X-Ray Structures of the Hexameric Building Block of the HIV Capsid

Owen Pornillos,1,5 Barbie K. Ganser-Pornillos,1,5 Brian N. Kelly,2 Yuanzi Hua,1 Frank G. Whitby,2 C. David Stout,3 Wesley I. Sundquist,2 Christopher P. Hill,2 and Mark Yeager1,4,5,*

1Department of Cell Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
2Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84112, USA
3Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
4Division of Cardiovascular Diseases, Scripps Clinic, La Jolla, CA 92037, USA
5Department of Molecular Physiology and Biological Physics, University of Virginia Health System, Charlottesville, VA 22908, USA
*Correspondence: yeager@scripps.edu
DOI 10.1016/j.cell.2009.04.063

SUMMARY

The mature capsids of HIV and other retroviruses organize and package the viral genome and its associated enzymes for delivery into host cells. The HIV capsid is a fullerene cone: a variably curved, closed shell composed of approximately 250 hexamers and exactly 12 pentamers of the viral CA protein. We devised methods for isolating soluble, assembly-competent CA hexamers and derived four crystallographically independent models that define the structure of this capsid assembly unit at atomic resolution. A ring of six CA N-terminal domains form an apparently rigid core, surrounded by an outer ring of C-terminal domains. Mobility of the outer ring appears to be an underlying mechanism for generating the variably curved lattice in authentic capsids. Hexamer-stabilizing interfaces are highly hydrated, and this property may be key to the formation of quasi-equivalent interactions within hexamers and pentamers. The structures also clarify the molecular basis for capsid assembly inhibition and should facilitate structure-based drug design strategies.

INTRODUCTION

The ribonucleoprotein genomic complex of human immunodeficiency virus type 1 (HIV-1) is encased within the mature capsid, a predominately cone-shaped shell assembled from ~1,500 copies of the viral CA protein (recently reviewed by Ganser-Pornillos et al., 2008). The HIV-1 capsid is a fullerene cone (Ganser et al., 1999; Jin et al., 1999; Li et al., 2000), with a body composed of a curved two-dimensional (2D) array of ~250 CA hexamers. To form a closed shell, the ends of the cone are capped by exactly 12 pentamers, with seven at the broad end and five at the narrow end. The fullerene model and its generality across retroviruses are now widely accepted (e.g., see Heymann et al., 2008).

HIV-1 CA is a highly helical protein with two independently folded domains, the N-terminal domain (NTD) and C-terminal domain (CTD), which are flexibly linked. Published high-resolution CA structures include isolated domains and full-length monomers from HIV-1 and other retroviruses (Berthet-Colominas et al., 1999; Campos-Olivas et al., 2000; Cornilescu et al., 2001; Gamble et al., 1997; Gitti et al., 1996; Jin et al., 1999; Khorasanzadeh et al., 1999; Mortuza et al., 2004, 2009). These structures collectively demonstrate that retroviral CA proteins share a common tertiary fold despite having widely divergent amino acid sequences and that, as a corollary, mature retroviral capsids are likely to be stabilized by similar quaternary interactions. The hexagonal capsid lattice is composed of three different types of interfaces: a six-fold symmetric NTD-NTD interface that creates hexameric rings, an intermolecular interface between the two domains (NTD-CTD) that reinforces the hexamer, and a homodimeric CTD-CTD interface that links the hexameric building blocks into an infinite hexagonal lattice (Bowzard et al., 2001; Gamble et al., 1997; Ganser-Pornillos et al., 2004, 2007; Lanman et al., 2003, 2004; Li et al., 2000; von Schwedler et al., 1998, 2003). This lattice architecture was unambiguously established by an electron cryomicroscopy (cryoEM) structure of 2D crystals of HIV-1 CA hexamers, albeit at moderate resolution (~9 Å) (Ganser-Pornillos et al., 2007). Moreover, crystal structures of isolated NTD and CTD have provided atomic models for two of the different types of interfaces in the hexagonal lattice: the dimeric CTD-CTD interface of HIV-1 CA (Worthylake et al., 1999) and the hexameric interface formed by the isolated NTD of murine leukemia virus (MLV) CA (Mortuza et al., 2004).

Despite steady progress in elucidating the structure of the retroviral capsid lattice, high-resolution crystal structures of hexagonal arrays of full-length retroviral CA proteins have not yet been reported. This is presumably due to the low intrinsic stability of CA hexamers and the challenges of preparing discrete oligomeric CA assemblies. Here, we describe engineered HIV-1 CA proteins that form homogenous populations of stable, soluble hexamers, which are functional for assembly in vitro. The X-ray crystal structures of these CA proteins extend our understanding of the hexameric capsid assembly unit to atomic resolution.

