Supporting Online Materials

Materials and Methods

Molecular biology for generation of *Spodoptera frugiperda* (Sf9) and mammalian expressed CXCR4-WT and CXCR4-T4L-ΔC constructs. The CXCR4-WT DNA was synthesized by DNA2.0 with flanking restrictions sites AscI at the 5' end and FseI at the 3' end. The expression vector designated pFastBac1-830220 is a modified pFastBac1 vector (Invitrogen) containing an expression cassette with an HA signal sequence followed by a FLAG tag at the N-terminus and a PreScission protease site followed by a 10xHis tag at the C-terminus. The components of the expression cassette were introduced using standard PCR based site-directed mutagenesis. The expression cassette also contains restriction sites for AscI and FseI allowing for the standard restriction digest and subsequent ligation of the synthesized CXCR4-WT DNA.

The CXCR4-1 gene, based on wild type CXCR4 and T4L sequences, included the following five additional features: (a) Asn2-Tyr161 of T4L were inserted between Ser229 and Lys230 within the CXCR4 ICL3 region. (b) A two amino acid Gly-Ser linker was inserted at both termini of T4L (T4L1). (c) C-terminal residues 326-352 of CXCR4 were truncated (dC326). (d) Two exogenous restriction sites were added, AscI at the 5' termini and FseI at the 3' termini. And (e) the gene encoding Lys225-Leu226 was modified to AAG CTT to introduce an endogenous restriction site, HindIII. The CXCR4-1 gene was further modified by introducing a thermostabilizing mutation $L125^{3.41}W$, using standard QuickChange PCR, then sub-cloned into the pFastBac1-830220 vector using the aforementioned restriction sites. Construction of CXCR4-2 was completed by modifying the CXCR4-1 gene. The C terminus was further truncated to residue 320 eliminating a total of 33 amino acids (dC320), using PCR with primer pairs 5'-AAG

ACC TCC GCA CAA CAC GCT TTG ACC AGT GGC CGG CCT CTG GAA GTT CTG TTC CAG GGG-3' and 5'-CCC CTG GAA CAG AAC TTC CAG AGG CCG GCC ACT GGT CAA AGC GTG TTG TGC GGA GGT CTT-3'. In the CXCR4-3 construct, the T4L was inserted between His228 and Gly231. Two residues (Ser229 and Lys230) were truncated and a Ser-Gly-Ser linker was added at the C terminus of T4L (T4L2). This CXCR4-3 construct was amplified using PCR primers encoding exogenous restriction sites (*Hind*III at the 5' end, GCG AAG CTT TCA CAC AAC ATC TTC GAG, and *Afl*II at the 3' end, GCG CTT AAG AGC TTT ACG CTT TTG GTG TCC TGA ACC TGA GTA AGC GTC CCA), and subsequently ligated into the corresponding restriction sites between Lys225 and Lys239 in the CXCR4-2 construct.

The CXCR4-WT DNA expression cassette was subcloned into the mammalian expression vector pACMV-TetO (*1*). CXCR4-T4L-ΔC constructs were subcloned into the mammalian expression vector pcDNA3.1(-) using the XhoI and EcoRI restriction sites. For the CXCR4-ΔC constructs, the T4L fusion and $L125^{3.41}W$ mutation were subsequently removed using standard QuickChange PCR.

Expression and purification of Sf9-expressed CXCR4 constructs for crystallization. High-titer recombinant baculovirus (>10^8 viral particles per ml) was obtained using the Bac-to-Bac Baculovirus Expression System (Invitrogen). Briefly, recombinant baculoviruses were generated by transfecting 5 μg of recombinant bacmid containing the target gene sequence into Sf9 cells using 3 μl of FuGENE HD Transfection Reagent (Roche) and Transfection Medium (Expression Systems). Cell suspensions were incubated for 4 days while shaking at 27 °C. P0 viral stocks were isolated after 4 days and used to produce high-titer baculovirus stocks. Viral titers were performed by flow cytometric methods by staining cells with gp64-PE (Expression Systems) (*2*).

Sf9 cells at cell density of 2-3 × 10^6 cells/ml were infected with P2 virus at MOI of 5. Cells were harvested by centrifugation at 48 hours post infection and stored at -80 °C until use.

