Assemblins as maturational proteases in herpesviruses

Martin Zühlsdorf and Winfried Hinrichs*

Abstract

During assembly of herpesvirus capsids, a protein scaffold self-assembles to ring-like structures forming the scaffold of the spherical procapsids. Proteolytic activity of the herpesvirus maturational protease causes structural changes that result in angularization of the capsids. In those mature icosahedral capsids, the packaging of viral DNA into the capsids can take place. The strictly regulated protease is called assemblin. It is inactive in its monomeric state and activated by dimerization. The structures of the dimeric forms of several assemblins from all herpesvirus subfamilies have been elucidated in the last two decades. They revealed a unique serine-protease fold with a catalytic triad consisting of a serine and two histidines. Inhibitors that disturb dimerization by binding to the dimerization area were found recently. Additionally, the structure of the monomeric form of assemblin from pseudorabies virus and some monomer-like structures of Kaposi’s sarcoma-associated herpesvirus assemblin were solved. These findings are the proof-of-principle for the development of new anti-herpesvirus drugs. Therefore, the most important information on this fascinating and unique class of proteases is summarized here.

INTRODUCTION

Herpesviruses are amongst the oldest and biggest viruses known to date. Still, they are not fully understood in various ways, although they have been subject to extensive research for decades [1, 2]. A large number of herpesviruses infect hosts ranging from mammals and birds to reptiles, but these viruses are quite host specific [3, 4]. An important aspect of the herpesviruses is their ability to establish a latent infection in their natural hosts [5]. The balance between the immune response and the latent or persistent infection is a typical feature of herpesvirus infection. If the immune response is disordered, this balance is lost and the viruses cause severe disease [6].

The order Herpesvirales consists of three families: Alloherpesviridae, Malacoherpesviridae and Herpesviridae. The latter is mostly referred to as herpesviruses and is divided into three subfamilies: alpha-, beta- and gammaherpesvirinae. There are eight herpesviruses that are pathogenic to man. The human herpesvirus 1, human herpesvirus 2 and human herpesvirus 3 are better known as herpes simplex virus 1 (HSV-1), herpes simplex virus 2 (HSV-2) and varicella zoster virus (VZV). These viruses are alphaherpesviruses and cause cold sores, genital herpes, and chickenpox and shin- gles, respectively. Human herpesvirus 5 (better known as human cytomegalovirus (HCMV), human herpesvirus 6 and human herpesvirus 7 are betaherpesviruses. Infections with these viruses are mostly asymptomatic in healthy individuals. Human herpesvirus 4 and human herpesvirus 8 are better known as Epstein–Barr virus (EBV) and Kaposi’s sarcoma-associated herpesvirus (KSHV), respectively. They are the causative agents of glandular fever and Kaposi’s sarcoma. KSHV and EBV are known to cause cancer, especially in immune-suppressed individuals, e.g. AIDS patients.

One of the essential steps in herpesvirus replication involves the assembly and processing of the viral capsids in the host’s nuclei. The icosahedral capsid of mature herpesviruses consists of several different proteins. The most important and frequent protein is the major capsid protein (MCP). It forms ring-like pentamers and hexamers, the so-called pentons and hexons, respectively [7–10]. The 12 vertices of the icosahedral capsid consist of 11 pentamers and the portal. The portal protein constitutes the portal as a ring-like dodecamer [11]. Detailed knowledge on how the singularity of the portal is ensured is not available, but there is evidence that it is incorporated at initiation or at an early step of capsid formation [12, 13]. The edges and planes of the icosahedron comprise 150 hexons [14, 15].

The assembly of capsids occurs in the nuclei of host cells. The MCP and the portal protein are neither able to enter the nucleus nor able to form icosahedral capsids on their own. Therefore, there are two proteins that support the MCP and the portal protein in these processes [16–19]. On
a genetic level, these two proteins are encoded by two in-frame, co-terminal genes, which are conserved in all herpesviruses [20–23]. The coded viral proteins are referred to as protease precursor (pPR, derived from the longer gene) and assembly protein precursor (pAP, derived from the shorter gene and identical to the C-terminal domain of pPR). Both carry nuclear localization signals (NLSs) and their identical C-termini can bind to the MCP and the portal protein (Fig. 1a). They translocate both proteins into the nucleus and direct them to the sites of capsid assembly [16–19]. These sites are called ‘assemblons’ [24].

The longer protein (pPR) consists of two domains: an N-terminal serine-protease domain (assemblin) and a C-terminal scaffold domain. The shorter protein (pAP) is similar to the C-terminal scaffold domain of pPR. At the assemblons, pPR and pAP auto-assemble into hemispheres and spheres with the approximate diameter of the capsid via their scaffold domain (pAP without MCP/portal: 60 nm) [25–28]. There is evidence that this self-association is mediated by coiled-coil interactions of the scaffold domains. The same region is also supposed to be important for the interaction with the MCP [28]. Self-associated pAP and pPR constitute the protein scaffold of a sphere with the MCP on its surface. Several additional viral proteins participate in assembling this fragile procapsid (about 124 nm in diameter [29]) by binding to the MCP and/or the portal protein [25–28].