RESULTS AND DISCUSSION

Isolation of HIV-1 CA Hexamers for Crystallization

Discrete HIV-1 CA hexamers were stabilized by two independent methods: thiol crosslinking and template-directed
hexamerization. In the former case, the cryoEM-based coordinate model of the HIV-1 CA lattice (PDB code 3dik) (Ganser-Pornillos et al., 2007) was examined to identify residue pairs that appeared to be in close contact across the NTD-NTD interface. Cysteines were introduced at these positions, and the mutant CA proteins were assembled under reducing conditions into cylinders, which are known to faithfully mimic the hexagonal portion of the HIV-1 capsid (Briggs et al., 2003; Campbell and Vogt, 1995; Ehrlich et al., 1992; Ganser-Pornillos et al., 2004; Gross et al., 1997; Li et al., 2000). Our expectation was that, within the assembled cylinders, one or more of the engineered cysteine pairs would be positioned optimally to form intermolecular disulfide bonds and, upon oxidation, create a covalently linked hexamer. This approach also provided a functional check for the mutant proteins.

As shown in Figure 1A, a construct with cysteine substitutions for A14 and E45 assembled into cylinders that were morphologically very similar to those formed by the wild-type protein (compare Figures 1A and 1B). Upon oxidative crosslinking, the CA subunits from these assemblies migrated almost exclusively as hexamers in nonreducing SDS-PAGE gels (Figure 1D, lane 5). Indeed, complete crosslinking was observed even in the presence of a 500-fold molar excess of reducing agent (data not shown), demonstrating that disulfide bond formation was driven by the specific noncovalent protein-protein interactions within the cylinders. Not surprisingly, the crosslinked cylinders were extraordinarily stable, and remained intact under conditions in which wild-type cylinders readily disassembled (e.g., low salt, or in the presence of the proline rotamase cyclophilin A) (data not shown).

To obtain isolated hexamers for crystallization, we therefore introduced two additional mutations (W184A and M185A) at the CTD-CTD interface, to weaken the extended lattice (Gamble et al., 1997; von Schwedler et al., 2003) while leaving the hexamer-stabilizing interfaces unchanged further. This CA construct assembled less efficiently but still formed cylinders (Figure 1C). More important, it crosslinked into discrete, soluble hexamers with ~100% efficiency (Figure 1D, lane 6, and Figure 1E).

Our second approach for creating discrete hexamers was to fuse the CcmK4 protein to the C-terminal end of HIV-1 CA. CcmK4 forms stable hexameric rings in solution and has accessible termini (Kerfeld et al., 2005), making it an attractive template for driving CA hexamerization. Various linker lengths and sequences were tested, and all of the CA-CcmK4 fusion proteins that expressed solubly were also hexameric in solution, as analyzed by analytical equilibrium sedimentation (data not shown). However, different constructs eluted with different retention times by gel filtration, and we assumed that only the constructs that behaved as apparent hexamers (late eluters) were correctly folded (data not shown). The successful construct consisted of CA residues 1–226, a two-residue linker, and the full-length CcmK4 sequence followed by the remains of an affinity tag. As in the crosslinked construct, it was necessary to introduce the W184A and M185A mutations to the CTD region to prevent the hexamers from polymerizing into insoluble aggregates.

Architecture of the CA Hexamer

Four crystallographically independent models of the HIV-1 CA hexamer were derived. The first model was obtained from crystals of CcmK4-templated CA, and was determined to 7 Å resolution (Figure S1 available online; see Table 1 for crystallographic statistics). The remaining three hexamer models were determined from two different crystal forms of crosslinked CA, which diffracted to 1.9 and 2.7 Å resolution (Figures S2 and S3; Table 1).

The quaternary organization of the crosslinked and templated hexamers are identical (to the limit of their respective resolutions).
and mutually validate the two oligomer-stabilizing strategies. The hexamers are also very similar to the cryoEM structure (Ganser-Pornillos et al., 2007) (Figure S4) and are consistent with other structural and biochemical data (Bartonova et al., 2008; Bowzard et al., 2001; Cardone et al., 2009; Gamble et al., 1997; Ganser-Pornillos et al., 2004, 2007; Lanman et al., 2003, 2004; Li et al., 2000; Mortuza et al., 2004; von Schwedler et al., 1998, 2003). As shown in Figure 2, the CA hexamer is composed of two concentric rings, with the NTD and CTD forming the inner and outer rings, respectively. Intramolecular interactions between the two domains of each protomer are minimal, but both pack against the neighboring NTD subunit. Indeed, the NTD-NTD and NTD-CTD interfaces essentially merge into one contiguous hexamerization interface (Figure 3), emphasizing the degree to which the NTD and CTD interactions cooperate to create the hexameric assembly.