Insect cell membranes were disrupted by thawing frozen cell pellets in a hypotonic buffer containing 10mM HEPES, pH 7.5, 10 mM $MgCl_2$, 20 mM KCl and protease inhibitor cocktail (Roche). Extensive washing of the raw membranes was performed by repeated centrifugation in the same hypotonic buffer (two - three times), and then in a high osmotic buffer containing 1.0 M NaCl, 10 mM HEPES, pH 7.5, 10 mM $MgCl_2$, 20 mM KCl, and protease inhibitor cocktail (three - four times), followed by Dounce homogenization to resuspend the membranes in fresh wash buffer thereby separating soluble and membrane associated proteins from integral transmembrane proteins. Highly purified membranes were resuspended in 10 mM HEPES, pH 7.5, 10 mM $MgCl_2$, 20 mM KCl, 30% (v/v) glycerol, and protease inhibitor cocktail, then flash-frozen with liquid nitrogen and stored at -80 °C until further use.

Purified membranes were thawed on ice in the presence of 200 μM CXCR4 compound (IT1t or CVX15) and EDTA-free protease inhibitor cocktail (Roche), and incubated at 4 °C for 1 hour. The membranes were then solubilized in 50 mM HEPES, pH 7.5, 500 mM NaCl, 0.5% (w/v) n-dodecyl-β-D-maltopyranoside (DDM, Anatrace), 0.1% (w/v) cholesteryl hemisuccinate (CHS) (Sigma), and 100 μM CXCR4 compound (IT1t or CVX15) for three hours at 4 °C. The supernatant was isolated by centrifugation at 160,000 × g for 40 minutes, and incubated in 5 mM buffered imidazole, 800 mM NaCl, with TALON IMAC resin (Clontech) overnight at 4 °C. Typically, 2 ml of resin per 1 L of original culture volume was used. After binding, the resin was washed with ten column volumes of 50 mM HEPES, pH 7.5, 800 mM NaCl, 10% (v/v) glycerol, 0.1% (w/v) DDM, 0.02% (w/v) CHS, 20 mM imidazole and 100 μM CXCR4 compound (IT1t or CVX15), followed by ten column volumes of 25 mM HEPES, pH 7.5, 500 mM NaCl, 10% (v/v)

glycerol, 0.05% (w/v) DDM, 0.01% (w/v) CHS, 10 mM $MgCl_2$, 5 mM ATP (Sigma) and 100 µM CXCR4 compound (IT1t or CVX15), and five column volumes of 25 mM HEPES, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM, 0.01% (w/v) CHS and 100 µM CXCR4 compound (IT1t or CVX15). The protein was then eluted with 4 column volumes of 25 mM HEPES, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM, 0.01% (w/v) CHS, 300 mM imidazole and 500 µM CXCR4 compound (IT1t or CVX15). PD MiniTrap G-25 column (GE healthcare) was used to remove imidazole and increase the compound concentration to 1 mM. The protein was then treated overnight with His-tagged PreScission protease (home-made) and His-tagged PNGase F (home-made) to remove the C-terminal His-tag and deglycosylate the receptor. The compound concentration was increased to 2 mM in this step. PreScission protease and PNGase F were removed by Ni-NTA superflow resin (Qiagen), which was incubated at 4 °C for 1 hour. The His-tag cleaved receptor was collected in the Ni-NTA column flow through. The receptor was then concentrated to 60-70 mg/ml with a 100 kDa molecular weight cut-off Vivaspin concentrator (Vivascience). Protein purity and monodispersity was tested by SDS-PAGE and analytical size-exclusion chromatography (aSEC). Typically, the protein purity exceeded 95%, and the aSEC profile showed a single peak, indicative of receptor monodispersity.

