Title: Structural insights into CXCR4 modulation and oligomerization

Authors

Kei Saotome[1]*, Luke L. McGoldrick[1]*, Jo-Hao Ho[1], Trudy F. Ramlall[1], Sweta Shah[1], Michael J. Moore[1], Jee Hae Kim[1], Raymond Leidich[1], William C. Olson[1], Matthew C. Franklin[1]*

[1]Regeneron Pharmaceuticals, Inc. Tarrytown, NY 10591

*Correspondence: kei.saotome@regeneron.com (K.S.); luke.mcgoldrick@regeneron.com (L.L.M); matthew.franklin@regeneron.com (M.C.F)

Abstract

Activation of the chemokine receptor CXCR4 by its chemokine ligand CXCL12 regulates diverse cellular processes. CXCR4 also serves as a key target for diseases such as cancer and HIV. Previously reported crystal structures of CXCR4 bound to antagonists revealed the architecture of an inactive, homodimeric receptor. However, many structural aspects of CXCR4 remain poorly understood, including its activation by CXCL12, as well as its assembly into higher-order oligomers. Here, we use cryoelectron microscopy (cryoEM) to investigate various modes of CXCR4 regulation in the presence and absence of G_i protein. CXCL12 activates CXCR4 by inserting its N-terminus deep into the CXCR4 orthosteric pocket. The binding of FDA-approved antagonist AMD3100 is stabilized by electrostatic interactions with acidic residues in the 7 transmembrane helix bundle. A potent antibody blocker, REGN7663, binds across the extracellular face of CXCR4 and inserts its CDR-H3 loop into the orthosteric pocket. Trimeric and tetrameric structures of CXCR4 reveal, to our knowledge, previously undescribed modes of GPCR oligomerization. Remarkably, CXCR4 adopts distinct subunit conformations in trimeric and tetrameric assemblies, highlighting how oligomerization could allosterically regulate chemokine receptor function.

Main

Chemokine receptors are a family of Class A G-protein coupled receptors (GPCRs) that mediate cell migration in response to binding of chemokine ligands[1]. CXCR4 is a well-studied chemokine receptor that is activated by the chemokine ligand CXCL12 (also called stromal cell-derived factor 1, or SDF-1) and signals primarily through coupling with G_i protein[2], regulating cell migration in hematopoeisis, neovascularization, angiogenesis and various other physiological processes[3]. CXCR4 is involved in numerous diseases, including roles as a cancer marker implicated in tumor proliferation[4] and as a coreceptor for X4-tropic HIV strains[5]. Mutations in CXCR4 that result in enhanced and prolonged signaling result in a rare immune disorder called WHIM (warts, hypogammaglobulinemia, infections, and myelokathexis) syndrome[6]. The significant roles of CXCR4 in health and disease have made the receptor an intensely investigated drug target[7]. The small molecule CXCR4 antagonist AMD3100 (plerixafor), initially developed as an HIV entry inhibitor[8], was FDA-approved as a hematopoietic stem cell mobilizer for autologous transplantation in patients with Non-Hodgkin's lymphoma or multiple myeloma[9,10]. Numerous additional CXCR4-targeting therapeutics have been developed[7], notably including monoclonal antibodies with improved pharmacokinetic properties and thus potentially greater efficacy compared to small molecules and peptides[11-13].

Structural studies of Class A GPCRs have focused on isolated monomeric forms of the receptors bound to various ligands, pharmacological modulators, and transducer proteins[14]. However, increasing evidence suggests GPCRs can form dimers and higher order oligomers in the plasma membrane, with implications for signaling and therapeutic action[15]. Chemokine receptors are no exception; a multitude of studies have indicated the existence of homo- and hetero- oligomers[16-18], including crystal structures of antagonist-bound CXCR4 consistently revealing homodimeric forms[19,20]. Interestingly, CXCR4 has also shown a propensity to form higher order oligomers using a mechanism that can be separated from dimerization[21].

Despite its critical roles in health and disease, many mechanistic aspects of CXCR4 remain poorly understood, owing to a lack of structural information. These include its activation by CXCL12, binding mode of AMD3100, coupling to G_i protein, inhibitory action of antibodies, and mechanisms of higher order oligomerization. Here, we address these open questions by reporting a series of cryoelectron microscopy (cryoEM) structures of CXCR4 complexes.

Results

Structural basis of CXCL12 and AMD3100 action on CXCR4

To stabilize active state signaling complexes and improve protein yields we made the following modifications to wild type CXCR4: we replaced the N-terminal methionine with an HA (hemagglutinin) signal peptide[22], included a previously characterized constitutively active mutation (N119S)[23], and fused monomeric eGFP[24] and FLAG tag to the receptor C-terminus (Extended Data Fig. 1a). We refer to this construct as $CXCR4_{EM}$. We also employed a $G\alpha_i$ construct harboring dominant negative mutations[25] to facilitate isolation of receptor/G_i complexes in the absence of stabilizing antibody fragments[26]. Fluorescence detection size exclusion chromatography (FSEC)[27] experiments indeed indicated complex formation between $CXCR4_{EM}$ and G_i in the absence of agonist (Extended Data Fig. 1b). We prepared detergent-solubilized $CXCR4_{EM}/G_i$ complexes and first determined cryoEM structures in apo, CXCL12-bound, and AMD3100-bound states at overall resolutions of 2.7, 3.3, and 3.2 Å, respectively (Fig. 1, Extended Data Figs. 2, 3). Each of the structures shows a prototypical arrangement of an active receptor coupled to heterotrimeric G protein, including a hallmark kink of TM6 relative to previously reported crystal structures of antagonist-bound CXCR4[19,20] (Extended Data Fig. 4a). We therefore refer to the CXCR4 conformation in these structures as active.

Our cryoEM reconstruction of $CXCR4_{EM}/G_i$/CXCL12 revealed clear signal for the chemokine bound at the extracellular side of the receptor (Fig. 1a). Density for the chemokine N-terminus (res. 1-12) was sufficiently resolved to build side chains (Fig. 1b), whereas the remainder of the ligand was less resolved due to flexibility and only permitted main chain tracing (Extended Data Fig. 3i). Consequently, interactions between the chemokine N-terminal region and receptor orthosteric pocket (chemokine recognition site 2) were readily discernible, while interactions between the globular portion of the ligand and the N-terminus of CXCR4[28] (chemokine recognition site 1) were unclear. CXCL12 is known to exist in monomeric and dimeric forms that have been shown to yield distinct signaling outcomes upon CXCR4 binding[29,30]. Weak signal corresponding to a second protomer of the CXCL12 dimer could be observed in our cryoEM reconstruction, consistent with the notion that dimeric forms of CXC ligands act on single receptor subunits[20,31] (Extended Data Fig 3i).

The binding mode of CXCL12 onto CXCR4 is overall similar to those found in published structures of CC and CXC chemokine/chemokine receptor complexes[31-36] (Fig. 2, Extended Data Fig. 4b). However, the CXCL12 binding pose observed in our structure notably differs from that of CXCL12 bound to atypical chemokine receptor 3 (ACKR3, formerly known as CXCR7)[37], a promiscuous receptor that has been suggested to function as a chemokine “scavenger” and has approximately 10-fold higher affinity for CXCL12 than CXCR4[38,39]. The CXCL12 C-terminal α helix is rotated ~70° when bound to ACKR3 relative to CXCR4 (Extended Data Fig. 4c). Correspondingly, the 40s loop of CXCL12 is situated proximal to the N-terminal region in CXCR4, while it is nearby ECL3 in ACKR3. In addition to the distinct overall chemokine/receptor docking orientations, the binding geometries of the CXCL12 N-terminus within the orthosteric pockets of each receptor are also unique (Extended Data Fig. 4d).

Mutations at the distal N-terminus of CXCL12 can convert the chemokine into an antagonist[40], highlighting its importance for receptor activation. Our structure shows how the CXCL12 N-terminus protrudes into the orthosteric pocket of CXCR4 and makes extensive contacts with the TM core (Fig. 2a). The distal CXCL12 N-terminus is positioned overall deeper into the pocket than that of the antagonistic viral chemokine vMIP-II[20] (Extended Data Fig. 4e), consistent with their respective ligand functions. $P2_{CXCL12}$ penetrates deepest into the orthosteric pocket, contacting the side chain of Y116$^{3.32}$ (Ballesteros-Weinstein numbering[41] in superscript). The side chain of K1$_{\mathrm{CXCL12}}$ projects upward from the TM core to the extracellular side of the receptor and is positioned to interact electrostatically with D97$^{2.63}$ and possibly D187$^{\mathrm{ECL2}}$. S4$_{\mathrm{CXCL12}}$ makes an apparent hydrogen bond interaction with E288$^{7.39}$. L5$_{\mathrm{CXCL12}}$ packs onto a mainly hydrophobic surface composed of L41$^{1.35}$, Y45$^{1.39}$, W94$^{2.60}$, A98$^{2.64}$. R8$_{\mathrm{CXCL12}}$ appears poised to make a charge-charge interaction with D262$^{6.58}$, as predicted previously based on charge-swap experiments[28]. Several of the CXCR4 residues mentioned above (W94$^{2.60}$, D97$^{2.63}$, Y116$^{3.32}$, D187$^{\mathrm{ECL2}}$, E288$^{7.39}$) have been shown to be important for CXCL12/CXCR4 signaling[28,42,43], underscoring the functional relevance of the interactions observed in our cryoEM structure. We expand on the structural basis of CXCL12 activation of CXCR4 in a following section.

We observed unambiguous density for the bilobed AMD3100 molecule in our cryoEM reconstruction of CXCR4$_{\mathrm{EM}}$/G$_i$/AMD3100 (Fig. 1c). Although it has primarily been described as an antagonist[44], our observation that AMD3100 binds to the active CXCR4$_{\mathrm{EM}}$/G$_i$ complex without disrupting G protein coupling is consistent with the compound acting as a weak partial agonist on constitutively active mutants of CXCR4. AMD3100 binds the orthosteric pocket using a diagonal orientation and directly blocks CXCL12 docking, although its overall binding mode is shifted toward TM5/6 relative to the CXCL12 N-terminus (Fig. 2b,c). Each of the two positively charged cyclam rings of AMD3100[45] is stabilized electrostatically by an acidic side chain pointed toward the center of the ring; the cyclam moiety closer to the extracellular side interacts with D262$^{6.58}$ while the cyclam proximal to the transmembrane core interacts with E288$^{7.39}$. The closely matched spacings between the side chains of D262$^{6.58}$ and E288$^{7.39}$ residues and the cyclam rings therefore appears be the main binding determinant of AMD3100 and other bicyclam analogues. Consistent with our structure, a previous study showed that that D262N and E288A mutants each reduced the affinity of AMD3100 to CXCR4 by more than 50-fold[45]. The central aromatic ring of the phenylenebis(methylene) linker connecting the two cyclam moieties makes hydrophobic contacts with I284$^{7.35}$, which is positioned directly in between D262$^{6.58}$ and E288$^{7.39}$ in the orthosteric pocket. This interaction may contribute to the increased potency of bicyclams with an aromatic linker rather relative to those with an aliphatic linker[46]. The rigidity imposed by the aromatic linker on the relative positions of the two cyclam moieties may also play a role in stabilizing the binding pose of AMD3100.

Antagonism of CXCR4 by REGN7663 mAb

Antibody-based therapeutics against CXCR4 and other GPCRs are a promising alternative to small molecules due to their high specificity to the target, opportunity for Fc-effector functions, and favorable pharmacokinetic properties[11,13,47,48]. REGN7663 is a fully human anti-CXCR4 monoclonal antibody (mAb) generated using VelocImmune mice[49,50]. We showed using a CRE-Luciferase reporter assay that REGN7663 is a potent blocker (IC50=2.72 nM) of CXCL12-induced signaling in HEK293 cells engineered to overexpress CXCR4 (Fig. 3a). Further, in the absence of CXCL12, REGN7663 decreased the apparent basal activity (EC50=1.71 nM), indicating inverse agonism in the setting of CXCR4 overexpression (Fig. 3b). To understand how REGN7663 binds and inhibits CXCR4, we determined a 3.4 Å resolution cryoEM structure of REGN7663 Fab in complex with $CXCR4_{EM}/G_i$ (Fig. 3c, Extended Data Fig. 5a-d). The structure revealed that REGN7663 binds directly onto the extracellular face of CXCR4, antagonizing the receptor by steric blockade of CXCL12 binding. Most of the REGN7663 epitope resides at the extracellular N-terminal region and ECL2 (Extended Data Fig. 5e,f). The REGN7663 heavy chain dominates the binding interactions, burying significantly more surface area (~1100 $Å^2$) than the light chain (~300 $Å^2$). Although the overall architecture of the complex is similar to the apo, CXCL12-bound, and AMD3100-bound $CXCR4_{EM}/G_i$ structures, REGN7663 binding induces distinct conformations of the N-terminus and ECL2, suggesting their flexibility is important for specific mAb binding (Extended Data Fig. 5g). Heavy chain complementarity-determining regions (CDRs) 1 and 2 of REGN7663 are oriented toward the extracellular ends of TM4 and TM5, while light chain CDRs are oriented extracellular to TM1 and TM2 (Fig. 3d). Remarkably, the CDR-H3 loop of REGN7663 wedges between the CXCR4 N-terminus and ECL2, exhibiting a partial insertion into the CXCR4 orthosteric pocket. The side chain of REGN7663 residue R105 protrudes deepest into the orthosteric pocket, making an apparent charge-charge interaction with E288$^{7.39}$(Fig. 3e). The insertion of CDR-H3 loop, though not activating in the case of REGN7663, is reminiscent of how the CDR3 loop of the single domain antagonist antibody JN241 occupies the orthosteric pocket of apelin receptor[51]. Taken together with the finding that JN241 was converted into a full agonist through subtle engineering of CDR3[51], our structure of REGN7663/CXCR4 complex illustrates the potential for full antibodies (containing light and heavy chains) functionally modulating GPCRs by inserting CDR loop(s) into the orthosteric pocket.

Conformational changes associated with CXCR4 activation and Gα_i protein docking

We next sought to assess the conformational changes associated with CXCR4 activation. Available crystal structures of inactive, antagonist-bound CXCR4 contain construct modifications, namely T4 lysozyme (T4L) inserted at ICL3 and a thermostabilizing mutation in TM3, that could confound comparison with our current structures. We therefore determined a 3.1 Å resolution cryoEM structure CXCR4$_{EM}$ in the absence of G$_i$ protein, utilizing REGN7663 Fab as a fiducial mark (Fig. 4a, Extended Data Fig. 5h-k). Structural alignment of the REGN7663 Fab/CXCR4$_{EM}$/G$_i$ structure with the G$_i$-free REGN7663 Fab/CXCR4$_{EM}$ structure showed nearly identical conformations at the REGN7663 epitope/paratope regions but distinct conformations at the intracellular half of the receptor, including the characteristic movement of TM6 underlying receptor activation (Fig. 4b, Extended Data Fig. 5l). Additional conformational changes upon activation/G$_i$-binding include movement of TM5 toward TM6, subtle displacement of TM2 outward, an inward kink of TM7, and loss of ordered structure in H8. We note that H8 was also unresolved in previously determined antagonist-bound CXCR4 crystal structures[19,20], suggesting this is a consistent feature of the inactive receptor.

We further compared the conformations of the inactive and CXCL12-bound structures to analyze how CXCL12 binding results in activation (Fig. 4c). Binding of CXCL12 N-terminal coil to the orthosteric pocket requires structural changes to the inactive state pocket. Residues P2 and S4 at the CXCL12 N-terminus push E288$^{7.39}$ outward and toward the cytoplasmic side, while V3$_{\text{CXCL12}}$ forces a downward displacement of Y255$^{6.51}$. The movements of E288$^{7.39}$ and Y255$^{6.51}$ are in turn transmitted to F292$^{7.43}$, which has been previously implicated in CXCR4 signal transmission[43], and conserved toggle switch residue[52] W252$^{6.48}$, respectively. Together, these conformational changes trigger further structural rearrangements that ultimately stabilize the active, G$_i$-bound conformation of CXCR4. Furthermore, a close comparison revealed that due to binding of the CXCL12 N-terminus in the orthosteric pocket, E288$^{7.39}$ side chain reorients, along with a small, ~0.7-1 Å outward movement of the extracellular half of TM7 helix relative to our AMD3100/CXCR4$_{\text{EM}}$/G$_i$, REGN7663 Fab/CXCR4$_{\text{EM}}$/G$_i$, and apo CXCR4$_{\text{EM}}$/G$_i$ structures (Extended Data Fig 6a). This slight conformational difference at TM7 induced by CXCL12 may explain why it is full agonist, while the other ligands are not. Similar structural mechanisms of chemokine activation to that described above for CXCL12 have been observed for the CCR2/CCL2 complex[32] and CCR5/MIP-1α complex[34].

Like other class A GPCRs, coupling of Gα_i to CXCR4 is mediated by insertion of the Gα_i α5 helix and C-terminal “wavy hook" into the cytoplasmic-facing core of the receptor TM domain (Fig. 4d). “Wavy hook” residues L353 and F354 bury deepest into CXCR4 and contact R134$^{3.50}$, Q233$^{\text{ICL3}}$, K236$^{6.32}$, A237$^{6.33}$, T240$^{6.36}$, and A307 mainly via van der Waals and hydrophobic interactions. Gα_i α5 helix makes numerous additional contacts with TM2, TM3, ICL2, TM5, ICL3, and TM6. Salt bridge interactions between D341(Gα_i)/K234$^{6.30}$ and E28(Gα_i)/K149 probably play an important role in stabilizing the docking of G$_i$ protein onto CXCR4. Although the overall G$_i$ binding mode of CXCR4 and other chemokine receptors is shared, the angle at which the Gα_i α5 helix docks into the TM bundle differs slightly (Extended Data Fig. 6b). While CXCR4, CXCR1[36], and CXCR2[31] show highly similar α5 docking angles, the docking angles in CCR1[33], CCR2[32], and CCR5[34] are similar to each other and shifted relative to CXCR4, owing to distinct intracellular loop conformations and receptor interactions with $G\alpha_i$ (Extended Data Fig. 6c). More specifically, in the CC chemokine receptors, $G\alpha_i$ α5 helix is shifted toward ICL2 and further from ICL3. Available data therefore indicate that CXC and CC chemokine receptors have slightly different G_i docking geometries.