**Interactions between Hexamers**

The CcmK4-templated crystal and the hexagonal crystal of crosslinked CA were composed of stacked sheets, with each sheet corresponding to a flattened version of the mature hexameric lattice (Figure 2C, Figures S1B and S2B), as also seen in the HIV-1 CA cryoEM structure (Ganser-Pornillos et al., 2007). These structures therefore recapitulate all three relevant CA-CA interfaces. Within each sheet, neighboring hexamers are connected exclusively by CTD contacts made through the CTD-CTD dimerization interface (Figure 2C). The crystallographically distinct dimer interactions appear to be very similar to each

---

**Table 1. Crystallographic Statistics**

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>Crosslink</th>
<th>Crosslink</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam line</td>
<td>APS 22-BM</td>
<td>SSRL BL7-1</td>
<td>SSRL BL7-1</td>
</tr>
<tr>
<td>Space group</td>
<td>P6</td>
<td>P21,21,21</td>
<td>C2</td>
</tr>
<tr>
<td>Cell dimensions, Å</td>
<td>a = 91.0</td>
<td>a = 136.1</td>
<td>a = 90.6</td>
</tr>
<tr>
<td></td>
<td>b = 91.0</td>
<td>b = 136.7</td>
<td>b = 156.4</td>
</tr>
<tr>
<td></td>
<td>c = 56.8</td>
<td>c = 208.4</td>
<td>c = 196.6</td>
</tr>
<tr>
<td>Resolution range, Å</td>
<td>50–1.90 (1.97–1.90)</td>
<td>50–2.70 (2.80–2.70)</td>
<td>50–7.0 (7.25–7.0)</td>
</tr>
<tr>
<td>Rsym, %</td>
<td>10.0 (52.0)</td>
<td>8.5 (45.3)</td>
<td>11.4 (32.3)</td>
</tr>
<tr>
<td>Mean</td>
<td>I</td>
<td>&lt;</td>
<td>I&gt;</td>
</tr>
<tr>
<td>Completeness, %</td>
<td>99.9 (100)</td>
<td>96.0 (92.4)</td>
<td>99.6 (98.5)</td>
</tr>
<tr>
<td>Average redundancy</td>
<td>10.5 (10.5)</td>
<td>4.2 (3.7)</td>
<td>7.5 (7.0)</td>
</tr>
<tr>
<td>Average mosaicity, °</td>
<td>0.46</td>
<td>0.57</td>
<td>0.60</td>
</tr>
<tr>
<td>Wilson B factor, Å²</td>
<td>23.5</td>
<td>66.7</td>
<td></td>
</tr>
</tbody>
</table>

**Refinement**

| Resolution range        | 27–1.90 (1.98–1.90) | 35–2.70 (2.75–2.70) | 50–7.0 (7.95–7.0) |
| Number of unique reflections | 21,334 (2,489) | 102,181 (4,871) | 4,070 (1,105) |
| Reflections in free set  | 1,060 (149) | 2,565 (171) | 407 (128) |
| Rwork, %                | 23.2 (24.1) | 23.8 (31.3) | 28.2 (31.2) |
| Rfree, %                | 26.9 (30.7) | 26.3 (33.4) | 32.3 (35.9) |
| Number of nonhydrogen atoms | Protein | 1,553 | 19,082 | 10,218 |
|                          | Water     | 193 |
| Average B factor, Å²     | Protein | 25.85 | 56.43 |
|                          | Water     | 30.88 |
| Coordinate deviations from ideal geometry | Bond lengths, Å | 0.005 | 0.007 |
|                          | Bond angles, ° | 0.775 | 0.915 |
| PROCHECK/MOLPROBITY distribution of phi and psi angles, % | Favored | 95.2/99.5 | 95.1/98.3 |
|                          | Allowed   | 4.8/0.5 | 4.5/1.7 |
|                          | Disallowed | 0/0 | 0.3/0 |

Values in parentheses are for the highest resolution shell. $R_{sym} = \frac{\Sigma|I| - \langle|I|\rangle}{\Sigma|I|}$, where $I$ is the measured intensity of each reflection, and $\langle|I|\rangle$ is the intensity averaged from symmetry equivalents. $R_{work} = \frac{\Sigma|F_o - F_c|}{\Sigma|F_o|}$, where $F_o$ and $F_c$ are observed and calculated structure factors, respectively. $R_{free}$ is the cross-validation $R$ factor calculated for reflections in the free set, which were not used in refinement.
other, and resemble the X-ray structure of the isolated full-affinity CTD dimer (pdb code 1a43) (Worthylake et al., 1999), although the agreement in the relative orientations of the CTDs about the dyad is not exact (Figure 2D). This dimer interface is distinctly different from those seen in a shorter CTD construct (2a8o) (Gamble et al., 1997), or in the presence of an assembly inhibitor (2buo) (Ternois et al., 2005), or upon mutation-induced domain swapping (2ont) (Ivanov et al., 2007; Ternois et al., 2005; Wong et al., 2008).

