Structures of the CCR5 N Terminus and of a Tyrosine-Sulfated Antibody with HIV-1 gp120 and CD4

Chih-chin Huang,1*, Son N. Lam,2*, Priyamvada Acharya,3 Min Tang,1 Shi-Hua Xiang,3 Syed Shahzad-ul Hussan,2 Robyn L. Stanfield,4 James Robinson,3 Joseph Sodroski,3 Ian A. Wilson,4 Richard Wyatt,1 Carole A. Bewley,2 Son N. Lam,2* Priyamvada Acharya,3 Min Tang,1 Shi-Hua Xiang,3 Syed Shahzad-ul Hussan,2 Robyn L. Stanfield,4 James Robinson,3 Joseph Sodroski,3

The CCR5 co-receptor binds to the HIV-1 gp120 envelope glycoprotein and facilitates HIV-1 entry into cells. Its N terminus is tyrosine-sulfated, as are many antibodies that react with the co-receptor binding site on gp120. We applied nuclear magnetic resonance and crystallographic techniques to analyze the structure of the CCR5 N terminus and that of the tyrosine-sulfated antibody 412d in complex with gp120 and CD4. The conformations of tyrosine-sulfated regions of CCR5 (α-helix) and 412d (extended-loop) are surprisingly different. Nonetheless, a critical sulfotyrosine on CCR5 and on 412d induces a rearrangement of nuclear magnetic resonance (NMR) and x-ray structures (Fig. 1C). Structure calculations were carried out on the ordered region comprising residues 7 to 15. A total of 70 distance restraints (corresponding to 35 intraresidue and 35 interresidue NOEs), and 56 dihedral angle restraints were included in the final round of structure calculations, which gave rise to an ensemble of 40 structures with a backbone root-mean-square deviation (rmsd) of 0.46 Å and an rmsd of 1.39 Å for all atoms in the ordered region (residues 9 to 14) (table S2). Superpositions of the final ensemble defined a helical conformation for residues 9 to 15, which was variable in length, and three extracellular loops (ECLs) (Fig. 1A). The structure of the co-receptor has not been determined, but some insight has come from the crystal structures of other family members (4).

Elements critical to interactions with HIV-1 are located in the co-receptor N terminus and around its second extracellular loop (ECL2) (5–8). The co-receptor N terminus interacts with a highly conserved 4-stranded bridging sheet in gp120, which assembles upon CD4 binding, whereas the ECL2 region of the co-receptor interacts with the tip of the immunodominant V3 loop in gp120. Considerable distance separates these two interactive regions, which suggests that they are independent (9–12).

The N-terminal interaction of co-receptor with HIV-1 requires an unusual posttranslational modification, O-sulfation of tyrosine (13). On CCR5, tyrosines at residues 3, 10, 14, and 15 may be O-sulfated, but sulfations at residues 10 and 14 are sufficient to facilitate interaction with HIV-1 (14). Interestingly, many CD4-induced antibodies that react with the bridging sheet region are also modified by O-sulfation (15). To define structurally the interaction of HIV-1 with the N terminus of CCR5 and to understand the molecular details of the mimicry of this interaction by CD4-induced antibodies, we used a combination of nuclear magnetic resonance (NMR) and x-ray crystallography to determine the structures of the N terminus of CCR5 and of a functionally sulfated antibody, 412d, in complex with HIV-1 gp120. Analysis of these structures, combined with molecular docking and saturation transfer difference NMR, identified a conserved site on gp120, which recognizes sulfotyrosine with high selectivity.

We used NMR techniques that exploit the transfer of information from bound to ligand-free states (16, 17) to analyze the interactions of a 14-residue peptide (CCR52-15), which consisted of residues 2 to 15 of CCR5 with sulfotyrosine (Tys) at positions 10 and 14 (Fig. 1) (18). We collected two-dimensional (2D) nuclear Overhauser enhancement spectroscopy (NOESY) spectra of solutions containing CCR52-15 either free or in the presence of gp120, CD4, or a gp120-CD4 complex (peptide:protein ratio of 40:1). Whereas spectra containing free CCR52-15 and CCR52-15 with either gp120 or CD4 contained few cross peaks, CCR52-15 in the presence of the gp120-CD4 complex gave rise to high-quality spectra containing numerous NOEs (Fig. 1B and fig. S1). Complete 1H, 13C, and 15N assignments of CCR52-15 (table S1) were made on the basis of standard 2D homonuclear and heteronuclear NMR experiments that measure scalar and dipolar couplings.

The NOESY data of CCR52-15 in the presence of gp120-CD4 (Fig. 1B) were sufficient for calculating a high quality ensemble of NMR structures (Fig. 1C). Structure calculations were carried out on the ordered region comprising residues 7 to 15. A total of 70 distance restraints (corresponding to 35 intraresidue and 35 interresidue NOEs), and 56 dihedral angle restraints were included in the final round of structure calculations, which gave rise to an ensemble of 40 structures with a backbone root-mean-square deviation (rmsd) of 0.46 Å and an rmsd of 1.39 Å for all atoms in the ordered region (residues 9 to 14) (table S2). Superpositions of the final ensemble defined a helical conformation for residues 9 to 15, which deviated from the ideal by a backbone rmsd of only 0.26 Å (Fig. 1D). Sulfotyrosines 10 and 14 extended from the same face of the helix, with sulfate moieties separated by ~10 Å and an ~90° rotation around the helix axis.

We were unable to obtain crystals of CCR52-15 in complex with HIV-1 gp120-CD4, and the size and glycosylation of the ternary complex hindered direct determination by NMR. We were, however, able to obtain ~3.5 Å diffraction from crystals of the antigen-binding fragment (Fab) of the 412d antibody, in complex with gp120 (core with V3, CCR5-dependent isolate YU2) and CD4. The 412d antibody is functionally tyrosine-sulfated, binds to a CD4-induced epitope that overlaps the site of co-receptor binding on HIV-1 gp120, and recognizes preferentially CCR5-dependent strains of HIV-1 gp120 (15). Moreover, the tyrosine-sulfated region of 412d can be sub-
stituted for the tyrosine-sulfated region of CCR5 to create a chimeric 412d/CCR5 receptor that supports HIV-1 entry (19).