Oligomeric structures of CXCR4

Although GPCRs are generally understood to function as monomeric units, numerous studies have reported that chemokine receptors form dimers and higher order oligomers at the cell surface as expression levels increase[53-56]. Homo- and hetero-oligomerization have been proposed to add complexity to chemokine receptor function, perhaps through allosteric communication between interacting subunits[57,58]. Multiple structures of CXCR4 from different crystal forms showed the same homodimeric architecture[19,20], demonstrating that the detergent-solubilized receptor has the propensity to dimerize using specific intersubunit interactions mainly involving TM5 and TM6. Our size exclusion chromatography (SEC) data of $CXCR4_{EM}$ consistently showed multiple peaks with different elution volumes, including peaks corresponding to oligomeric species larger than monomeric $CXCR4_{EM}$ or $CXCR4_{EM}/G_i$ (Extended Data Fig. 1b, 2a). Wild type CXCR4 fused to GFP showed a similar FSEC profile to $CXCR4_{EM}$, indicating that the apparent oligomerization was not specific to the constitutively active N119S mutation present in $CXCR4_{EM}$.

We isolated and characterized a presumed oligomeric SEC peak (Extended Data Fig. 2a) of $CXCR4_{EM}$ using cryoEM. Initial cryoEM data yielded clear top/bottom views of trimeric and tetrameric species, but preferred orientation precluded structure determination. After screening various sample preparation conditions, we ultimately employed stage-tilted data collection[59] to obtain 3.4 Å resolution reconstructions of $CXCR4_{EM}$ homotrimers and homotetramers in complex with REGN7663 Fab (Fig. 5, Extended Data Fig. 7a-j). According to 3D classification, our data contained a roughly 1 to 3 ratio of trimers to tetramer particles (Extended Data Fig. 7k). We did not observe 2D or 3D class averages consistent with dimeric CXCR4, excepting non-physiological antiparallel dimers in our samples prepared in the presence of G_i (Extended Data Fig. 7i). The trimer and tetramer both show CXCR4 subunits arranged symmetrically around a cavity at the central axis, at first glance evoking structural similarity to homomeric ion channels, though CXCR4 has no known channel function. In the case of the CXCR4 oligomers, we found evidence for numerous bound lipids at the central axis in the cryoEM maps (Fig. 5 c,f, Extended Data Fig. 8). Due to matching shape features, we tentatively built three phosphatidic acids and three cholesterol molecules in the trimeric map central cavity and four phosphatidic acids and eight cholesterols in the tetrameric cavity (Extended Data Fig. 8d,h). Although the presumed cholesterol molecules could in principle correspond to exogenously added cholesteryl hemisuccinate present in the purification buffers, the EM density we have modeled as phosphatidic acid strongly resembles a phospholipid, and not the LMNG detergent used for purification. This implies that the central cavity lipids were carried over from the cell membrane and remained stably bound through purification, indicating that the oligomeric structures reported here are representative of species present in the CXCR4-expressing cells used in this study and not an artifact of the purification process. The presence of ordered lipids plugging the central axis of CXCR4 oligomers is reminiscent of microbial channelrhodopsin trimers, though the quarternary arrangement of the 7-TM protomers differs[60,61].

The comparable interprotomer interfaces of trimeric and tetrameric CXCR4 are composed of TM5, TM6, and TM7 of one protomer interacting with TM1 and TM7 of its neighboring protomer (Fig. 5c,f). A ~20° rotation in the angle between neighboring subunits underlies the distinct oligomeric states (Fig. 5h). This oligomeric interface does not overlap with the dimeric interface observed in crystal structures of CXCR4[19,20] (Fig. 5h), speculatively allowing for “super-clustering” of CXCR4 protomers mediated by a combination of trimeric/tetrameric and dimeric interfaces

(Extended Data Fig. 9a,b). Structural superposition indicates that steric clash caused by the T4L fusion in the crystallization construct may have precluded the assembly of trimers or tetramers observed in our data (Extended Data Fig. 9c,d), thus suggesting why homodimer formation was favored for the T4L-fused receptor.

The trimeric interface is characterized by a buried surface area of ~1150 Å^2 and is primarily mediated by crisscrossing of TM6 and TM1 of neighboring protomers near the midpoint of the membrane (Fig. 6a). The diagonal orientation of TM6 results in interprotomer contacts with the cytoplasmic half of TM7. TM1 of the neighboring protomer makes additional interprotomer contacts with cytoplasmic end of TM5 and the extracellular tip of TM7. As expected from interactions between transmembrane helices, most of the residues involved are hydrophobic. As noted above, the tetramer interface is similar to the trimer interface (Fig. 6b). However, close inspection revealed a remarkable difference in the tetramer: a sterol-shaped density that we tentatively built as cholesterol present at the cytoplasmic half of the bilayer sandwiched between TM5/TM6 of one protomer and TM1/TM7 of its neighbor (Fig. 6b,c). To make space for sterol binding at the tetrameric interface, the intracellular portion of TM6 splays away from the interface and TM1 of the neighboring protomer rotates relative to their conformations in the trimeric interface (Fig. 6d). The TM1 rotation is concurrent with the rotation of the entire CXCR4 protomer, which in turn allows space for the additional subunit present in the tetrameric assembly (Fig. 5h). Our structures therefore imply that the absence or presence of lipid at the CXCR4 interprotomer interface may drive the assembly of trimers and tetramers, respectively. These findings provide a structural example supporting the idea that cholesterol regulates chemokine receptor oligomerization[62].

A super-resolution microscopy study reported that the simultaneous introduction of 3 mutations (K239E/V242A/L246A) within TM6, and located at the oligomerization interface observed in our structures, resulted in reduced higher order oligomerization of CXCR4[21]. We used FSEC to examine the effect of this triple mutant and other mutations at the oligomeric interface on the oligomerization behavior of the detergent-extracted receptor, using CXCR4$_{EM}$ as the background construct (Extended Data Fig. 10a). The K239E/V242A/L246A and K239E/V242W/L246W triple mutants both showed a reduced propensity to form oligomers relative to monomers, determined from the FSEC peak-area ratio of oligomer to monomer for each mutant (Extended Data Fig. 10b). We found that the single mutant V242W showed similarly reduced oligomerization, likely by introducing steric hindrance at the oligomerization interface. On the other hand, L246W increased apparent oligomerization and reduced monomer levels, possibly by augmenting the hydrophobic interactions between subunits. A mutation at a TM1 residue (L58W) that faces TM5 of the neighboring subunit also showed reduced oligomer/monomer ratio. Other TM1, TM6, and TM7 mutants showed no significant change in oligomer/monomer ratio (T51W) or did not have clearly interpretable FSEC chromatograms, presumably due to impacts on expression level or stability of the receptor in detergent. Overall, these biochemical data corroborate the oligomeric interface observed in our structural data.

We next examined the conformations of the individual protomers within the CXCR4 trimer and tetramer. As noted above, a striking difference is the kink at TM6 associated with sterol binding (Fig. 6e,f). TM6 of the trimeric protomer is kinked outward relative to that of the tetrameric protomer, suggesting a more active-like conformation. Indeed, the structure of the trimeric protomer matches closely with active CXCR4 in complex with REGN7663 Fab and G$_i$ while the tetrameric protomer aligns well with the inactive CXCR4/REGN7663 Fab complex in the absence of G$_i$ (Fig. 6g,h). A noteworthy distinction between the trimeric CXCR4 protomer and active, G$_i$-bound CXCR4 is the conformation of ICL3, TM7 and H8; in the trimer, ICL3 is pushed away from the cytoplasmic-facing core, C-terminal end of TM7 is tucked inward, effectively blocking G$_i$ binding, and H8 is not visible in the cryoEM map (Fig. 6f,g). Therefore, while trimeric CXCR4 is composed of protomers with an active-like conformation, they are not structurally competent for G$_i$-coupling and as such cannot be deemed fully active. This structural observation agrees with FSEC data showing that the presence of G$_i$ did not result in a shift of the oligomeric peak

(Extended Data Fig. 1b). Overall, these oligomeric structures demonstrate that distinct protomeric conformations underpin the trimeric and tetrameric arrangements of CXCR4. Lipids found at the central axis and at the tetrameric interface appear to be important for oligomeric assembly.

Discussion

A longstanding drug target for HIV, cancer, and immune disorders, CXCR4 is one of the most well-studied chemokine receptors, and was the first to be crystallized. However, critical structures of CXCR4 remained missing. We have presented here a thorough investigation of CXCR4 structure using cryoEM. Our structure of active CXCR4 bound to CXCL12 shows how the chemokine N-terminus buries deep into the orthosteric pocket to activate the receptor. Mutations at the distal CXCL12 N-terminus[40] likely diminish its agonistic activity by disrupting the interactions between chemokine and receptor at the TM domain that are required for activation. Due to the flexibility of the complex, we were unable to resolve interactions between the receptor N-terminus and chemokine (chemokine recognition site 1). Therefore, further studies are necessary to visualize this important determinant of CXCL12/CXCR4 affinity.

Like CXCL12, the FDA-approved drug AMD3100 uses electrostatic interactions, namely between its two positively charged lactam rings and acidic residues in the CXCR4 TM domain, to stabilize a diagonal binding mode. We have also shown how a potent antibody inhibitor, REGN7663, blocks CXCL12 by binding across the extracellular face of CXCR4 and partially inserting its CDR-H3 loop into the orthosteric pocket. The structures of REGN7663/CXCR4 complexes do not provide a clear answer as to why this mAb has apparent inverse agonist activity in the setting of CXCR4 overexpression. Stable binding of REGN7663 to active state CXCR4/G_i, which might be unexpected for an inverse agonist mAb, was possibly enabled by the constitutively active N119S mutation present in our construct that shifts the conformational equilibrium of the receptor. While it is tempting to speculate that inverse agonism is related to interactions between REGN7663 and the CXCR4 TM domains, inverse agonist antibodies raised against the MC4R N-terminus have been reported[63], suggesting TM domain interactions are not a prerequisite for GPCR inverse agonist mAbs.

Though the functional relevance of chemokine receptor oligomerization *in vivo* awaits confirmation, CXCR4 oligomerization has been reported in various experimental settings, including crystal structures of parallel homodimers[17]. In this study, we have observed that detergent-solubilized CXCR4 exists in various oligomeric states, and determined structures of receptor trimers and tetramers. The parallel orientation of the protomers as well as the encapsulation of lipids at the central axis support the notion that these oligomeric species are present at the cell surface of insect cells overexpressing CXCR4 prior to detergent solubilization. Nonetheless, whether these species correspond to cell membrane oligomers observed previously[16,55] or are representative of *in vivo* CXCR4 requires further investigation. Interestingly, super-resolution microscopy experiments implicated three TM6 residues (K239, V242, L246) located at the oligomerization interface observed in our structures as being important for higher order oligomerization but not dimerization of CXCR4 in Jurkat cells[21]. Furthermore, the oligomerization-defective K239E, V242A, L246A triple mutant showed decreased chemotaxis in response to CXCL12 *in vitro*[21]. These previously reported data provide a link between our oligomeric structures of detergent-solubilized receptor and CXCR4 function in T cells.

Finally, we observed that oligomeric state and specifically, the binding of lipid at the oligomeric interface, are correlated with distinct conformations of CXCR4 protomers. While the individual protomers of trimeric CXCR4 exhibited an active-like conformation characterized by outward-kinking of TM6, the positioning of intracellular-facing structural elements (ICL3, TM7, and H8) appear to preclude the docking of G_i. Therefore, additional conformational changes would be required for the oligomeric CXCR4 entities observed here to participate directly in G protein-mediated cellular signaling.

Overall, our structures build on previous crystallographic studies[19,20] to provide a foundation for understanding how peptides, small molecules, chemokines, and antibody bind and affect the function of CXCR4 in diverse ways. Our data also provide a structural perspective on oligomerization as a potential mode of GPCR regulation, adding a layer of complexity to studies that have focused on monomers as the functional units in physiology and disease.

Methods

FSEC-based Construct Screening

Expression constructs (shown in Extended Data Fig. 1a) were codon optimized, synthesized, and cloned into pFastBac1 or pFastbac Dual vectors by Genscript. Second generation baculoviruses (P1) encoding human CXCR4, CXCR4$_{EM,}$ Gα_i, or Gβ_1/Gγ_2 (expressed together using pFastBac Dual) were generated in ExpiSf9 cells (ThermoFisher), titered, and adjusted to approximately 2.5x10^8 ivp/ml. The titering assay was performed using flow cytometry to detect envelope protein gp64 displayed on the surface of infected cells. ExpiSf9 cells at ~5x10^6 cells/ml were infected with either CXCR4 alone (1:11 viral dilution), or with Gα_i (1:22 viral dilution) and Gβ_1/Gγ_2 (1:22 viral dilution). Cells were harvested by centrifugation after 72 hr growth (120 rpm shaking, 27°C, 125 ml flat-bottom flask, Innova 44 shaker). After freeze-thaw (-80°C), cell pellets, each from 1 ml of culture, were resuspended in 200 µl lysis buffer (25 mM Tris pH 7.5, 50 mM NaCl, 2 mM $MgCl_2$, cOmplete (EDTA-free) protease inhibitor, 5 mM $CaCl_2$, 50 mU/mL Apyrase) and rotated at 4°C for 1 hr. For the samples to which G$_i$ was added, G$_i$ containing pellets were first suspended in 200 µl lysis buffer. 200 µl of G$_i$ slurry was then used to resuspend the receptor containing pellets. After 1 hr, 200 µl of solubilization buffer (25 mM Tris pH 7.5, 50 mM NaCl, 2 mM $MgCl_2$, 5 mM $CaCl_2$, ~2% LMNG, ~0.2% CHS, cOmplete (EDTA-free) protease inhibitor, 50 mU/ml Apyrase) was added and the mixture was rotated at 4°C for an additional 1 hr at 4°C. Insoluble material was removed by centrifugation and each sample was subjected to FSEC (buffer: 25 mM Tris pH 7.5, 150 mM NaCl, 2 mM MgCl2, 0.01% LMNG, 0.001% CHS). A Zenix-C SEC-300 3 µM 300 Å 4.6x300mm column (flow rate: 0.35 ml/min) was used for the data shown in Extended Data Fig. 1b. For the data shown in Extended Data Fig. 10, a Zenix-C SEC-300 3 µM 300 Å 7.8x300mm column (flow rate: 0.75 ml/min) was used and the baculovirus used was not titered. FSEC data were collected using a Shimadzu LC system using LabSolutions v5.111 software.

Expression and Purification of CXCR4 and G_i proteins

ExpiSf9 cells at ~5×10^6 cells/ml were infected with P1 baculovirus encoding either CXCR4$_{EM}$ or Gα_I and Gβ_1/Gγ_2 as described above. Cells were harvested by centrifugation (3000 x g, 10 min, 4°C) after 72 hr growth (120 rpm shaking, 27°C, 2 L flat-bottom flask, Innova 44 shaker). Cell pellets were washed in ice-cold DPBS with cOmplete (EDTA-free) protease inhibitor and then subjected to freeze-thaw (-80°C) then resuspended in lysis buffer (25 mM Tris pH 7.5, 50 mM NaCl, 2 mM $MgCl_2$. 1x cOmplete (EDTA-free) protease inhibitor, 5 mM $CaCl_2$, 50 mU/mL Apyrase). Crude lysates containing CXCR4$_{EM}$ and G$_i$ were then combined and stirred at 4°C. After 1 hr, an equal volume (1 ml for every 1 ml of lysis buffer) of solubilization buffer (25 mM Tris pH 7.5, 50 mM NaCl, 2 mM $MgCl_2$, 5 mM $CaCl_2$ 2% LMNG, 0.2% CHS) was added to the slurry and the mixture was stirred at 4°C for 1 hr. Insoluble material was removed by centrifugation (100,000 x g, 4°C, 30 min). Anti-FLAG M2 Affinity Gel (Sigma cat# A2220) was used to capture CXCR4$_{EM}$-containing species. The protein-loaded resin was washed with SEC buffer (25 mM Tris pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 0.01% LMNG, 0.001% CHS) and protein was eluted in SEC buffer containing 0.15 mg/ml 3x FLAG peptide. The eluate was concentrated to approximately 0.5 ml and subjected to SEC. A tandem column was used to improve separation of different CXCR4$_{EM}$ species: a Superose 6 Increase 10/300 GL column was connected upstream of a Superdex 200 Increase 10/300 GL column. Fractions containing CXCR4$_{EM}$/G$_i$ protein complex were selected, pooled, concentrated, and mixed with either Fab', CXCL12, or AMD3100 prior to cryoEM grid making.

A comparable procedure was used for the production CXCR4$_{EM}$ to which G$_i$ was not added. In this case, SEC peaks corresponding to oligomeric and monomeric CXCR4$_{EM}$ were separately harvested and were each mixed with Fab' prior to cryoEM grid making.