Previous studies have suggested conformational plasticity in the tertiary fold of the CTD (Alcaraz et al., 2007; Bartonova et al., 2008; Berthet-Colominas et al., 1999; Battacharya et al., 2008; Ivanov et al., 2007; Ternois et al., 2005; Wong et al., 2008), and this flexibility is also evident in our structures of crosslinked CA. In one crystal form, the N-terminal two-thirds of helix 9 and the preceding loop are poorly ordered, and in the second crystal form this region adopts two distinct conformations. In one conformation, the loop trajectory places it close to the neighboring NTD, allowing polar side chains to participate in the hydrogen bond network at the NTD-CTD interface (Figure S5A). This CA CTD configuration has not been observed previously. In the second conformation, the loop resembles the position seen in the 1a43 monomer and does not contact the NTD (Figure S5B). We identified three additional areas of structural variability (data not shown): the major homology region (MHR) hairpin in two subunits appeared to have an alternative conformation with a noncanonical hydrogen-bonding network, which we did not attempt to model because of ill-defined density; the native C198-C218 disulfide bond appeared to exist in both the reduced and oxidized forms, with varying occupancies for each protomer, and helix 10 was characterized by variable and ill-defined densities across the different subunits, indicative of positional disorder and mobility. It is likely that the range of conformations seen in the crystals reflects both the natural plasticity of the CTD and amplified flexibility arising from the W184A and M185A mutations within helix 9.

Hexamer-Stabilizing Interactions between NTDs

The first three helices of CA contain the NTD-NTD-interacting residues, which form a loose 18 helix barrel at the center of the hexamer (Figure 2B). This interface contains a small hydrophobic core of aliphatic side chains, which include M39 and A42 (indicated in Figure 3B). These residues were previously shown by mutagenesis to be critical for both CA assembly in vitro and viral infectivity in vivo (Ganser-Pornillos et al., 2004; von Schwedler et al., 1998, 2003). Despite these limited hydrophobic interactions, the bulk of the interface is created by hydrophilic contacts. In particular, numerous ordered water molecules bridge polar side chain and backbone atoms throughout the entire interface, forming a pervasive hydrogen-bonding network (Figure 3C). Bridging waters were also observed in the hexameric X-ray structure of the isolated NTD of MLV CA (Mortuza et al., 2004), suggesting that heavily solvated interfaces may be a general property of retroviral CA hexamer interfaces. To our knowledge, similar water-rich interfaces have not been seen previously in nonretroviral capsid assemblies.

A very recent cryoEM study showed that the CA hexamer and pentamer of Rous sarcoma virus are quasi-equivalent, with pentamers formed by simply removing a protomer from the hexamer and closing the ring (Cardone et al., 2009). By analogy, helices...
1–3 in HIV-1 CA therefore switch from an 18 helix barrel in the hexamer to a 15 helix barrel in the pentamer, with concomitant adjustments in intermolecular contacts. Although details of the pentameric interactions are not yet known, the highly hydrated character of the hexameric interface is compatible with this quasi-equivalent switching mechanism, because water molecules should be particularly adept at repositioning to accommodate altered orientations in hydrogen bonding and side chain packing geometries.

**Hexamer-Stabilizing NTD-CTD Contacts**

Intermolecular NTD-CTD interactions are made primarily by extended side chains from helix 8 of the CTD, which pack against the C-terminal end of helix 3, the intervening loop, and the...
N-terminal end of helix 4 of the NTD (Figure 3). Additional contacts are also made between helix 11 (CTD) and the C-terminal end of helix 7 (NTD). As in the NTD-NTD interface, NTD-CTD contacts also include a hydrophobic component (Figure 3B), and mutations in two such residues (Y169 and L211) were recently shown to disrupt mature capsid assembly and viral infectivity (Bartonova et al., 2008). Once again, however, polar and water-mediated interactions are prevalent in this region (Figure 3C). In this case, the most striking feature is a series of interdomain helix-capping interactions. As shown in Figure 3D, the R173 side chain (CTD) forms a C cap for helix 3 (NTD); D166 (CTD), an N cap for helix 4 (NTD); and E71 (NTD), an N cap for helix 11 (CTD). In some protomers, Q219 (CTD) also forms an intermolecular C cap for helix 7 (NTD), and R143 (NTD) forms an intramolecular C cap for helix 8 (CTD). Thus, every helix in the NTD-CTD region has a corresponding cap that is provided by the other domain, and the capping residue has an extended and inherently flexible side chain. We speculate that the helix caps can serve as pivot points, allowing relative motions between the NTD and CTD to accommodate the varying degrees of surface curvature required to create a conical lattice.

