Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain

Marta Nadala, Philippe J. Mas, Alexandre G. Blanco, Carme Arnan, Maria Solá, Darren J. Hart, and Miquel Coll

*Institute for Research in Biomedicine, Barcelona Science Park, Baldiri Reixac 10, 08028 Barcelona, Spain; ‡Institut de Biologia Molecular de Barcelona (CSIC), Barcelona Science Park, Baldiri Reixac 10, 08028 Barcelona, Spain; and ‡European Molecular Biology Laboratory, Grenoble Outstation (UMI3265 UJF-EMBL-CNRS), BP 181, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

Edited* by Robert Huber, Max Planck Institute for Biochemistry, Planegg-Martinsried, Germany, and approved July 27, 2010 (received for review May 22, 2010)

During viral replication, herpesviruses package their DNA into the procapsid by means of the terminase protein complex. In human cytomegalovirus (herpesvirus 5), the terminase is composed of subunits UL89 and UL56. UL89 cleaves the long DNA concatamers into unit-length genomes of appropriate length for encapsidation. We used ESPRIT, a high-throughput screening method, to identify a soluble purifiable fragment of UL89 from a library of 18,432 randomly truncated ul89 DNA constructs. The purified protein was crystallized and its three-dimensional structure was solved. This protein corresponds to the key nuclease domain of the terminase and shows an RNase H/integrase-like fold. We demonstrate that UL89-C has the capacity to process the DNA and that this function is dependent on Mn²⁺ ions, two of which are located at the active site pocket. We also show that the nuclease function can be inactivated by raltegravir, a recently approved anti-AIDS drug that targets the HIV integrase.

Results

Expression of UL89 with a Library-Based Construct Screen. The two exons of the HCMV ul89 gene were cloned as a single DNA construct and initially tested for protein expression in several Escherichia coli strains and conditions. No protein obtained from these assays was stable enough to withstand purification. Similar results were obtained when the full-length gene was expressed in insect or mammalian cells. Extensive trials with refolding protocols were also unsuccessful. A number of constructs for each putative

Author contributions: M.N., M.S., D.J.H., and M.C. designed research; M.N., P.J.M., A.G.B., C.A., and M.C. performed research; M.N., M.S., D.J.H., and M.C. analyzed data; and M.N., M.S., D.J.H., and M.C. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The atomic coordinates for the UL89-C protein structures have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3N4P and 3N4Q).

1To whom correspondence should be addressed. E-mail: miquel.coll@irbbarcelona.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1007144107/-/DCSupplemental.
domain were designed, based on secondary structure prediction, globularity and disorder, but none of them expressed soluble protein in any system or condition assayed.

Subsequently, to find soluble domains, we used the combinatorial library method ESPRIT (23, 24), which generates comprehensive libraries of 5' or 3' truncated genetic constructs of the target. Both libraries were synthesized from the ul89 gene. We then screened 9,216 clones for each library form, corresponding to an approximate four-fold oversample of all possible domain boundaries, for expression of soluble protein. The two libraries were arrayed onto the same nitrocellulose membrane, and colonies were screened for putative soluble protein expression in colony format using measurements of in vivo biotinylation efficiency of a C-terminal biotin acceptor peptide by fluorescent streptavidin hybridization (23). Although a relatively large number of clones exhibited positive signals in the 3' truncation library, small-scale 4 mL liquid expression trials yielded only marginally soluble uninteresting fragments of less than 20 kDa in size. In contrast, the 5' truncation library yielded several partially soluble, purifiable constructs of similar size (approximately 37 kDa), from which the 48K22 construct was selected as showing the best behavior following scale-up testing (Fig. S1). This construct was only partially soluble (estimated at 5% of total UL89 protein), but was stable through scale-up to 12 L culture volumes and yielded approximately 1 mg of purifiable monodisperse protein per liter of culture. Other similar-sized constructs identified as partially soluble in small-scale testing did not maintain solubility during subsequent scale-up steps. Subsequent DNA sequencing and mass spectrometry fingerprinting identified construct 48K22 as a C-terminal fragment of UL89 (residues 418 to 674; Fig. S1), hereafter termed UL89-C. This fragment falls inside the predicted C-terminal nuclease domain, encoded in exon 2.