There is evidence that this self-association is mediated by coiled-coil interactions of the scaffold domains. The same region is also supposed to be important for the interaction with the MCP [28]. Self-associated pAP and pPR constitute the protein scaffold of a sphere with the MCP on its surface. Several additional viral proteins participate in assembling this fragile procapsid (about 124 nm in diameter [29]) by binding to the MCP and/or the portal protein [25–28]. The ratio of pPR and pAP in the procapsid is 1:105. The function of pAP can be taken over by pPR, but with reduced efficiency as demonstrated with pAP knock-out mutants [36].

In the procapsid, the serine-protease domain of pPR is activated by forming a homodimer [37–39]. It cleaves pAP and pPR near their C-termini (Fig. 1a, b), which leads to the release of the scaffold from the MCP and the portal. This causes conformational changes and the labile, spherical procapsid matures to the stable, icosahedral capsid [31, 33, 40]. Consequently, this cleavage site is called maturation site (M-site) [41]. The conformational changes lead to the establishment of the final conformation of the portal vertex as well [13, 42]. Untimely activation of pPR is proposed to be prevented by the bound MCP [43].

Additionally, the serine-protease domain cleaves another site exclusively in pPR. This site is situated between the protease domain and the scaffold domain; thus, releasing assemblin from the scaffold. This site is referred to as release site (R-site) [41]. The C-terminal cleavage product (scaffold) of pPR is not identical to pAP though, as there is a linker between assemblin and the scaffold domain, which remains at the scaffold domain (Fig. 1a, b). Unfortunately, in the literature both proteins are often claimed to be identical, which is not correct.

After proteolytic cleavage at their M-sites and the associated release of the scaffold proteins from the procapsid, packaging of the viral DNA into the capsids and simultaneous cleavage of the concatameric viral DNA to genome length can occur [44–47]. This is realized by the terminase complex, which is bound to the portal [11, 48, 49]. Capsids that contain DNA are called nucleocapsids. They lack scaffold, but it is not known whether the scaffold is expelled during or prior to DNA packaging.

In contrast to the scaffold, assemblin remains in the nucleocapsids of some herpesviruses like HSV-1, VZV or pseudorabies virus (PrV) after packaging of the viral DNA, for unknown reasons [25, 50, 51]. Furthermore, assemblin remains in capsids after infection of cells and it is not injected into host nuclei, as shown recently by assemblin-GFP fusion proteins [52]. Overexpression of pPR or deletion of the pAP gene does not increase the amount of assemblin in the mature capsids of these viruses. It has been discussed whether there are distinct binding sites for pPR and pAP in capsids, and whether the protease domain either prevents pPR from binding to pAP positions or directs pPR to pPR-specific sites [25]. In other herpesviruses, e.g. HCMV, neither assemblin nor fragments of it have been detected in virions. It has been discussed that this might be due to the additional space that is necessary for the HCMV genome, as it is 51 % longer than that of HSV-1 [53].

The structure and assembly of herpesvirus capsids are similar to those of capsids of bacteriophages with dsDNA, e.g. T4, T7 or λ phages [54, 55]. More detailed and complete information on capsid assembly and the replication cycle of herpesviruses can be found elsewhere [1, 2].

**ASSEMBLINS**

Assemblins are a family of serine proteases found in all herpesviruses [225–256 aa residues or 24.5–7 kDa (Uniprot [56])]. According to the MEROPS classification, assemblins are part of the SH superfamily constituting the family S21 [57]. Their catalytic triad is unique and consists of a serine residue as the nucleophile that is activated by two histidine side chains.

Superposition of the catalytic triad of assemblins with those of classical serine-proteases (Ser–His–Asp) revealed that the positions of the amino acid residues of the triads are nearly identical (Fig. 2) [58, 59]. The stabilization of the oxyanion hole is realized in a similar manner in all serine proteases as well.

The best characterized assemblin is that of HCMV. Thus, in this article the amino acid numbering of HCMV-assemblin (UniProtKB accession number P16753) is used unless stated otherwise.

**CLEAVAGE SITES AND RECOGNITION SEQUENCES**

While M- and R-sites are present in all pPR homologues, additional cleavage sites were found in assemblins from HCMV, simian cytomegalovirus (SCMV) and KSHV. One of these is the internal cleavage site (I-site) in the cytopemaloviruses, which allows cleavage after Ala-143 (HCMV
[60]) and Ala-127 (SCMV [61]), as shown in (Fig. 3a). In recombinant HCMV-assemblin, I-site cleavage results in a two-chain protease with retained activity [62].

There are two other cleavage sites exclusively in HCMV-assemblin. The cryptic site (C-site) is situated between Ala-209 and Ala-210 [63]. Neither I- nor C-site cleavage is essential for replication of HCMV, but both enhance infectivity. Blocking either site reduces virus titres and blocking both at once synergistically reduces virus titre by 90 %. This observation suggests that cleavage at these sites has different functions [64]. Furthermore, it was shown that cleavage of the C-site does not occur after cleavage at the I-site, although the C-site is present in one of the resulting fragments (Fig. 3a). This means the three-dimensional structure of that fragment is altered and the C-site is no longer accessible.

The second site unique in HCMV-pPR and HCMV-pAP is a site near the C-terminal tail (between Ala-675 and Ala-676, HCMV-pPR numbering); thus, called the tail-site (T-site). But cleavage at this site is negligible in soluble (active) and inclusion body (inactive) preparations of pPR, and not observed in purified denatured pPR recovered by immobilized metal-ion affinity chromatography (IMAC). It is supposed that cleavage at this site is caused or enhanced by refolding from a denatured state, when the catalytic site regains activity faster than the C-terminal domain refolds [65]. Therefore, this site is neglected in this article.