Fab' Production

REGN7663 IgG was diluted to 2 mg/ml in 20 mM HEPES pH 7.4, 150 mM NaCl. IdeS, an IgG-specific protease, was added to cleave off Fc region thereby producing F(ab')$_2$. 10 µg concentrated IdeS per 1 mg antibody (1:100) was added and the cleavage reaction was carried out at 37°C for 30 min. F(ab')$_2$ was reduced using approximately 88 mM cysteamine hydrochloride at 37°C for 10 min, in the presence of approximately 18 mM EDTA. Reduced Fab' was dialyzed against 20 mM HEPES pH 7.4, 150 mM NaCl overnight at 4°C. Fab' was further purified by IMAC (negative-pass to remove His-tagged IdeS) and CaptureSelect IgG-Fc (Multispecies) Affinity Matrix (negative-pass to remove Fc fragment.) F(ab') was treated with 20 mM iodoacetamide at room temperature, in the dark, for 30 min to alkylate the reduced hinge cysteines. Fab' was purified further via SEC (HighLoad 16/600 Superdex 75 pg column equilibrated to 25 mM Tris pH 7.5, 150 mM NaCl), and concentrated before use.

CRE-Luciferase CXCR4 functional assay

HEK293 cell lines were generated to stably express full-length human CXCR4 (hCXCR4; amino acids 1-352 of accession number NP_003458.1) along with a luciferase reporter cAMP response element (CRE, 4X)-luciferase-IRES-GFP. For CXCR4 CRE-Luciferase assay, HEK293/CRE-Luc/hCXCR4 cells were plated in Opti-MEM media (Invitrogen, cat# 31985-070) supplemented with 0.1% FBS (Seradigm, Cat#1500-500) at 37°C with 5% CO_2 for overnight. The cells were then incubated with 5uM of Forskolin (Sigma, cat# F6886) and serially diluted CXCL12 (Tocris, Cat# 350-NS) for activation of CXCR4 or pre-incubated with REGN7663 or control antibody for 30 minutes before adding 5uM of Forskolin without or with 500pM of SDF for inhibition of CXCR4 basal activity or SDF-induced CXCR4 activation. Cells were incubated for 5.5 hours at 37°C with 5% CO_2. At the conclusion of the incubations, the luciferase activity was detected using OneGlo (Promega, Cat# E6130) and luminescence was recorded by an EnVision Plate reader using

EnVision Manager v1.14 (Perkin Elmer). Results were analyzed using nonlinear regression (4-parameter logistics) with Prism 6 software (GraphPad) to obtain EC_{50} and IC_{50} values.

CryoEM grid preparation and data collection

$CXCR4_{EM}$ (G_i-bound complex, monomer, or oligomer) were concentrated to ~1 to ~5 mg/mL and left as is (“apo”, G_i-bound complex sample) or mixed with 0.5 mg/mL CXCL12 (Recombinant Human/Rhesus Macaque/Feline CXCL12/SDF-1 alpha, R&D Systems Catalog #: 350-NS-050/CF), or 1 mM AMD3100 (AMD 3100 octahydrochloride, R&D Systems Catalog #: 3299), or ~1 to 1.5 mg/mL REGN7663 Fab and incubated on ice for ~1 hour. Samples were pipetted onto freshly hydrogen/oxygen plasma cleaned UltrAuFoil 0.6/1 300 mesh grids and blotted then plunge frozen into liquid ethane using a Vitrobot Mark IV and stored in liquid nitrogen prior to data collection.

Samples were inserted into a Titan Krios G3i (ThermoFisher) microscope equipped with a BioQuantum K3 (Gatan) imaging system or a Glacios microscope equipped with a Falcon 4i camera and Selectris energy filter (ThermoFisher). Data were collected at nominal magnifications of 105 kx (0.85 Å/pixel) or 165 kx (0.696 Å/pixel) and energy filters were inserted with slit widths of 20 ev and 10 ev on the Titan Krios and Glacios microscopes, respectively. Automated data collections were carried using EPU v2.12 with an applied defocus range of -1.0 to -2.2 μM. A 40° stage tilt was applied during collection of the oligomeric $CXCR4_{EM}$/REGN7663 Fab complex sample to overcome preferred particle orientations. Additional details regarding data collection are shown in Extended Data Table 1.

CryoEM image processing

CryoEM data processing for apo $CXCR4_{EM}/G_i$, $CXCR4_{EM}/G_i$/AMD3100 $CXCR4_{EM}/G_i$/REGN7663Fab, $CXCR4_{EM}$/REGN7663 Fab trimer, and $CXCR4_{EM}$/REGN7663 Fab tetramer was carried out within the cryoSPARC v3.3.2 pipeline[64]. Patch motion correction and

Patch CTF estimation were used to align movie frames and estimate CTF parameters, respectively. Particle images were picked using 2D template based picker or TOPAZ v0.2.5[65] then extracted and subjected to multiple rounds of 2D classification, ab initio reconstruction and heterogeneous refinement to obtain a homogenous subset of particles with well resolved features corresponding to the target complex. Final map calculations were carried out using the Local Refinement job type. C3 and C4 symmetry were applied for refinement of the trimeric and tetrameric reconstructions of $CXCR4_{EM}$/REGN7663 Fab, respectively. Refinements of oligomeric CXCR4 conducted without applied symmetry yielded similar structures to the symmetric refinements, but at lower resolution.

Initial processing steps for the $CXCR4_{EM}/G_i$/CXCL12 and $CXCR4_{EM}$/REGN7663 Fab monomeric complexes were carried out in RELION-3[66]. CTF parameters were calculated using gctf[67] and CTFFIND4[68]. Particles were picked using TOPAZ[65], then sorted by 2D and 3D classification. Initial 3D refinements of the $CXCR4_{EM}/G_i$/CXCL12 complex had very weak density for the ligand. To improve signal for the bound ligand, successive rounds of alignment-free focused 3D classification was conducted, applying a mask around CXCL12. Selected particle images were then subjected to Bayesian polishing and then imported into cryoSPARC for final map refinements. For the $CXCR4_{EM}$/REGN7663 Fab complex, signal from constant region of the Fab was subtracted prior to final local refinement in cryoSPARC. Additional data processing details are listed in Extended Data Table 1.

Model building, structure refinement, and visualization

Model building was initiated by docking starting models into the cryoEM maps using the fit in map function in Chimera[69], followed by rounds of manual adjustment in coot 0.8.9[70] and real space refinement in Phenix 1.19[71]. Published structures of CXCR4 (PDB 4RWS[20]), G_i heterotrimer (PDB 7T2G), and an internal Fab structure were used as initial models to build the $CXCR4_{EM}/G_i$/REGN7663 Fab complex. $CXCR4_{EM}$ and G_i from this structure was then used as starting models for the other structures in this study. A crystal structure of CXCL12 (PDB 3HP3[72]) was used as an initial model for the chemokine. Side chains for CXCL12 residues 13-65 (excepting disulfide bonds) were truncated to C_β due to weak density. The REGN7663 Fab constant regions were omitted from the $CXCR4_{EM}$/REGN7663 Fab (without G_i), $CXCR4_{EM}$/REGN7663 Fab trimer, and $CXCR4_{EM}$/REGN7663 Fab tetramer models due to weak density. The eLBOW program[73] in Phenix was used to generate ligand coordinates and restraints for AMD3100. Structures were validated using Phenix and MolProbity[74]. Buried surface areas were calculated using PISA[75]. Pymol[76], Chimera version1.16[69], and ChimeraX version 1.2.5[77] were used to visualize structural data and generate figures.

Data and materials availability

Regeneron materials described in this manuscript may be made available to qualified, academic, noncommercial researchers through a materials transfer agreement upon request at https://regeneron.envisionpharma.com/vt_regeneron/. For questions about how Regeneron shares materials, use the email address preclinical.collaborations@regeneron.com. Atomic coordinates and cryoEM maps have been deposited into the Protein Data Bank (PDB) and Electron Microscopy Data Bank (EMDB) under the respective accession codes 8U4N and 41888 (Apo $CXCR4_{EM}$/G_i), 8U4O and 41889 ($CXCR4_{EM}$/G_i/CXCL12), 8U4P and 41890 ($CXCR4_{EM}$/G_i/AMD3100), 8U4Q and 41891 ($CXCR4_{EM}$/G_i/REGN7663Fab), 8U4R and 41892 ($CXCR4_{EM}$/REGN7663 Fab), 8U4S and 41893 ($CXCR4_{EM}$/REGN7663 Fab Trimer), 8U4T and 41894 ($CXCR4_{EM}$/REGN7663 Fab Tetramer).

References

1 Hughes, C. E. & Nibbs, R. J. B. A guide to chemokines and their receptors. *FEBS J* **285**, 2944-2971, doi:10.1111/febs.14466 (2018).

2 Busillo, J. M. & Benovic, J. L. Regulation of CXCR4 signaling. *Biochim Biophys Acta* **1768**, 952-963, doi:10.1016/j.bbamem.2006.11.002 (2007).

3 Britton, C., Poznansky, M. C. & Reeves, P. Polyfunctionality of the CXCR4/CXCL12 axis in health and disease: Implications for therapeutic interventions in cancer and immune-mediated diseases. *FASEB J* **35**, e21260, doi:10.1096/fj.202001273R (2021).

4 Guo, F. *et al.* CXCL12/CXCR4: a symbiotic bridge linking cancer cells and their stromal neighbors in oncogenic communication networks. *Oncogene* **35**, 816-826, doi:10.1038/onc.2015.139 (2016).

5 Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872-877, doi:10.1126/science.272.5263.872 (1996).

6 Heusinkveld, L. E., Majumdar, S., Gao, J. L., McDermott, D. H. & Murphy, P. M. WHIM Syndrome: from Pathogenesis Towards Personalized Medicine and Cure. *J Clin Immunol* **39**, 532-556, doi:10.1007/s10875-019-00665-w (2019).

7 Caspar, B. *et al.* CXCR4 as a novel target in immunology: moving away from typical antagonists. *Future Drug Discov* **4**, FDD77, doi:10.4155/fdd-2022-0007 (2022).

8 Donzella, G. A. *et al.* AMD3100, a small molecule inhibitor of HIV-1 entry via the CXCR4 co-receptor. *Nat Med* **4**, 72-77, doi:10.1038/nm0198-072 (1998).

9 DiPersio, J. F., Uy, G. L., Yasothan, U. & Kirkpatrick, P. Plerixafor. *Nat Rev Drug Discov* **8**, 105-106, doi:10.1038/nrd2819 (2009).

10 DiPersio, J. F. *et al.* Phase III prospective randomized double-blind placebo-controlled trial of plerixafor plus granulocyte colony-stimulating factor compared with placebo plus granulocyte colony-stimulating factor for autologous stem-cell mobilization and transplantation for patients with non-Hodgkin's lymphoma. *J Clin Oncol* **27**, 4767-4773, doi:10.1200/JCO.2008.20.7209 (2009).

11 Ghobrial, I. M. *et al.* A Phase Ib/II Trial of the First-in-Class Anti-CXCR4 Antibody Ulocuplumab in Combination with Lenalidomide or Bortezomib Plus Dexamethasone in Relapsed Multiple Myeloma. *Clin Cancer Res* **26**, 344-353, doi:10.1158/1078-0432.CCR-19-0647 (2020).

12 Kashyap, M. K. *et al.* Targeting the CXCR4 pathway using a novel anti-CXCR4 IgG1 antibody (PF-06747143) in chronic lymphocytic leukemia. *J Hematol Oncol* **10**, 112, doi:10.1186/s13045-017-0435-x (2017).

13 Peng, S. B. *et al.* Inhibition of CXCR4 by LY2624587, a Fully Humanized Anti-CXCR4 Antibody Induces Apoptosis of Hematologic Malignancies. *PLoS One* **11**, e0150585, doi:10.1371/journal.pone.0150585 (2016).

14 Gusach, A., Garcia-Nafria, J. & Tate, C. G. New insights into GPCR coupling and dimerisation from cryo-EM structures. *Curr Opin Struct Biol* **80**, 102574, doi:10.1016/j.sbi.2023.102574 (2023).

15 Milligan, G., Ward, R. J. & Marsango, S. GPCR homo-oligomerization. *Curr Opin Cell Biol* **57**, 40-47, doi:10.1016/j.ceb.2018.10.007 (2019).

16 Ward, R. J. *et al.* Chemokine receptor CXCR4 oligomerization is disrupted selectively by the antagonist ligand IT1t. *J Biol Chem* **296**, 100139, doi:10.1074/jbc.RA120.016612 (2021).

17 Munoz, L. M. *et al.* Receptor oligomerization: a pivotal mechanism for regulating chemokine function. *Pharmacol Ther* **131**, 351-358, doi:10.1016/j.pharmthera.2011.05.002 (2011).

18 Sohy, D. *et al.* Hetero-oligomerization of CCR2, CCR5, and CXCR4 and the protean effects of "selective" antagonists. *J Biol Chem* **284**, 31270-31279, doi:10.1074/jbc.M109.054809 (2009).

19 Wu, B. *et al.* Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066-1071, doi:10.1126/science.1194396 (2010).

20 Qin, L. *et al.* Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* **347**, 1117-1122, doi:10.1126/science.1261064 (2015).

21 Martinez-Munoz, L. *et al.* Separating Actin-Dependent Chemokine Receptor Nanoclustering from Dimerization Indicates a Role for Clustering in CXCR4 Signaling and Function. *Mol Cell* **71**, 873, doi:10.1016/j.molcel.2018.08.012 (2018).

22 Guan, X. M., Kobilka, T. S. & Kobilka, B. K. Enhancement of membrane insertion and function in a type IIIb membrane protein following introduction of a cleavable signal peptide. *J Biol Chem* **267**, 21995-21998 (1992).

23 Zhang, W. B. *et al.* A point mutation that confers constitutive activity to CXCR4 reveals that T140 is an inverse agonist and that AMD3100 and ALX40-4C are weak partial agonists. *J Biol Chem* **277**, 24515-24521, doi:10.1074/jbc.M200889200 (2002).

24 von Stetten, D., Noirclerc-Savoye, M., Goedhart, J., Gadella, T. W., Jr. & Royant, A. Structure of a fluorescent protein from Aequorea victoria bearing the obligate-monomer mutation A206K. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**, 878-882, doi:10.1107/S1744309112028667 (2012).

25 Liang, Y. L. *et al.* Dominant Negative G Proteins Enhance Formation and Purification of Agonist-GPCR-G Protein Complexes for Structure Determination. *ACS Pharmacol Transl Sci* **1**, 12-20, doi:10.1021/acsptsci.8b00017 (2018).

26 Zhang, X. *et al.* Evolving cryo-EM structural approaches for GPCR drug discovery. *Structure* **29**, 963-974 e966, doi:10.1016/j.str.2021.04.008 (2021).

27 Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673-681, doi:10.1016/j.str.2006.01.013 (2006).

28 Stephens, B. S., Ngo, T., Kufareva, I. & Handel, T. M. Functional anatomy of the full-length CXCR4-CXCL12 complex systematically dissected by quantitative model-guided mutagenesis. *Sci Signal* **13**, doi:10.1126/scisignal.aay5024 (2020).

29 Ray, P. *et al.* Secreted CXCL12 (SDF-1) forms dimers under physiological conditions. *Biochem J* **442**, 433-442, doi:10.1042/BJ20111341 (2012).

30 Drury, L. J. *et al.* Monomeric and dimeric CXCL12 inhibit metastasis through distinct CXCR4 interactions and signaling pathways. *Proc Natl Acad Sci U S A* **108**, 17655-17660, doi:10.1073/pnas.1101133108 (2011).

31 Liu, K. *et al.* Structural basis of CXC chemokine receptor 2 activation and signalling. *Nature* **585**, 135-140, doi:10.1038/s41586-020-2492-5 (2020).

32 Shao, Z. *et al.* Molecular insights into ligand recognition and activation of chemokine receptors CCR2 and CCR3. *Cell Discov* **8**, 44, doi:10.1038/s41421-022-00403-4 (2022).

33 Shao, Z. *et al.* Identification and mechanism of G protein-biased ligands for chemokine receptor CCR1. *Nat Chem Biol* **18**, 264-271, doi:10.1038/s41589-021-00918-z (2022).

34 Zhang, H. *et al.* Structural basis for chemokine recognition and receptor activation of chemokine receptor CCR5. *Nat Commun* **12**, 4151, doi:10.1038/s41467-021-24438-5 (2021).

35 Wasilko, D. J. *et al.* Structural basis for chemokine receptor CCR6 activation by the endogenous protein ligand CCL20. *Nat Commun* **11**, 3031, doi:10.1038/s41467-020-16820-6 (2020).

36 Ishimoto, N. *et al.* Structural basis of CXC chemokine receptor 1 ligand binding and activation. *Nat Commun* **14**, 4107, doi:10.1038/s41467-023-39799-2 (2023).

37 Yen, Y. C. *et al.* Structures of atypical chemokine receptor 3 reveal the basis for its promiscuity and signaling bias. *Sci Adv* **8**, eabn8063, doi:10.1126/sciadv.abn8063 (2022).

38 Saaber, F. *et al.* ACKR3 Regulation of Neuronal Migration Requires ACKR3 Phosphorylation, but Not beta-Arrestin. *Cell Rep* **26**, 1473-1488 e1479, doi:10.1016/j.celrep.2019.01.049 (2019).

39 Gustavsson, M., Dyer, D. P., Zhao, C. & Handel, T. M. Kinetics of CXCL12 binding to atypical chemokine receptor 3 reveal a role for the receptor N terminus in chemokine binding. *Sci Signal* **12**, doi:10.1126/scisignal.aaw3657 (2019).

40 Crump, M. P. *et al.* Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1. *EMBO J* **16**, 6996-7007, doi:10.1093/emboj/16.23.6996 (1997).