Overall Structure of UL89-C. UL89-C displays a wedged shape with dimensions 40 x 35 x 46 Å. A central eight-stranded mixed β-sheet, with parallel and antiparallel strands, is flanked by helices α on both sides (Fig. 1A). At one side, hydrophobic interactions pack α2 and α3 against the C-sheet. At the other side, helices α1, α4, α5, and α6 form a bunch that interacts with the β-sheet by hydrophobic interactions from one side of α5 and α6 and by hydrophilic contacts made by the α1 and α4 C-terminal ends. Two 3_10 helices, n1 and n2, at loops connecting β1 to β2 and α6 to β10, border one end of the β-sheet. The strand order in the central sheet is 1, 4, 2, 3, 5, 6, and 10 with topology +4, −1, −1, +3x, +1x, −5x, +6 (25) (Fig. 1C and Fig. S2). At both lateral edges of the β-sheet, β1 and β10 form short strands of only three amino acids each. At one end of the β-sheet, long loops surround a cleft that typically harbors the active site in proteins sharing this fold. One of these loops flanking the active site cavity folds in a twisted β-hairpin, formed by β7 and β8 (Fig. 1A).

UL89-C belongs to the RNase H-Like Superfamily. A search for structurally similar proteins revealed that UL89-C has the characteristic fold of the RNase H-like superfamily of nucleases and polynucleotidyl transferases (26). The closest structural relatives to UL89-C are the recently reported nuclease domains of the large terminase subunits of bacteriophages, RB49 and T4 gp17 (15) (RMSD 2.6 Å for 158 equivalent Cα and 2.7 Å for 159 equivalent Cα, respectively) and SPP1 G2P (16) (RMSD 3.0 Å for 146 equivalent Cα). The Holliday junction resolution RuvC (27) (RMSD 2.6 Å for 115 equivalent Cα), the HIV-integrase (28, 29) (RMSD 2.6 Å for 78 equivalent Cα) and the avian sarcoma virus integrase (30) (RMSD 2.9 Å for 85 equivalent Cα). The crystal structures of all these proteins and other members of the superfamily display the same basic fold but vary in length and show almost no amino acid sequence identity (i.e., 7.7% identity between UL89-C and the closest structural relative, the nuclease domain of RB49 gp17, after structural alignment).

The structural homology between these enzymes can be well described from the structural pattern of human RNase H1 (Hs-RNase H1) (31), which consists of a five-stranded β-sheet surrounded by α helices on both sides. The order and orientation of the strands within the β-sheet is conserved: 3, 2, 1, 4, and 5, one of them being antiparallel to the other four (1[111]). These strands are equivalent to the UL89-C β-strands 4, 3, 2, 5, and 6, respectively, whereas helices αA, αB, and αE Hs-RNase H1 correspond to helices α2, α3, and α6 of UL89-C. All these elements are arranged similarly in all proteins of the superfamily, except for α6 (αE in Hs-RNase H1), which runs in the opposite direction in UL89-C, RB49 gp17, SPP1 G2P, and RuvC with respect to the other members of the superfamily (Fig. S3). UL89-C (257 aa) is larger than the bacteriophage homologous proteins gp17 (206 aa) (15) and G2P (178 aa) (16). It is also larger and more complex than RNase H, integrase or resolvase nuclease domains, with the central β-sheet composed of 8 strands rather than 5, and further α helices and other secondary structure elements. Other members of the superfamily, like Tn5 transposase (32, 33) and Piwi–Argonaute (34), also have additional structural elements around the basic RNase H fold, although they are quite different from those found in UL89, thus reflecting their diverse functions, substrates and interactions with other proteins.