We solved the 412d-gp120-CD4 structure by molecular replacement. Despite less than optimal resolution and completeness, initial unbiased maps showed clear definition of important antibody features (fig. S2). Structure refinement resulted in an R_	ext{cryst} of 20% (R_	ext{free} 27%) (Fig. 2, table S3, and fig. S3). The increased 412d interaction surface is due primarily to an increase in buried surface associated with its CDR H3. Comparison of free (20) and bound structures of 412d shows that extensive ordering occurs in CDR H3 when bound to gp120 (fig. S5).

The two sulfotyrosines in the CDR H3 region of 412d bind to gp120 in quite different ways (Fig. 2). The sulfotyrosine at residue 100 of 412d (Tys 100412d) [Kabat numbering (21)] is mostly exposed, with its aromatic ring making p-cation interactions with the guanidinium of Arg 327gp120 and its sulfate group making only peripheral electrostatic interactions. By contrast, the side-chain of Tys 100c412d is mostly buried, with Ile 322gp120 and Ile 326gp120 embracing one face of the tyrosine ring, while the aliphatic base of Arg 440gp120 supports the other. Together, the two sulfotyrosines account for about 20% of the total buried surface on 412d, with almost 100 Å² derived from Tys 100c412d.

To facilitate interactions with the sulfotyrosines in 412d, the V3 stem is rearranged. The conserved Arg 298bp120 and Pro 299bp120 at the base of the V3 loop are mostly unchanged, but the subsequent Asn residues at 301gp120 and 302gp120 shift ~7 Å to form one wall of the Tys 100c412d sulfate-binding pocket. Residue 301bp120 is N-glycosylated, but the glycan faces solvent, and its presence should have little impact on the ability of the binding pocket to form. Meanwhile, in the returning strand (22), Ile...
Fig. 2. Structure of the tyrosine-sulfated antibody 412d in complex with HIV-1 gp120 and CD4. (A) Ribbon representation. CD4 is yellow, the heavy chain of Fab 412d is dark blue, the light chain is cyan, and gp120 is gray, except for the V3 loop, which is orange. The CDR H3 loop of 412d is red, with sulfo tyrosines depicted in stick representation. (B) Close-up, with molecular surface of gp120 in gray and sulfotyrosines of 412d (red labels) and select residues of gp120 (black labels) in stick representation. Dotted lines represent coordinating hydrogen bonds between gp120 and the sulfate group of Tys100c412d. The sulfate of Tys 100c412d makes a full complement of ionic interactions: a salt bridge to Arg 298gp120 and hydrogen bonds to the side-chain nitrogen of Asn 302gp120, the side-chain hydroxyl of Thr 303gp120, and the main-chain amides of 302gp120, 303gp120, and 441gp120 (34).

Fig. 3. Interaction of the N terminus of CCR5 with HIV-1 gp120-CD4. (A) Molecular docking. The 20 lowest energy structures (black) from 200 docking runs of CCR5-15 are shown in stick representation. Despite initial random orientations, all favorable docking solutions had Tys 14 binding at the bridging sheet-V3 interface; none had Tys 10 at this cleft. Ribbon representations illustrate CD4 in yellow, gp120 in gray (with V3 in orange), and the lowest energy structure of CCR5-15 in purple. (B) Close-up, with molecular surface of gp120 in gray and select residues of gp120 (black labels) and CCR5 (purple labels) in stick representation. (C) Saturation transfer difference NMR spectrum of CCR5-15 in the presence of gp120-CD4 (red) overlaid on a control 1H spectrum (black). Experimental conditions were identical to those used for NOE experiments, except that the carrier was set at –1 and 50 parts per million for on- and off-resonance saturation, respectively. The intensities of the most strongly enhanced peaks (Tys 14 and Tyr 15) have been normalized to the corresponding signals in the control spectrum. Peak assignments made by 2D NMR (table S1) appear above their corresponding doublet signals. Tys 14 and Tyr 15 show strong saturation transfer difference effects, whereas Tys 10 shows a medium effect and Tyr 3 a very weak effect. These effects correlate directly with the buried surface area of each tyrosine ring in the docked structure. See fig. S9 for overlaid spectra employing 1 to 7 s saturation. (D) Effect of CCR5 2–15 on the proteolytic sensitivity of the V3. Electrophoresis on an 8% to 25% gradient SDS polyacrylamide gel shows the results of thrombin digestion on gp120 (core with V3; YU2 R5 strain of HIV-1) alone, or in the presence of sCD4 or sCD4 and CCR5-15 (35). (E) Structural intermediates of HIV-1 entry. At far left, a single monomer of unliganded gp120 (gray) is shown with separated β-hairpins. The threefold axis, from which gp41 interacts in the functional oligomer, is labeled with the number 3. In the CD4-bound state, the bridging sheet assembles, and the V3 (orange) is exposed and flexible. The next state involves either (upper pathway) the interaction of the CCR5-ECL2 region with the V3 tip or (lower pathway) the interaction of the CCR5 N terminus, which induces rigidification of the V3 stem. Engagement of CCR5 at both N terminus and ECL2 region triggers additional conformational changes leading to HIV-1 entry.
getically favorable interaction (orientations, multiple runs of the excised CDR would recapitulate the 412d-gp120 crystal structure. We also observed good correlation between interface results in formation of a more rigid V3.

By employing molecular docking and saturation transfer difference NMR, we sought to use the 412d-gp120-CD4 structure to ascertain how gp120 interacts with the N terminus of CCR5. We first tested whether docking (Autodock 3.0) of the CDR H3 loop of 412d to gp120 would recapitulate the 412d-gp120 crystal structure. Starting from random initial positions and orientations, multiple runs of the excised CDR H3 loop (residues 97 to 1000) produced an energetically favorable interaction (~16.04 kcal/mol), which closely resembled its location and contacts in the crystal structure (Cx rmsd between crystal and docked CDR H3 was 1.03 Å) (fig. S6). We next docked the NMR structure of the CCR5 N terminus to the crystal structure of gp120-CD4. Multiple runs produced a cluster of energetically favorable solutions (~17.60 kcal/mol for the optimal solution), which placed CCR52-15 at the bridging sheet-V3 interface (fig. 3, A and B). The top 10% of the solutions (20 best solutions from 200 runs) had rmsds of 1.04 Å (Cx) and 2.24 Å (all atoms).

