Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists

Beili Wu,[1] Ellen Y. T. Chien,[1] Clifford D. Mol,[1] Gustavo Fenalti,[1] Wei Liu,[1] Vsevolod Katritch,[2] Ruben Abagyan,[2] Alexei Brooun,[3] Peter Wells,[3] F. Christopher Bi,[3] Damon J. Hamel,[2] Peter Kuhn,[1] Tracy M. Handel,[2] Vadim Cherezov,[1] Raymond C. Stevens[1]*

Chemokine receptors are critical regulators of cell migration in the context of immune surveillance, inflammation, and development. The G protein–coupled chemokine receptor CXCR4 is specifically implicated in cancer metastasis and HIV-1 infection. Here we report five independent crystal structures of CXCR4 bound to an antagonist small molecule IT1t and a cyclic peptide CVX15 at 2.5 to 3.2 angstrom resolution. All structures reveal a consistent homodimer with an interface including helices V and VI that may be involved in regulating signaling. The location and shape of the ligand-binding sites differ from other G protein–coupled receptors and are closer to the extracellular surface. These structures provide new clues about the interactions between CXCR4 and its natural ligand CXCL12, and with the HIV-1 glycoprotein gp120.

Chemokine receptors are G protein–coupled receptors (GPCRs) that, together with their small protein ligands, regulate the migration of many different cell types, most notably leukocytes (*1*–*3*). CXCR4, one of 19 known human chemokine receptors, is activated exclusively by the chemokine CXCL12 (also known as stromal cell–derived factor–1, SDF-1) and couples primarily through G_i proteins. Targeted deletion of CXCR4 or CXCL12 in mice confers embryonic lethality and leads to defects in vascular and CNS development, hematopoiesis, and cardiogenesis (*4*, *5*). CXCR4 has been associated with more than 23 types of cancers, where it promotes metastasis, angiogenesis, and tumor growth or survival (*6*–*10*). Furthermore, T-tropic HIV-1 uses CXCR4 as a co-receptor for viral entry into host cells (*11*). Thus, the discovery that endogenous CXCL12 inhibits HIV-1 entry suggested the therapeutic potential of targeting CXCR4 to block viral infection (*12*, *13*). Despite a wealth of data related to CXCR4 and GPCRs in general, many aspects of ligand binding and signaling are poorly understood at the molecular level. For instance, CXCR4 has a propensity to form hetero- and homooligomers (*14*, *15*), and such oligomerization could play a role in the allosteric regulation of CXCR4 signaling (*16*). Although structural understanding of GPCRs has benefited from a number of recent breakthroughs (*17*–*20*), coverage of the superfamily's phylogenetic tree is incomplete, and a structure of a GPCR that is activated by a protein ligand has not been reported.

Protein engineering, ligand selection, and structure determination. Here we report the crystal structures of human CXCR4 in complex with a small-molecule antagonist at 2.5 Å resolution and with a cyclic peptide inhibitor at 2.9 Å resolution. Three stabilized constructs (CXCR4-1, CXCR4-2, and CXCR4-3) (table S1) expressed in baculovirus-infected *Spodoptera frugiperda* (Sf9) insect cells were selected for structural studies on the basis of thermal stability, monodispersity, and lipid matrix diffusion. Similar to the previously determined high-resolution structures of the β_2-adrenergic receptor (β_2AR) (*17*, *21*) and A_{2A} adenosine receptor (A_{2A}AR) (*18*), the CXCR4 constructs contain a T4 lysozyme (T4L) fusion inserted between transmembrane (TM) helices V and VI at the cytoplasmic side of the receptor. In addition, all three constructs contain a thermostabilizing L125$^{3.41}$W mutation (*22*–*24*). The constructs differ in the precise T4L junction site, the position of the C-terminal truncation, as well as a T240$^{6.36}$P mutation in CXCR4-3, and required further stabilization with ligands to facilitate purification and crystallization. Two antagonists were selected for crystallization trials on the basis of ligand solubility, binding affinity, and induced protein thermostability (tables S2 and S3) and using a small, druglike, isothiourea derivative (IT1t) (*25*) and CVX15, a 16-residue cyclic peptide analog of the horseshoe crab peptide polyphemusin, which was previously characterized as an HIV-inhibiting and antimetastatic agent (*26*–*28*).

Before crystallization trials, the effects of the protein engineering on CXCR4 function were evaluated using radioligand-binding and calcium flux assays. CXCR4-WT (wild type) expressed in Sf9 cells binds a [^{3}H]bis(imidazolylmethyl) amine analog (BIMA) with an affinity similar to that of the same construct expressed in human embryonic kidney (HEK) 293 cells (dissociation constant, K_d = 3.5 ± 1.5 and 3.7 ± 1.4 nM, respectively). All other constructs expressed in Sf9 cells also show similar binding affinity to BIMA and IT1t (table S3). However, CXCR4-1 and CXCR4-2 display lower binding affinities for the CVX15 peptide compared with CXCR4-WT and CXCR4-3. Calcium flux assays demonstrated the expected result that these constructs do not activate G proteins (fig. S1), because of the T4L insertion in the third intracellular loop, which is critical for G protein interactions. Assays with the same constructs lacking T4L confirmed that the stabilizing L125$^{3.41}$W mutation, as well as the various C-terminal truncations, did not adversely affect calcium release, whereas the T240$^{6.36}$P mutation, which is present only in the CXCR4-3 construct, abolished signaling.

After extensive optimization of crystallization conditions in lipidic mesophase, five distinct crystal structures were obtained (table S4). CXCR4-1, CXCR4-2, and CXCR4-3 were cocrystallized with IT1t (two crystal structures for CXCR4-2); crystals

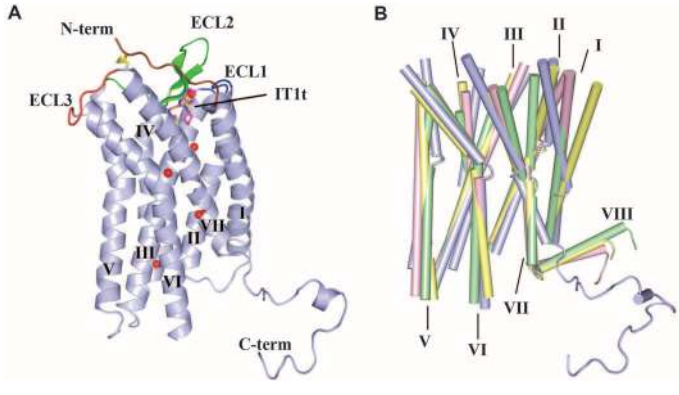

Fig. 1. Overall fold of the CXCR4-IT1t complex and comparison with other GPCR structures. (**A**) Overall fold of the CXCR4-2–IT1t. The receptor is colored blue. The N terminus, ECL1, ECL2, and ECL3 are highlighted in brown, blue, green, and red, respectively. The compound IT1t is shown in a magenta stick representation. The disulfide bonds are yellow. Conserved water molecules (*68*) are shown as red spheres. (**B**) Comparison of TM helices for CXCR4 (blue); β_2AR (PDB ID: 2RH1; yellow); A_{2A}AR (PDB ID: 3EML; green); and rhodopsin (PDB ID: 1U19; pink).