Inhibitors of Capsid Assembly

The HIV-1 capsid performs an essential role early in the viral replication cycle, and inhibition of capsid assembly by small molecules is therefore being pursued as a therapeutic strategy for the treatment of HIV/AIDS. Indeed, even molecules that simply alter capsid stability may have therapeutic benefit because CA mutations that either enhance or reduce stability (and presumably, the rate of disassembly) lead to dramatic decreases in viral infectivity (Forshey et al., 2002). The NTD-CTD interface appears to be a particularly attractive site for inhibition, because two experimental inhibitors of HIV-1 capsid assembly (CAP-1 and CA-I) both appear to target this set of interactions.

CAP-1 is a small molecule that binds within a hydrophobic pocket at the bottom of the NTD, through an induced-fit mechanism (Kelly et al., 2007; Tang et al., 2003). In the CA hexamer, the aromatic side chain of F32 fills the pocket, forming orthogonal ring-stacking interactions with H62 and Y145 (Figure 4A). The stack is reinforced by the guanidinium group of R162 from the neighboring CTD. Comparison with 2pr, the NTD-CA-I structure (Kelly et al., 2007), shows that the inhibitor displaces the three aromatic side chains from the pocket (arrows in Figure 4B) and disrupts the polar network that stabilizes the helix 3/4 loop conformation. This analysis confirms the idea that CAP-1 acts as an allosteric inhibitor of capsid assembly, by accessing an adjacent hidden pocket and altering the local geometry required to make the NTD-CTD interface (Kelly et al., 2007).

CA-I and its progeny are small peptides that bind to the CTD (Bartonova et al., 2008; Bhattacharya et al., 2008; Sticht et al., 2005; Ternois et al., 2005) and were proposed to disrupt capsid assembly by competitively inhibiting formation of the NTD-CTD interface (Ganser-Pornillos et al., 2007; Sticht et al., 2005; Ternois et al., 2005). Superposition of the NTD-CTD interface with 2buo, the CTD/CA-I cocrystal structure (Ternois et al., 2005), confirms that the helical peptide occludes binding of the NTD to the CTD (Figure 4C), although the binding modes of the two “ligands” are completely different, and their respective interactions are mediated by distinct sets of residues. Specifically, binding of helix 4 to the CTD is mediated primarily by polar contacts (as discussed above), whereas the CA-I peptide occupies a highly hydrophobic pocket that is exposed by a slight opening of the domain (Bartonova et al., 2008; Bhattacharya et al., 2008; Sticht et al., 2005; Ternois et al., 2005).

CA-I induces a different CTD-CTD dimer (Bartonova et al., 2008; Ternois et al., 2005), in terms of the relative orientations of the domains, compared to those observed in this study and in 1a43, the unbound isolated domain (Worthylake et al., 1999) (data not shown). These observations lend support to the proposal that CA-I allosterically induces an “inert” CTD structure that is incapable of extending the capsid lattice (Bartonova et al., 2008). The two possible mechanisms are not exclusive, and CA-I may inhibit assembly by disrupting both NTD-CTD and CTD-CTD interactions.
Mechanisms for Generating Capsid Curvature

Conical fullerene shells are highly curved, and this curvature varies in two different ways. Specifically, the dihedral angles between adjacent hexamer planes vary gradually and continuously along lines of hexamers within the body of the cone, but change more drastically at the ends, at sites immediately adjacent to the pentamers (Li et al., 2000). By analogy to other characterized viral capsids, curvature may be accommodated by structural plasticity in the tertiary folds of the individual subunits, and/or by alterations at the three different intersubunit interfaces.

Although our 3D crystal lattices are not curved, we identified regions in which conformational variation can occur, by comparing our multiple crystallographically distinct CA structures (four distinct hexamers and 19 distinct monomers). On the tertiary level, least-squares superposition of the NTD models from the high-resolution structures illustrates that this domain is essentially invariant, except for the cyclophilin-binding loop, which does not form capsid-stabilizing contacts and is dispensable for assembly (Ganser-Pornillos et al., 2004) (Figure 5A). On the quaternary level, the six-NTD core is also invariant across our four different hexamers (Figure 5C), indicating that distortion of the hexamer ring is not a major mechanism for generating lattice curvature. Nevertheless, the water-mediated NTD-NTD interfaces seen in our structures are likely to facilitate repacking into the pentamer, which is the dominant factor in closing the capsid shell.
In contrast to the NTD, the globular fold of the CTD is more flexible, and the greatest variability is observed in a 15 residue stretch that includes the dimerization helix (Figures 5B and 5S). The structural variation seen in this region could, in principle, affect packing geometry at the CTD-CTD dimerization interface. However, the observed conformations, although intriguing, cannot be unambiguously interpreted in functional terms because they occur at sites where we mutated the protein to prevent formation of hyperstable lattices.