41 Juan, A. B. & Harel, W. in *Receptor Molecular Biology* Vol. 25 (ed C. Sealfon Stuart) 366-428 (Academic Press, 1995).

42 Brelot, A., Heveker, N., Montes, M. & Alizon, M. Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J Biol Chem* **275**, 23736-23744, doi:10.1074/jbc.M000776200 (2000).

43 Wescott, M. P. *et al.* Signal transmission through the CXC chemokine receptor 4 (CXCR4) transmembrane helices. *Proc Natl Acad Sci U S A* **113**, 9928-9933, doi:10.1073/pnas.1601278113 (2016).

44 Wang, J., Tannous, B. A., Poznansky, M. C. & Chen, H. CXCR4 antagonist AMD3100 (plerixafor): From an impurity to a therapeutic agent. *Pharmacol Res* **159**, 105010, doi:10.1016/j.phrs.2020.105010 (2020).

45 Rosenkilde, M. M. *et al.* Molecular mechanism of AMD3100 antagonism in the CXCR4 receptor: transfer of binding site to the CXCR3 receptor. *J Biol Chem* **279**, 3033-3041, doi:10.1074/jbc.M309546200 (2004).

46 Bridger, G. J. *et al.* Synthesis and structure-activity relationships of phenylenebis(methylene)-linked bis-tetraazamacrocycles that inhibit HIV replication. Effects of macrocyclic ring size and substituents on the aromatic linker. *J Med Chem* **38**, 366-378, doi:10.1021/jm00002a019 (1995).

47 Bobkov, V. *et al.* Antibodies Targeting Chemokine Receptors CXCR4 and ACKR3. *Mol Pharmacol* **96**, 753-764, doi:10.1124/mol.119.116954 (2019).

48 Hutchings, C. J., Koglin, M., Olson, W. C. & Marshall, F. H. Opportunities for therapeutic antibodies directed at G-protein-coupled receptors. *Nat Rev Drug Discov* **16**, 787-810, doi:10.1038/nrd.2017.91 (2017).

49 Murphy, A. J. *et al.* Mice with megabase humanization of their immunoglobulin genes generate antibodies as efficiently as normal mice. *Proc Natl Acad Sci U S A* **111**, 5153-5158, doi:10.1073/pnas.1324022111 (2014).

50 Macdonald, L. E. *et al.* Precise and in situ genetic humanization of 6 Mb of mouse immunoglobulin genes. *Proc Natl Acad Sci U S A* **111**, 5147-5152, doi:10.1073/pnas.1323896111 (2014).

51 Ma, Y. *et al.* Structure-guided discovery of a single-domain antibody agonist against human apelin receptor. *Sci Adv* **6**, eaax7379, doi:10.1126/sciadv.aax7379 (2020).

52 Filipek, S. Molecular switches in GPCRs. *Curr Opin Struct Biol* **55**, 114-120, doi:10.1016/j.sbi.2019.03.017 (2019).

53 Hernanz-Falcon, P. *et al.* Identification of amino acid residues crucial for chemokine receptor dimerization. *Nat Immunol* **5**, 216-223, doi:10.1038/ni1027 (2004).

54 Percherancier, Y. *et al.* Bioluminescence resonance energy transfer reveals ligand-induced conformational changes in CXCR4 homo- and heterodimers. *J Biol Chem* **280**, 9895-9903, doi:10.1074/jbc.M411151200 (2005).

55 Hamatake, M. *et al.* Ligand-independent higher-order multimerization of CXCR4, a G-protein-coupled chemokine receptor involved in targeted metastasis. *Cancer Sci* **100**, 95-102, doi:10.1111/j.1349-7006.2008.00997.x (2009).

56 Lao, J. *et al.* Single-Molecule Imaging Demonstrates Ligand Regulation of the Oligomeric Status of CXCR4 in Living Cells. *J Phys Chem B* **121**, 1466-1474, doi:10.1021/acs.jpcb.6b10969 (2017).

57 Stephens, B. & Handel, T. M. Chemokine receptor oligomerization and allostery. *Prog Mol Biol Transl Sci* **115**, 375-420, doi:10.1016/B978-0-12-394587-7.00009-9 (2013).

58 Martinez-Munoz, L., Villares, R., Rodriguez-Fernandez, J. L., Rodriguez-Frade, J. M. & Mellado, M. Remodeling our concept of chemokine receptor function: From monomers to oligomers. *J Leukoc Biol* **104**, 323-331, doi:10.1002/JLB.2MR1217-503R (2018).

59 Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796, doi:10.1038/nmeth.4347 (2017).

60 Tucker, K., Sridharan, S., Adesnik, H. & Brohawn, S. G. Cryo-EM structures of the channelrhodopsin ChRmine in lipid nanodiscs. *Nat Commun* **13**, 4842, doi:10.1038/s41467-022-32441-7 (2022).

61 Morizumi, T. *et al.* Structures of channelrhodopsin paralogs in peptidiscs explain their contrasting K(+) and Na(+) selectivities. *Nat Commun* **14**, 4365, doi:10.1038/s41467-023-40041-2 (2023).

62 Legler, D. F. *et al.* Modulation of Chemokine Receptor Function by Cholesterol: New Prospects for Pharmacological Intervention. *Mol Pharmacol* **91**, 331-338, doi:10.1124/mol.116.107151 (2017).

63 Peter, J. C. *et al.* Antibodies against the melanocortin-4 receptor act as inverse agonists in vitro and in vivo. *Am J Physiol Regul Integr Comp Physiol* **292**, R2151-2158, doi:10.1152/ajpregu.00878.2006 (2007).

64 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

65 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153-1160, doi:10.1038/s41592-019-0575-8 (2019).

66 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, doi:10.7554/eLife.42166 (2018).

67 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12, doi:10.1016/j.jsb.2015.11.003 (2016).

68 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).

69 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

70 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

71 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544, doi:10.1107/S2059798318006551 (2018).

72 Murphy, J. W., Yuan, H., Kong, Y., Xiong, Y. & Lolis, E. J. Heterologous quaternary structure of CXCL12 and its relationship to the CC chemokine family. *Proteins* **78**, 1331-1337, doi:10.1002/prot.22666 (2010).

73 Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **65**, 1074-1080, doi:10.1107/S0907444909029436 (2009).

74 Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S. & Richardson, D. C. New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink "waters," and NGL Viewer to recapture online 3D graphics. *Protein Sci* **29**, 315-329, doi:10.1002/pro.3786 (2020).

75 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).

76 Schrodinger, L. The PyMOL Molecular Graphics System.

77 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).

Acknowledgements

We thank various Regeneron scientists including Yi Zhou, Micah Rapp, and Drew Murphy for discussions, Linda Molla and Samira Chandwani for project management, and Regeneron cloud/HPC teams for supporting cryoEM data storage and processing.

Author contributions

K.S., L.L.M., J.H., M.M., J.H.K. and M.C.F. conceptualized the studies. L.L.M. and T.R. expressed and purified proteins for cryoEM. K.S. conducted cryoEM experiments and analyzed structural data, with contributions from M.C.F.. J.H. and S.S. conducted Cre-Luciferase assays. K.S., L.L.M., J.H., M.M., J.H.K., R.L., W.C.O., and M.C.F. analyzed data and supervised the overall project. K.S. and L.L.M. drafted the manuscript with input from J.H. and M.C.F. The manuscript was finalized by all authors.

Competing interests

Regeneron authors own options and/or stock of the company. W.C.O. is an officer of Regeneron.

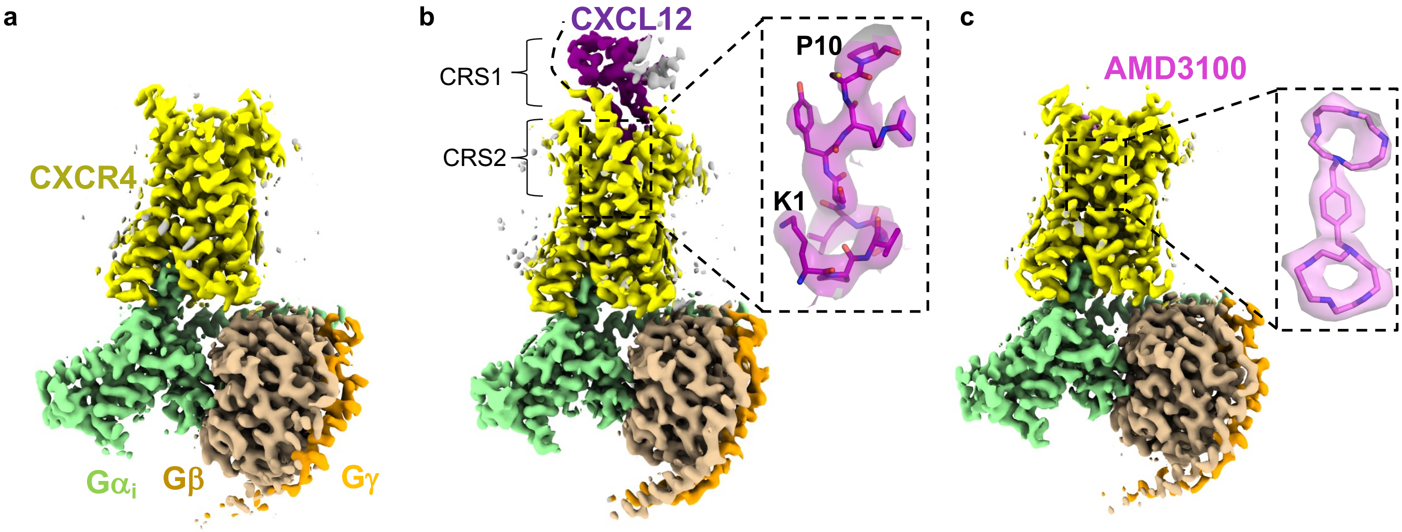


Figure 1. CryoEM reconstructions of CXCR4/G$_i$ complexes. a, apo CXCR4/G$_i$ complex. **b,** CXCR4/G$_i$/CXCL12 complex. Inset shows fit of CXCL12 N-terminal tail (res. 1-10) in cryoEM map, shown as semitransparent surface. Locations of chemokine recognition sites 1 and 2 are labeled. Curved dotted line represents missing density for distal N-terminus of CXCR4, which has been reported to interact with CXCL12. **c**, CXCR4/G$_i$/AMD3100 complex. Inset shows fit of AMD3100 compound in cryoEM map.

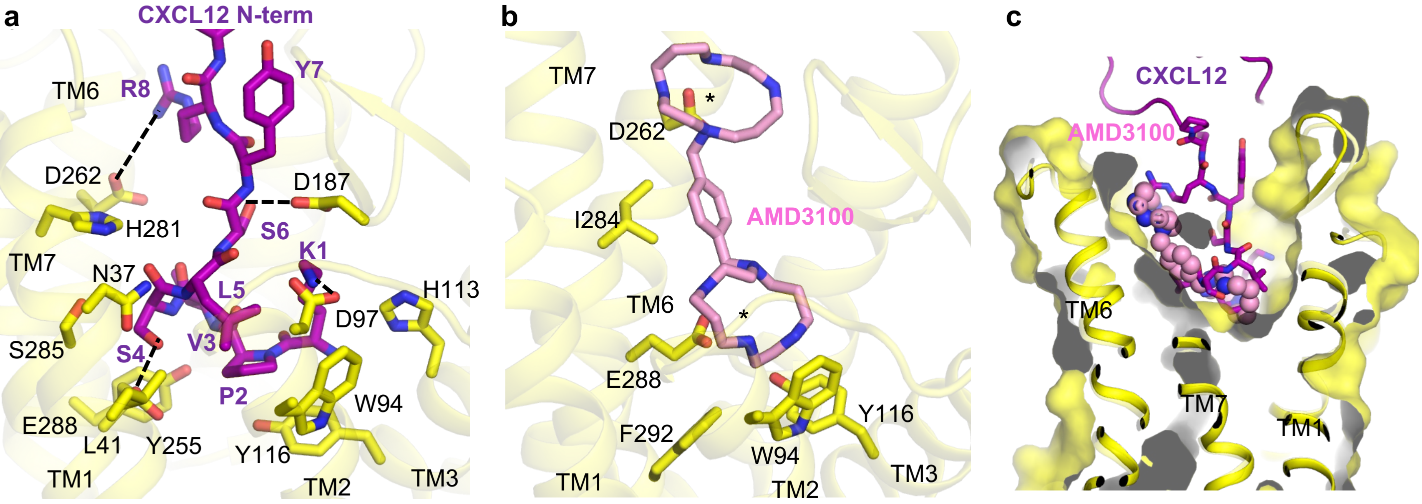


Figure 2. Interactions between CXCR4 and ligands. a, expanded view of interaction between CXCL12 N-terminal tail and CXCR4 orthosteric pocket. Hydrogen-bonding and electrostatic interactions are depicted as dashed lines. **b,** expanded view of AMD3100 binding at CXCR4 orthosteric pocket. Asterisks indicate positions of the two lactam rings, each of which interact with acidic residues. **c,** cutaway surface view of CXCR4 orthosteric pocket. CXCL12 N-term is shown as sticks and AMD3100 is shown as spheres to illustrate their relative binding positions in the orthosteric pocket.

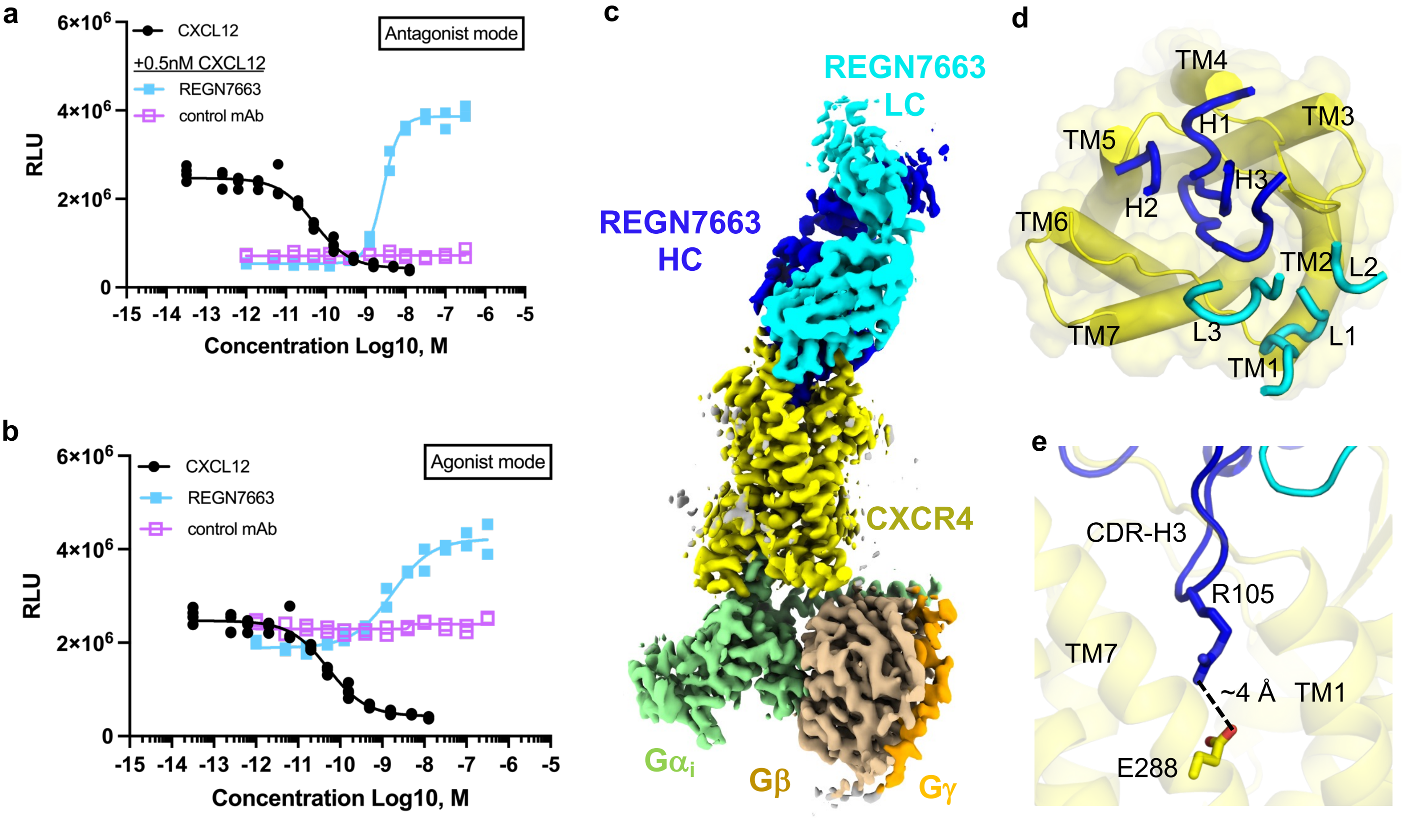


Figure 3. CXCR4 antagonism by REGN7663 mAb. a, CRE-Luciferase reporter assay showing CXCL12-dependent decrease in signal and block of CXCL12 activity (at 0.5 nM CXCL12) by REGN7663. **b,** REGN7663 shows concentration-dependent increase in signal relative to baseline in the absence of CXCL12, demonstrating inverse agonism. In (a) and (b), data from two replicate experiments are shown for the REGN7663 and control mAb, and four replicates are shown for CXCL12 (the same data for CXCL12 are shown in (a) and (b) to allow comparison with mAb data). **c,** cryoEM reconstruction of CXCR4$_{EM}$/G$_i$/REGN7663 Fab complex, with each polypeptide chain colored differently. **d,** top-down view of CXCR4 (yellow) with CDR loops of bound REGN7663 shown (blue = heavy chain, cyan = light chain). **e,** electrostatic interaction between CDR-H3 of REGN7663 and CXCR4 orthosteric pocket-facing residue E288.