To validate the docked CCR5-gp120 structure, we performed saturation transfer difference NMR (17 on CCR52-15 in the presence of gp120-CD4. Control and difference spectra are shown in fig. 3C. Contact surfaces of Tys and Tyr residues of CCR5 in the docked orientation correlated well with saturation transfer difference enhancements (fig. 3C). We also observed good correlation between interacting residues in the docked gp120-CCR5 interface and gp120 and CCR5 substitutions (9, 25–27) that affect gp120-CCR5 binding (fig. S7).

The N terminus of CCR5 approaches from the same face of gp120 as CD4 but binds to an orthogonal surface at the intersection of the bridging sheet and the V3 loop (fig. 3). The first CCR5 residues (Ser 7 and Pro 8) that are ordered in the NMR structure interact with the V3 stem. In the helix (residues 9 to 15), Tys 10 interacts with the gp120 core and forms a salt bridge with Arg 327gp120. Asp 11 forms an ionic interaction with Arg 440gp120, and Tys 14 is completely sequestered in the crevice between V3 and the bridging sheet, and the aromatic ring of Tyr 15 packs against Ile 439gp120 on the bridging sheet. The structural rearrangements required to form the Tys 14 binding pocket would be expected to rigidify the V3 stem. We tested V3-proteolytic susceptibility (fig. 3D). CD4 enhances V3-proteolytic susceptibility to thrombin (28, 29), whereas the combination of CD4 and CCR52-15 reduced proteolytic susceptibility (fig. 3D), consistent with CCR5-rigidification of V3.

Overall, the gp120 recognition surface for CCR52-15 is much more highly conserved for CCR5-dependent isolates compared with those that use CXCR4. Good electrostatic complementarity is found between the acidic CCR52-15 and gp120, where the negatively charged C-terminus helix dipole is oriented toward the basic bridging sheet (fig. S8). The docked structure provides an explanation for the observed lack of order at the N terminus of CCR52-15, where CCR5 appears to extend away from gp120. At the C terminus, Tyr 15 points toward the target cell membrane where, in five residues, a disulfide would normally be made between the N terminus (Cys 20) and the third extracellular loop (Cys 269).

Despite the highly divergent tyrosine-sulfated structures of 412d and CCR5, a single sulfotyrosine (residue 100c in 412d and residue 14 in CCR5) is recognized in a similar manner by gp120 (fig. 4). We used mutagenesis to probe the degree of similarity in this recognition (fig. S10). The alteration of a single nitrogen in a contact residue (Asn302Asp) in the conserved binding pocket ablates recognition of both 412d and CCR5, whereas a similar substitution (Asn300Asp), just outside the binding pocket, had little effect (30).

The observed convergence of recognition likely reflects the high selectivity of this site for sulfotyrosine (a 7 Å deep pocket, with hydrophobic walls and a cationic floor, which is unlikely to interact favorably with other nonmodified amino acids). Such selectivity and favorable energetics bode well for design of therapeutics targeted at this site, because the gp120 residues that line the sulfotyrosine binding pocket are highly conserved for co-receptor binding.

The structure of the CCR5 N terminus with gp120-CD4 provides a further snapshot of the HIV-1 entry pathway (fig. 3E). Before binding CD4, the bridging sheet is not formed and the V3 loop is occluded. Binding of CD4 induces bridging sheet assembly and V3 exposure. At this stage, the V3 is flexible and poised close to the target cell membrane. Subsequent interactions with CCR5 are still being elucidated. We show structural details for one: engagement by gp120 of the CCR5 N terminus, which requires formation of a conserved pocket for sulfotyrosine binding and converts the flexible V3 stem into a rigid β-hairpin. It will be interesting to integrate the order and timing of the rearrangements revealed here into the HIV-1 entry mechanism.

References and Notes

www.sciencemag.org

SCIENCE VOL 317 28 SEPTEMBER 2007

1933
The Slit Receptor EVA-1 Coactivates a SAX-3/Robo–Mediated Guidance Signal in *C. elegans*

Kazuko Fujisawa, Jeffrey L. Wrana, Joseph G. Culotti

The SAX-3/roundabout (Robo) receptor has Shiga-like toxin 1 (SLT-1)/Slit–dependent and –independent functions in guiding cell and axon migrations. We identified enhancer of ventral-axon guidance defects of *unc-40* mutants (EVA-1) as a *Caenorhabditis elegans* transmembrane receptor for SLT-1. EVA-1 has two predicted galactose-binding ectodomains, acts cell-autonomously for SLT-1/Slit–dependent axon migration functions of SAX-3/Robo, binds to SLT-1 and SAX-3, colocalizes with SAX-3 on cells, and provides cell specificity to the activation of SAX-3 signaling by SLT-1. Double mutants of *eva-1* or *slt-1* with *sax-3* mutations suggest that SAX-3 can (when *slt-1* or *eva-1* function is reduced) inhibit a parallel-acting guidance mechanism, which involves UNC-40/deleted in colorectal cancer.

The UNC-6/netrin guidance cue and its neuronal receptors, UNC-5 and UNC-40/deleted in colorectal cancer (DCC), are used in different combinations to guide growing axons toward (by attraction) or away (by repulsion) from the ventral nerve cord (VNC) of *Caenorhabditis elegans* (1). The incomplete penetrance of pioneer-axon guidance defects observed in *unc-6/netrin* and *unc-40* single- and double-null mutants (Table 1) suggests that other mechanisms act in parallel with netrin signaling to guide axons toward the VNC. One such mechanism involves the Shiga-like toxin 1 (SLT-1)/Slit guidance cue, a large secreted protein with several predicted N- and O-glycosylation sites (2), and its receptor SAX-3, a homolog of the transmembrane (TM) roundabout (Robo) receptor (3–6). Both *Drosophila* and vertebrate slit bind to Robo receptors (3, 7). *C. elegans* SLT-1/Slit is expressed predominantly by dorsal body-wall muscles and repels SAX-3/Robo–expressing AVM and PVM pioneer axons toward the VNC (2), concomitant with UNC-40–mediated attraction of these same axons toward the VNC by ventral sources of UNC-6 (1).

In *C. elegans*, *slt-1* and *sax-3* mutations affect the guidance of several of the same pioneer axons (8). For example, the pioneer axon of the lateral AVM sensory neuron in the anterior body extends toward and then along the VNC in wild-type (WT) animals (Fig. 1, A and B), but in *slt-1* and *sax-3* mutants, the AVM axon frequently grows directly toward the head (Fig. 1C). Cell-specific rescue experiments have demonstrated that *sax-3(s)*–dependent guidance of AVM axons is cell-autonomous (6, 8). Although SAX-3/Robo is the only previously known receptor for SLT-1, *slt-1* mutants of *C. elegans* do not exhibit the nerve-ring and epithelial defects of *sax-3(s)/robo* mutants, suggesting that SAX-3/Robo has both Slit-dependent and -independent functions in development (2).