The main pathway of cleavage of HCMV-pPR is the following: 1, M-site cleavage; 2, R-site cleavage; 3, I-site cleavage [63]. Cleavage at the C-site inactivates the protease.

Cleavage at the dimer-disruption site (D-site) in KSHV-assemblin (between Ala-203 and Ser-204, KSHV

Fig. 1. Properties and functions of the pPR and pAP. (a) Schematic overview of pPR and pAP and their processing by assemblin. The residues of the catalytic triad are shown in purple and the start methionine of pAP is shown in red. The regions of the NLSs and MCP-binding are marked in green and blue, respectively. The release site (R-site) and maturational site (M-site) for assemblin are conserved in all pPRs and pAPs, and are indicated by arrows. (b) Schematic overview of the core of herpesvirus capsids before and after cleavage by assemblin. After cleavage, the scaffold is expelled during or prior to packaging of the viral DNA into the capsids. (a) and (b) are taken and adapted from [75] and [69], respectively.
numbering) inactivates the protease (Fig. 3b). Its homologue from murine herpesvirus 68 has a similar site, which is likely to inactivate the respective assemblin after cleavage at this site as well [66]. These D-sites are close to the same region as the C-site in HCMV assemblin (Fig. 3).

Assemblins are highly specific proteases. The cleaved bonds are always between two small amino acids (Ala–Ser/Ala/Thr/Asn), with a strong preference for Ala–Ser. Since the active site is quite narrow, bulky amino acid side chains are not able to bind to the active site. Sequence alignments lead to the recognition sequences presented in Table 1 [63, 66, 67].

The recognition sites are quite similar, but not each assemblin is able to cleave each of the above sites. Assemblins, pPR and pAP from HCMV, SCMV and HSV-1 were expressed and studied with respect to their proteolytic activities. The HCMV and SCMV proteases cleaved the M-sites of all three homologous and heterologous substrates and an SCMV R-site substrate. The HSV-1 protease cleaved only its own substrate [68].

The predicted M-site for the scaffold of PrV is P–V–Q–A ↓ S (488, PrV-pPR numbering [67]). However, there is some evidence that either this cleavage site is not correct or there is a second cleavage site at the C-terminus of the PrV scaffold [69]. (i) The predicted M-site for the scaffold of PrV is the only predicted or proven cleavage site for an assemblin, where the P4 position is occupied by a proline residue. (ii) This site is in a repeat region of the scaffold, which mainly consists of alanines and prolines. Thus, it is very likely intrinsically disordered without typical secondary conformation. (iii) In a sequence alignment, most M-sites are found further towards the C-termini of their respective scaffolds (Fig. 4). The M-site of HCMV scaffold is the only exception besides the currently predicted PrV-scaffold M-site. (iv) There is another site that fits the consensus sequence of cleavage sites for assemblins: T–I–D–A ↓ S (503, PrV-pPR numbering). This site is (a) very similar to other known and predicted M-sites from closely related herpesvirus scaffolds [T–V–

---

**Fig. 2.** Superposition of the catalytic triads of different serine proteases. Green, dimeric HCMV assemblin (PDB entry 1cmv); cyan, seleno-subtilisin from *B. subtilis* (PDB entry 1sel); black, bovine alpha-chymotrypsin (PDB entry 4cha); orange, human trypsin 1 (PDB entry 1trn). Numbers are distances in Å, labelled residues are from HCMV assemblin. The peptide-backbone nitrogen of Arg-165 stabilizes the oxyanion hole. This figure was created using PyMOL version 1.7.1.3 (2014) molecular graphics system (Schrödinger).

---

**Fig. 3.** Additional cleavage sites and fragments in assemblins from HCMV (a) and KSHV/murine herpesvirus 68 (b). This figure was created using GIMP version 2.8.10 (available at www.gimp.org).
D–A S (EHV-1, Equine herpesvirus 1; BoHV-1, Bovine herpesvirus 1 and GaHV-1, Gallid herpesvirus 1) and H–I–D–A S (SaHV-2, Saimiriine herpesvirus 1), see Table 2) and to the D-sites of assemblins from KSHV and murine herpesvirus 68 (Table 1); (b) located further towards the C-terminus and near the M-sites of other scaffolds in a sequence alignment (Fig. 4, Table 2); and (c) outside the intrinsically disordered repeat region of the scaffold. Thus, this second site seems to be a more suitable M-site of the scaffold of PrV or both sites are cleaved, but no experimental data is available.