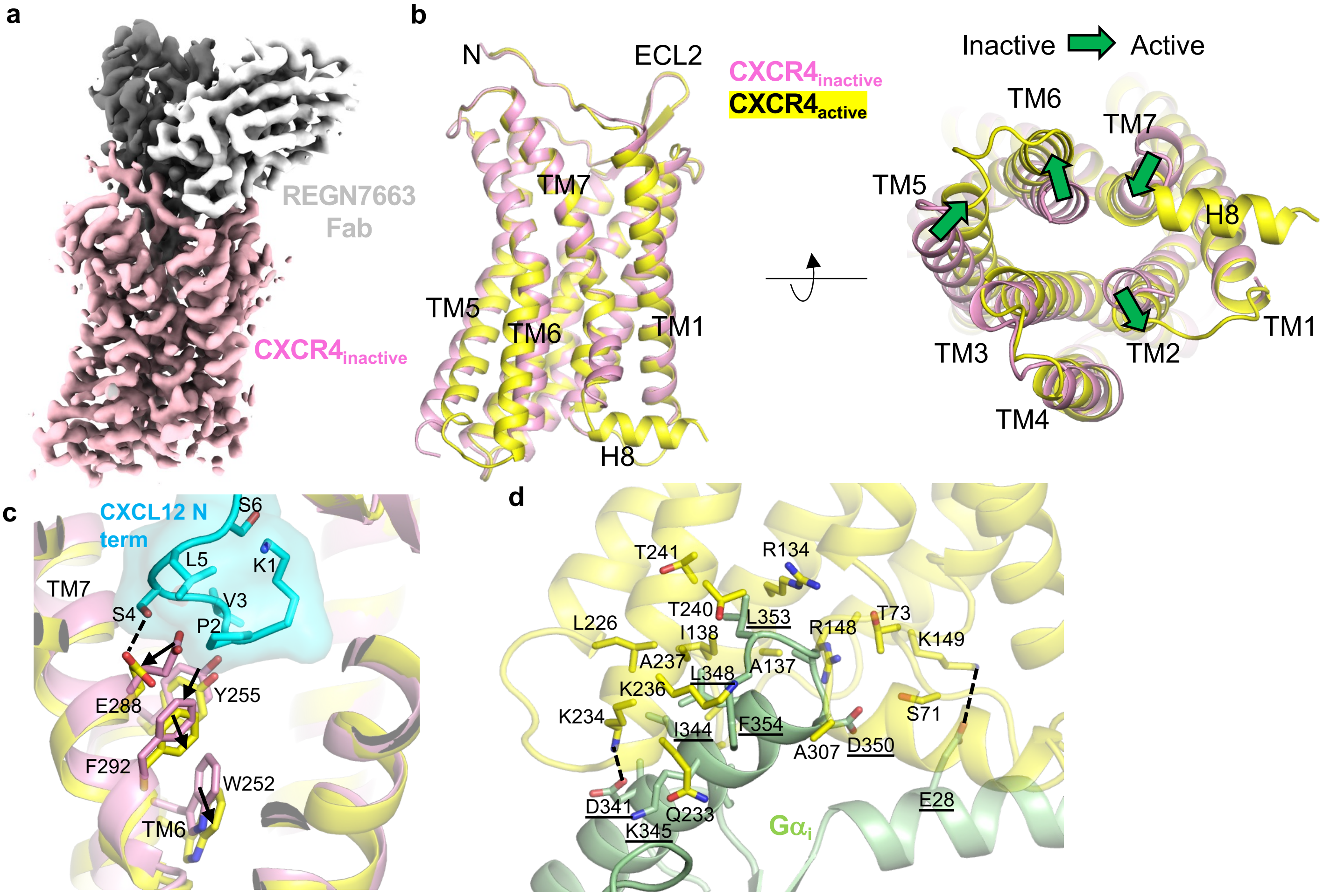


Figure 4. Inactive CXCR4 structure and structural bases of activation. a, cryoEM reconstruction of inactive CXCR4$_{EM}$/REGN7663 Fab complex (CXCR4=pink, REGN7663 heavy chain=gray, REGN7663 light chain=white). **b,** structural alignment of inactive CXCR4 (pink) and active CXCR4 (yellow, CXCR4$_{EM}$/G$_i$/REGN7663 Fab complex was used for alignment). Side view is shown on left and bottom-up view is shown on right. Green block arrows depict conformational transitions from inactive to active CXCR4. **c,** expanded view showing CXCL12 N-term (cyan) binding to active CXCR4 (yellow). Inactive CXCR4 (pink) is shown for comparison and residues important for transmitting chemokine binding into activation are shown as sticks. **d,** expanded view of Gα_i (light green) binding to active CXCR4 (yellow). Residues participating in interaction are shown as sticks and labeled (Gα_i residue labels underlined). Electrostatic interactions are highlighted with dashed lines.

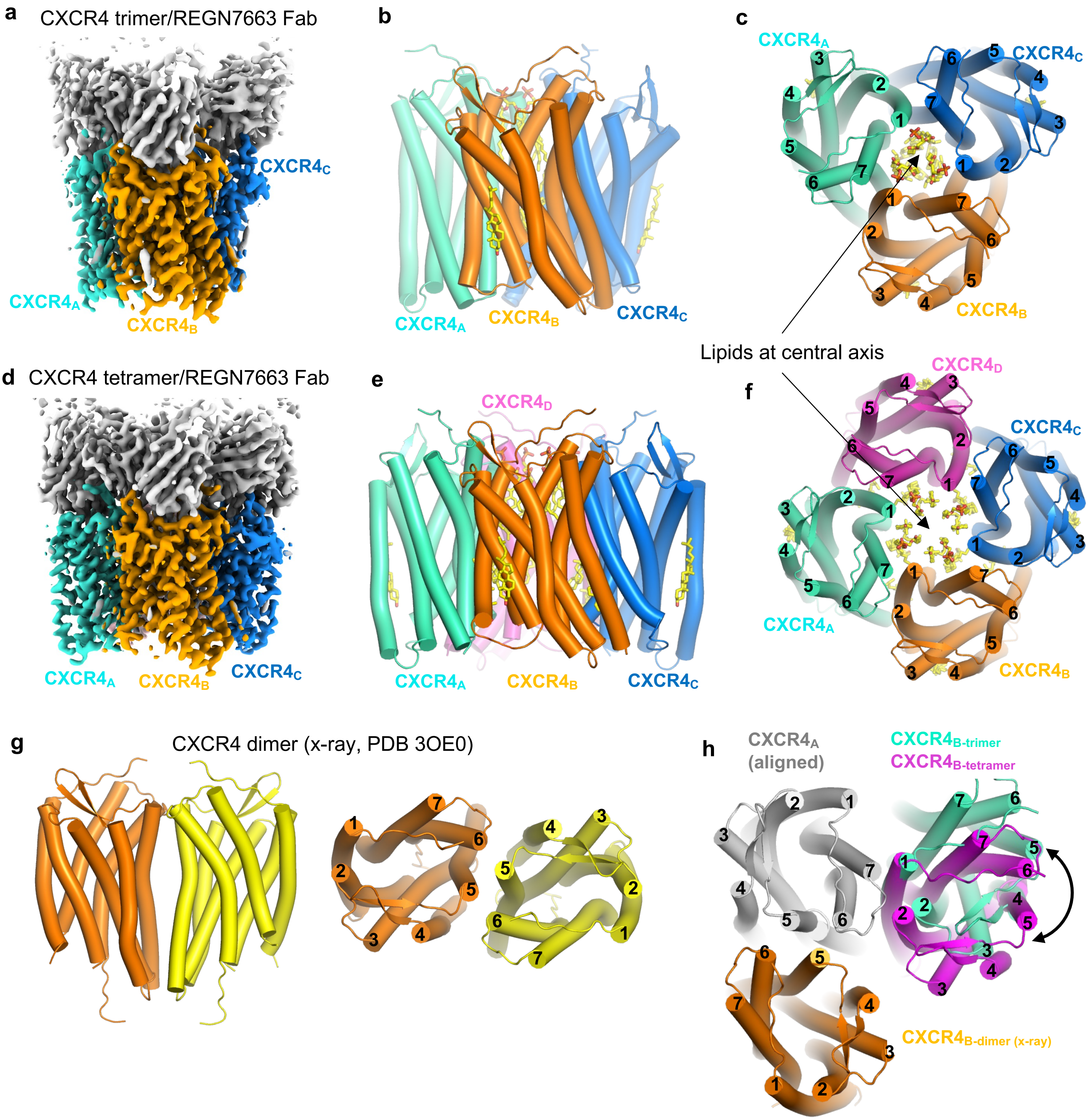


Figure 5. Oligomeric CXCR4 structures. a, cryoEM reconstruction of CXCR4 trimer in complex with REGN7663 Fab. **b,c,** side (b) and top-down (c) views of CXCR4 trimer structure. TM helices are shown as cylinders and bound lipids are shown as sticks**.** Fab molecules are omitted for clarity. **d,** cryoEM reconstruction of CXCR4 tetramer in complex with REGN7663 Fab. **e, f,** side (e) and top-down (f) views of CXCR4 tetramer structure. **g**, side (left) and top (right) views of previously reported dimeric crystal structure of CXCR4. **h**, Top-down view of a CXCR4 protomer (gray) showing positions of neighboring subunits from dimer (orange), trimer (cyan) and tetramer (magenta).

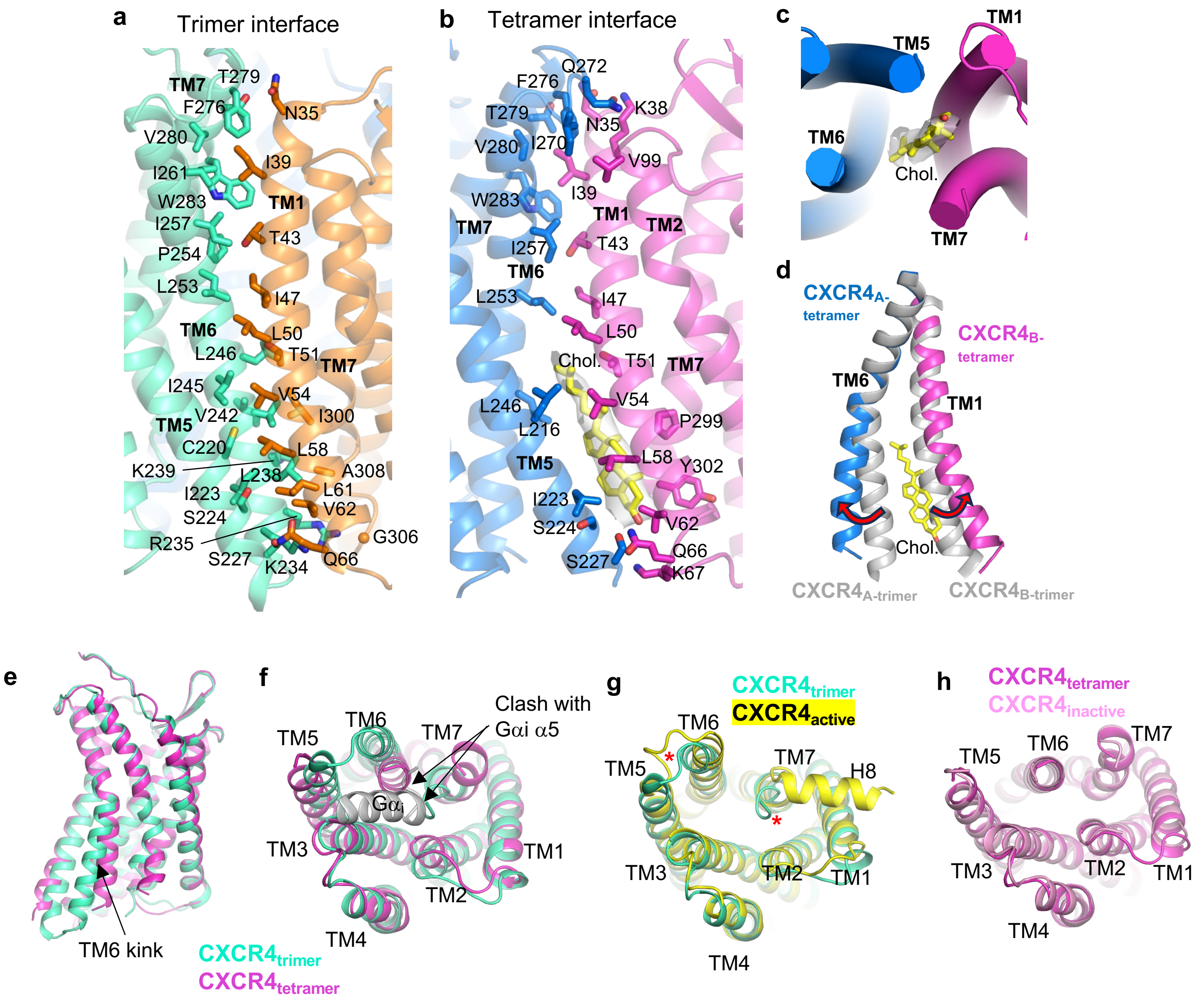


Figure 6. Oligomeric interfaces and protomer conformations. a, interprotomer interface of CXCR4 trimer. Interface residues are shown as sticks and labeled. **b,** interprotomer interface of CXCR4 tetramer. Interface residues and modeled cholesterol shown as sticks. Density corresponding to cholesterol is shown as transparent gray surface. **c,** bottom-up view showing position of cholesterol at the tetramer interface. **d,** structural alignment of TM6 and TM1 at the trimer (gray) and tetramer (blue and magenta, with cholesterol in yellow sticks). **e,f,** side (e) and bottom-up (f) views of protomeric structures of trimeric (cyan) and tetrameric (magenta) CXCR4. Binding of Gα_i α5 helix (gray) is prevented by steric clash. **g,** structural alignment of trimeric CXCR4 protomer (cyan) and active CXCR4 protomer (yellow). Red asterisks highlight the distinct positions of ICL3 and TM7. **h**, structural alignment of tetrameric CXCR4 protomer (magenta) and inactive CXCR4 (pink).

a

CXCR4$_{EM}$ (UniProt ref.: P61073)

MKTIIALSYIFCLVFAGAPEGISIYTSDNYTEEMGSGDYDSMKEPCFREENANFNKIFLPTIYSIIFLTGIVGNGLVILVMGYQKKLR
SMTDKYRLHLSVADLLFVITLPFWAVDAVANWYFGNFLCKAVHVIYTVSLYSSVLILAFISLDRYLAIVHATNSQRPRKLLAEKVV
YVGVWIPALLLTIPDFIFANVSEADDRYICDRFYPNDLWVVVFQFQHIMVGLILPGIVILSCYCIIISKLSHSKGHQKRKALKTTVILIL
AFFACWLPYYIGISIDSFILLEIIKQGCEFENTVHKWISITEALAFFHCCLNPILYAFLGAKFKTSAQHALTSVSRGSSLKILSKGKRG
GHSSVSTESESSSFHSSGRPLEVLFQGPGGGGSVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT
GKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKED
GNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDH
MVLLEFVTAAGITLGMDELYKDYKDDDDK

HA signal peptide – linker – CXCR4 (residues 2-252 + N119S) – linker – 3C site – linker – monomeric eGFP (+H231L) - FLAG

Gα$_{i1\text{-}DM}$ (UniProt ref.: P63096)

MHHHHHHGGGGSGCTLSAEDKAAVERSKMIDRNLREDGEKAAREVKLLLLGAGESGKCTIVKQMKIIHEAGYSEEECKQYKA
VVYSNTIQSIIAIIRAMGRLKIDFGDSARADDARQLFVLAGAAEEGFMTAELAGVIKRLWKDSGVQACFNRSREYQLNDSAAYYL
NDLDRIAQPNYIPTQQDVLRTRVKTTGIVETHFTFKDLHFKMFDVTAQRSERKKWIHCFEGVTAIIFCVALSDYDLVLAEDEEMN
RMHASMKLFDSICNNKWFTDTSIILFLNKKDLFEEKIKKSPLTICYPEYAGSNTYEEAAAYIQCQFEDLNKRKDTKEIYTHFTCST
DTKNVQFVFDAVTDVIIKNNLKDCGLF

Methionine – 6x His – linker - Gα$_{i1\text{-}DN}$ (residues 2-354 + S47C, G202T, G203A, E245A, A326S)

Gβ$_1$ (UniProt ref.: P62873)