[1]Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. [2]Skaggs School of Pharmacy and Pharmaceutical Sciences, and San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093, USA. [3]Pfizer Worldwide Research and Development, 10770 Science Center Drive, San Diego, CA 92121, USA.

*To whom correspondence should be addressed. E-mail: stevens@scripps.edu

of CXCR4-3 were also obtained with CVX15. Data collection and refinement statistics for all five crystal structures are shown in table S1 (*29*).

Overall architecture of CXCR4. The overall structure of CXCR4 bound to the small-molecule antagonist IT1t is conserved in all crystal structures with a Cα root mean square deviation (RMSD) in a TM bundle of less than 0.6 Å. Binding of the CVX15 cyclic antagonist peptide induced conformational differences relative to IT1t in the CXCR4-3/CVX15 structure (TM Cα RMSD of 0.9 Å). For clarity, we focus on the highest-resolution crystal structure of CXCR4-2/IT1t (2.5 Å, monomer A) for discussion of the CXCR4 structural features and comparison with other GPCR structures. The final model includes 293 residues (27 to 319) of the 352 residues of CXCR4 and residues 2 to 161 of T4L. The 26 N-terminal residues of CXCR4 did not have interpretable density and are presumed to be disordered. The main fold of CXCR4 consists of the canonical bundle of seven-TM α helices (Fig. 1A), which shows about the same level of structural divergence from seven-TM helical bundles of previously solved GPCR structures (Cα RMSDs ≈ 2.0 to 2.2 Å) (Fig. 1B). The most striking differences in the disposition of the TM helices of CXCR4 are the following: (i) The extracellular end of helix I is shifted toward the central axis of the receptor by 9 Å compared with β_2AR and by more than 3 Å compared with A_{2A}AR. (ii) Helix II makes a tighter helical turn at Pro92$^{2.58}$, resulting in ~120° rotation of its extracellular end compared with other GPCR structures (this rotation essentially introduces a one-residue gap in the sequence alignment that would result in wrong residues facing the ligand-binding pocket in a homology model that did not account for the rotation). (iii) Both intracellular and extracellular tips of helix IV in CXCR4 substantially deviate (~5 and ~3 Å, respectively) from their consensus positions in other GPCRs. (iv) The extracellular end of helix V in CXCR4 is about one turn longer. (v) Helix VI has a similar shape in all structures and is characterized by a sharp kink at the highly conserved residue, Pro254$^{6.50}$; however, its extracellular end is shifted by ~3 Å in CXCR4 relative to β_2AR and A_{2A}AR. Finally (vi), the extracellular end of helix VII in CXCR4 is two helical turns longer than in other GPCR structures. These two extra turns place Cys274$^{7.25}$ at the tip of helix VII in a strategic position to form a disulfide bond with Cys28 in the N-terminal region. Taken together, these multiple differences suggest that accurate homology modeling of even the CXCR4 TM bundle, let alone the entire structure, would be challenging.

The extracellular interface of CXCR4 consists of 34 N-terminal residues; extracellular loop 1 (ECL1, residues 100 to 104) linking helices II and III; ECL2 (residues 174 to 192) linking helices IV and V; and ECL3 (residues 267 to 273) linking helices VI and VII (Fig. 1A). Clear density starts at Pro27, adjacent to Cys28, which pins the base of the N-terminal segment to Cys274$^{7.25}$ at the tip of helix VII via a disulfide bond; these two cysteines are conserved in all chemokine receptors except CXCR5 and CXCR6 (fig. S2). Another disulfide links Cys109$^{3.25}$ with Cys186 of ECL2, which is the largest extracellular loop in CXCR4. Although ECL2 length, sequence, and secondary structure vary dramatically in GPCRs, the disulfide connecting ECL2 with the extracellular end of helix III is highly conserved in chemokine receptors and most other class A GPCRs. Both disulfide bonds at the extracellular side of CXCR4 are critical for ligand binding (*30*), and the crystal structure shows that they function by constraining ECL2 and the N-terminal segment (residues 27 to 34), which shapes the entrance to the ligand-binding pocket.

The intracellular side of CXCR4 contains intracellular loop 1 (ICL1, residues 65 to 71) linking