Active Site Cleft. The active site is located at one end of the central β-sheet in a cleft formed by conserved residues, four of them acidic (Fig. 1B). In all structures with a RNase-like fold, the active site is located at a topologically equivalent position, at one end of the β-sheet where two parallel β-strands (β2 and β5) separate in a fork-like manner. Asp463, Glu534, and Asp651 coordinate two metal cations (see above). Asp463 is located at the C-terminal end of β2 whereas Glu534 is present at the end of β5. Asp651 is found at the beginning of α6, the last α-helix in the structure, which lies diagonally to the two β-strands on one of the faces of the central β-sheet (Fig. 1). These three acidic amino acids are likely to play a role in the catalytic mechanism of the RNase-like superfamily.
acids are fully conserved and confer a strong electronegative character to the active site (Fig. 2 and Fig. S2). A further conserved aspartate residue, Asp650, is located close to the active site cleft (Fig. 1B) but does not interact directly with any of the metal ions. Comparison of the active site of UL89-C with other RNase-like nucleases (Fig. S3) shows that the presence of several acidic residues coordinating metal ion is a signature of the superfamily, in particular the central residue is always an aspartate (Asp463 in UL89-C). The other residues coordinating the metal may vary. For example in bacteriophage SPP1 G2P (16), one of the closest structural relatives to UL89, an aspartate residue (Asp521) occupies a position equivalent to Glu534 in UL89-C and a histidine residue (His400) that of Asp561. However, in T4 and RB49 gp17 (16) the residues coordinating the metal ion are identical to those of UL89-C (Fig. S3). Moreover, in these structures an aspartate residue occupies the equivalent position of Asp650 in UL89-C, although two additional acidic residues of the active pocket present in the bacteriophage structures, Asp406 and Glu401 in RB49 gp17, are not present in UL89-C.

The Active Site Accommodates Two Cations. In the crystal not soaked with MnCl₂, one metal ion was clearly identified at the active site. This corresponds to metal B, as defined by Nowotny and Yang (35). In molecule D of this crystal, an electron density peak initially assigned as a water molecule (the strongest peak of the water list) could also correspond to another metal ion located at a second position with low occupancy. Indeed, an anomalous difference map calculated from diffraction data from a crystal soaked with MnCl₂ showed two peaks at these two positions (Fig. S4). The first one is coordinated by Asp463 and Glu534 and the second by Asp463 and Asp 651 (Fig. 1B). Asp463, Glu534 and Asp651 (and the closest residues Pro464, Ala465, Gly535, Asn536, and Asp650) are fully conserved among human herpesvirus terminases (Fig. S2), thereby suggesting that they are essential for cation coordination and thus for catalysis. Indeed, Mg²⁺ or Mn²⁺ cations are required for the functioning of these enzymes and a two-metal catalysis has been proposed for their enzymatic mechanism (35, 36).

In Vitro Nuclease Assays and Mutants. An in vitro assay demonstrated that UL89-C has the capacity to degrade linear and circular DNA and this function is strongly activated by Mn²⁺ (Fig. 3A and B). In the presence of this cation, UL89-C converts supercoiled circular plasmid DNA to nicked open circular DNA, subsequently to linear DNA and finally to completely degraded DNA (Fig. S5). Similarly, UL89-C also degrades linear DNA (Fig. 3A and C). Similar behavior was previously described for the UL89 full-length protein (5). The reaction performed in the same conditions but in the presence of Mg²⁺ instead of Mn²⁺ converted only supercoiled circular plasmid to nicked open circular DNA. With Ca²⁺, the DNA degradation was even less efficient (Fig. 3A and B). To verify that the residues of the structurally inferred active site were truly involved in the nuclease activity of the protein, we designed a set of single and double mutants and tested their activity. The single mutants D463A, D651A, and the double mutant D463A/E534A showed only residual nuclease function. Mutant proteins show only residual activity (Fig. 3C and D). These results confirmed that UL89-C harbors the nuclease activity critical for the function of the full-length protein.

Inactivation by Raltegravir. The structural similarity between the herpesvirus terminase nuclease domain and the HIV integrase prompted us to test the inhibitory properties of integrase inhibitors on UL89-C. One of these integrase inhibitors, raltegravir (MK0518), was approved by the FDA in 2007 for the treatment of AIDS (37). Raltegravir turned out to be a strong inhibitor of
the nucleolytic activity of UL89-C (Fig. 4). A recent structure of the prototype foamy virus integrase in complex with DNA and the inhibitor shows that raltegravir binds at the active site, directly coordinating the metal ions (38). Presumably, it would bind in a similar way to UL89-C. In contrast to raltegravir, another integrase inhibitor, elvitegravir (GS9137), had no inhibitory effect on UL89-C under similar conditions.