We identified a TM protein, enhancer of ventral-axon guidance defects of *unc-40* mutants (EVA-1), that is required to guide the AVM pioneer axon to the VNC (Fig. 1, A and B) by acting as a receptor for SLT-1. EVA-1 acts cell-autonomously, and ectopic expression of EVA-1 in SAX-3–expressing cells confers SLT-1 sensitivity to their migration. Thus, EVA-1 is predicted to be a receptor for SLT-1 that acts in conjunction with SAX-3 (as a likely co-receptor) to provide cell specificity for the activation of SAX-3 signaling by SLT-1. We also discovered a previously unknown in vivo function for SAX-3/Robo, which is to inhibit a signaling mechanism that normally functions in parallel to SLT-1 to guide pioneer axons along the dorsal/
Structure of Full-Length HIV-1 CA: A Model for the Mature Capsid Lattice

Barbie K. Ganser-Pornillos,1 Anchi Cheng,1 and Mark Yeager1,2,3,*
1The Scripps Research Institute, Department of Cell Biology, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
2Scripps Clinic, Division of Cardiovascular Diseases, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA
3University of Virginia Health System, Department of Molecular Physiology and Biological Physics, PO Box 800736, Charlottesville, VA 22908, USA
*Correspondence: yeager@scripps.edu
DOI 10.1016/j.cell.2007.08.018

SUMMARY

The capsids of mature retroviruses perform the essential function of organizing the viral genome for efficient replication. These capsids are modeled as fullerene structures composed of closed hexameric arrays of the viral CA protein, but a high-resolution structure of the lattice has remained elusive. A three-dimensional map of two-dimensional crystals of the R18L mutant of HIV-1 CA was derived by electron cryocryotography. The docking of high-resolution domain structures into the map yielded the first unambiguous model for full-length HIV-1 CA.

Three important protein-protein assembly interfaces are required for capsid formation. Each CA hexamer is composed of an inner ring of six N-terminal domains and an outer ring of C-terminal domains that form dimeric linkers connecting neighboring hexamers. Interactions between the two domains of CA further stabilize the hexamer and provide a structural explanation for the mechanism of action of known HIV-1 assembly inhibitors.

INTRODUCTION

During the late stages of their replication cycle, retroviruses assemble as “immature” virions composed of 2000–4000 copies of the virally encoded Gag polyprotein (Briggs et al., 2004, 2006; Parker et al., 2001). Concomitant with release from an infected cell, virions undergo a process called “maturation,” during which Gag is proteolytically cleaved into three major proteins: the MA protein remains associated with the viral membrane, the NC protein recruits and packages the viral genome, and the CA protein oligomerizes to form a closed “mature” capsid that surrounds this NC/RNA complex (Göttlinger, 2001; Kräusslich, 1996). Cleavage of Gag and assembly of the mature capsid are essential for infectivity.

Mature retroviral capsids can be conical, spherical, or cylindrical (Vogt, 1997). Despite this variation in capsid morphology, it is now established that retroviral CA proteins adopt the same tertiary structure, and that the different capsid shapes arise from common design principles. Retroviral CA proteins are composed of two independently folded domains, the N-terminal domain (NTD) and the C-terminal domain (CTD), connected by a flexible linker (Berthet-Colominas et al., 1999; Campos-Olivas et al., 2000; Cornilescu et al., 2000; Gamble et al., 1996, 1997; Gitti et al., 1996; Jin et al., 1999; Khorasanizadeh et al., 1999; Kingston et al., 2000; Momany et al., 1996; Mortuza et al., 2004; Tang et al., 2002; Worthinglake et al., 1999). Both domains of CA have predominantly α-helical secondary structure: the HIV-1 CA NTD is shaped as an arrowhead, with seven α helices (numbered 1–7) and an amino-terminal β-hairpin (Berthet-Colominas et al., 1999; Gamble et al., 1996, 1997; Gitti et al., 1996), while the CTD is more globular, with a single-turn 310-helix and four short α helices (numbered 8–11) (Berthet-Colominas et al., 1999; Gamble et al., 1997; Worthinglake et al., 1999). The prevailing model is that all mature capsids are built on hexameric lattices of CA and closed by incorporating 12 pentameric declinations. The relative distribution of the declinations within the hexameric lattice defines the shape of the capsid: tubular capsids are formed by introducing 6 pentamers at each end of a cylindrical tube, conical capsids result from an asymmetric distribution of the pentamers at the two ends of a conical assembly, and “spherical” capsids arise from a more even distribution of the pentamers throughout the hexameric lattice (Ganser et al., 1999; Ganser-Pornillos et al., 2004; Jin et al., 1999; Li et al., 2000).

This general model is supported by extensive structural, biochemical, and genetic data. For example, low-resolution EM studies of in vitro assemblies of CA, such as cylinders and two-dimensional (2D) crystals, as well as bona fide mature capsids have shown that retroviral CA proteins form a hexameric lattice, with the NTD apparently forming the hexamers, and the CTD forming dimeric linkers that connect neighboring hexamers (Briggs et al., 2003; Li et al., 2000; Mayo et al., 2002, 2003). In addition, X-ray crystallographic studies of
truncated CA proteins have revealed an atomic resolution model for the hexamer formed by the NTD of N-tropic murine leukemia virus (Mortuza et al., 2004) and several possible models for the CTD dimer (Gamble et al., 1997; Ivanov et al., 2007; Jin et al., 1999; Termon et al., 2005; Worthylake et al., 1999). Finally, biochemical and genetic experiments suggest the existence of NTD-CTD interactions that are essential for capsid formation (Bowzard et al., 2001; Ganser-Pornillos et al., 2004; Lanman et al., 2003, 2004). In spite of these substantial results, a high-resolution structure of full-length, hexameric CA has never been visualized.