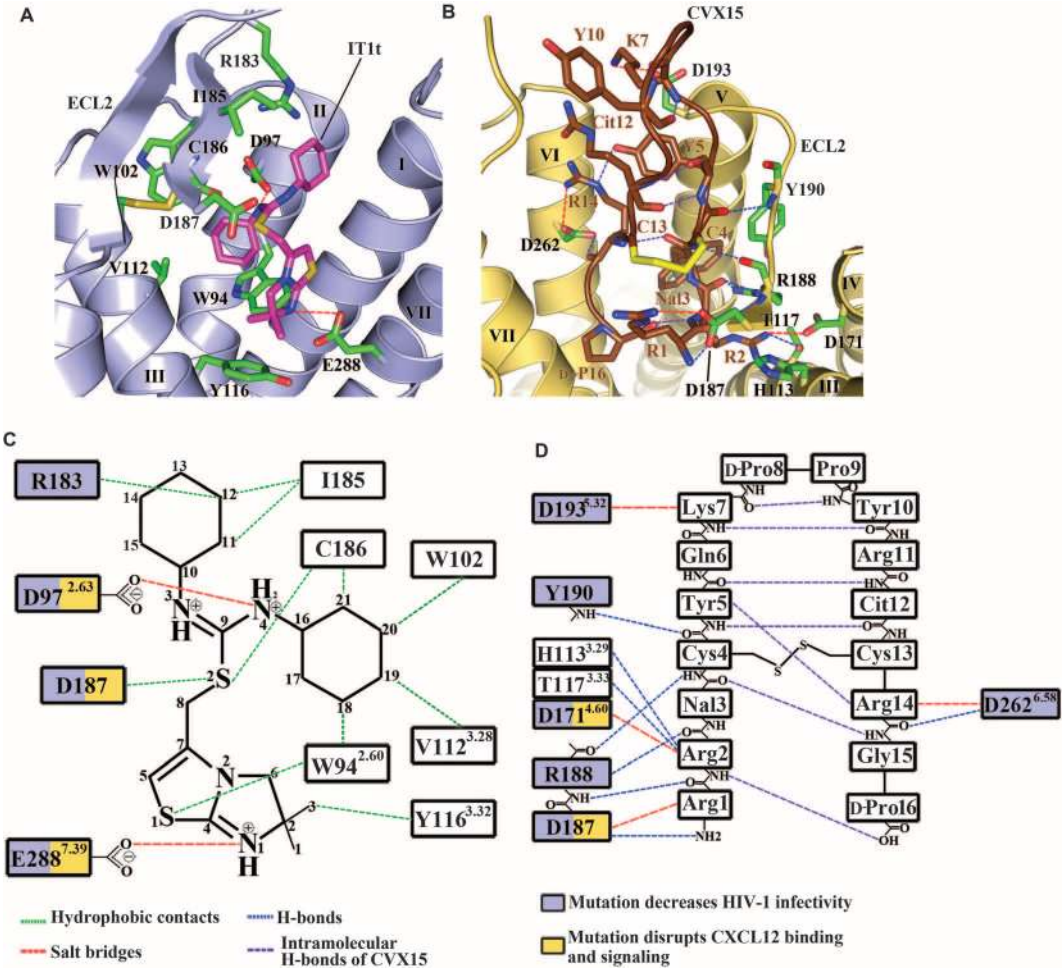

Fig. 2. CXCR4 ligand-binding cavities for the small molecule IT1t and the cyclic peptide CVX15. (**A**) CXCR4 ligand-binding cavity for the small molecule IT1t. IT1t (magenta) and the residues of the receptor (green) involved in the ligand interactions are shown in stick representation. Nitrogen atoms are blue and sulfur atoms are yellow. Key for dashed lines is shown below. Only the helices involved in the receptor-ligand interaction and part of ECL2 are shown. (**B**) CXCR4 ligand-binding cavity for the peptide CVX15. The residues of CVX15 (brown) and the residues of the receptor (green) involved in receptor-ligand polar interactions are shown in stick representation. The Cys4-Cys13 disulfide bridge in CVX15 is shown as a yellow stick. (**C**) Schematic representation of selected interactions between CXCR4 and IT1t in the ligand-binding pocket. Mutations reported to decrease HIV-1 infectivity and to disrupt CXCL12 binding and signaling are indicated with blue and yellow squares, respectively (*57*, *69*). (**D**) Schematic representation of selected interactions between CXCR4 and CVX15 in the ligand-binding pocket.

helices I and II; ICL2 (residues 140 to 144) linking helices III and IV; ICL3 (residues 225 to 230) linking helices V and VI; and the C terminus. ICL3 also contains T4L inserted between Ser229 and Lys230 and flanked by short linkers (GS-T4L-GS). Structural alignment of CXCR4 with high-resolution GPCR structures indicates that the intracellular half of the TM bundle is structurally more conserved (Cα RMSDs with β_2AR, A_{2A}AR, and rhodopsin are 1.8, 1.9, and 1.4 Å, respectively) than the extracellular half (2.6, 2.2, and 2.2 Å, respectively). Therefore, it comes as a surprise that in all five CXCR4 structures, helix VII is about one turn shorter at the intracellular side, ending just after the GPCR-conserved NPxxY motif, and that all structures lack the short α helix VIII (Fig. 1B). The C-terminal part of CXCR4 beyond Ala303$^{7.54}$ adopts an extended conformation and participates in a number of crystal contacts with the extracellular side of a symmetry-related molecule in the highest-resolution crystal structure, CXCR4-2–IT1t (fig. S4A). Because of its structural persistence and common α-helical sequence motif [F(RK)xx(FL)xxx(LF)], helix VIII was thought to be a regular structural element of all class A GPCRs. However, CXCR4 contains only a par-

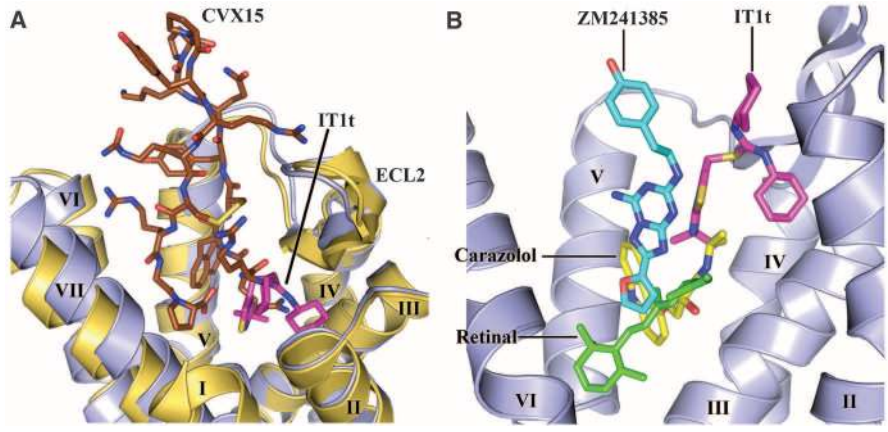

Fig. 3. CXCR4 ligand-binding modes and comparison with other GPCR structures. (**A**) Comparison of the ligand-binding modes for IT1t and CVX15. CXCR4 molecules in the CXCR4-2–IT1t and CXCR4-3–CVX15 complexes are colored blue and yellow, respectively. IT1t (magenta) and CVX15 (brown) are shown as sticks. (**B**) Comparison of the small-molecule ligand-binding modes for CXCR4, β_2AR (PDB ID: 2RH1), A_{2A}AR (PDB ID: 3EML), and rhodopsin (PDB ID: 1U19). Only CXCR4 helices are shown (blue). The ligands IT1t (for CXCR4, magenta), carazolol (for β_2AR, yellow), ZM241385 (for A_{2A}AR, cyan), and retinal (for rhodopsin, green) are shown in stick representation.