Discussion

A Powerful Construct Screening Technique to Obtain UL89-C. UL89, like other herpesvirus DNA packaging proteins, is scarcely expressed in a soluble, purifiable form. Even using insect or mammalian eukaryotic expression systems, we and others were unable to purify a soluble form of this protein in sufficient amounts for crystallographic or even limited proteolysis studies. The ESPRIT (23, 24) analysis reported here permitted the oversampling of all possible domain boundaries as hexahistidine tag fusion positions. However, even this approach resulted in a very low number of soluble expression constructs. This finding is indicative of the challenging nature of UL89. The resulting construct encoding the UL89 C-terminal domain expressed protein that was partially soluble, but the purifiable material was monodisperse and well-behaving through subsequent concentration and crystallization steps. The identification of this otherwise obscure expression-compatible construct is illustrative of the power of this technique to find rare soluble forms of difficult proteins, and this approach appears particularly effective for viral proteins with uncertain domain boundaries (39).

DNA Binding. UL89-C cleaves dsDNA in vitro (Fig. 3), as reported previously for the full-length protein (5). This domain should bear the structural determinants for DNA binding. An electrostatic surface calculation indicates that a number of positively charged residues, located in different loops, surround the active site cleft (Fig. 2B). From this calculation, the shape of the surface, and superpositions with Bacillus halodurans (40) and human RNase H structures in complex with a DNA/RNA hybrid (31) and Tn5 transposase in complex with DNA (32), we manually built a model for dsDNA bound to UL89-C (Fig. S6). In this model, the loops Lβ2-β3 and Lβ5-α3 fit into the major groove of the DNA, whereas positively charged side chains appear in close proximity to the phosphates. The sugar-phosphate backbone enters the active site but does not get close enough to the metal ion positions. A distortion (bent) from the regular straight B-DNA conformation (albeit probably distorted) where the minor groove is too narrow to permit the entrance of loops Lβ2-β3 and Lβ5-α3. Therefore, interaction with the bases, if present, would be performed through the major groove. Indeed, loop Lβ5-α3 is longer and more protruding than its RNase H equivalent. The shallow RNA/DNA hybrid minor groove cannot accommodate this loop, but a deeper B-DNA major groove would fit (Fig. S6).

UL89-C Within the Terminase Complex. In the crystal structure, UL89-C shows four protein molecules in the asymmetric unit, A, B, C, and D. Molecules A and B interact with each other about a local two-fold axis, as do molecules C and D (Fig. S7). The interaction surface is at the edge of the central β-sheet so that the sheet extends from one protein to its neighbor. Although UL89 dimers have been detected by cross-linking and gel filtration of the full-length protein (8), UL89-C eluted as a monomer in the size-exclusion chromatography. Thus, with the data available, it is unclear whether the dimer observed in the crystal structure has any physiological relevance or whether it is due to crystal packing. Furthermore, phage and herpesvirus terminases are believed to form toroidal structures and assemble as such like the 12-fold portal protein (5, 10), for which a dimer like that observed in the crystal structure of UL89-C would not fit. It has been demonstrated by cryo-EM that the phage T4 gp17, a homolog of UL89, forms pentamers (15). In the present structure there is no evidence of ring formation and it is likely that the oligomerization determinants for such an arrangement are outside the UL89-C domain.

UL89 interacts with the UL56 subunit of the terminase. On the basis of results from deletion experiments, the amino acids of UL89 proposed to be involved in this interaction span from residues 580–600 (8). This segment corresponds to the exposed helix α4 (Fig. 1 and Fig. S2) and is thus suitable for interaction with UL89 partners. The segment includes three residues that are fully conserved among human herpesvirus, namely Lys583, Ala586, and Asn595. This observation suggests a similar interaction scheme within the family. Furthermore, helix α4 has no counterpart in RNase H or integrases, which are enzymes that do not interact with any protein equivalent to UL56.