An alternative approach to derive a molecular model of full-length CA is to determine a moderate resolution structure by electron crystallography of 2D crystals and dock NMR or X-ray crystal structures of the individual domains. Here, we report the three-dimensional structure of hexameric arrays of full-length HIV-1 CA at 9 Å resolution using electron cryomicroscopy (cryoEM) and image analysis. Docking the high-resolution structures of the two HIV-1 CA domains into the cryoEM density map enabled visualization of a pseudoatomic model of full-length CA hexamers. The lattice parameters of the 2D crystals closely match the packing of CA in authentic mature virions (Briggs et al., 2003; Yeager et al., 1998). Therefore, we believe that these crystals recapitulate the hexameric lattice of HIV-1 CA in mature infectious particles.

RESULTS AND DISCUSSION

Assembly and Structure Determination of Full-Length HIV-1 CA Hexamers

Wild-type full-length HIV-1 CA assembles in vitro into long cylinders (Figure 1A) (Ehrlich et al., 1992). In contrast, a single point mutation, R18A, shifts the in vitro assembly phenotype into spheres, cones, and cylinders – the three mature capsid shapes observed in retrovirions (Ganser-Pornillos et al., 2004). We have examined other amino acid substitutions of Arg18 and discovered that replacement with a large hydrophobic residue (i.e., Val, Ile, Leu, Phe) generated CA proteins that assembled predominantly as spheres, as illustrated in Figure 1B for the mutant R18L. Optimization of the assembly conditions yielded cylinders, cones, and very large spheres up to ~2.5 μm in diameter. When applied to a continuous carbon EM grid, these large spheres collapsed and flattened, and behaved as well-ordered 2D crystals (Figure 1C), exhibiting diffraction to beyond 10 Å resolution (Figure 1D). As expected, optical diffraction analysis revealed two non-overlapping hexagonal lattices arising from the top and bottom layers of the flattened spheres (Figure 1D). Analysis of low-dose images of frozen-hydrated crystals at a series of tilt angles yielded a three-dimensional map of full-length HIV-1 CA at 9.0 Å resolution parallel to the plane of the flattened sphere and 20 Å perpendicular to it (Figure 2A and Figure S1 in the Supplemental Data available with this article online; Table 1).

Figure 1. Assembly and Initial Characterization of the HIV-1 CA Hexameric Lattice

(A) Negative stain electron micrograph of recombinant wild-type HIV-1 CA assembled into long, hollow tubes in vitro. The scale bar represents 100 nm.

(B) Under the same conditions used in panel (A), R18L HIV-1 CA assembles into small spheres.

(C) After optimizing the assembly conditions, R18L HIV-1 CA assembles into large spheres, which behaved as 2D crystals when flattened on continuous carbon EM grids.

(D) Fourier transform of an image of a flattened sphere preserved in vitreous ice, collected at 0° tilt (arrow points to a reflection at 9.8 Å⁻¹). The transform displays two superimposed hexameric lattices that arise from the top and bottom layers of the flattened sphere.

Description of the CryoEM Map and Fitting of the NTD and CTD

Full-length HIV-1 CA packs as a hexamer within the flattened spheres (p6 plane group; a = b = 92.7 Å) (Figure 2A, Tables 1 and S1). The CA hexamers have an exterior diameter of ~90 Å, and the protein shell is ~60 Å thick. Given the dimensions of the NTD (18 × 35 × 45 Å) and CTD (29 × 36 × 28 Å), this is a surprisingly compact structure, indicating that the two domains of CA are in close contact. There are several well-resolved lobes of density within each monomer, some of which appear rod-like and can be interpreted as individual α helices (for example, see Figures 3A and 4).

To generate an unbiased molecular model from the experimental density map, we used a correlation-based program to perform automated rigid-body fitting of ordered residues from high-resolution structures of the isolated CA NTD (residues 1–148; pdb code 1gwp) and CTD (residues 149–219; pdb code 1a43) (see Experimental Procedures) (Chacón and Wrighers, 2002; Gitti et al., 1996; Worthylake et al., 1999; Wrighgers et al., 1999). As shown in Figures 2–5, a pseudoatomic model based on docking the high-resolution NTD and CTD structures
Figure 2. Structure of the HIV-1 CA Hexamer

(A) Slab view of the experimental map (contoured at 1.5 \( \sigma \)) and pseudoatomic model with the NTD in green and CTD in blue. The hexamer is outlined in red, and the p6 unit cell is outlined in yellow. The six-, three-, and two-fold symmetry axes are indicated by hexagons, triangles, and ellipses, respectively.

(B) Side view of the pseudoatomic monomer model of full-length CA, with the secondary structural elements labeled. The carboxy terminus of the NTD and amino terminus of the CTD (\( \sim 4 \) Å apart) are indicated by yellow asterisks.

(C) Top view of one hexamer, colored as in (A). For clarity, one monomer is outlined in white.

(D) Top view of one hexamer with each CA monomer colored differently.

(E and F) Mapping of alanine scanning mutagenesis data of HIV-1 CA onto the pseudoatomic model. Shown are a top view in (E) and slabbed side view in (F) of one hexamer, plus CTDs from adjacent hexamers. Mutations that did not affect assembly in vitro are shown in green or blue, matching their locations in the NTD or CTD, respectively. Mutations that diminished or abolished assembly are colored based on their positions relative to the three protein-protein interfaces: NTD-NTD (orange), CTD-CTD (yellow), and NTD-CTD (red). Arg18, which was mutated to leucine for this study, is colored magenta.
showed excellent correspondence with the experimental density. We did not observe density corresponding to the last 12 residues of CA, which are typically disordered in X-ray structures of the CTD. The cyclophilin A-binding loop, between helices 4 and 5 in the NTD, is also not resolved in our map, consistent with the observation that this loop is not required for CA assembly in vitro (Ganser-Pornillos et al., 2004).

A top view of the pseudoatomic model corresponds to a view of the outer surface of the capsid (Figure 2A) and shows that six CA molecules associate laterally to form the hexameric unit, with the NTDs (colored green) close to the surface. The CTDs (colored blue) are located adjacent to and beneath the NTDs (Figure 2F), and form homodimers that connect neighboring hexamers. Although the NTDs and CTDs were fitted separately as individual domains, the pseudoatomic model clearly defines the full-length CA monomer because the carboxy terminus of each NTD is located immediately adjacent to the amino terminus of a neighboring CTD (indicated by asterisks in Figures 2B–2D). Interestingly, the intramolecular contacts between the two domains of one CA monomer appear to be fairly minimal. Instead, the NTD and CTD participate in extensive intermolecular contacts, which undoubtedly stabilize the hexameric lattice (Figure 2D).