The most significant structural variation we observed is in the relative orientations of the two domains. This variability is best described as rigid-body motions of the CTD subunits relative to the NTD core of the hexamer (Figure 5C). The mobility of the CTD subunits is restricted primarily by their interactions with neighboring NTD subunits, because intramolecular contacts between the two domains of each protomer and intermolecular contacts between adjacent CTDs within each hexamer are minimal. The intermolecular NTD-CTD interface must therefore have inherent plasticity that can accommodate the observed rigid-body motions. A superposition of the independent NTD-CTD interfaces in our high-resolution CA hexamers is shown in Figure 5D. It illustrates that although the two CA domains can move relative to one another, critical direct protein-protein interactions (such as the helix caps) are preserved, because they act as pivot points for the rigid-body motions. At the periphery of the interface, polar contacts are mediated by water molecules (Figure 3C), which can presumably adjust slightly to maintain energetically favorable positions (as discussed above). As with the NTD-NTD contacts, the use of flexible side chains and solvent-mediated interactions for NTD-CTD interactions probably facilitates the formation of the quasi-equivalent pentamer.

The picture that emerges is that the CA hexamer is organized as an apparently rigid ring of NTD subunits, surrounded by a belt of relatively mobile CTD subunits. As interactions between adjacent hexamers are mediated only by the CTD-CTD dimerization interface, variations in NTD-CTD angles within the hexamer effectively change the tilt of the dimerization interface relative to the hexamer plane. This provides an intuitively straightforward mechanism for modulation of dihedral angles between adjacent hexamers to generate curvature in the HIV capsid. Whether this is augmented by variation in CTD-CTD packing at the dimerization interface remains an open question.

**EXPERIMENTAL PROCEDURES**

**Design and Crystallization of CcmK4-Templated CA Hexamers**

A variety of CA-CcmK4 fusion constructs were assembled from coding regions of HIV-1A4/E3 CA and CcmK4 from *Synechocystis* sp. PCC 6803 and verified by DNA sequencing. CA-CcmK4 constructs with different CA end points (residues 221–228) and linker sequences (two to five residues) were constructed, all with C-terminal polyhistidine tags. Protein expression followed the autoinduction method (Studier, 2005). Cell pellets were lysed in buffer (20 mM sodium phosphate [pH 7.4], 500 mM NaCl, 20 mM imidazole) supplemented with protease inhibitors. The soluble fraction was applied to a gravity column packed with MOLREP (Vagin and Teplyakov, 1997) revealed a strong six-fold noncrystallographic axis parallel to the c* axis of the C2 space group, and consideration of likely solvent content was consistent with the presence of six CA-CcmK4 subunits in the asymmetric unit. The cryoEM-based CA hexamer model was positioned with EPMR (Kissinger et al., 1999). Due to the high degree of noncrystallographic symmetry (NCS) in this crystal form, the test set was selected in thin resolution shells with DATAMAN (Kleywegt and Jones, 1996). With the limited resolution, refinement was restricted to treating the separate NTDs and CTDs as rigid bodies with PHENIX (Adams et al., 1999) to yield an R_work of 34% and an R_ref of 36% (Table 1).

Despite extensive effort, we were not able to define a precise location for the CcmK4 portion of the fusion protein. Some density is seen aligned with the CA hexamer six-fold, and fills a volume that must be occupied by CcmK4 in order to complete the crystal lattice, but it was not possible to position a CcmK4 hexamer into this density, and multiple molecular replacement calculations failed to find a convincing solution. Analysis of washed crystals on SDS-PAGE indicated that the fusion protein was intact (data not shown), and our preferred explanation is that CcmK4 can occupy multiple conformations. One extreme possibility is that the CcmK4 hexamers are oriented 50% up and 50% down with respect to the CA hexamer. This is suggested by the location of the N termini on the outer rim of the CcmK4 hexamer, and the apparently equal probability that the CA hexamer might nucleate on either side of the CcmK4 hexamer.

**Design and Characterization of Double-Cysteine Mutants**

Cysteine mutants were based on a pET11a (Novagen) construct harboring HIV-1A4/E3 CA under the control of the T7 promoter. Mutations were introduced with the Quikchange method (Stratagene) and verified by DNA sequencing. The two native cysteines in the CTD (C198 and C218) were retained, since the cryoEM model indicated a very low likelihood of spurious crosslinking with these residues. Proteins were expressed and purified as previously described (Yoo et al., 1997), with the addition of 200 mM j-mercaptoethanol (jME) to all buffers. CA proteins were assembled in vitro either by direct dilution (von Schwedler et al., 1998) or by overnight dialysis (Gross et al., 1997) into assembly buffer (50 mM Tris [pH 8], 1 mM NaCl) containing 20–200 mM jME. Final protein concentrations were 0.1–30 mg/mL. Assembled particles were visualized by transmission EM, as previously described (Ganser-Pornillos et al., 2004). Crosslinking was achieved by subsequent dialysis into assembly buffer with the jME concentration reduced to 20 mM or lower. The extent and efficiency of crosslinking was assessed by nonreducing SDS-PAGE.