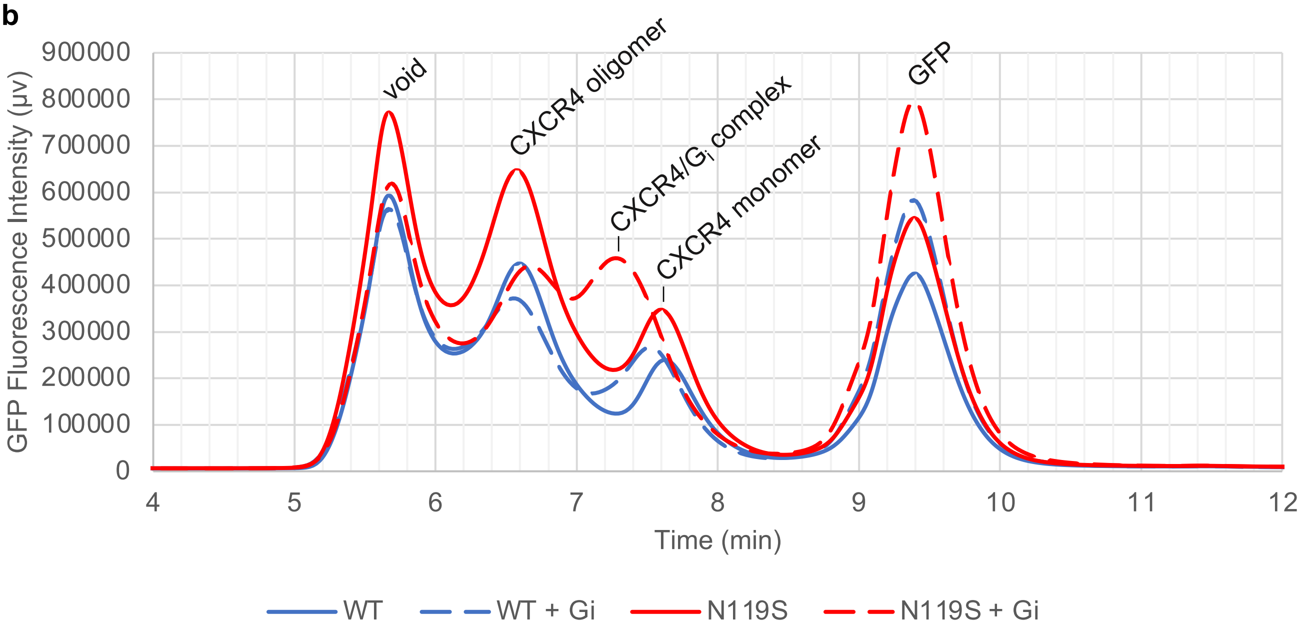
MHHHHHHGSSGSELDQLRQEAEQLKNQIRDARKACADATLSQITNNIDPVGRIQMRTRRTLRGHLAKIYAMHWGTDSRLLVSA
SQDGKLIIWDSYTTNKVHAIPLRSSWVMTCAYAPSGNYVACGGLDNICSIYNLKTREGNVRVSRELAGHTGYLSCCRFLDDNQI
VTSSGDTTCALWDIETGQQTTTFTGHTGDVMSLSLAPDTRLFVSGACDASAKLWDVREGMCRQTFTGHESDINAICFFPNGN
AFATGSDDATCRLFDLRADQELMTYSHDNIICGITSVSFSKSGRLLLAGYDDFNCNVWDALKADRAGVLAGHDNRVSCLGVTD
DGMAVATGSWDSFLKIWN

Methionine – 6x His – Gβ$_1$ (residues 2-340)

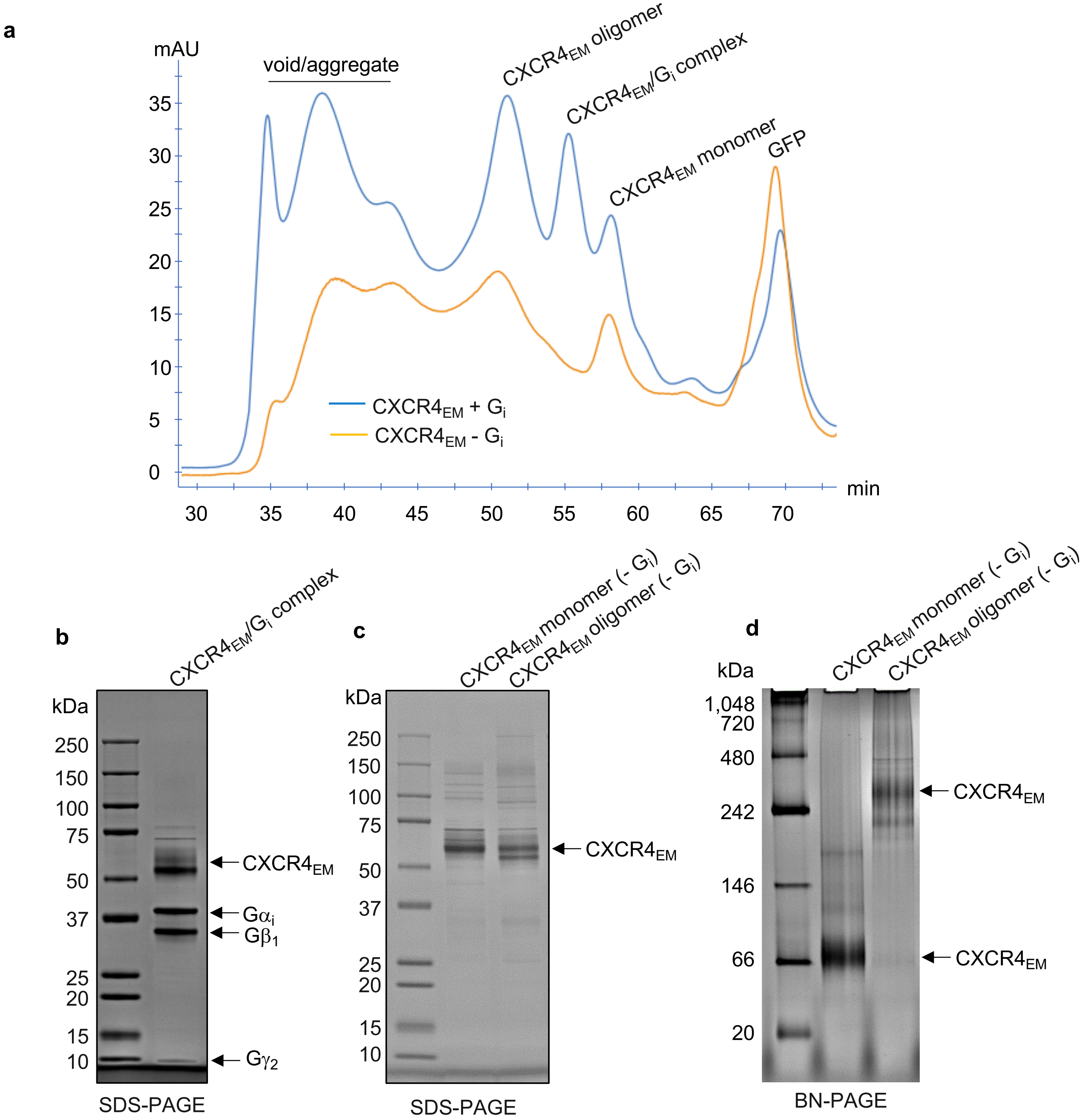
Gγ$_2$ (UniProt ref.: P59768)

MASNNTASIAQARKLVEQLKMEANIDRIKVSKAAADLMAYCEAHAKEDPLLTPVPASENPFREKKFFCAIL

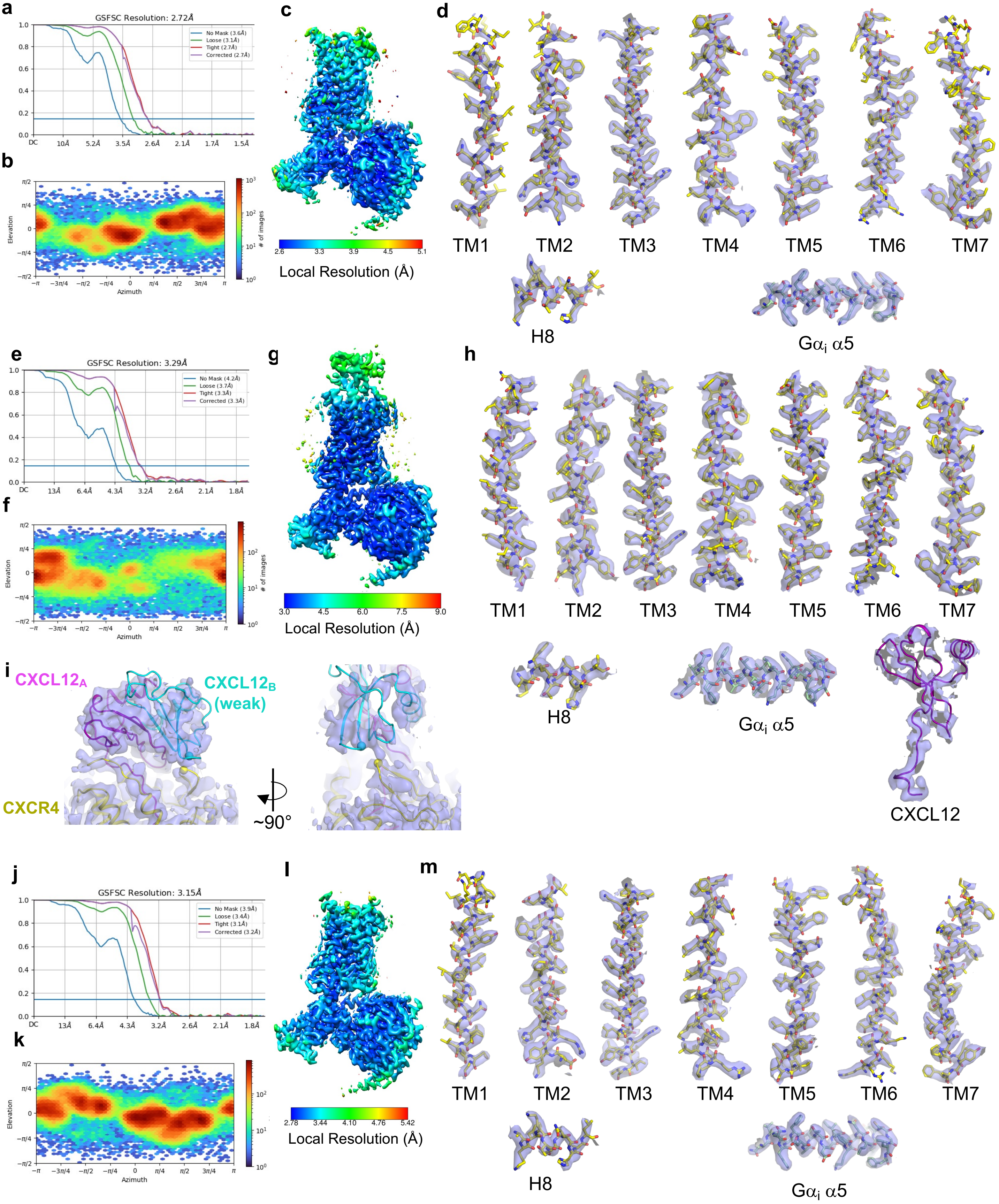
b



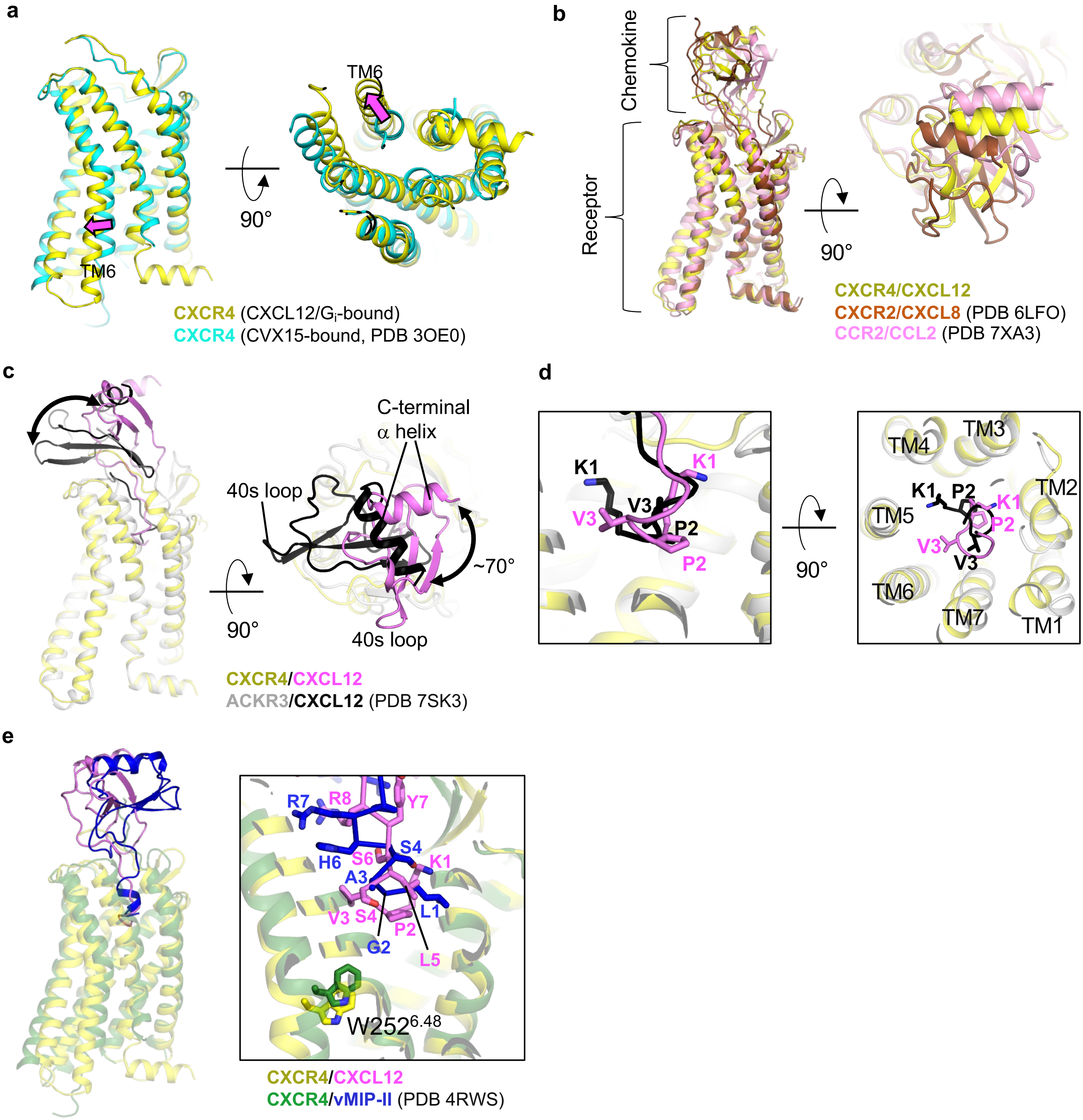
Extended Data Fig. 1. Protein constructs and FSEC-based protein screening. a, Primary structures of protein constructs used in structural studies. **b,** Fluorescence-detection size exclusion chromatography (FSEC) screening of wild type CXCR4 (blue) and N119S-containing CXCR4 (referred to as CXCR4$_{EM}$, red) in the presence (solid lines) or absence (dashed lines) of added G$_i$. The nominally wild type CXCR4 construct was identical to that shown in (a) without the N119S point mutation. Chromatograms are annotated with presumed peak positions of various species present in the samples.



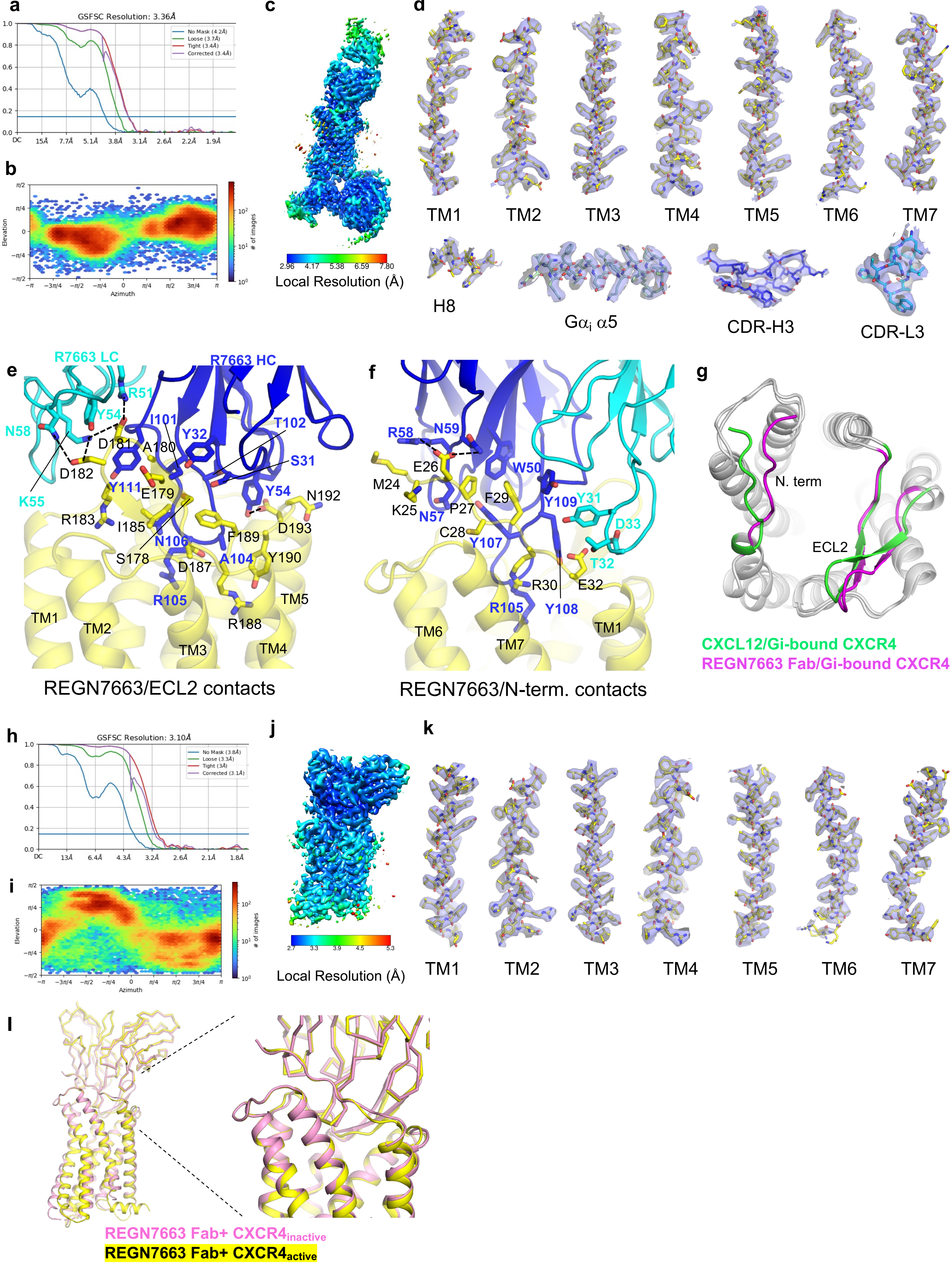
Extended Data Fig. 2. Purification of CXCR4 complexes. a, Size-exclusion chromatography (SEC) traces for a CXCR4$_{EM}$ purification with added G$_i$ (blue), and for a purification prepared in the absence of exogenously added G$_i$ (orange). **b,c** SDS-PAGE (4-20% Tris-Glycine) showing the subunit content and purity of prepared cryoEM samples for CXCR4$_{EM}$/G$_i$ complex (b) and CXCR4$_{EM}$ prepared in the absence of added G$_i$. CXCR4$_{EM}$/G$_i$ complex sample is representative of multiple purifications performed. 2% (v/v) 2-Mercaptoetanol was present in the SDS-PAGE samples prior to loading. **d,** Blue native (BN) PAGE (4-16% Bis-Tris) of SEC-purified CXCR4$_{EM}$ monomer and oligomer samples prepared in the absence of added G$_i$.