UL89 as a Drug Target. Viral DNA encapsidation machinery has no counterpart in the mammalian cell, thus implying that the proteins involved in this process represent promising selective targets for antiviral therapy. Several studies have reported that inhibitors of DNA packaging in herpesviruses specifically target UL89 and UL56, although the binding sites of the proteins have not been elucidated (18–22). Our study demonstrates that the UL89 C-terminal domain of HCMV and the equivalent domains in all herpesviruses bear the essential nuclease function of the terminase for DNA packaging (Fig. 3). We reveal the three-dimensional structure of this domain in detail and describe the essential residues for the nuclease function, which we demonstrate can be inhibited by raltegravir, an HIV integrase inhibitor approved by the FDA for AIDS treatment in October 2007 (37). This study therefore opens a way for the design of further optimized inhi-
bitor against UL89-C that may be useful for the development of unique antiviral drugs.

Materials and Methods

Identification of the UL89-C Soluble Construct from a Complete 5' and 3' Gene Truncation Library. The UL89 gene from the HHV-5 genome comprises two exons; these were amplified, cloned separately, and subsequently ligated together. The library was constructed as described (23). Briefly, for the 5' deletion library, the gene was cloned into a pET9a-derived vector out of frame with a tobacco etch virus cleavable N-terminal hexahistidine tag (MGHHHHHHDDYITPPNTENQFGQ) and in frame with a short linker and C-terminal biotin acceptor peptide (SNNGGSGLLDIEAKEYQWHE). The presence of AatII and Ascl sites between the hexahistidine tag encoding DNA and the UL89 gene permitted unidirectional truncation of the 5' end of the gene using an AatII and AsclI. Hexahistidine tag fusions of the truncated gene were generated following recircularization of the plasmid with T4 DNA ligase. Following transformation, the plasmid library was harvested from the E. coli cloning strain (Omnimax T1; Invitrogen) and used to transform BL21-CodonPlus-RIL (Stratagene). Robotic processing of the library to identify putative soluble expression constructs was done as described (23, 24). Briefly, 18,432 colonies comprising 9,216 for the 5' and 3' deletion libraries were picked robotically into microtiter plates of TB broth and grown overnight. These were gridded robotically onto nitrocellulose membranes over LB agar plates and then robotically picked into microtiter plates of TB broth and grown overnight. The 96 most intense positive clones of each library were isolated from the fluorimager was used to identify colonies expressing biotinylated proteins. These were gridded robotically onto nitrocellulose membranes over LB agar plates and then robotically picked into microtiter plates of TB broth and grown overnight. The 96 most intense positive clones of each library were isolated from the fluorimager and were expressed and purified with similar efficiencies.

Wild-type and mutant proteins were expressed and purified with similar efficiencies. Purified wild-type and mutant UL89-C domains (final concentration 2 μM) were incubated with 200 ng of circular and linear (digested with HindIII) pUC18 plasmid (2,686 bp) in a reaction containing 30 mM Tris pH 8 and 50 mM NaCl for 1 h at 37 °C. The effect of several metal ions was studied by adding 3 mM (final concentration) MgCl₂, CaCl₂, or MnCl₂. The activity was terminated by adding EDTA to a final concentration of 30 mM. The samples were analyzed by agarose gel electrophoresis with ethidium bromide staining. For the inhibitory assay, a range of concentrations of raltegravir (Chemietek) were added to the reaction. A stock solution of 5 mM raltegravir was prepared at 50% DMSO and was further diluted with 30 mM Tris pH 8, 50 mM NaCl to obtain the final concentration.

Crystallization and Heavy-Atom Derivatization. Protein UL89-C was crystallized by mixing 2 μL of protein solution containing 10 mg/mL of UL89C, 30 mM Tris buffer (pH 8), 50 mM NaCl and 5 mM EDTA with 2 μL of precipitant solution containing 10% (w/v) polyethylene glycol 8000, 150 mM calcium acetate hydrate and 100 mM Mes (pH 6.8), using the sitting drop vapor diffusion method.