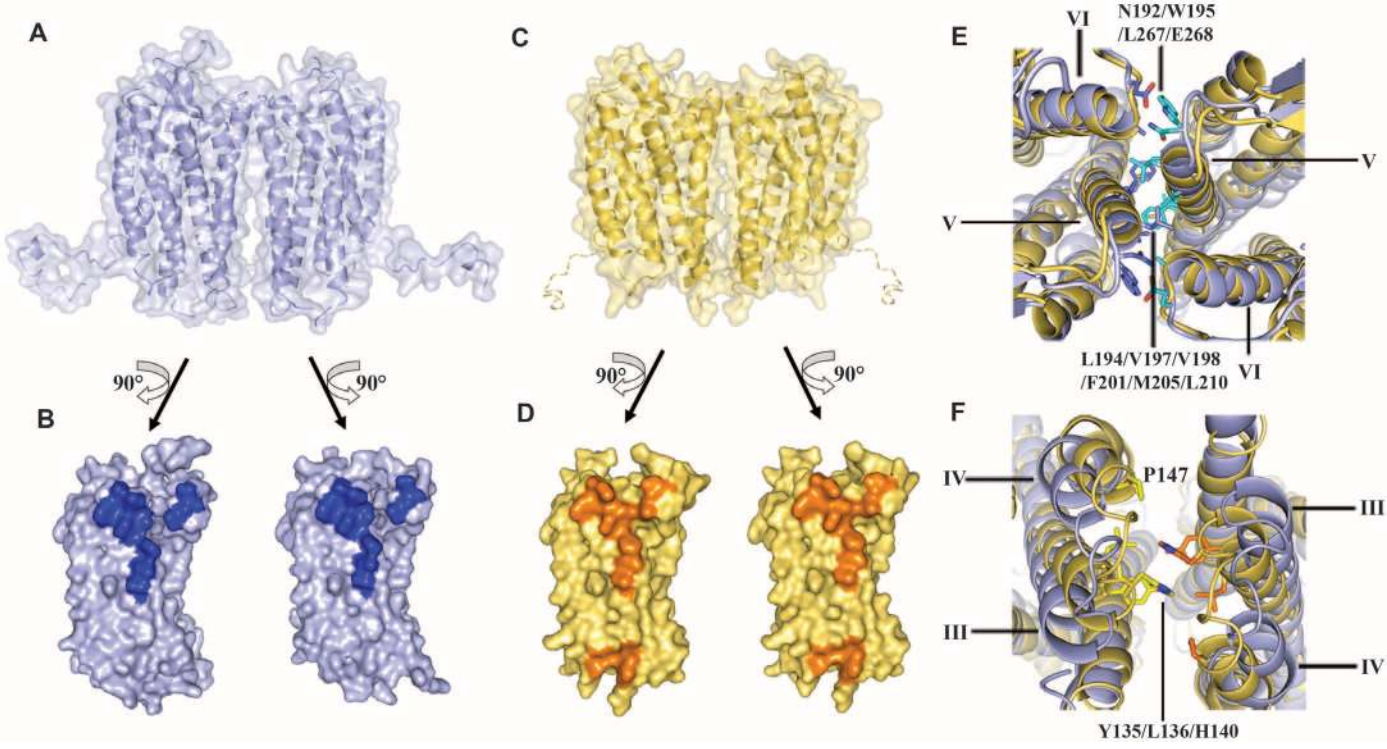

Fig. 4. Dimer interactions in CXCR4-2–IT1t and CXCR4-3–CVX15. (**A**) Molecular surface representation of the CXCR4 dimer in CXCR4-2–IT1t (blue). (**B**) Dimer interface in CXCR4-2–IT1t. The surface involved in dimerization is highlighted in dark blue. (**C**) Molecular surface representation of the CXCR4 dimer in CXCR4-3–CVX15 (yellow). A hypothetical path of the C terminus, which is not observed in the CXCR4-3–CVX15 structure, is shown as a dashed curve. (**D**) Dimer interface in CXCR4-3–CVX15. The surface involved in dimer interaction is highlighted in orange. (**E**) Top view of the extracellular side of the dimers. Two structures show similar interactions via helices V and VI. Residues of CXCR4-2–IT1t involved in the dimer interaction are shown in stick representation and are colored blue in molecule A, cyan in molecule B. (**F**) Bottom view of the intracellular side of the dimers. Contacts can only be observed at the intracellular tips of helices III and IV, and ICL2 in CXCR4-3–CVX15. The residues of CXCR4-3–CVX15 involved in the dimer interaction are shown in stick representation and are colored yellow and orange. These interactions are not present in the CXCR4–IT1t complex.

tially conserved motif FKxxAxxxL, and although it may be capable of forming an α helix under certain conditions, this helix would be less stable because of the replacement of Phe or Leu with Ala. In addition, CXCR4 lacks a putative palmitoylation site at the end of helix VIII, which anchors to the lipid membrane in many GPCRs.

Construct CXCR4-3 contains a T240$^{6.36}$P mutation near the intracellular side of helix VI, which results in retention of ligand-binding affinity but abolishes signaling (table S3 and fig. S1). Comparison of the CXCR4-3 structure with CXCR4-1 and CXCR4-2 reveals that the only effect of the T240$^{6.36}$P mutation is the disruption of a short section of helix VI between Gly231$^{6.27}$ and Pro240$^{6.36}$. Because helix VI is thought to be one of the major players in the signaling mechanism (*31*, *32*), disruption of its structure would likely affect G protein binding and activation. Thus, T240$^{6.36}$P represents a novel structure-based uncoupling mutation.

Molecular recognition of the small molecule IT1t and the cyclic CVX15 peptide by CXCR4. Strong electron density was observed for IT1t in the binding cavity of both subunits of the CXCR4 homodimer (fig. S3A). Compared with previous GPCR structures, the cavity is larger, more open, and located closer to the extracellular surface (Fig. 2, A and C; Fig. 3B; and table S5). The IT1t ligand occupies part of the pocket defined by side chains from helices I, II, III, and VII, but makes no contact with helices IV, V, and VI, in stark contrast to ligands in previous GPCR structures. The nitrogens of the symmetrical isothiourea group are both protonated with a net positive resonance charge; one of them (N4) forms a salt bridge (2.9 Å) with the Asp97$^{2.63}$ side chain. Note that the electron density does not preclude the existence of a very similar ligand conformation with a flipped thiourea group, in which the N3 nitrogen forms a salt bridge to Asp97$^{2.63}$, and the N4 nitrogen makes a polar interaction with main-chain carbonyl of Cys186 in ECL2. The importance of both nitrogens is supported by a reduction in binding affinity of ~100-fold upon methylation of one of them (*25*). Both cyclohexane rings fit into small subpockets and so make hydrophobic contacts with CXCR4. Connected by a short flexible linker, the imidazothiazole ring system is the only part of the ligand that contacts helix VII, in particular, by making a salt bridge (2.8 Å) between the protonated imidazothiazole N1 and Glu288$^{7.39}$ (*33*).