a
GSFSC Resolution: 2.72Å
No Mask (3.6Å)
Loose (3.1Å)
Tight (2.7Å)
Corrected (2.7Å)
1.0
0.8
0.6
0.4
0.2
0.0
DC
10Å
5.2Å
3.5Å
2.6Å
2.1Å
1.7Å
1.5Å
b
Elevation
Azimuth
π/2
π/4
0
−π/4
−π/2
−π
−3π/4
−π/2
−π/4
0
π/4
π/2
3π/4
π
10³
10²
10¹
10⁰
# of images
c
2.6
3.3
3.9
4.5
5.1
Local Resolution (Å)
d
TM1
TM2
TM3
TM4
TM5
TM6
TM7
H8
Gαi α5
e
GSFSC Resolution: 3.29Å
No Mask (4.2Å)
Loose (3.7Å)
Tight (3.3Å)
Corrected (3.3Å)
1.0
0.8
0.6
0.4
0.2
0.0
DC
13Å
6.4Å
4.3Å
3.2Å
2.6Å
2.1Å
1.8Å
f
Elevation
Azimuth
π/2
π/4
0
−π/4
−π/2
−π
−3π/4
−π/2
−π/4
0
π/4
π/2
3π/4
π
10²
10¹
10⁰
# of images
g
3.0
4.5
6.0
7.5
9.0
Local Resolution (Å)
h
TM1
TM2
TM3
TM4
TM5
TM6
TM7
H8
Gαi α5
CXCL12
i
CXCL12A
CXCL12B (weak)
CXCR4
~90°
j
GSFSC Resolution: 3.15Å
No Mask (3.9Å)
Loose (3.4Å)
Tight (3.1Å)
Corrected (3.2Å)
1.0
0.8
0.6
0.4
0.2
0.0
DC
13Å
6.4Å
4.3Å
3.2Å
2.6Å
2.1Å
1.8Å
k
Elevation
Azimuth
π/2
π/4
0
−π/4
−π/2
−π
−3π/4
−π/2
−π/4
0
π/4
π/2
3π/4
π
10²
10¹
10⁰
l
2.78
3.44
4.10
4.76
5.42
Local Resolution (Å)
m
TM1
TM2
TM3
TM4
TM5
TM6
TM7
H8
Gαi α5

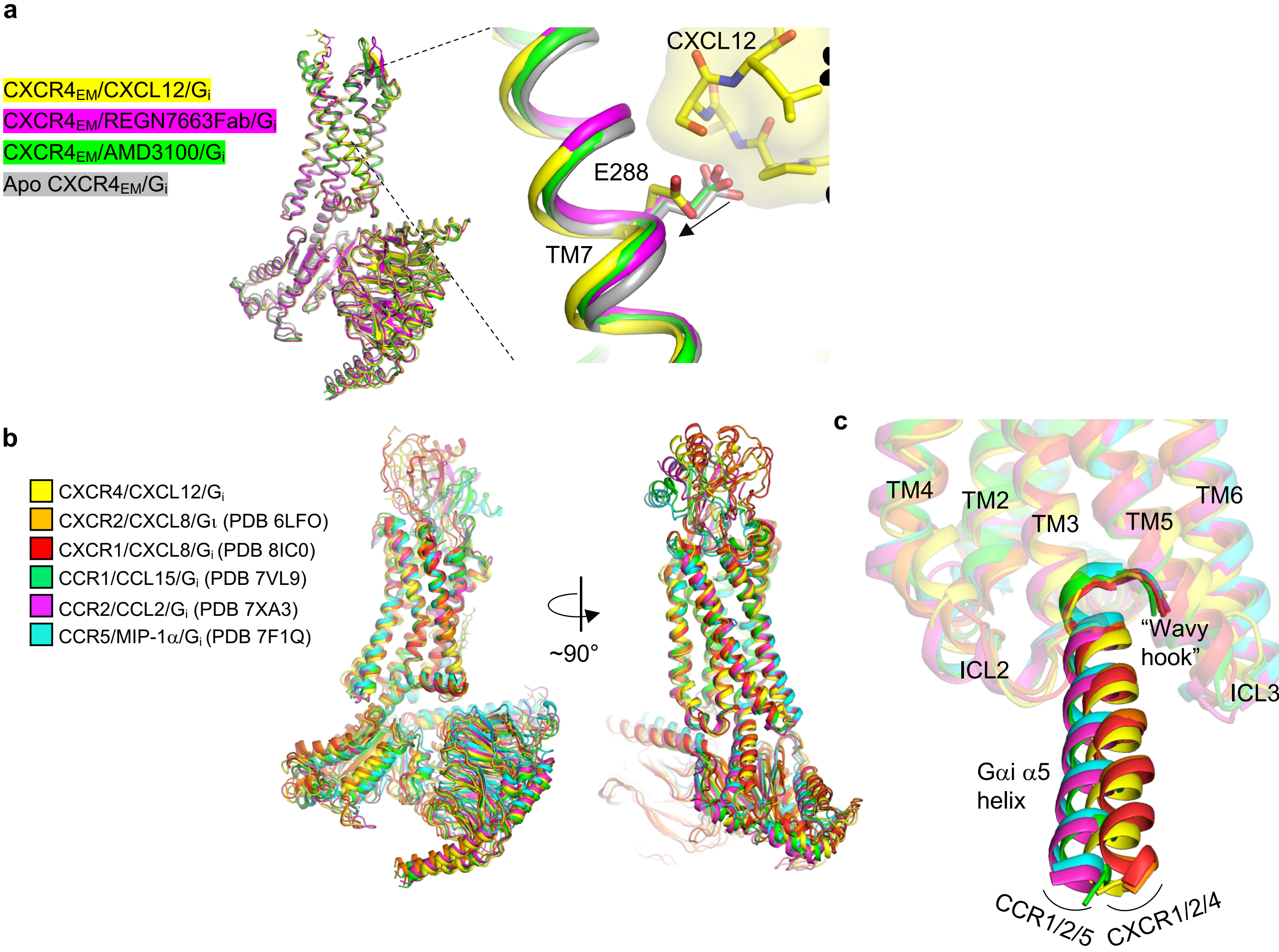
Extended Data Fig. 3. CryoEM reconstruction of Apo CXCR4$_{EM}$/G$_i$, CXCL12-bound CXCR4$_{EM}$/G$_i$, AMD3100-bound CXCR4$_{EM}$/G$_{i.}$ a-d, FSC curve (a), particle angular distribution plot (b), local resolution map calculated in cryoSPARC (c), and map/model fits of selected regions (d) for Apo CXCR4$_{EM}$/G$_i$. **e-h**, FSC curve (e), particle angular distribution plot (f), local resolution map calculated in cryoSPARC (g), and map/model fits of selected regions (h) for CXCL12-bound CXCR4$_{EM}$/G$_i$. **i**, two views showing fit of a CXCL12 dimer (arranged on the basis of PDB 3GV3) into cryoEM map, shown at 4 sigma in pymol. **j-m,** FSC curve (j), particle angular distribution plot (k), local resolution map calculated in cryoSPARC (l), and map/model fits of selected regions (m) for AMD3100-bound CXCR4$_{EM}$/G$_i$.



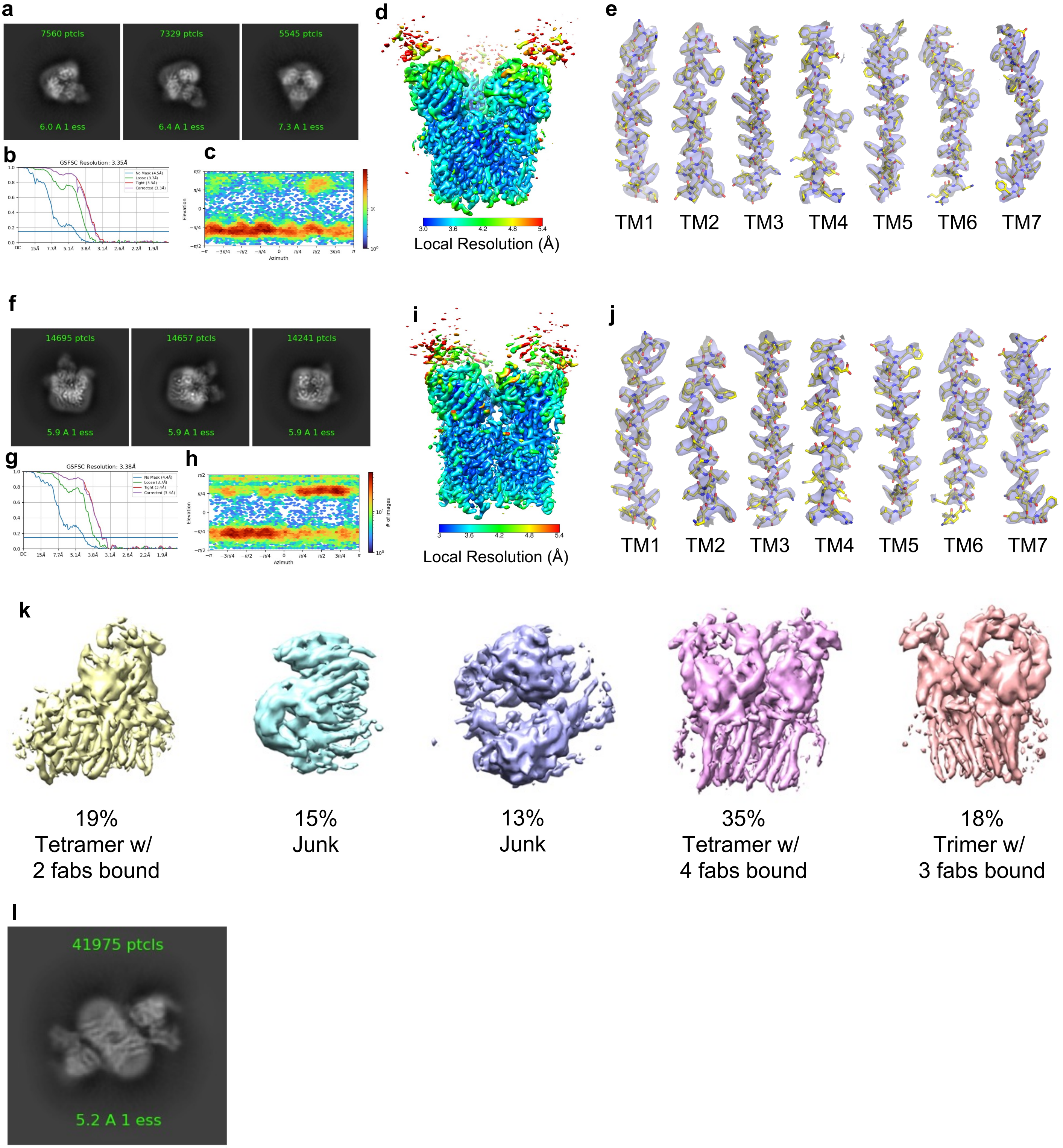
Extended Data Fig. 4. Structural comparisons of chemokine receptor structures. a, structural alignment of active CXCR4 (yellow, this study, CXCL12/G$_i$-bound complex) and inactive, antagonist-bound CXCR4 (cyan, PDB 3OE0). Magenta block arrows depict movement of TM6. **b**, Alignment of CXCR4/CXCL12 complex (yellow) with CXCR2/CXCL8 complex (brown, PDB 6LFO), and CCR2/CCL2 complex (pink, PDN 7XA3). G protein models are omitted for clarity. **c**, Receptor-based alignment of CXCR4/CXCL12 (yellow/pink) with ACKR3/CXCL12 (gray/black, PDB 7SK3). Arrow highlights different docking orientations of CXCL12 onto the two receptors. **d**, expanded views of showing different binding modes of CXCL12 N-termini (pink in CXCR4 complex and black in ACKR3 complex) in CXCR4 and ACKR3. **e**, alignment of CXCR4/CXCL12 complex (yellow/pink) and CXCR4/vMIP-II (green/blue). Inset shows expanded views of chemokine N-terminal positions within orthosteric pocket, highlighting the positions of toggle switch residue W252 in sticks.


a
GSFSC Resolution: 3.36Å
No Mask (4.2Å)
Loose (3.7Å)
Tight (3.4Å)
Corrected (3.4Å)
DC 15Å 7.7Å 5.1Å 3.8Å 3.1Å 2.6Å 2.2Å 1.9Å
b
Elevation
Azimuth
# of images
c
2.96 4.17 5.38 6.59 7.80
Local Resolution (Å)
d
TM1 TM2 TM3 TM4 TM5 TM6 TM7
H8
Gα$_i$ α5
CDR-H3
CDR-L3
e
R7663 LC
R7663 HC
R51
Y54
N58
K55
I101
Y32
T102
S31
D181
A180
D182
Y111
E179
Y54
N192
R183
I185
D193
F189
N106
S178
Y190
D187
A104
R105
TM5
TM1
TM2
R188
TM3
TM4
REGN7663/ECL2 contacts
f
R58
N59
E26
M24
W50
K25
N57
P27
F29
Y109
Y31
D33
C28
Y107
T32
R30
E32
R105
Y108
TM6
TM7
TM1
REGN7663/N-term. contacts
g
N. term
ECL2
CXCL12/Gi-bound CXCR4
REGN7663 Fab/Gi-bound CXCR4
h
GSFSC Resolution: 3.10Å
No Mask (3.8Å)
Loose (3.3Å)
Tight (3Å)
Corrected (3.1Å)
DC 13Å 6.4Å 4.3Å 3.2Å 2.6Å 2.1Å 1.8Å
i
Elevation
Azimuth
# of images
j
2.7 3.3 3.9 4.5 5.3
Local Resolution (Å)
k
TM1 TM2 TM3 TM4 TM5 TM6 TM7
l
REGN7663 Fab+ CXCR4$_{inactive}$
REGN7663 Fab+ CXCR4$_{active}$