### The NTD-NTD Hexamerization Interface

Consistent with previous EM studies at 20–30 Å resolution (Li et al., 2000; Mayo et al., 2002), the NTD of HIV-1 CA forms the inner ring of the hexamers (Figures 2A–2C). NTD-NTD interactions are mediated through helices 1, 2, and 3, which associate as an 18-helix bundle in the center of the hexamer (Figure 3B). This arrangement provides a molecular understanding of mutagenesis and biochemical experiments showing that alanine substitutions of surface-exposed residues in helices 1 and 2 disrupt HIV-1 CA assembly in vitro and in vivo (orange spheres in Figures 2E and 2F) (Ganser-Pornillos et al., 2004; von Schwedler et al., 2003) and that amide protons in helices 1, 2, and 3 are protected from deuterium exchange upon assembly in vitro (Lanman et al., 2003). Interestingly, the ribbon model of the NTD structure deviates somewhat from the boundary of the experimental density, particularly for helices 4, 5, and 7 (Figure 3A). These deviations may indicate the occurrence of assembly-induced conformational changes, although they may equally be artifacts due to the limited $z$-resolution of the cryoEM map.

The amino terminal $β$-hairpin, which forms in the context of the mature CA protein, is ordered in our structure...
Biochemical and genetic experiments of this region of CA have been puzzling: although the $\beta$-hairpin forms after Gag proteolysis and is essential for infectivity, CA proteins missing the entire hairpin can still assemble into cylinders in vitro (Tang et al., 2002; Gross et al., 1998, 2000; von Schwedler et al., 1998). It has been suggested that the six $\beta$-hairpins might form extended interactions within the mature CA hexamer (Mortuza et al., 2004), but the map shows that the individual strands are too far separated to form a $\beta$-barrel type structure (Figures 2C and 2D). It is therefore more likely that the $\beta$-hairpin stabilizes assembly-competent conformations of important residues in helix 1 and 2, and/or that $\beta$-hairpin formation disrupts interactions that stabilize the immature lattice, as suggested previously (Kelly et al., 2006; Mortuza et al., 2004; Tang et al., 2002).

The limited resolution of our map precludes a precise description of side chain conformations. Nevertheless, the pseudomolecular model defines a local probability sphere wherein each individual residue is most likely located. Thus, we can make some informed guesses as to the positions of these side chains and their possible interactions. For example, Arg18, which was mutated to leucine for this study, is located at an annulus, where helices 1 from the six NTDs are at their closest approach. It appears that the guanidinium sidechain of Arg18 is well placed to form annular interactions (colored magenta in Figures 2E and 2F), which may explain the dramatic effects on the morphology of in vitro assembled particles when this residue is mutated (Ganser-Pornillos et al., 2004).

The HIV-1 CA NTD hexamer is essentially identical to the hexamer formed by the NTD of N-tropic murine leukemia virus CA that was crystallized without its CTD (Figure 3B) (Mortuza et al., 2004). Note that there is no detectable primary sequence homology between these two retroviral CA proteins, which also form capsids of different shapes (HIV-1 is conical and N-MLV is spherical). This result establishes that the CA hexamer is conserved across at least two widely divergent, and likely all, retroviruses. Importantly, the unit cell spacing of both the HIV-1 and N-MLV NTD hexamers (92.7 Å) corresponds to the CA spacings measured in EM studies of mature virions (90–110 Å) (Briggs et al., 2003; Yeager et al., 1998), supporting the idea that these crystals represent the hexameric lattice in mature capsids. Interestingly, low-resolution EM studies further indicate that CA also forms hexamers in the immature retroviral capsid, i.e., as a domain of the viral Gag polyprotein. These hexamers appear smaller, with a spacing of 80 Å, and may therefore have a different packing arrangement (Briggs et al., 2004, 2006; Yeager et al., 1998).

**The CTD-CTD Dimerization Interface**

HIV-1 CA forms soluble dimers (Gamble et al., 1997; Rosé et al., 1992; von Schwedler et al., 1998). Dimer formation in solution is mediated solely through the CTD, and biochemical and structural studies have mapped the dimerization interface primarily to helix 9 (Gamble et al., 1997). Notably, alanine substitutions of W184 and M185, located in the middle of helix 9, diminish dimerization in vitro and infectivity in vivo (shown as yellow spheres in Figures 2E, 2F, 4A, and 4B) (Gamble et al., 1997; von Schwedler et al., 1998, 2003). This interface is therefore required for efficient assembly of both mature and immature capsids. Note that HIV-1 CA is the only retroviral CA protein to form...
dimers in solution, among the several that have been characterized in detail. Therefore, a measurable dimerization affinity in solution is not critical for assembly; nevertheless, we expect that all retroviral CA CTDs use the same helix 9 interface for capsid lattice assembly.

To date, five X-ray crystal structures of wild-type HIV-1 CA CTD have been reported, which in the aggregate suggest four possible configurations of the CTD dimer (Berthet-Colominas et al., 1999; Gamble et al., 1997; Ivanov et al., 2007; Ternois et al., 2005; Worthylake et al., 1999). Three of these dimers form via parallel pair-wise packing of helix 9 from each monomer, but the crossing angles of helix 9 are different in each structure (Figures 4B–4D, left panels). Each two-fold axis in the cryoEM map is surrounded by symmetry-related helical densities that are almost parallel. This most closely matches the helix 9 packing geometry in 1a43, the X-ray structure of the apo CA148-231 dimer (Gamble et al., 1997; Worthylake et al., 1999), and indeed, 1a43 is the only structure that can be fitted reasonably as a dimeric unit into our map (Figure 4B). Automated rigid-body fitting of the two other CTD dimers (1a8o and 2buo) failed to provide meaningful solutions (not shown). Figures 4C and 4D illustrate the results of manual fitting of the 1a8o and 2buo dimers, respectively, wherein one monomer was matched to the cryoEM density: there is a clear mismatch with the second monomer. The density match for the fitted 1a43 dimer was still not optimal, however (Figure 4B), and the best fit was obtained by fitting the CTD as a monomeric unit (Figure 4A). These observations indicate that there may be conformational changes in the CTD dimer during assembly. For example, a minor adjustment in the dimer packing geometry appears to occur, in comparing the 2D lattice (Figure 4A) and 3D crystal (Figure 4B). We speculate that this may involve sliding or twisting motions of the helix 9 pairs across the dyad, although the atomic details of such small movements cannot be visualized at the current resolution. Additionally, changes in the tertiary structure of the CTD itself appear required to accommodate binding of the NTD, through an intermolecular NTD-CTD interface that was not resolved in previous low-resolution EM structures of full-length CA hexamers (see below).