### NOMENCLATURE

Depending on the context, the word assemblin can refer to either a protein domain, a class of proteins or – by addition of a species name – to a specific protein (e.g. HCMV-assemblin). This results from the fact that assemblins are a domain of the pPR proteins, but after release the assemblin domain itself is a functional protein too. The name assemblin is derived from its function in cleaving the pAP [41]. Other terms that can be found for assemblin are ‘maturation protease’ or simply ‘protease’. Both terms are not

### Table 1. Recognition sequences and cleavage sites for assemblins, and in which proteins and herpesviruses these are found

<table>
<thead>
<tr>
<th>Cleavage site</th>
<th>Protein and virus</th>
<th>P4</th>
<th>P3</th>
<th>P2</th>
<th>P1</th>
<th>P1’</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-site</td>
<td>pPR from all herpesviruses</td>
<td>Y</td>
<td>V/L</td>
<td>K/Q</td>
<td>A</td>
<td>S/N/T</td>
</tr>
<tr>
<td>M-site</td>
<td>pPR and pAP from all herpesviruses</td>
<td>–</td>
<td>V/L/I</td>
<td>N/D/Q/E</td>
<td>A</td>
<td>S</td>
</tr>
<tr>
<td>I-site</td>
<td>Assemblins from cytomegaloviruses</td>
<td>D</td>
<td>I/V</td>
<td>N/E</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>C-site</td>
<td>HCMV-assemblin only</td>
<td>A</td>
<td>V</td>
<td>D</td>
<td>A</td>
<td>S</td>
</tr>
<tr>
<td>D-site</td>
<td>Assemblins from KSHV and murine herpesvirus 68</td>
<td>A</td>
<td>I</td>
<td>D</td>
<td>A</td>
<td>S</td>
</tr>
<tr>
<td>General properties</td>
<td>–</td>
<td>–</td>
<td>Aliphatic</td>
<td>Hydrophilic</td>
<td>Small</td>
<td>Small, mostly with hydroxyl group</td>
</tr>
</tbody>
</table>
appropriate for the specific assemblin, because there is at least one additional protease in herpesviruses [70].

Of course, the naming of proteins and genes from herpesviruses developed historically. Thus, the nomenclature of the proteins and genes involved in the function of assemblin is quite diverse, and sometimes used incorrectly because of the lack of conventions. Assemblin is often used synonymously with pPR or PR in the literature, which is not correct. Only the N-terminal domain of pPR should be referred to as assemblin to prevent misunderstanding.

As described above, the 5'–ends of the genes encoding pPR overlap in frame with the gene encoding pAP. Originally, these genes were proposed to be referred to as assembly protein nested genes (APNGs) with genes encoding pPR and pAP named APNG1 and APNG.5, respectively [20]. Unfortunately, these genes are named individually for the most relevant human herpesviruses (Table 3). Genes from herpesviruses that are of inferior relevance and were sequenced later are sometimes named after their more relevant homologues. Most abbreviations contain 'VP' for 'viral protein', 'ICP' for 'infected cell polypeptide' or 'UL' as well as 'U' for 'unique long' (derived from the genome structure).

Additionally, the cleavage products must be named. Depending on the specific assemblin, 5–9 different cleavage products occur (Figs 1a and 3a, b, the C-terminal tail is omitted). Both pPR and pAP lose their precursor status when they are cleaved at their M-sites. After cleavage at these sites, they are referred to as PR and AP, respectively. Thus, the most significant cleavage products that are common in all herpesvirus scaffold proteins are already named generically (pPR, pAP, PR, AP and assemblin). The short C-terminal part of the scaffold, which remains at the MCP after cleavage at the M-site, is called tail [71], which is a quite unspecific name. The scaffold part of pPR after cleavage at the R- and M-site is sometimes called PRc [71].

Due to the lack of conventions, we suggest the following generic naming scheme for scaffold-related proteins. X-Y(-Z) with X being the source organism (e.g. HCMV), Y being the protein (pPR, pAP, PR, AP or assemblin), and Z in case of other cleavage products: Z1 and Z2 being the abbreviations of the N- and C-terminal cleavage sites, respectively. For the actual N- and C-termini of the precursors, we recommend using α and ω to avoid misunderstandings with the C-sites. The most complex example would be the fragments of HCMV protease-precursor that was cleaved at the I-, R- and M-sites, i.e. after Ala-143, -256 and -643, respectively (cleavage at the C-site does not occur after I-site cleavage [63]):

HCMV-assemblin-α-I (corresponding to HCMV-pPR amino acid residues 1–143)
HCMV-assemblin-I-R (corresponding to HCMV-pPR amino acid residues 144–256)
HCMV-pPR-R-M (corresponding to HCMV-pPR amino acid residues 257–643)
HCMV-pPR-M-ω (corresponding to HCMV-pPR amino acid residues 644–708).

**ACTIVITY**

Assemblins are approximately 1300- to 3000-fold less active than classical serine proteases (Table 4). The active site contains only the dyad of Ser and His of the three components of typical Ser-His-Asp(Glu) triads from other serine hydrolases
Source organisms are given in parentheses. This list is not exhaustive. HHV, Human herpesvirus.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Alternative protein name</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCP</td>
<td>pPL19, pPL86</td>
<td>UL19 (HSV-1 and PrV), UL86 (HCMV)</td>
</tr>
<tr>
<td>Assemblin</td>
<td>Pr, PR, PR2, ICP35ab,</td>
<td>See pPR</td>
</tr>
<tr>
<td>pPR</td>
<td>pUL26, pUL80, pUL80a</td>
<td>UL26 (HSV-1 and PrV), ORF33 (VZV), ORF35 (EHV-1), UL80 or UL80a (HCMV), U53 (HHV-6 and -7), BVRF2 (EBV), ORF17 (KSHV)</td>
</tr>
<tr>
<td>pAP</td>
<td>Pre-VP22a, ICP35, ICP35cd, pUL26.5, pUL80.5</td>
<td>UL26.5 (HSV-1 and PrV), ORF33.5 (VZV), UL80.5 (HCMV), ORF17.5 (KSHV)</td>
</tr>
</tbody>
</table>