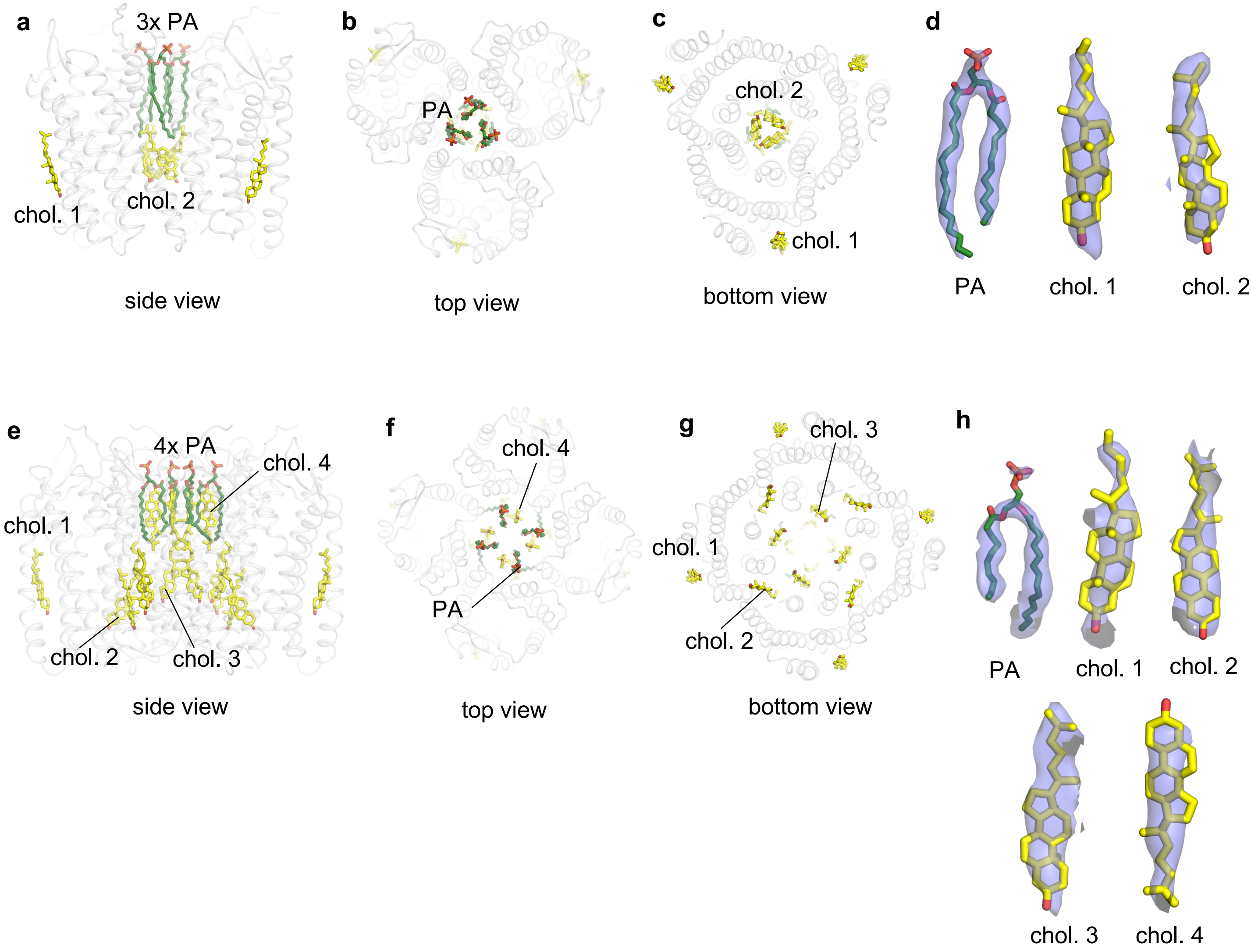
Extended Data Fig. 5. CryoEM reconstruction of REGN7663 Fab/CXCR4$_{EM}$/G$_i$ and REGN7663 Fab/CXCR4$_{EM}$. a-d, FSC curve (a), particle angular distribution plot (b), local resolution map calculated in cryoSPARC (c), and map/model fits of selected regions (d) for REGN7663 Fab/CXCR4$_{EM}$/G$_i$. **e,f,** Expanded view of contacts between REGN7663 Fab (light chain in cyan, heavy chain in blue) and CXCR4 ECL2 (e) and N-term (f). Epitope and paratope residues are shown as sticks and labeled, and apparent salt bridges/hydrogen bonds between mAb and receptor are shown as dashed lines. **g**, structural alignment of CXCR4 bound to CXCL12 and REGN7663 Fab. N-term. and ECL2 are colored green (CXCL12-bound) or magenta (REGN7663 Fab-bound) to highlight their different positions. **h-k**, FSC curve (h), particle angular distribution plot (i), local resolution map calculated in cryoSPARC (j), and map/model fits of TM helices (k) for REGN7663 Fab/CXCR4$_{EM}$ without G$_i$. **i**, aligned structures of CXCR4/REGN7663 Fab complex in the inactive (pink) and active (yellow, Gi-bound) conformations. Note the REGN7663 Fab variable region and cytoplasmic half of the CXCR4 domain are mostly superimposable.



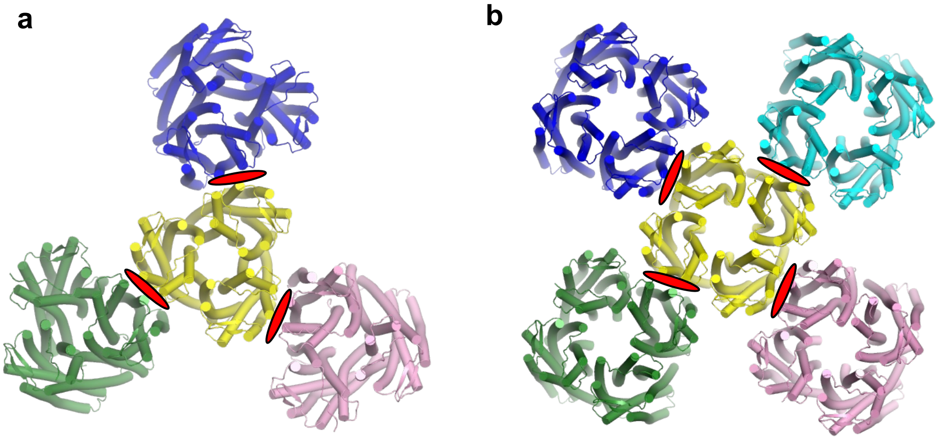
Extended Data Fig. 6. Gα$_i$ interactions in chemokine receptors. a, alignment of CXCR4/Gi complexes bound to CXCL12 (yellow), REGN7663 Fab (magenta), AMD3100 (green) or in the absence of ligand (apo, gray). Inset shows expanded view around E288 residue. Bound CXCL12 is shown as yellow transparent surface and sticks, highlighting how it enforces a rotameric change of E288 and slight shift of TM7 in the CXCL12-bound complex. **b,** receptor-based alignment showing architecture of various chemokine/chemokine receptor/Gi complexes. **c,** expanded view showing docking of Gαi α5 helix into cytoplasmic pocket of chemokine receptors. Note that the Gαi α5 helix is positioned closer to ICL2 in CC chemokine receptor complexes, while it is closer to ICL3 in CXC chemokine receptor complexes.



Extended Data Fig. 7. CryoEM of CXCR4 oligomers. a-e, example 2D class averages obtained from tilted data collection (a), FSC curves (b), angular distribution plot (c), local resolution map (d) and model/map fits of TM helices (e) of trimeric CXCR4$_{EM}$/REGN7663 Fab complex. **f-j,** example 2D class averages obtained from tilted data collection (f), FSC curves (g), angular distribution plot (h), local resolution map (i) and model/map fits of TM helices (j) of tetrameric CXCR4$_{EM}$/REGN7663 Fab complex. **k**, Output maps from ab initio reconstruction conducted on the oligomeric CXCR4$_{EM}$/REGN7663 Fab particles. Particles belonging to classes of tetramer with 4 fabs bound or trimer with 3 fabs bound were selected for further processing. **l**, An example 2D class average showing an anti-parallel dimer of CXCR4$_{EM}$/G$_i$.

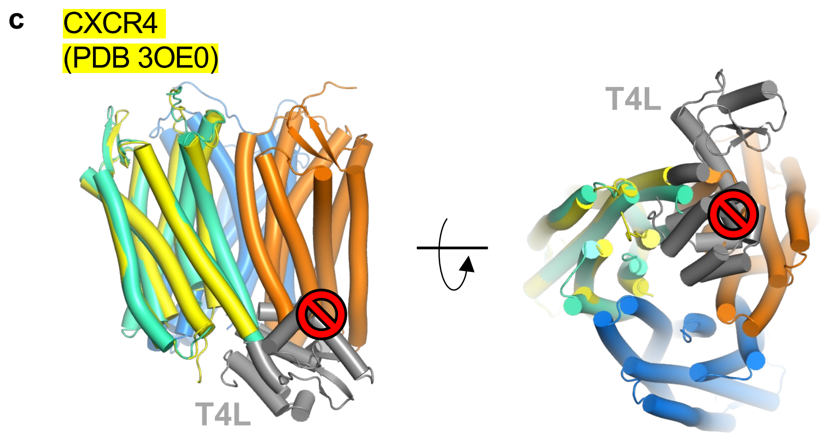


Extended Data Fig. 8. Lipids resolved in oligomeric structures of CXCR4. a-c, side (a), top-down (b) and bottom-up (c) views of trimeric CXCR4, highlighting positions of built lipid molecules. Cholesterols (chol.) are shown as yellow sticks and phosphatidic acid (PA) are shown as green sticks. **d,** fit of lipid molecules (shown as sticks) to map (transparent blue surface) in trimeric CXCR4. Chol. 1 and chol 2. refer to lipids labeled in a-c**. e-g,** side (e), top-down (f) and bottom-up (g) views of tetrameric CXCR4, highlighting positions of built lipid molecules. **h,** fit of lipid molecules (shown as sticks) to map (transparent blue surface) in tetrameric CXCR4.

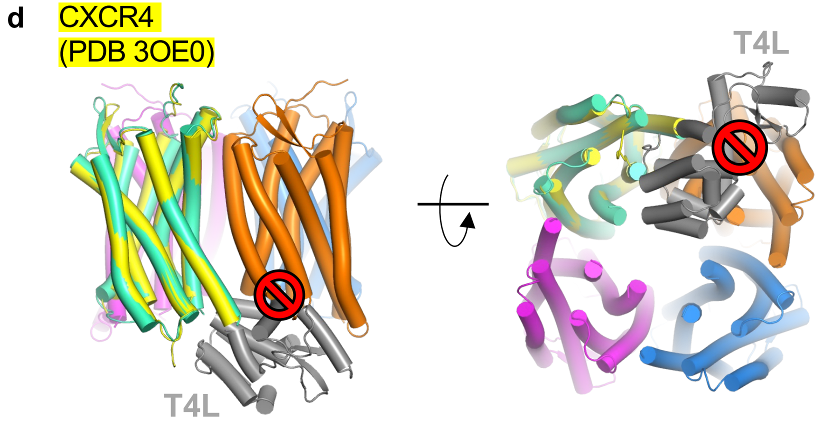


Hypothetical clustering of CXCR4 trimers

Hypothetical clustering of CXCR4 tetramers

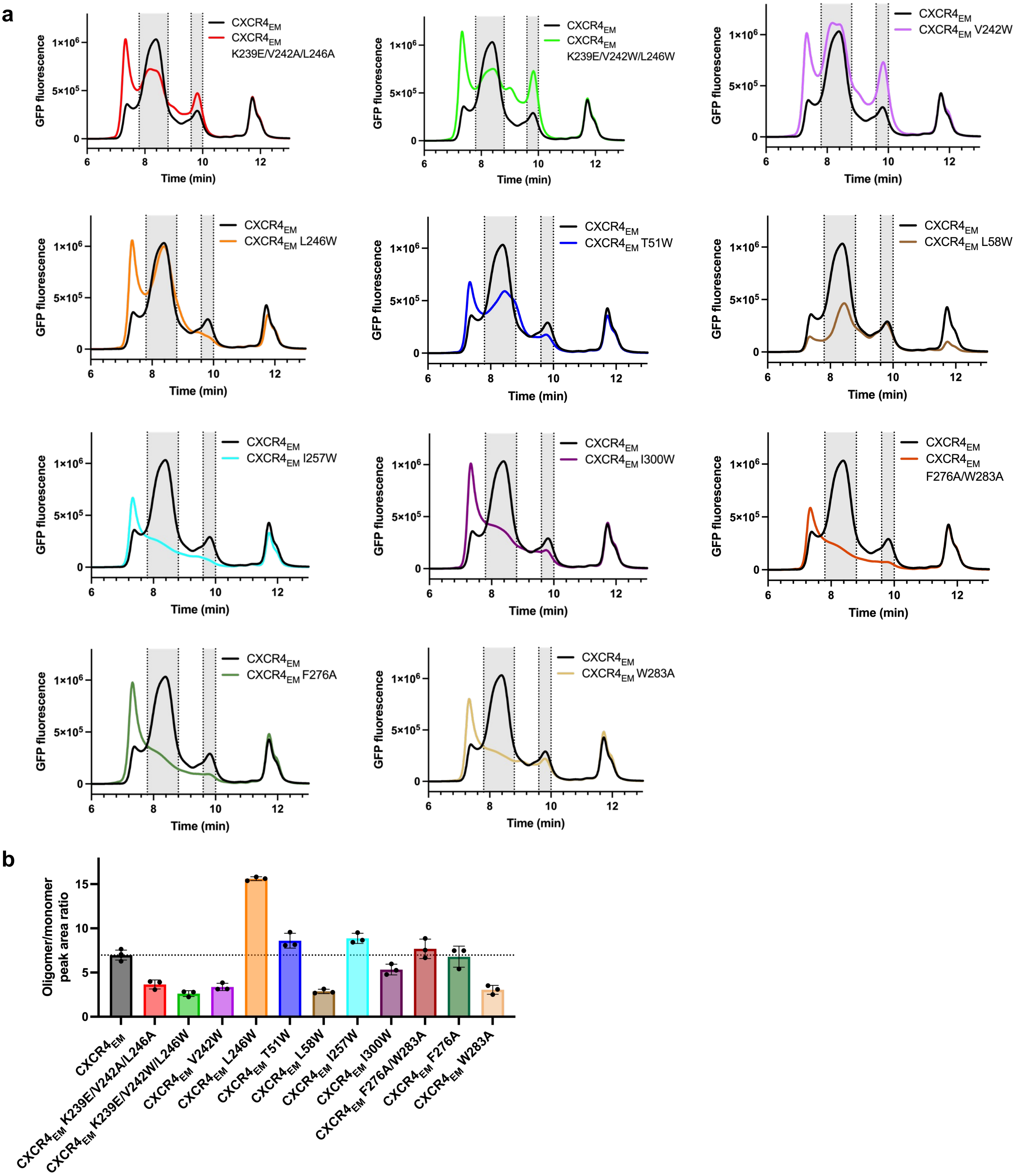


Alignment of CXCR4-T4L with trimeric CXCR4$_{EM}$



Alignment of CXCR4-T4L with tetrameric CXCR4$_{EM}$

Extended Data Fig. 9. Structural analysis of dimeric and oligomeric structures of CXCR4. a,b, top-down view of hypothetical models of four CXCR4 trimers (a) or five CXCR4 tetramers (b) clustered via dimeric interfaces (red ovals) observed in crystal structures. **c,d,** a single subunit from dimeric x-ray structure of CXCR4 (receptor in yellow, ICL3-fused T4L in gray) aligned to trimeric CXCR4 (c) or tetrameric CXCR4 (d). Side and top views are shown. Red symbols indicate steric clash between T4L and neighboring protomers that would prevent trimer or tetramer assembly. The steric hindrance caused by fused T4L may explain why dimeric CXCR4 was favored over trimeric or tetrameric CXCR4 in previous crystallographic studies.



Extended Data Fig. 10. FSEC analysis of CXCR4 oligomerization interface mutants. a, FSEC chromatograms of control construct CXCR4$_{\mathrm{EM}}$ (black trace) and its mutants (colored traces) tracking GFP fluorescence. The same chromatogram for the control are shown in each for comparison. Gray shaded regions indicate elution times corresponding to CXCR4 oligomer (7.8-8.8 min) and monomer (9.6-10 min). **b**, ratio of oligomer to monomer peak areas, calculated according to shaded regions in **a**. horizontal dotted line corresponds to mean value for CXCR4$_{\mathrm{EM}}$. Column heights indicate mean values, and error bars show standard deviations calculated from N=3 or 4 FSEC experiments using two independently generated baculoviruses for each construct. Note that several mutants (I257W, I300W, F276A/W283A, F276A, W283A) showed poor chromatographic behavior overall, presumably due to poor expression or stability in detergent.

	CXCR4$_{EM}$/ G$_i$	CXCR4$_{EM}$/ G$_i$/CXCL12	CXCR4$_{EM}$/ G$_i$/ AMD3100	CXCR4$_{EM}$/ G$_i$/ REGN7663 Fab	CXCR4$_{EM}$/ REGN7663 Fab	CXCR4$_{EM}$/ REGN7663 Fab Trimer	CXCR4$_{EM}$/ REGN7663 Fab Tetramer
Data collection and processing							
Magnification	165,000	105,000	105,000	105,000	105,000	105,000	105,000
Voltage (kV)	200	300	300	300	300	300	300
Electron exposure (e^{-}/Å^2)	~50	~40	~40	~40	~40	~40	~40
Defocus range	-1.0 to -2.0	-1.2 to -2.2	-1.2 to -2.2	-1.2 to -2.2	-1.2 to -2.2	-1.2 to -2.2	-1.2 to -2.2
Pixel size (Å)	0.696	0.85	0.85	0.85	0.85	0.85	0.85
# of Movies	9,693	11,922	10,013	10,346	5,827	5,329	5,329
Initial number of particles	5.0M	13.0M	8.1M	7.1M	6.2M	3.0M	3.0M
Particles selected after 2D classification	615K	805K	803K	1.03M	639K	585K	585K
Final selected particles	183,399	87,963	193,216	148,703	102,810	27,104	31,775
Symmetry imposed	C1	C1	C1	C1	C1	C3	C4
Map resolution (Å)	2.72	3.29	3.15	3.36	3.10	3.35	3.38
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Refinement							
Initial Model used	This study	3HP3, this study	This study	4RWS, 7T2G	This study	This study	This study
Model composition							
Non-hydrogen atoms	6,963	7,401	7,006	10,409	4,107	12,684	16,968
Protein residues	889	966	892	1,339	513	1,557	2,056
Ligands	1	1	2	1	1	9	20
R.m.s. deviations							
Bond lengths (Å)	0.005	0.002	0.002	0.005	0.002	0.003	0.004
Bond angles (°)	0.653	0.493	0.457	0.642	0.629	0.715	0.700
Validation							
MolProbity score	1.74	1.56	1.41	1.69	1.50	1.55	1.75
Rotamer outliers (%)	0.00	0.00	0.00	0.00	0.00	0.67	0.00
Clash score	8.65	5.56	4.24	6.54	4.99	7.06	8.53
Ramachandran plot							
Favored (%)	96.02	96.12	96.71	95.31	96.42	97.08	95.85
Allowed (%)	3.98	3.88	3.29	4.46	3.58	2.92	4.15
Disallowed (%)	0.00	0.00	0.00	0.23	0.00	0.00	0.00
Deposition ID							
PDB	8U4N	8U4O	8U4P	8U4Q	8U4R	8U4S	8U4T
EMDB	41888	41889	41890	41891	41892	41893	41894

Extended Data Table 1. CryoEM data, structure refinement, and validation. Note that CXCR4$_{EM}$/REGN7663 Fab trimer and tetramer structures were obtained from the same dataset.