Figure 5. The NTD-CTD Intermolecular Interface Is Formed by Insertion of the NTD Helix 4 into a Groove in the CTD

(A) Stereo view of the NTD-CTD interface. The experimental density is contoured at 1.8 σ. Note that the NTD-CTD interface is physically linked to the CTD-CTD dimerization interface by helix 9, which may provide a structural basis for cooperativity.

(B) Identical stereo view showing the positions of K70 and K182 (red spheres), which can be crosslinked in vitro in assembled tubes of wild-type HIV-1 CA. The cyan helix represents the position occupied by the peptide CA-I in 2buo, the cocrystal structure with the isolated CTD.

(C) Identical stereo view in surface representation. The pseudoatomic model shows apparent steric clashes between the NTD and CTD, indicated by red asterisks. We presume that conformational adjustments must occur in these regions to optimize residue packing.
As described in a recent report, an HIV-1 CA CTD construct containing a single amino-acid deletion forms a “domain-swapped” dimer (Ivanov et al., 2007). This dimer was proposed to be an assembly intermediate in the formation of the immature Gag lattice, but a role in mature CA capsid assembly could not be ruled out. The domain-swapped CTD structure (pdb code: 2ont) has a distinct packing geometry and could not be fitted as a dimeric unit into our cryoEM density map (not shown). Additionally, a second distinguishing feature of the domain-swapped dimer is the absence of a kink in helix 9 (Ivanov et al., 2007). This kink is clearly present in our cryoEM map and in all previous X-ray structures of the CTD (Figure S2). These observations suggest that a domain-swapped CTD dimer may not be present in the assembled hexameric lattice of mature HIV-1 CA.

The NTD-CTD Interface
The existence of an NTD-CTD interaction was originally inferred from studies of Rous sarcoma virus CA, which showed that an assembly incompetent CTD mutant (R170Q, helix 9) could be partially rescued by an additional mutation in the NTD (P65Q, helix 4) (Bowzard et al., 2001). Chemical crosslinking experiments in HIV-1 CA cylinders also revealed that K182 (helix 9, CTD) is in close proximity to K70 (helix 4, NTD), further supporting this idea and suggesting that the interface is conserved (see Figure 5B) (Lanman et al., 2003). Importantly, the crosslinking experiments suggested that the interaction is intermolecular and not between the NTD and CTD of the same CA molecule. In an elegant series of experiments, Lanman, et al. (2003, 2004) used a deuterium exchange protection assay to establish that the NTD-CTD interaction occurs both in vitro and in vivo. The interaction surfaces were mapped to helix 4 in the NTD and helices 8 and 9 in the CTD.

Our structure now reveals that in forming the mature lattice, the amino-terminal end of helix 4 of the NTD inserts into a groove in the CTD (Figure 5). The pseudoatomic model identifies potential contacts between helix 4 and helices 8 and 9 on one side of the groove, and the loop connecting helices 10 and 11 on the other (Figures 5B and 5C). To accommodate helix 4, some rearrangements in the above regions must occur, presumably by an induced fit mechanism. This is evident from apparent steric clashes between the NTD and CTD (Figure 5C), which were not adjusted during rigid-body fitting of the two domains. The pseudoatomic model also places portions of helices 8 and 11 in close proximity to helix 7 and the loop between helices 3 and 4 at the “bottom” of the NTD, suggesting that the NTD-CTD interface may extend to this region. The interactions between neighboring CA molecules in the hexameric lattice are therefore more extensive than originally inferred from low-resolution models.

Therapeutic Implications
The discovery of two inhibitors that appeared to target the NTD-CTD interface underscores its importance in the formation of the mature CA lattice. One inhibitor, called CA-I, is a 12-residue peptide that was discovered by Kraüsslich and colleagues through phage display methods. CA-I abrogates both mature and immature capsid assembly in vitro (Sticht et al., 2005). An X-ray crystal structure of CA-I in complex with the isolated HIV-1 CA CTD (2buo) revealed that the peptide binds between helices 8, 9, and 11, and the authors proposed a competitive mode of action in mature capsid assembly (Ternois et al., 2005). Indeed, superposition of 2buo with our pseudoatomic model shows that CA-I binds in the same groove occupied by helix 4 (cyan helix in Figure 5B). A second inhibitor, called CAP-1, is a small molecule that has been shown to inhibit CA assembly in vitro and HIV-1 replication in tissue culture (Tang et al., 2003). Although CAP-1 was originally identified as a possible inhibitor of β-hairpin formation at the “top” of the NTD, subsequent NMR chemical shift footprinting studies suggested that instead, it binds at the “bottom” of the NTD. Residues that appear to form part of the NTD-CTD interface are among those that exhibited the greatest chemical shift changes (Tang et al., 2003). Taken together, these studies indicate that both CA-I and CAP-1 inhibit mature CA assembly by disrupting the complementary surfaces that form the NTD-CTD interface, and importantly, show that the HIV-1 CA lattice is a legitimate therapeutic target.

Implications for Capsid Formation
As noted above, the lattice spacing of the CA hexamers in the 2D crystals match the paracrystalline lattice in mature HIV-1 virions (Briggs et al., 2003; Yeager et al., 1998). Because our structure was solved from large, flattened spheres of HIV-1 CA, the 2D lattice would recapitulate the mature capsid lattice at its planar limit. Indeed, flat faces are apparent in EM images of natural HIV-1 capsids, particularly at the broad end of the cone (Accola et al., 2000; Kotov et al., 1999; Welker et al., 2000). However, it is clear that a fullerenic cone capped on both ends exhibits a predominantly curved surface, and that therefore, the hexameric lattice must be inherently flexible. Cone models indicate that there are two distinct types of curvature within the conical HIV-1 capsid. One type exists around the 12 pentameric declinations required to close the cone. In these regions, there must be local five-fold symmetry, which the current structure does not readily explain. The second type of curvature is found in the conical body of the capsid. In these regions, the direction and degree of curvature changes continuously along the hexameric lattice. Low-resolution studies of CA tubes of different diameters and pitch indicate that the hexamers are preserved across these distinct tubes (Li et al., 2000), suggesting that their differing curvatures are accommodated, not by large global changes, but by more subtle local changes in protein tertiary and/or quaternary structure.