In the CXCR4-3–CVX15 complex, the bulky 16-residue ligand fills most of the binding-pocket volume (Fig. 2, B and D; fig. S3B; and table S5). The peptide forms a disulfide-stabilized (Cys4 to Cys13) β hairpin, with D-Pro8-Pro9 at the tip of the turn exposed to the extracellular milieu. The N-terminal part of the peptide backbone from Arg1 to Cys4 forms hydrogen bonds to CXCR4 backbone residues Asp187 to Tyr190 and so adds a partial third strand to the ECL2 β hairpin. The core-specific interactions are formed by two arginines at the peptide N terminus: Arg1 makes polar interactions with Asp187 (3.1 and 3.4 Å); Arg2 interacts with Thr117$^{3.33}$ (2.7 Å) and Asp171$^{4.60}$ (3.1 Å) and may form an additional hydrogen bond with His113$^{3.29}$ (3.1 Å), depending on its protonation state. The bulky naphthalene ring of Nal3 is anchored in a hydrophobic region bordered by helix V. Arg14 makes a salt bridge with Asp262$^{6.58}$ (3.5 Å) and an intramolecular hydrogen bond to the Tyr5 side chain, which in turn makes hydrophobic contacts with helix V side chains. Finally, the C-terminal D-proline is buried in the pocket next to the N terminus of the peptide and so makes a water-mediated interaction with the Asp288$^{7.39}$ side chain of CXCR4. The importance of the above interactions is supported by analyses of structure-activity relations of a series of CVX15 analogs (*26*).

The small-molecule and peptide ligand-binding sites substantially overlap (Fig. 3A). As CVX15 fills the entire pocket, some conformational variations between the two complexes are not surprising. CVX15 binding induces major deviations in the base of the receptor N terminus (residues 29 to 33), as well as a minor adjustment of extracellular tips of helices VI (~1.5 Å inward), VII (~1.5 Å tangential), and V (~0.5 Å outward). Major differences observed between binding of IT1t and CVX15 to CXCR4 compared with ligand-binding modes in β_2AR, A_{2A}AR, and rhodopsin (Fig. 3B) highlight the structural plasticity of GPCR binding sites.

Receptor dimerization. CXCR4 has been previously shown to homo- and heterodimerize, constitutively and upon ligand binding, by many different experimental methods (*14*, *15*, *34*–*40*). Although the functional importance of dimerization remains incompletely characterized, a considerable body of data suggests that it has important in vivo pharmacological effects. For example, WHIM syndrome (warts, hypogammaglobulinemia, infections, and myelokathexis syndrome) has been linked to mutations in the C terminus of CXCR4 and results in truncated variants that exhibit enhanced signaling and fail to desensitize and internalize upon CXCL12 stimulation. As a primarily heterozygous disease in which truncated CXCR4 is coexpressed with the wild-type receptor, dimerization has been proposed as the most likely mechanism to explain the dominance of mutant CXCR4 over the wild-type receptor (*41*, *42*). The structures presented here corroborate the concept of CXCR4 dimerization and define the dimer interface for a human GPCR with substantial buried surface area (850 Å^2). A similar, parallel, symmetric dimer of CXCR4 is observed in all five crystal structures (Fig. 4 and fig. S4), which suggests that these contacts represent a biologically relevant homodimer interface.

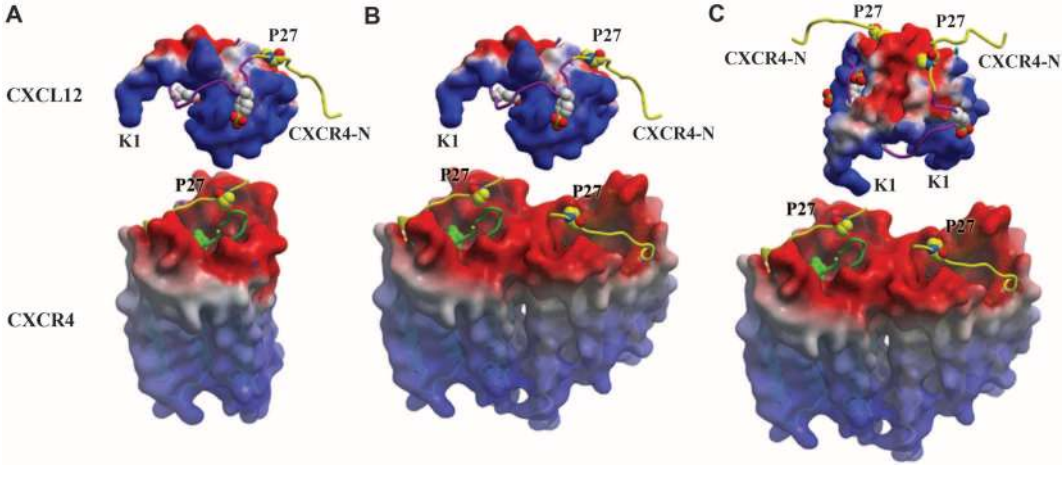

Fig. 5. Stoichiometry of possible CXCR4–CXCL12 binding or signaling complexes. No information on the orientation of CXCL12 with respect to CXCR4 is implied from the models presented. (**A**) Monomeric CXCR4 binding monomeric CXCL12, (**B**) dimeric CXCR4 binding monomeric CXCL12, (**C**) dimeric CXCR4 binding dimeric CXCL12 at either one or both orthosteric sites on each protomer. Alternatively, the 2:2 complex could involve two CXCL12 monomers binding dimeric CXCR4 (not shown). Both CXCR4 and CXCL12 surfaces are colored according to their electrostatic potential from red (negative) to blue (positive), highlighting the charge complementarity of these proteins. The portion of the CXCR4 N-terminal domain (CXCR4-N) present in both the CXCL12 complex (PDB ID: 2K05) and crystal structures of this study is colored yellow, while the remainder is purple (site 1). Pro27 and the three sulfotyrosines from the CXCR4 N terminus are represented with space-filling models. The CVX15 peptide (green ribbon) is shown in one CXCR4 receptor per panel and suggests the binding site for Lys1 and the rest of the flexible N-terminal region of CXCL12, which is critical for receptor activation (site 2). Figures were prepared using ICM software (www.Molsoft.com).

In dimers of CXCR4 bound to IT1t, the monomers interact only at the extracellular side of helices V and VI, leaving at least a 4 Å gap between the intracellular regions, which is presumably filled by lipids (Fig. 4, A and B, and table S6). Dimer association is driven mostly by hydrophobic interactions involving Leu194$^{5.33}$, Val197$^{5.36}$, Val198$^{5.37}$, Phe201$^{5.40}$, Met205$^{5.44}$, and Leu210$^{5.49}$ contacts. A substantial role is also played by a Trp195$^{5.34}$-Leu267$^{6.63}$ contact, which includes both side-chain stacking and a hydrogen bond from Trp195$^{5.34}$ (NE1) to the main-chain carbonyl oxygen of Leu267$^{6.63}$. Another specific polar interaction includes a hydrogen-bonding network between the side chains of Asn192 and Glu268 in opposing receptors, which also involves the main-chain carbonyl oxygens of Leu266$^{6.62}$ and Trp195$^{5.34}$. Pro191 in ECL2 likely plays a role in this network by stabilizing the Trp195$^{5.34}$ side-chain conformation. As these contacts persist throughout all five crystal structures, they are likely genuine, rather than artifacts of crystallization (Fig. 4E).