Table 3. Aliases and gene names for MCP, assemblin, pPR and pAP

Table 4. Activity data for selected serine proteases and some variants of the third residue of the catalytic triad

<table>
<thead>
<tr>
<th>Serine protease (variant)</th>
<th>Reference</th>
<th>Substrate</th>
<th>(K_{cat}) [s(^{-1})] *</th>
<th>(K_m) [M]</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCMV-assemblin</td>
<td>[73]</td>
<td>4-(4′-dimethylaminophenylazo)-benzoyl-Arg-Gly-Val-Val-Ala-Ser-Ser-Arg-Leu-Ala-(2′-aminooethyl)amino-naphthalene-1-sulfonic acid</td>
<td>3.3 \times 10^{-2}</td>
<td>1.01 \times 10^{-5}</td>
</tr>
<tr>
<td>HCMV-assemblin (H157A)</td>
<td>[73]</td>
<td>4-(4′-dimethylaminophenylazo)-benzoyl-Arg-Gly-Val-Val-Ala-Ser-Ser-Arg-Leu-Ala-(2′-aminooethyl)amino-naphthalene-1-sulfonic acid</td>
<td>3.3 \times 10^{-3}</td>
<td>2.18 \times 10^{-5}</td>
</tr>
<tr>
<td>HSV-1-assemblin</td>
<td>[105]</td>
<td>Ala-Gly-His-Thr-Tyr-Leu-Gln-Ala-Ser-Glu-Lys-Phe-lys-Met-Trp-Gly</td>
<td>(10)*</td>
<td>(0.5)</td>
</tr>
<tr>
<td>Subtilisin</td>
<td>[106]</td>
<td>N-Succinyl-Ala-Ala-Pro-Phe-p-nitroanilide</td>
<td>2.0 \times 10^{-2}</td>
<td>NA</td>
</tr>
<tr>
<td>Subtilisin (S24C, D32A)</td>
<td>[106]</td>
<td>N-Succinyl-Ala-Ala-Pro-Phe-p-nitroanilide</td>
<td>2.3 \times 10^{-3}</td>
<td>4.8 \times 10^{-4}</td>
</tr>
<tr>
<td>Trypsin</td>
<td>[107]</td>
<td>Z-Gly-Pro-Arg-7-amino-4-methyl-coumarin</td>
<td>(19 100)</td>
<td>(0.4)</td>
</tr>
<tr>
<td>Trypsin</td>
<td>[107]</td>
<td>Insulin peptide, β chain</td>
<td>4.7</td>
<td>1.4 \times 10^{-4}</td>
</tr>
<tr>
<td>Trypsin (D102N)</td>
<td>[107]</td>
<td>Z-Gly-Pro-Arg-7-amino-4-methyl-coumarin</td>
<td>2.17 \times 10^{-2}</td>
<td>2.7 \times 10^{-6}</td>
</tr>
<tr>
<td>Trypsin (D102N)</td>
<td>[107]</td>
<td>Insulin peptide, β chain</td>
<td>(2400)</td>
<td>(3.6)</td>
</tr>
<tr>
<td>Trypsin (D102N)</td>
<td>[107]</td>
<td>Insulin peptide, β chain</td>
<td>1.1 \times 10^{-3}</td>
<td>9.0 \times 10^{-5}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(38 000)</td>
<td>(1.6)</td>
</tr>
</tbody>
</table>

*Numbers in parentheses are the ratios between wild-type and the respective activity values of the variants.

In assemblins, the orientation of His-157 is sterically ensured by a proline (Pro-58), which is conserved in all assemblins (representative examples in Fig. 5a). The pyrroline ring traps the imidazole of His-157 in the active orientation, while other residues would allow alternative rotamers, which would decrease the assemblin activity (Fig. 5b, c). Additionally, some conserved water molecules near the active site were identified that are likely to influence catalysis [73, 75].

Interestingly, studies of HCMV-assemblin showed that the M-site is cleaved faster than the R-site and biological substrates are cleaved more efficiently by full-length pPR than by assemblin [71].

C-terminal modifications potentially have a substantial impact on the activity of assemblins. Extensions of the C-terminus reduce the activity in some cases [40, 76, 77]. Shortening the C-terminus of HCMV-assemblin by 3 or 8 aa residues leads to incapability of the assemblin to cleave its I-site, but a substrate carrying an M-site was cleaved [60]. A deletion of its C-terminal alanine did not affect the affinity of PrV-assemblin towards an inhibitor (disisopropyl-fluorophosphat) [75].

REGULATION OF ACTIVITY

Processing of pPR by assemblin releases the binding site of the MCP from the scaffold domain (M-site) and assemblin...
from the scaffold domain (R-site). Processing of pAP by assemblin releases the MCP from the scaffold protein (M-site). There is a NLS in the scaffold domain of pPR and pAP [42]. In scaffold-protein-deficient mutants of HSV-1, it was shown that the MCP translocation into the nucleus of infected cells is drastically reduced [78] and viruses lacking the NLS replicated to titres reduced by 1000-fold [42]. Therefore, premature processing at the M-site would prevent the MCP from being transported into the nucleus and to the site of capsid assembly, i.e. capsid formation would be impeded. Prematurely cleaved pPR at its R-site leads to assemblin not being transported to the nucleus and into the capsids. Consequently, M-site cleavage cannot occur, which would result in fully assembled capsids containing scaffold in their cores (B-capsids). The scaffold is trapped in the capsids and prevents the viral DNA from being packed into the capsids. Thus, activation of assemblins must be strictly regulated.

