Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain

Marta Nadala,b, Philippe J. Mas1, Alexandre G. Blanconb, Carme Arnanab, Maria Solab, Darren J. Hartc, and Miquel Colla,b,1

Abstract
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 concatemers into unit-length genomes of appropriate length for encapsidation. We used ESFRIT, 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 nucleic function can be inactivated by raltegravir, a recently approved anti-AIDS drug that targets the HIV integrase.

H
uman cytomegalovirus (HCMV) is a member of the herpes family of viruses or Herpesviridae. This group includes the human pathogens herpes simplex virus type 1 and 2, varicella zoster virus, Epstein–Barr virus, cytomegalovirus, roseolovirus, and Kaposi sarcoma-associated herpesvirus. Among these, HCMV, which belongs to the Betaherpesviridae subfamily, is widespread throughout the human population and causes the most morbidity and mortality. HCMV infection is rarely serious for the host cells and propagates to other individuals. In contrast, infection or reactivation of HCMV is a major cause of life-threatening complications in immunocompromised individuals, such as organ transplant recipients and leukemia or AIDS patients, and is the most significant viral cause of birth defects in industrialized countries (1).

The HCMV genome consists of linear dsDNA of 230 kb with the highest coding capacity among Herpesviridae. HCMV, like all other herpesviruses, replicates its genomic DNA into high molecular mass head-to-tail concatamers. The newly synthesized multicopy chains of DNA are then excised into unit-length genomes and each genome is packaged singly into one viral procapsid (2). Maturation of unit-length genome molecules involves viral DNA recognition and cleavage at the site-specific pac motifs, which are redundant motifs found at both the 5′ and 3′ genomic termini (2, 3). The dsDNA endonucl ease and packaging activities are performed by a protein complex, the terminase, composed by subunits UL56 and UL89 (4, 5). UL56 has been reported to recognize the pac motif (6).

After cleaving one end of the DNA, the terminase translocates the viral DNA into the procapsid, deriving energy for this process through ATP hydrolysis. The procapsid is filled with the DNA molecule and the terminase performs a second dsDNA cleavage, thereby concluding the translocation (3). UL89 has predicted ATPase activity and is most probably the molecular motor for helicase-like DNA translocation (7). In addition, UL89 binds and cleaves DNA molecules, and this activity is enhanced when UL56 is present (5). In agreement, other studies have shown that UL89 interacts specifically with the C-terminal part of UL56 (8). UL56 has the capacity to bind linearized DNA but only upon addition of UL89 is the DNA cut into smaller fragments. This observation indicates that these proteins mediate a concerted reaction of DNA recognition and cleavage (5).

Proteins homologous to UL89 are known in all herpesviruses, as are the other terminase subunits, thus indicating that the DNA packaging mechanism is highly conserved (9). The ul89 ORF includes two exons separated by a 3,902 bp intron. It encodes a two-domain 674 amino acid protein with predicted N-terminal ATPase and C-terminal nuclease activity (Fig. S1) (7). Bacteriophages translocate shorter DNA molecules into their capsids by similar packaging systems (9). Their terminases have been intensively studied, in particular those of phages T4 and RB49 (10), λ (11), SPP1 (12), P22 (13), and Sf6 (14), and the structure of the nuclease domains of the large terminase subunits gp17 of RB49 (15) and G2P of SPP1 (16) and the full-length large terminase subunit gp17 of T4 (15) have been determined. A theoretical model for the structure of the C-terminal domain of UL89 has been proposed recently (17).

Several studies report inhibitors that prevent the formation of new virions through blockage of the termination system (18–22). However, the structural and functional characterization of herpes packaging proteins, which could assist further discovery and development of antiviral molecules, has been hindered by the difficulties in expressing enough soluble material for structural analysis. We have overcome these problems by using ESFRIT (23, 24), a combinatorial library method for defining soluble constructs through random gene truncation and expression screening. This approach yielded a single active soluble construct corresponding to the C-terminal nuclease domain of UL89, the structure of which we report herein at a resolution of 2.15 Å.

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 Excherichia 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
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 × 35 × 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 C2, 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 C4 C-terminal ends. Two 310 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, 9, 4, 3, 2, 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 Cβ and 2.7 Å for 159 equivalent Cβ, respectively) and SPP1 G2P (16) (RMSD 3.0 Å for 146 equivalent Cβ), the Holliday junction resolvase RuvC (27) (RMSD 2.6 Å for 115 equivalent Cβ), the HIV-integrase (28, 29) (RMSD 2.6 Å for 78 equivalent Cβ) and the avian sarcoma virus integrase (30) (RMSD 2.9 Å for 85 equivalent Cβ). 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 (11111). 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 also is 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 Ta5 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 below). 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 indicated by yellow spheres. UL56-interacting helix α4 is highlighted in blue, α helices in cyan, 310 helices in green, and β strands in magenta. (B) Detailed view of the active site area. One Mn2+ ion is shown coordinated by Asp463 and Glu534 and a second one is coordinated by Asp463 and Glu651. (C) Topology diagram of UL89. Color code as in (A).
In molecule soaked with MnCl₂ difference map calculated from diffraction data from a crystal at a second position with low occupancy. Indeed, an anomalous water list could also correspond to another metal ion located initially assigned as a water molecule (the strongest peak of Mg⁴⁺ 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. 3 A 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 activity (Fig. 3 C 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 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. 3 A 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. 3 A and C). Similar behavior was previously described for the UL89 full-length protein (5). The reaction performed in the presence of the full-length protein.