In addition, dimers of CXCR4 bound to CVX15 are stabilized by interactions at the intracellular ends of helices III and IV and at ICL2, controlled largely by hydrophobic interactions of Tyr135$^{3.51}$, Leu136$^{3.52}$, His140, and Pro147 side chains (~400 Å^2 buried) (Fig. 4, C, D, and F, and table S6). It appears that binding of the bulky CVX15 peptide induces a small tilt in the extracellular part of helix V, which brings the intracellular parts of opposing receptors into close contact. This type of ligand-induced conformational change could explain the cooperative binding observed with certain CXCR4 ligands, as well as the effects of allosteric modulators. Specifically, binding of a ligand to one receptor could induce a structural change in helix V of the second receptor, which could modify the ligand-binding affinity to the second receptor, resulting in either negative or positive cooperativity. Extending this concept to chemokine receptor heterodimers, CXCR4 has been reported to dimerize with CCR2 and CCR5, and both complexes show negative binding cooperativity with their ligands, not only in vitro but also in vivo (*37*, *39*), an observation that may have implications for drug efficacy.

The CXCR4 dimer is strikingly different from previous models of GPCR dimerization, which suggested contacts through either helix I or helices IV and V (*43*–*47*) and implied contacts throughout the length of the TM bundle. It is also notable that with the exception of Trp195$^{5.34}$ (conservation ~70%), little sequence conservation is found among chemokine receptors for the residues that constitute the dimerization site, even though many receptors have been shown to oligomerize (*40*). The specific nature of the interactions may facilitate the ability of CXCR4 to heterodimerize with other chemokine receptors (*37*, *39*, *48*), as well as GPCRs outside of the chemokine family (*49*), although one cannot discount the possibility that many modes of oligomerization may exist.

Implications for the two-site model of chemokine binding and complexes of CXCR4 with CXCL12 and gp120. The known structures of chemokines, including CXCL12, feature a disordered N-terminal domain that largely controls receptor signaling and is hypothesized to penetrate the receptor helical bundle (*50*, *51*). The chemokine N terminus is followed by a core globular domain, which is thought to bind to the receptor N terminus and ECLs, which form an interaction site that confers affinity and specificity (*52*). The separation of the binding and signaling functions has led to the so-called "two-site" model of receptor binding, with the chemokine core domain being the "site one" docking domain and the chemokine N terminus being the "site two" signaling trigger (*50*, *53*, *54*). The nuclear magnetic resonance (NMR) structure of CXCL12 in complex with a 38-residue, sulfotyrosine-containing peptide derived from the CXCR4 N terminus has been determined [Protein Data Bank (PDB) ID: 2K05] (*55*). This structure is thought to represent at least part of the "site one" complex and reveals important interactions between CXCL12 and residues, including three sulfated tyrosines, that are absent from the CXCR4 receptor structure.

The peptide and small-molecule complexes of CXCR4 identify the likely "site two" of the chemokine-signaling trigger. The IT1t compound and CVX15 peptide have both been characterized as competitive inhibitors of CXCL12, and many of the receptor-ligand contacts in the cocrystal structures presented are important for CXCL12 binding, including the acidic Asp187, Glu288$^{7.39}$, and Asp97$^{2.63}$ (Fig. 2) (*56*, *57*). The CVX15 peptide, rich in basic residues, may trace, to some extent, the path of the N-terminal signaling peptide of CXCL12 (KPVSLSYR), and the binding site of IT1t may point to the major anchor region for this domain. Furthermore, our preliminary modeling studies suggest that Lys1, the most critical residue in CXCL12 for receptor activation, could reach into the CXCR4 pocket and interact with one of these acidic residues (Fig. 5). The extensive binding site mapped out by the CVX15 peptide also clarifies how progressive shortening of the CXCL12 N terminus leads to a gradual loss of binding affinity (*50*). Taken together, these data suggest that the small molecule and cyclic peptide block ligand binding by acting as orthosteric competitors of the CXCL12 N-terminal signaling trigger, providing strong support for the two-site model of binding. Along these lines, a recent NMR study showed that the CXCR4 antagonist AMD3100 could displace the CXCL12 N terminus from the receptor without displacing the chemokine core domain (*58*).