A central part of the regulation of assemblin activity is a monomer–dimer equilibrium [37]. The dimer is weakly associated with dissociation constants (K_D) in the micromolar range [79]. It is the active species, while the monomer is almost inactive. The equilibrium is shifted towards the dimer by higher concentrations of assemblin, salts and co-solvents like glycerol [37, 75, 79]. Dimerization induces refolding of a loop (residues 162 to 171) near the active site. In its conformation in dimeric assemblin, this loop stabilizes the oxyanion hole [75, 80–82]. This mechanism is essential to activate the assemblin. Untimely proteolytic activity might be prevented by binding of the MCP to the scaffold [43].

Another probable mechanism of regulation is the processing of assemblin at additional sites in some herpesviruses like HCMV (I- and C-sites) and KSHV (o-site). Furthermore, it is assumed that the capsid environment itself enhances dimerization and activity [39].

**STRUCTURE**

The first crystal structures of an assemblin were published by three groups in 1996 [58, 59, 83]. All three structures are of dimeric HCMV-assemblin. Crystal structures of dimeric
assemblins from HSV-2, VZV, PrV, HCMV, KSHV and EBV are available in the protein data bank (PDB). The only monomeric assemblin structure available is of PrV-assemblin and was recently published by us [75]. In addition to the HCMV-assemblin numbering of amino acid residues and secondary structure elements, the PrV-assemblin numbering is added in braces in the following paragraphs when appropriate.

Monomeric structure

The PrV-assemblin for crystal structure analyses was produced as a recombinant protein in Escherichia coli. It was expressed and crystallized with an N-terminal (His)₆-tag with a thrombin linker and without the C-terminal alanine (Ala-256 [225]). Both modifications have no influence on either the overall structure of the monomer or the activity of the PrV-assemblin as discussed by us [75]. However, the deletion of the C-terminal alanine results in a different conformation of the eight C-terminal residues. In all dimeric full-length assemblin structures, the C-termini are positioned identically and we proposed that this conserved conformation of the C-terminus occurs in full-length monomeric PrV-assemblin as well. Compared to its dimeric form, monomeric PrV-assemblin is missing one helix. For simplicity, the helix numbering is adapted to dimeric assemblin.

The core structure of the monomeric PrV-assemblin is a β-barrel that is formed by two β-sheets: a four-stranded anti-parallel sheet and a three-stranded mixed sheet. The β-barrel is surrounded by seven α-helices (Fig. 6). Atomic temperature factors suggest that the β-barrel is rigid and the α-helices are more flexible, especially the two C-terminal α-helices (α5 [7] and α6 [8]). In one of the two polypeptide chains of the asymmetric unit, helix α5 [7] is partially disordered. We assumed that those two C-terminal helices are similarly positioned in other alphaherpesviruses. In contrast, the two C-terminal helices of monomeric assemblins from beta- and gammaherpesviruses are completely disordered according to circular dichroism (CD) - and Nuclear Magnetic resonance (NMR)-spectroscopic data [38, 84]. We proposed that this difference results from several residues that are conserved in assemblins from alphaherpesviruses on the one hand and beta-/gammaherpesviruses on the other hand. These residues are part of a loop between β5 [6] and β6 [7] (referred to as the oxyanion-hole loop, OHL) and of the C-terminal helix. Additionally, we observed two slightly different conformations of the OHL of monomeric PrV-assemblin (Fig. 6).

The side chains of the catalytic triad are solvent exposed and in activating hydrogen bonding contact to each other (Fig. 6). The lack of activity of the monomeric assemblin results from the absence of an oxyanion hole near the active-site serine, because the OHL is not in an appropriate conformation, whereas the core of the protease including the residues of the catalytic triad is unchanged. This was predicted earlier and verified by our monomeric structure [75, 80, 82]. Additionally, there are some models of monomer-like structures of C-terminally truncated KSHV-assemblin (A222–256) with inhibitors that bind to the dimerization area [81, 84], see the discussion in the Inhibitors section below.

Dimeric structure

The interface area of the weakly associated assemblin dimers is about 1300 Å². Dimeric assemblins (Fig. 7) have a very similar tertiary core structure as the monomeric PrV-assemblin (Fig. 8). The rigid β-barrel is nearly identical in all assemblin structures, but the amount, length and orientations of α-helices varies slightly depending on the subfamily of the source virus (alpha- or beta-/gammaherpesvirus [85]) and on how α-helices are assigned. An unstructured part of the monomeric PrV-assemblin (amino acid residues 24–31 [14–19]) forms a short α-helix upon dimerization in all assemblins. It is not defined as an α-helix in all PDB files though. The OHL is an important task for the transition state in proteolytic catalysis. Its conformation is conserved in all active dimeric assemblins. There are two consecutive arginine residues in this loop (Arg-165 and -166 [136 and 137]), which are conserved throughout all assemblins. As the peptide backbone of Arg-165 [136] stabilizes the oxyanion hole in this conformation (Fig. 9) it is referred to as the OHL. The stabilization of the oxyanion hole results in the catalytic activity of the dimeric form of assemblins. Compared to the monomeric structure of PrV-assemblin the position of Arg-165 [136] is shifted by 11 Å, while the positions and orientations of the catalytic site residues are identical (Fig. 8). A more detailed comparison of the monomeric and the dimeric structures can be found in our publication of the structures of PrV-assemblin [75].