On the basis of the packing interactions in our pseudoatomic model of full-length HIV-1 CA and X-ray crystal
structures of the individual domains, we envision two possible mechanisms for inducing curvature within the hexameric lattice. In the first mechanism, each CA hexamer could act as a rigid structural unit, and lattice flexibility would be achieved primarily through alternative packing of the CTD-CTD interaction surfaces. This suggestion arises from the crystallographic observation that CTD dimers, although formed from the same interface, can have different packing geometries (Figure 1B). For example, it is possible that graduated rotations about this interface could produce the systematic, subtle changes in curvature required to produce the conical body of the capsid. Precedence for this idea comes from recent structural studies of the capsid protein of cowpea chlorotic mottle virus, which demonstrate that small changes in the packing angle of two monomers about a dimer interface (range of rotation from 38° to 45°) allowed the same protein to form two different types of closed capsids (T = 1 and T = 3 particles), as well as spirals (Tang et al., 2006). In the second mechanism, lattice flexibility could be achieved by changes in the CA tertiary structure. This would have the effect of “warping” the hexamer, allowing formation of a curved lattice while keeping the interaction interfaces fixed. Finally, we note that the two proposed mechanisms are not mutually exclusive, and one can invoke a scenario wherein a combination of both hexamer warping and dimer interface rotations may produce different degrees of lattice curvature.

The large range of curvatures available to the assembling CA lattice should, in theory, lead to high error rates and one would expect large percentages of aborted or spiralled structures that fail to close. Nevertheless, HIV-1 CA does form what appear to be closed (or very nearly closed) structures in vitro and in vivo with high frequency. For example, we found that almost all the small spheres formed by the R18L mutant appeared closed, despite ranging in size from 20–100 nm in diameter (i.e., each formed by the R18L mutant appeared closed, despite

Electron Cryomicroscopy and Image Analysis

For EM analyses, a freshly prepared carbon-coated molybdenum grid was placed on 8 μl of the crystallization solution for 90 s. The grid was then washed with several drops of 0.1 M KCl, blotted to near dryness, and plunged into a slurry of liquid ethane. The grids were transferred under liquid nitrogen to a Tecnai F20 electron cryomicroscope (Philips/FEI) operating at 120 kV. Images of crystals were recorded on Kodak SO-163 film using either the standard low-dose package or the automated Legoìnion software package (Suloway et al., 2005). Optical defraction was used to select the best images of crystals with minimal drift and astigmatism. The negatives were digitized using a Zeiss SIGMA densitometer at a step size of 0.05 μm and cropped to a size of 4000 × 4000 or 6000 × 6000 pixels with Adobe Photoshop. Image files were then converted to the MRC format, and Fourier transforms were calculated. The crystals were indexed manually using XIMDIP, and corrections for lattice distortions and effects due to the contrast transfer function were performed using the MRC program suite (Amos et al., 1982; Crowther et al., 1996; Henderson et al., 1986, 1990). Image data were merged as follows: First, the program ALLSPACE was used to determine the 2D plane group symmetry (Table S1) and calculate the appropriate phase origins used for image alignment (Valpuesta et al., 1994). The best (i.e., highest resolution) image recorded from an untilted sample was used as a reference to determine the phase origin of the remaining images. Special care was taken to determine the correct tilt geometry for images of tilted crystals. Lattice lines were then fitted to the 3D data, and an initial map was created using the CCP4 suite of programs. The phase origins and tilt geometries of each crystal were then refined against this initial model. Iterative refinements against subsequent models were continued until no further improvement in the overall phase residuals were observed. After obtaining the most self-consistent set of image parameters, the data set was again remerged and averaged, lattice lines were fitted and a final model was created. Data statistics are presented in Table 1, and a graphical representation of data quality is shown in Figure S1 (Cheng and Yeager, 2004).

Rigid Body Fitting

The NMR structure of the NTD (pdb code 1gwp; http://www.rcsb.org) was initially fitted into a map covering 4 unit cells (equivalent to 4 hexamers or 24 monomers) using the program COLORES (standard parameters) with the target resolution set to 9.0 Å (Chacón and Wriggers, 2002; Wriggers et al., 1999) (http://situs.biomachina.org). The program readily identified 24 unique, non-overlapping positions for the NTD. X-ray crystal structures of the CTD (1a43, 2buo, and 1a8o) were next fitted into the map as dimers. Only 1a43 yielded meaningful solutions. To fit the CTD as a monomer, a map was first calculated by subtracting out the NTD densities. Using this map, all three search models yielded the same 24 unique, non-overlapping solutions. The fits were then optimized to the full map, using the program COLACOR (Chacón and Wriggers, 2002; Wriggers et al., 1999). Note that our initial fitting was performed on a map with an applied B-factor of ~500. For clarity, the final figures were made from a map with an applied B-factor of ~1000. The automated fitting of both maps was identical (data not shown). The final pseudoatomic model for the HIV-1 CA monomer and hexamer was generated from the monomer fits of 1gwp (residues 1–148) and 1a43 (residues 149–219).

Figures

Structural representations were created using PymOL (Delano Scientific, Inc.).
**Supplemental Data**

Supplemental Data include two figures and one table and can be found with this article online at [http://www.cell.com/cgi/content/full/131/1/70/DC1/](http://www.cell.com/cgi/content/full/131/1/70/DC1/).

**ACKNOWLEDGMENTS**

We thank W.I. Sundquist, K. Baker, and K. Dryden for discussions and critical reading of the manuscript; K. Altieri, F. Dawood, and Y. Hua for excellent technical assistance; and J. Quispe for help with the Lego

...n software package. Some of the electron microscopy experiments were conducted at the National Resource for Automated Molecular Microscopy (NRAMM), which is supported by the NIH through the National Center for Research Resources' P41 program (RR17573). This work was supported by the NIH through Immunology Training Grant T32-AR07244 (B.G.-P.) and R01 Grant GM66087 (M.Y.). B.G.-P. is a postdoctoral fellow of the George W. Hewitt Foundation for Medical Research.

Received: May 15, 2007

Revised: July 19, 2007

Accepted: August 15, 2007

Published: October 4, 2007

**REFERENCES**