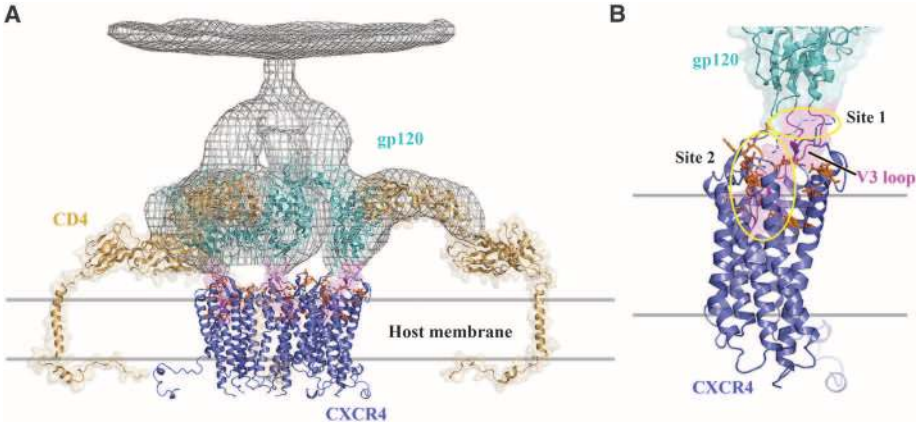

Fig. 6. Model of early stages of the HIV-1 entry process. (**A**) Viral entry begins with binding of envelope spikes consisting of a heterotrimer $(gp120)_3$ $(gp41)_3$ [wire, adapted from density map of gp120/CD4/17b Fab complex derived by cryo–electron tomography of intact HIV-1 spikes (*70*); PDB ID: 3DNO] to CD4 on the surface of host target cells. Glycoprotein gp120 (core structure, cyan, PDB ID: 2QAD) interacts with CD4 (tan, PDB ID: 1WIP and 2KLU). This interaction triggers conformational changes in gp120 that increase the exposure of the third variable loop V3 (magenta) and a region of gp120 between inner and outer domains. CCR5 or CXCR4 (blue) is then recruited as a co-receptor. The number of spikes involved in viral entry and the number of molecules of CD4 or CXCR4 binding to a single spike are unknown; here, three CD4 molecules are represented, which results in the close approach of gp120 molecules to the host cell membrane where the interaction with three CXCR4 molecules is depicted. (**B**) By analogy to a two-site model based on CCR5 (*65*), the N terminus of CXCR4 containing sulfotyrosines (site 1, circled in yellow) binds first to the base of the V3 loop, which induces further conformational changes in gp120 that enable V3 to bind to the extracellular side of CXCR4, primarily ECL2, ECL3, and the ligand-binding cavity (site 2, circled in yellow). CXCR4 residues previously shown to affect gp120 binding are shown as sticks with carbons colored in orange. A hypothetical path of the CXCR4 N terminus, which is not observed in the current structure, is shown as a blue dashed curve. Only CXCR4 monomers are shown for clarity, although dimers are also possible. Figures 1, 2, 3, 4, and 6 were prepared using PyMOL.

Chemokines are able to bind their receptors as monomers in order to activate cell migration (*59*). However, chemokine oligomers, including CXCL12, appear to be functional and to induce alternative signaling responses, such as cellular activation or signals to halt migration (*55*, *60*, *61*), which suggests the concept that these complexes dynamically change their stoichiometries and structures as part of their functional regulation. Given the oligomeric nature of CXCR4 and the complementary electrostatic surfaces of the ligand and receptor, one can envision CXCL12 binding the receptor as a 1:1, 1:2, or 2:2 ligand:receptor complex (Fig. 5). Additional experiments will be necessary to fully define the relevance and functional implications of different chemokine: receptor stoichiometries and structures. Nevertheless, the current CXCR4 structures are compatible with emerging concepts of signaling diversity induced by alternative binding modes of the ligands.

CXCR4 and the related CCR5 serve as co-receptors for HIV-1 viral particles, facilitating their entry into cells. Structures have been reported for the other key components of the entry complex, HIV-1 glycoproteins gp120 and gp41, and the host leukocyte glycoprotein receptor CD4 (*62*–*64*). The N termini of CXCR4 and CCR5, including sulfated tyrosine residues, have been implicated in gp120 binding, analogous to CXCL12 recognition (*65*). Other structural features critical to the interaction involve the gp120 V3 loop, which becomes exposed on CD4 binding (*66*) and then interacts with CXCR4 ECL2 and ECL3. The basic character of the protruding V3 loop along with acidic residues in the CXCR4 binding pocket have been reported to be important for HIV-1 infectivity (Fig. 2, C and D) (*57*, *67*), which suggests that the loop could also penetrate the pocket (Fig. 6). Thus, the CXCR4 structures suggest testable hypotheses regarding interaction of CXCR4 with its natural ligand and with HIV-1 gp120. The real challenge will be in understanding the dynamic changes in these complexes that result in signal transduction and viral fusion. As further details of these interactions are resolved, new opportunities for drug discovery efforts targeting specific functional states of the receptor will likely emerge.