Crystals were flash-cooled in 12 % polyethylene glycol 400 as cryoprotectant. To prepare Mn²⁺-derivatised crystals, native crystals were soaked for 1 h in the crystallization solution enriched with 50 mM MnCl₂. To prepare Hg heavy-atom derivatives, native protein crystals were soaked for 24 h in the crystallization solution enriched with 0.5 mM ethylmercuritiochisalicylic acid sodium salt.

Structure Solution and Refinement. A native dataset was collected at the ESRF ID14-2 beamline to a resolution of 2.15 Å. Crystals belonged to space group P2₁2₁2₁ with cell dimensions a = 82.8, b = 87.9, c = 189.4 Å and α = β = γ = 90 °C. A dataset from a native crystal soaked with Mn²⁺ was collected at ESRF ID29; the crystals belonged to the same space group with the similar cell dimensions. In addition, data for an Hg-derivative were collected at ESRF BM16, at a wavelength of 1.00726 Å (Hg-bound absorption edge). Native and derivative diffraction data were processed using XDS (41), and then scaled, reduced and merged with XSCALE (41) (Table S1). Phases were obtained by single isomorphous replacement anomalous scattering (SIRAS). SHARP (42) was used to determine the positions of 5 Hg atoms using data to 3.5 Å, and phasing the data to 2.15 Å. The resulting map was of insufficient quality for automatic tracing and most of the polypeptide chain had to be built manually using Coot (43). The crystals contained four UL89-C molecules per asymmetric unit. Atomic positions and their associated B-factors were refined with Refmac5 (44) using noncrystallographic symmetry restraints. The model was improved by alternating cycles of automatic refinement and interactive model building (Fig. S5). The final refinement cycles included TLS refinement. The Mn²⁺-soaked crystal structure showed two strong electron density peaks at the active site corresponding to the metal ions (Fig. S4). These ions were included and the structure refined with Refmac5. The quality of the stereochemistry of the two structures was assessed with Procheck (45) (Table S1).

In Vitro Nuclease Assay. Purified wild-type and mutant UL89-C domains (final concentration 2 μM) were incubated with 200 ng of circular and linear (digested with HindIII) pUC18 plasmid (2,686 bp) in a reaction containing 30 mM Tris pH 8 and 50 mM NaCl for 1 h at 37 °C. The effect of several metal ions was studied by adding 3 mM (final concentration) MgCl₂, CaCl₂, or MnCl₂. The activity was terminated by adding EDTA to a final concentration of 30 mM. The samples were analyzed by agarose gel electrophoresis with ethidium bromide staining. For the inhibitory assay, a range of concentrations of raltegravir (Chemietek) were added to the reaction. A stock solution of 5 mM raltegravir was prepared at 50% DMSO and was further diluted with 30 mM Tris pH 8, 50 mM NaCl to obtain the final concentration.

Miscellaneous. Search for folding relatives was performed with MATRAS (46). Structural alignments and RMSD calculations were performed with SSM (47). Fig 1 and Figs. 53, 54, 56, 57, and 58 were drawn with Pymol (48). Fig. 2 was generated with GRASP (49) and Pymol.

ACKNOWLEDGMENTS. We thank Ms. Jenny Colom for help during protein purification, Dr. Jordi Bernués for help during nuclease assays, Ms.uzzana Kaczmarska for elvitegravir inhibition assays and Prof. David Stuart for hosting M.N. to perform expression tests in eukaryotic systems. This study was supported by Ministerio de Ciencia e Innovación Grant BFU2008-0237/28/M.C., Generalitat de Catalunya Grant 2009 SGR 1309 (M.C.), and the European Commission (Spin2-Complexes LS6G-CT-2006-031220). Synchrotron data collection was supported by the European Synchrotron Radiation Facility and the European Union. Crystallographic screening and preliminary X-ray analysis were performed at the Automated Crystallography Platform, Barcelona.


Correction

BIOCHEMISTRY


The authors note that the author name Phillipe J. Mas should have appeared as Philippe J. Mas. The corrected author line appears below. The online version has been corrected.

Marta Nadal, Philippe J. Mas, Alexandre G. Blanco, Carme Arná, Maria Solà, Darren J. Hart, and Miquel Coll

www.pnas.org/cgi/doi/10.1073/pnas.1013165107