**Production and Crystallization of Crosslinked CA Hexamers**

Crosslinked CA A14C/E45C/W184A/M185A hexamers were prepared by sequential dialysis of 10–30 mg/mL protein into assembly buffer containing 200 mM jME, assembly buffer with 0.2 mM jME, and, finally, 20 mM Tris (pH 8). Each dialysis step was performed at 4 °C, for at least 8 hr. The soluble crosslinked hexamers were somewhat prone to aggregation, but remained competent for crystal formation even after storage at 4 °C for several days. The crosslinked hexamers readily formed several visually distinct crystal forms. The best crystals showed hexagonal and prism-like morphology, and were obtained with the same precipitant (10%–12% PEG 8,000) and protein-precipitant ratio (2:1), but at different pH and temperature (hexagonal = 100 mM sodium malonate [pH 6.5], 4 °C; prism = 100 mM Tris [pH 7.4], 20 °C). Crystals were crosslinked by soaking in mother liquor containing 30% glycerol or ethylene glycol for 10 min (in 10% increments).

**Structure Determination of CcmK4-Templated CA**

Data were collected at APS beamline 22-BM. The hexagonal crystals had unit cell parameters of a = b = 157.3 Å, c = 56.8 Å, α = β = 90°, and γ = 120°.
Data on the prism-like crystals were collected at SSRL beamline 7-1. Orthorhombic crystal form (which displayed no diffraction pathologies).}

Two-thirds of the reflections were systematically weak, indicative of translational pseudosymmetry. Strong reflections followed the selection rule \((h,k,l) \pm (3n,0,0)\), and were on average four times larger than the weak reflections. A Patterson map calculated with only the strong subset showed a peak of equal intensity to the origin at fractional coordinates \((0.67,0.33,0.0)\) (peak intensity was 80% of origin when calculated with all data). Using only the strong reflections, the data can therefore be indexed in space group \(P6\) with a smaller unit cell \((a' = b' = 91.0 \text{ Å}, c' = 56.8 \text{ Å}, R_{sym} = 10\%)\) (Table 1) containing one CA protein in the asymmetric unit. Note that the pseudo-cell dimensions closely match the dimensions of the 2D crystal lattice in the cryoEM structure (92.7 Å) (Ganser-Pornillos et al., 2007), and that the \(a'\) and \(c'\) edges in the pseudo-cell are related to the true cell \(a\) and \(c\) edges by the equations \(a' = a/\sqrt{3}\) and \(c' = c\), respectively. These indicated that the crystal was composed of stacked sheets of CA hexamers and that each sheet is a flattened version of the 2D CA lattice within the capsid (Figure S2).

Molecular replacement in the pseudo-cell setting was performed with MOLREP (Vagin and Teplyakov, 1997). To provide a check against possible model bias, we used crystal structures of the NTD and CTD in complex with assembly inhibitors (2pxr and 2buo, respectively). Our expectation was that the model-phased map would indicate different polypeptide conformations at the inhibitor-binding sites compared to the search models, and indeed, these were observed. The map also showed clearly defined density for regions that were absent from the search models. The merged intensities and rigid-body refined coordinates \((R_{merge} = 39\%)\) were submitted to the Bias Removal Server (http://tuna.tamu.edu/) for map calculation with the Shake\&wARP algorithm (Reddy et al., 2003) (Figure S2A). The full model was rebuilt manually into this map with COOT (Emsley and Cowtan, 2004). Positional and isotropic \(B\) factor refinement were performed in PHENIX (Adams et al., 2002), with simulated annealing and automated water-picking protocols. The current model has an \(R_{factor}\) of 23% and \(R_{merge}\) of 27% (against randomly selected 5% of the data), with good geometry and no residues in disallowed regions of the Ramachandran plot (Table 1).

Statistical analyses of the reflection intensities, test refinements, and real space considerations indicated that the true space group is most likely perfectly hemihedrally twinned \(P3\) (twin law = \("-h,-k,l\"\)), with one hexamer in the asymmetric unit. The combined pathologies of pseudosymmetry and twinning have made refinement in the true space group problematic. We therefore chose to simply report the structure in the pseudo-cell setting, with the understanding that it does not completely reflect the structural plasticity of the protein. The structure represents an average of both the pseudotranslationally related molecules (because only the strong reflections were used) and the two “twinned domains” (because the twin-related reflections were merged). Fortunately, the pseudosymmetry and twinning in this crystal form appeared to be mainly due to alternative conformations in a small proportion of the monomer, spanning ~15 residues at the helix 8/9 loop and the N-terminal end of helix 9. This region was characterized by ill-defined density in this crystal form and therefore not modeled. As illustrated in Figure S5, this same region was also variable in the orthorhombic crystal form (which displayed no diffraction pathologies).