References and Notes

1. M. Baggiolini, *Nature* **392**, 565 (1998).
2. B. Moser, M. Wolf, A. Walz, P. Loetscher, *Trends Immunol.* **25**, 75 (2004).
3. C. R. Mackay, *Nat. Immunol.* **2**, 95 (2001).
4. Q. Ma *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9448 (1998).
5. Y. R. Zou, A. H. Kottmann, M. Kuroda, I. Taniuchi, D. R. Littman, *Nature* **393**, 595 (1998).
6. A. Müller *et al.*, *Nature* **410**, 50 (2001).
7. F. Balkwill, *Semin. Cancer Biol.* **14**, 171 (2004).
8. A. Zlotnik, *J. Pathol.* **215**, 211 (2008).
9. A. M. Fulton, *Curr. Oncol. Rep.* **11**, 125 (2009).
10. B. A. Teicher, S. P. Fricker, *Clin. Cancer Res.* **16**, 2927 (2010).
11. Y. Feng, C. C. Broder, P. E. Kennedy, E. A. Berger, *Science* **272**, 872 (1996).
12. C. C. Bleul *et al.*, *Nature* **382**, 829 (1996).
13. E. Oberlin *et al.*, *Nature* **382**, 833 (1996).
14. G. J. Babcock, M. Farzan, J. Sodroski, *J. Biol. Chem.* **278**, 3378 (2003).
15. Y. Percherancier *et al.*, *J. Biol. Chem.* **280**, 9895 (2005).
16. J. Wang, M. Norcross, *Drug Discov. Today* **13**, 625 (2008).
17. V. Cherezov *et al.*, *Science* **318**, 1258 (2007).
18. V. P. Jaakola *et al.*, *Science* **322**, 1211 (2008).
19. T. Warne *et al.*, *Nature* **454**, 486 (2008).
20. J. H. Park, P. Scheerer, K. P. Hofmann, H. W. Choe, O. P. Ernst, *Nature* **454**, 183 (2008).
21. M. A. Hanson *et al.*, *Structure* **16**, 897 (2008).
22. C. B. Roth, M. A. Hanson, R. C. Stevens, *J. Mol. Biol.* **376**, 1305 (2008).
23. In Ballesteros-Weinstein numbering, a single most-conserved residue among the class A GPCRs is designated *x*.50, where *x* is the transmembrane helix number. All other residues on that helix are numbered relative to this conserved position.
24. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr; and x, any amino acid.
25. G. Thoma *et al.*, *J. Med. Chem.* **51**, 7915 (2008).
26. H. Tamamura *et al.*, *FEBS Lett.* **550**, 79 (2003).
27. S. J. DeMarco *et al.*, *Bioorg. Med. Chem.* **14**, 8396 (2006).
28. G. Moncunill *et al.*, *Mol. Pharmacol.* **73**, 1264 (2008).
29. Materials and methods are available as supporting material on *Science* Online.
30. H. Zhou, H. H. Tai, *Arch. Biochem. Biophys.* **373**, 211 (2000).
31. P. Scheerer *et al.*, *Nature* **455**, 497 (2008).
32. K. P. Hofmann *et al.*, *Trends Biochem. Sci.* **34**, 540 (2009).
33. Experimental measurement of the acidic dissociation constant, pK_a, by means of multiplexed capillary electrophoresis gave $pK_{a1} = 5.39$, $pK_{a2} = 7.89$, and $pK_{a3} = 9.92$, indicating that the conjugated acid of isothiourea N4 has a pK_a of 5.4, which is also consistent with the prediction ($pK_a = 4.84 \pm 0.60$) by pK_a database from ACD/Labs, and suggesting that N4 is unprotonated at the physiological condition of pH = 7.4. Crystallization of the CXCR4-2/IT1t complex was performed at pH 5.5, and the acidic microenvironment of $Glu288^{7.39}$ may further shift the equilibrium toward protonation.
34. A. J. Vila-Coro *et al.*, *FASEB J.* **13**, 1699 (1999).
35. K. E. Luker, M. Gupta, G. D. Luker, *FASEB J.* **23**, 823 (2009).
36. A. Levoye, K. Balabanian, F. Baleux, F. Bachelerie, B. Lagane, *Blood* **113**, 6085 (2009).
37. D. Sohy *et al.*, *J. Biol. Chem.* **284**, 31270 (2009).
38. J. Wang, L. He, C. A. Combs, G. Roderiquez, M. A. Norcross, *Mol. Cancer Ther.* **5**, 2474 (2006).
39. D. Sohy, M. Parmentier, J. Y. Springael, *J. Biol. Chem.* **282**, 30062 (2007).
40. C. L. Salanga, M. O'Hayre, T. Handel, *Cell. Mol. Life Sci.* **66**, 1370 (2009).
41. K. Balabanian *et al.*, *Blood* **105**, 2449 (2005).
42. B. Lagane *et al.*, *Blood* **112**, 34 (2008).
43. F. Fanelli, P. G. De Benedetti, *J. Comput. Aided Mol. Des.* **20**, 449 (2006).
44. M. Filizola, *Life Sci.* **86**, 590 (2010).
45. W. Nemoto, H. Toh, *Curr. Protein Pept. Sci.* **7**, 561 (2006).
46. P. H. Reggio, *AAPS J.* **8**, E322 (2006).
47. S. Vohra *et al.*, *Biochem. Soc. Trans.* **35**, 749 (2007).
48. N. Isik, D. Hereld, T. Jin, *PLoS ONE* **3**, e3424 (2008).
49. O. M. Pello *et al.*, *Eur. J. Immunol.* **38**, 537 (2008).
50. M. P. Crump *et al.*, *EMBO J.* **16**, 6996 (1997).
51. C. Dealwis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6941 (1998).
52. S. K. Gupta, K. Pillarisetti, R. A. Thomas, N. Aiyar, *Immunol. Lett.* **78**, 29 (2001).
53. I. Clark-Lewis *et al.*, *J. Leukoc. Biol.* **57**, 703 (1995).
54. T. N. Wells *et al.*, *J. Leukoc. Biol.* **59**, 53 (1996).
55. C. T. Veldkamp *et al.*, *Sci. Signal.* **1**, ra4 (2008).
56. A. Brelot *et al.*, *J. Virol.* **73**, 2576 (1999).
57. A. Brelot, N. Heveker, M. Montes, M. Alizon, *J. Biol. Chem.* **275**, 23736 (2000).
58. Y. Kofuku *et al.*, *J. Biol. Chem.* **284**, 35240 (2009).
59. C. D. Paavola *et al.*, *J. Biol. Chem.* **273**, 33157 (1998).
60. V. Appay, A. Brown, S. Cribbes, E. Randle, L. G. Czaplewski, *J. Biol. Chem.* **274**, 27505 (1999).
61. L. G. Czaplewski *et al.*, *J. Biol. Chem.* **274**, 16077 (1999).
62. R. Diskin, P. M. Marcovecchio, P. J. Bjorkman, *Nat. Struct. Mol. Biol.* **17**, 608 (2010).
63. M. Pancera *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1166 (2010).
64. P. D. Kwong *et al.*, *Nature* **393**, 648 (1998).
65. C. C. Huang *et al.*, *Science* **317**, 1930 (2007).
66. C. C. Huang *et al.*, *Science* **310**, 1025 (2005).
67. B. J. Doranz *et al.*, *J. Virol.* **73**, 2752 (1999).
68. T. E. Angel, M. R. Chance, K. Palczewski, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8555 (2009).
69. S. Tian *et al.*, *J. Virol.* **79**, 12667 (2005).
70. J. Liu, A. Bartesaghi, M. J. Borgnia, G. Sapiro, S. Subramaniam, *Nature* **455**, 109 (2008).

71. This work was supported in part by the Protein Structure Initiative grant U54 GM074961 for structure production, NIH Roadmap Initiative grant P50 GM073197 for technology development, NIH grants R21 RR025336 and R21 AI087189 to V.C., and Pfizer. T.M.H. acknowledges support from NIH R01 AI037113 and R01 GM081763, D.J.H. from F32 GM083463, and R.A. from R01 GM071872. The authors thank J. Velasquez for help on molecular biology, T. Trinh and K. Allin for help on baculovirus expression, I. Wilson and D. Burton for careful review and scientific feedback on the manuscript, W. Schief for electron microscopy models of gp120-gp41-CD4 complex, G. W. Han for evaluating the structure quality and preparation for PDB submission, and A. Walker for assistance with manuscript preparation. The authors acknowledge E. La Chapelle on chemistry tool compound synthesis; A. Rane on radiolabeling of [^{3}H]BIMA; M. Cui for help developing [^{3}H]BIMA binding assay; Y. Zheng, The Ohio State University, and M. Caffrey, Trinity College (Dublin, Ireland), for the generous loan of the in meso robot [built with support from the National Institutes of Health (GM075915), the NSF (IIS0308078), and Science Foundation Ireland (02-IN1-B266)]; and J. Smith, R. Fischetti, and N. Sanishvili at the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) beamline at the Advanced Photon Source for assistance in development and use of the minibeam and beamtime. The GM/CA-CAT beamline (23-ID) is supported by the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). R.C.S. is the founder and is a board member of Receptos, a biotech company focused on GPCR structure–based drug discovery. Transfer of IT1t, CVX15, and [^{3}H]BIMA will require a Material Transfer Agreement (MTA) with Pfizer, and transfer of all other constructs and biological materials requires an MTA with the Scripps Research Institute. Atomic coordinates and structure factors have been deposited in the Protein Data Bank with identification codes 3ODU (CXCR4-2–IT1t, $P2_1$), 3OE0 (CXCR4-3–CVX15, *C2*), 3OE8 (CXCR4-2–IT1t, *P1*), 3OE9 (CXCR4-3–IT1t, *P1*), and 3OE6 (CXCR4-1–IT1t, *I222*).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1194396/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S6
References

29 June 2010; accepted 2 September 2010
Published online 7 October 2010;
10.1126/science.1194396