TEMPERATURE-SENSITIVE VARIANT

A temperature-sensitive variant (ts1201) of HSV-1-assemblin was obtained by an UL26 double mutant (Y30F/A48V,
HSV-1-numbering) [86]. Most likely, these mutations destabilize the region of helix α1, which in turn destabilizes the OHL and results in the temperature-sensitive property of the assemblin of ts1201 [75]. This variant was used to show that procapsids can evolve into mature capsids by temperature-induced activation of assemblin [87]. Based on this information and the determined structure of dimeric and monomeric PrV-assemblin, we proposed the generation of a temperature-sensitive variant for PrV-assemblin by a Y13F/R24P/A30V or a Y13F/R24K/A30V mutant [75].

INHIBITORS

Most classical serine-protease inhibitors show comparably low potential to inhibit assemblins. Benzamidine, cystatin, aprotinin, chymostatin, pepstatin A, leupeptin, antipain, Naα-tosyl-Lys chloromethyl ketone and PMSF had no effect on the activity of HCMV-assemblin [88]. There is a structural model of an HCMV variant with bound PMSF though [73]. Disopropylfluorophosphate, Naα-tosyl-Phe chloromethylketone, iodoacetamide, N-ethylmaleimide and Zn2+ ions, however, can be used to inhibit assemblins [88].

Interruption of biochemical processes can be generated by blocking the assembly of specific quaternary structures of proteins [89], and complexes of proteins with nucleic acids [90]. A recent review discussed the inhibition of HIV-1 capsid assembly by small molecule compounds and their therapeutic potential [91]. This is an elegant option even for proteins that were supposed to be undruggable. An analogous mechanism of inhibition was examined for assemblins, instead of covalent inhibitor binding to the active-site serine. Activation can be prevented by blocking the dimerization area in monomeric assemblin. In this way, the assemblin is trapped in its monomeric, inactive state. Bioinformatic analyses showed that protein–protein interactions mostly rely on ‘hot spots’ [92]. These hot spots are mainly aromatic residues with a prevalence for Trp. In the dimerization area of assemblins, there is a conserved aromatic residue: Tyr128 in HCMV-assemblin, and Trp as the corresponding residue in EBV- and KSHV-assemblins. It is anticipated that this aromatic residue is a key residue for dimerization [84].

Inhibitors that bind to the dimerization area were found when screening an α-helical-mimetic library with KSHV-assemblin [93]. Kinetic IC50 values of the best inhibitor (DD2) and some derivatives (Table 5) with assemblins from KSHV, EBV, HCMV and HSV-2 were compared [81].
assemblins from KSHV, EBV (gammaherpesviruses) and HCMV (betaherpesvirus) were inhibited at similar rates with IC\(_{50}\) values for the most potent inhibitor DD2 of 1.5, 7.7 and 7.4 \(\mu\)M, respectively. KSHV-assemblin was inhibited with a fivefold higher affinity compared to others. Thus, virus-specific inhibitors are obviously possible.

HSV-2-assemblin (alphaherpesvirus) is inhibited considerably more weakly (IC\(_{50}\) value of 52 \(\mu\)M). This might be due to lower affinity of the inhibitor or due to differences in the experimental conditions, because of weaker dimer stability in line with lower proteolytic activity of HSV-2-assemblin. Consequently, a different experimental setup had to be used for HSV-2 [81].

Crystal structures of truncated KSHV-assemblin (\(\Delta\)222–256) that was inhibited with DD2 or DD2 derivatives were determined [81, 84]. In this KSHV-assemblin construct, the two C-terminal helices were removed by C-terminally truncating the assemblin. The rationale behind this is that these helices are disordered in monomeric KSHV-assemblin [84]. In the resulting structural models, the inhibitors are non-covalently bound near the hot spot residue Trp. The next-to-last helix of KSHV-assemblin is the major dimerization helix. Truncation of these residues, forming this helix, compelled an obligate monomeric construct [84]. Nonetheless, all the structural models of inhibited KSHV-assemblin (\(\Delta\)222–256) are artificial dimers, due to packing constrains. In each model, there are two inhibitor molecules bound between two monomers. Their orientation differs from the native dimer by rotation (~80\(^\circ\)) of one monomer around an axis perpendicular to the plane between the two monomers [75]. The OHL adopts a conformation which is different from that in the native and active dimeric state of KSHV-assemblin, because the essential allosteric conformational changes are not induced. This renders the protease inactive. Presumably, the inhibitor molecules are mimicking the missing dimerization helix as they are derived from screening an \(\alpha\)-helical-mimetic library [75, 93].

