extracellular immunoglobin-like domains of CD80 contain residues critical for binding T cell surface receptors CTLA-4 and CD28. *J. Biol. Chem.* **270**, 21181–21187.
44. Sorensen, P., Kussmann, M., Rosen, A., Bennett, K. L., Thrige Dda, G., Uvebrant, K. *et al*. (2004). Identification of protein-protein interfaces implicated in CD80-CD28 costimulatory signaling. *J. Immunol.* **172**, 6803–6809.
45. Schwartz, J. C., Zhang, X., Fedorov, A. A., Nathenson, S. G. & Almo, S. C. (2001). Structural basis for co-stimulation by the human CTLA-4/B7-2 complex. *Nature*, **410**, 604–608.
46. DeLano, W. L., Ultsch, M. H., de Vos, A. M. & Wells, J. A. (2000). Convergent solutions to binding at a protein–protein interface. *Science*, **287**, 1279–1283.
47. Leslie, A. G. (1999). Integration of macromolecular diffraction data. *Acta Crystallog. sect. D*, **55**, 1696–1702.
48. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
49. Taylor, H. O. & Leslie, A. W. G. (1998). A program to detwin merohedrally twinned data. *CCP4 Newletter on Protein Crystallog.* **35**, 9.

50. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallog. sect. D*, **55**, 849–861.
51. Terwilliger, T. C. (2002). Automated structure solution, density modification and model building. *Acta Crystallog. sect. D*, **58**, 1937–1940.
52. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallog. sect. D*, **60**, 2126–2132.
53. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321.
54. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
55. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P. *et al*. (2004). Developments in the CCP4 molecular-graphics project. *Acta Crystallog. sect. D*, **60**, 2288–2294.
56. Kleywegt, G. J. & Jones, T. A. (1994). A super position. *ESF/CCP4 Newsl.* **31**, 9–14.
57. DeLano, W. (2002). *The Pymol Molecular Graphics System* (0.95 edit.), DeLano Scientifics, San Carlos, CA, USA.
58. Jones, S. & Thornton, J. M. (1996). Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.

Edited by R. Huber

(Received 23 December 2005; received in revised form 25 March 2006; accepted 29 March 2006)
Available online 18 April 2006