![Figure 2](http://www.pnas.org/cgi/doi/10.1073/pnas.1007144107)

**Fig. 2.** UL89-C surface representations. (A) Conservation surface of UL89-C based on a Risler matrix calculation. Residue color is shown on the basis of conservation score within the eight human herpesvirus: orange, more than 80% similarity; green, full conservation. Metal ions are shown in yellow. (B) Electrostatic potential surface of UL89-C in the same view as (A). The positive potential is shown in blue whereas the negative potential is in red.

![Figure 3](http://www.pnas.org/cgi/doi/10.1073/pnas.1007144107)

**Fig. 3.** UL89-C in vitro nuclease assay. (A) Effect of divalent cations on UL89-C linear dsDNA nuclease activity. Lane 1: linear (digested with HindIII) pUC18 plasmid in the absence of the nuclease. Lane 2: nuclease reaction in the absence of divalent cations. Lane 3: nuclease reaction in the presence of Mg²⁺. Lane 4: nuclease reaction in the presence of Ca²⁺. Lane 5: nuclease reaction in the presence of Mn²⁺. (B) Effect of divalent cations on UL89-C circular dsDNA nuclease activity. Lane 1: circular pUC18 plasmid in the absence of the nuclease. Lane 2: linear (digested with HindIII) pUC18 plasmid in the absence of the nuclease. Lane 3: nuclease reaction in the absence of divalent cations. Lane 4: nuclease reaction in the presence of Mg²⁺. Lane 5: nuclease reaction in the presence of Ca²⁺. Lane 6: nuclease reaction in the presence of Mn²⁺. (C) Nuclease assays with the UL89-C WT and mutant proteins with linear dsDNA. Lane 1: linear (digested with HindIII) pUC18 plasmid in the absence of the nuclease. Lane 2: nuclease assay with the WT proteins. Lane 3: nuclease assay with the D463A mutant. Lane 4: nuclease assay with the D651A mutant. Lane 5: nuclease assay with the D463A/E534A mutant. (D) Nuclease assay with the UL89-C WT and mutant proteins with ds-circular DNA. Lane 1: circular pUC18 plasmid in the absence of the nuclease. Lane 2: nuclease assay with the WT protein. Lane 3: nuclease assay with the D463A mutant. Lane 4: nuclease assay with the D651A mutant. Lane 5: nuclease assay with the D463A/E534A mutant. Mutant proteins show only residual nuclease function. Labels: sc, supercoiled circular plasmid DNA; ni-oc, nicked open circular plasmid DNA; li, linear plasmid DNA.
the nuclease 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 crystalllographic 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 (33), 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 used in the model would be necessary for the scissile phosphate to reach the metal ions without clashes of the DNA with the active site surrounding loops. In the bacteriophage large terminase structures, some of these loops (i.e., β2-β3 and β5-α3) are shorter or less protruding, resulting in a slightly less deep active site. However, Smits et al. (16) reported that a protruding β-hairpin in the SPP1 G2P structure would clash with the DNA and suggested that the conformation of this loop changes upon DNA binding. This loop corresponds to loop α5-α6 in UL89-C, which was disordered and not visible in our structure, in agreement with its proposed flexibility.

In RNase H, the equivalent loops to UL89 Lβ2-β3 and Lβ5-α3 contact the minor groove of the RNA/DNA hybrid, instead of the major groove of the dsDNA, as in our DNA-UL89-C docking model. This observation is not contradictory because the conformation of the RNA/DNA hybrid is a mixture between A and B forms, where the minor groove is wider and the bases are accessible. However, the viral dsDNA is most likely in the B-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 at 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 550–600 (8). This segment corresponds to the exposed helix 4 (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 4 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’ Gene Truncation Library. The U89 gene from the HHV5 tone strain 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 (MGHHHHHHDDYITPPENLYFQG) and in frame with a shorter linker and C-terminal biotin acceptor peptide (SNNGGSGGLNDIFEAQKIEWHE). The presence of Aartf and AscI sites between the hexahistidine tag encoding DNA and the U89 gene permitted unidirectional truncation of the 5’ end of the gene using an AscI linker. 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-Codon-Plus-RIL (Stratagene). Robotically 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.

Crystallization and Heavy-Atom Derivatization. Wild-type and mutant proteins were expressed and purified with similar soluble expression constructs as described (23, 24). Briefly, Purified wild-type and mutant UL89-C domains were digested with HindIII and cloned into pUC18 plasmid (2,686 bp) in a reaction containing 300 mM Tris pH 8, 50 mM NaCl to obtain the final concentration. 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.

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, 50 mM NaCl, and 50 mM NaCl. Elution was achieved using a linear salt gradient of 50 mM-1 M NaCl. Fractions were analyzed by SDS-PAGE and those that showed higher purity were pooled, concentrated, and subjected to size-exclusion chromatography using a Superdex 75 10/300 GL column (GE Healthcare). The elution was performed at 10.5 mL, corresponding to 31 kDa. Wild-type and mutant proteins were expressed and purified with similar efficiencies. 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 Mesi (pH 8), using the sitting drop vapor diffusion method.

Crystals were flash-cooled in 12 % polyethylene glycol 400 as cryoprotectant. To prepare Mn+2 -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 ethylenediaminetetraacetic 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 and β = γ = 90 °C. A dataset from a native crystal soaked with Mn+2 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 5). 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) usingnoncrystallographic symmetry restraints. The model was improved by alternating cycles of automatic refinement and interactive model building (Fig. 5). The final refinement cycles included TLS refinement. The Mn+2 -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).


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\textsuperscript{a,b}, Philippe J. Mas\textsuperscript{c}, Alexandre G. Blanco\textsuperscript{a,b}, Carme Arnan\textsuperscript{a,b}, Maria Solà\textsuperscript{b}, Darren J. Hart\textsuperscript{c}, and Miquel Coll\textsuperscript{a,b,1}

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