Additional dimer-disruptors for KSHV-assemblin (indole and aminothiazole scaffolds; Table 5) were found recently [94]. An advantage in designing dimer-disruptors for assemblins are their relatively weak association constant of dimer formation [75]. Taken together, this information points towards small-molecule inhibitors with aromatic parts to bind to the hot spots of the dimerization area of assemblins. Hydrophobic interactions by \(\pi-\pi\) stacking of aromatic compounds with aromatic amino acid residues, but also polar interactions with the \(\pi\) system of aromatic

---

**Fig. 9.** Active site and OHL of inhibited dimeric PrV-assemblin. The OHL is coloured orange, the covalently bound diisopropylphosphate is coloured red and the catalytic triad is coloured purple. Dashed lines illustrate parts of the hydrogen-bond network (distances in the range of 2.8–3.4 Å). The strictly conserved Arg-165 (136) forms the oxyanion hole with its peptide backbone N-H. Diisopropylphosphate mimics the tetrahedral transition state of the natural substrates. This figure is reproduced from [75].
Table 5. Some representative inhibitors of assemblin dimerization

Information is derived from published work [81, 94]. Structural models for KSHV-assemblin (D222–256) inhibited by DD2, compound 2 and compound 3 are available in the PDB as entries 3njq, 4p3h and 4p2t, respectively [81, 84]. Inhibition by fragments 1–17 was determined indirectly by decreased dimer-to-monomer peak ratios of $^{13}$C-Met HSQC NMR, for details see [94]. Structures were drawn with EasyChem, version 0.6 (F-X Coudert; http://easychem.sourceforge.net/”).

<table>
<thead>
<tr>
<th>Structure</th>
<th>Name</th>
<th>IC$_{50}$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>KSHV</td>
</tr>
<tr>
<td>DD2</td>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td>Compound 2</td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>Compound 3</td>
<td></td>
<td>3.6</td>
</tr>
<tr>
<td>Fragment 1</td>
<td></td>
<td>51.2</td>
</tr>
<tr>
<td>Fragment 3</td>
<td></td>
<td>495</td>
</tr>
<tr>
<td>Fragment 4</td>
<td></td>
<td>187</td>
</tr>
<tr>
<td>Fragment 6</td>
<td></td>
<td>421</td>
</tr>
<tr>
<td>Fragment 8</td>
<td></td>
<td>81.7</td>
</tr>
<tr>
<td>Fragment 9</td>
<td></td>
<td>89.5</td>
</tr>
<tr>
<td>Fragment 12</td>
<td></td>
<td>445</td>
</tr>
<tr>
<td>Fragment 16</td>
<td></td>
<td>367</td>
</tr>
<tr>
<td>Fragment 17</td>
<td></td>
<td>235</td>
</tr>
</tbody>
</table>
moieties, were found to contribute significantly to the binding between a ligand and nearby protein residues.

**RELATIONSHIP TO THE PROHEAD PROTEASE OF BACTERIOPHAGES**

The structure of the prohead protease of bacteriophage T4 was published recently (PDB entry 5jbl) [55]. Comparison of its structure with others revealed that the closest structural relative is the assemblin from PrV (PDB entry 4v08). A superposition of these structural models has an root mean square deviation (r.m.s.d.) of 2.2 Å using 91 equivalenced Ca atoms with only 7 % sequence identity. In contrast to assemblins, this protease has a classical serine-protease triad (Ser-His-Asp), forms a pentamer and its activation mechanism relies on unblocking the active sites via interactions in the prohead [55].

The fold of the T4 prohead protease shows similarities to proteins forming the tubes in phage tails. The phage DNA is ejected through these tubes. But there are several other proteins in phages, as well as in tubes of the bacterial type VI secretion systems, which share structural similarity with this phage tail tube-like fold. Thus, this fold might have its origins in procapsids or in the tail of a primordial virus [55]. The vast number of phages in the biosphere (~10^{31} phages [95]) suggests that a polypeptide fold comparable to assemblins might be the most abundant protein folding motif on earth [55].

**CONCLUSIONS**

There are eight human herpesviruses and almost all adults are infected with at least one of these viruses [96–102]. Antiviral therapy is of great importance as herpesvirus infection in immunocompromised patients can evoke cancer or encephalitis and, therefore, death. Herpesvirus capsid assembly relies on the precise proteolytic activity of assemblins. Thus, the development of specific assemblin inhibitors for clinical application will be a promising way to disrupt the life-cycle of herpesviruses. Beside the conventional approach of screening for active-site inhibitors that mimic the substrate, another antiviral target seems to be suitable. The mechanism of activating the proteolytic properties relies on conformational changes induced by dimerization. The design and screening for surface-blocking drugs is a challenging task. The desired small-molecule drug must have a higher affinity to a much smaller part of a large surface to compensate the hydrophobic effect and the contribution of sequence-specific interactions. However, the weak association of the assemblin dimers and the conformational flexibility of the dimerization area in combination with its significance to the catalytic activity of these proteases allow this exception. The proof of principle is the assumed monomer-like structure of an assemblin fragment from the gammaherpesvirus KSHV (A222–256) stabilized by a non-covalent bound inhibitor [84].

A lot of great research has been conducted on assemblins in the last few decades. But still there are plenty of questions left. The knowledge of the exact mechanism triggering activation is limited and the relevance of additional cleavage sites in assemblins from several herpesviruses is unknown. Furthermore, there is little structural information on the pPR or the scaffold proteins. The structural changes in the capsids that induce assemblin activity are also poorly understood.

**Funding information**

The authors received no specific grant from any funding agency.

**Conflicts of interest**

The authors declare that there are no conflicts of interest.

**References**


Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.