**Structure Determination of Crosslinked CA, Orthorhombic Crystal Form**

Data on the prism-like crystals were collected at SSRL beamline 7-1. On the basis of systematic absences, the space group was identified as \(P2_12_12_1\) \((R_{sym} = 8.5\%)\), with two hexamers in the asymmetric unit (Figure S3). This crystal form was also solved by molecular replacement in MOLREP (Vagin and Teplyakov, 1997) \((R = 49\%)\), with a hexameric search model derived from the partially refined structure in the hexagonal pseudo-cell. The solution was deemed reliable by the appearance of unbiased density for regions that were deliberately deleted from the search model, and further confirmed with an anomalous difference density map derived from a selenomethionine data set collected at APS beamline 22-ID to 3.5 Å (densities for all 120 Se sites in the asymmetric unit were clearly visible at 2-15σ; data not shown). The two hexamers in the asymmetric unit are stacked head to head with approximately coincident six-fold axes. The self-rotation function of the molecular replacement solution was identical to the experimental.

The test set was selected in thin resolution shells with DATAMAN (Kleywegt and Jones, 1996). Initially, 22 domains in the asymmetric unit, omitting the NTD and CTD of one CA molecule, were refined as rigid bodies in REFMAC (Murshudov et al., 1997) \((R_{free} = 39\%)\). The domains were refitted into their corresponding omit maps with COOT (Emsley and Cowtan, 2004) and then copied onto the other molecules with the NCS transformation matrices. Simulated annealing and omit refinement were performed in PHENIX (Adams et al., 2002). Density for the NTD was of significantly better quality compared to the CTD. Subsequent rounds of model building used NCS-averaged maps calculated separately for the two domains, simulated annealing omit maps, and a SHAKE\&wARP map (Reddy et al., 2003). The highly variable regions at the CTD (residues 176–187) were left unmodeled until the last refinement cycle, to obtain the best unbiased maps for chain tracing (Figure S5). This region was completely modeled in chains C, E, F, J, and L, partially modeled in A, B, G, and I, and unmodeled in D, H, and K. Because of the relatively poor quality of the electron density, the chain traces for this region must be considered tentative. The quality of the density for helix 10 was also highly variable. We attempted to derive a model that would account for the observed flexibility in the protein while taking advantage of the 12-fold improvement in observation/parameter ratio afforded by NCS. The current best approach (adapted from ter Haar et al., 1998) is to define four segments of the CTD as separate NCS groups (residues 149–174, 175–189, 190–204, and 209–219). The global region of the NTD, the helix 6/7 loop, and \(\beta\)-hairpin were also defined as separate NCS groups. The current model has \(R_{factor}\) and \(R_{merge}\) of 24% and 26%, respectively, with good geometry (Table 1).

**ACCESSION NUMBERS**

Coordinates and structure factors are available from http://www.rcsb.org:4800/pdb/3h4e; crosslinked CA, hexagonal crystal form, 3h47; and cross-linked CA, orthorhombic crystal form, 3h4e.

**Supplemental Data**

Supplemental Data include five figures and can be found with this article online at http://www.cell.com/supplemental/S0092-8674(09)00580-7.

**ACKNOWLEDGMENTS**

We thank Kent Baker and V. Mitch Luna for technical assistance, Todd Yeates for crystallographic discussions, Jason Harless for clerical support, and the DNA synthesis and sequencing core facilities at The Scripps Research Institute and University of Utah. Some of the electron microscopy experiments were conducted at the National Resource for Automated Molecular Microscopy (NRAMM), which is supported by the National Institutes of Health (NIH) through the National Center for Research Resources (NCRR). X-ray diffraction data were collected at Stanford Synchrotron Radiation Laboratory (SSRL) beamline 7-1 and at Southeast Regional Collaborative Access Team (SERCAT) beemlines 22-BM and 22-ID at Advanced Photon Source (APS), Argonne National Laboratory. SSRL and APS are supported by the U.S. Department of Energy, Office of Biological and Environmental Research, The SSRL Structural Molecular Biology Program is also supported by the NIH through the NCRR, and the National Institute of General Medical Sciences. This study was funded by NIH grants R01-GM06067 (M.Y.) and P50-GM082545 (M.Y., C.P.H., and W.I.S.). B.G.-P. was supported by a postdoctoral fellowship from the George E. Hewitt Foundation for Medical Research.

Received: March 23, 2009
Revised: April 23, 2009
Accepted: April 30, 2009
Published online: June 11, 2009

**REFERENCES**