Lipidic cubic phase crystallization of CXCR4 constructs. Lipidic cubic phase (LCP) crystallization trials were performed using an *in meso* crystallization robot as previously described (*3*). 96-well glass sandwich plates were filled with 40-50 nl protein-laden LCP boluses overlaid by 0.8 µl of precipitant solution in each well and sealed with a glass coverslip. The protein-LCP mixture contained 40% (w/w) receptor solution, 54% (w/w) monoolein, and 6% (w/w) cholesterol. Crystallization set-ups were performed at room temperature (~20 °C). Plates were incubated and imaged at 20 °C using an automated incubator/imager (RockImager 1000, Formulatrix). Initial crystallization conditions (0.1 M HEPES, pH 7.0, 30% (v/v) PEG400, 0.4 M

sodium malonate or sodium citrate) were found by the LCP-FRAP assay (*4*). After extensive optimization, five distinct crystal forms have been obtained (Table S4). Crystals were harvested directly from LCP matrix using MiTeGen micromounts and flash frozen in liquid nitrogen.

Data collection and structure solution. X-ray data were collected on the 23ID-B/D beamline (GM/CA CAT) at the Advanced Photon Source, Argonne, IL using a 10 μm minibeam at an incidence wavelength of 1.0330 Å and a MarMosaic 300 CCD detector. Crystals were invisible after flash-freezing into liquid nitrogen, and a similar alignment and data-collection strategy was followed as previously described (*5*-*7*) for several hundred crystal samples. Most of the crystals diffracted to 2.3 – 3.5 Å resolution when exposed to 1 s of unattenuated beam using 1° oscillation, however, data collection was limited to 10 – 30 frames per crystal, due to the fast onset of radiation damage in the microcrystals. Data were integrated, scaled and merged using XDS and HKL2000 (*8*-*9*). A complete data set of CXCR4-2/IT1t (space group $P2_1$) was obtained by merging data collected from 2 crystals. Initial phase information was obtained by molecular replacement using a polyalanine model of the 7 TM α-helices of β_2AR (PDB ID: 2RH1) and a T4L model derived from the A_{2A}AR-T4L structure (PDB ID: 3EML). Data for CXCR4-3/CVX15 were processed at 2.9 Å resolution. All refinements were performed with the Refmac5 software suite (*10*) followed by manual examination and rebuilding of the refined coordinates in program X-fit (*11*) using both $|2F_o\text{-}F_c|$ sigma-A weighted and $|F_o\text{-}F_c|$ maps, as well as omit maps.

Data collection and refinement statistics for all five crystal forms are shown in Table S1. Electron density omit maps for IT1t and CVX-15 are shown in Fig. S3 and crystal packing diagrams for the different crystal forms are shown in Fig. S4.

Ligand binding assay and functional assays. *Ligand binding assay with Sf9 expressed receptor:* For the competition binding assay, the radio-labeled ligand [^{3}H]BIMA (51.2 Ci/mmol) was

prepared through the tritio-debromination of a potent CXCR4 antagonist, bis(imidazolylmethyl) amine analog (BIMA). For binding assays using the filter plate method, membrane preparations from Sf9 cells expressing CXCR4 were incubated for 120 minutes at room temperature with different concentrations of antagonists (BIMA, ITlt and CVX15) and 6nM [^{3}H]BIMA in 20 mM HEPES, pH 7.5, 1 mM $CaCl_2$, 5 mM $MgCl_2$. For CVX15 and BIMA competition assay, the assay buffer also contained 150 mM NaCl and 1% BSA. Unbound radioligand was removed by rapid filtration through a 96-well GF/C filter plate presoaked in 0.3% polyethylenimide (MultiScreen Harvest Plate, Millipore Corp.), and rinsed five times with 500 μL of ice-cold PBS. After drying the harvest plates, 30 μl of BetaScint scintillation liquid (Perkin-Elmer Life Sciences) were added per well. The bound radioactivity was measured using a Packard TopCount NXT. Nonspecific binding was determined in parallel reactions in the presence of an excess of 30 μM unlabelled BIMA. Scintillation proximity assays (SPA) were performed in isoplate-96 white frame clear well microplates (Perkin Elmer). Membranes containing expressed CXCR4 constructs were incubated at room temperature for 60 minutes with different concentrations of antagonists, 6 nM [^{3}H]BIMA, 250 μg of wheat germ agglutinin SPA beads (Perkin Elmer) in assay buffer (50 mM HEPES, pH 7.5, 1 mM $CaCl_2$ and 10 mM $MgCl_2$ final concentration). The signal from receptor bound [^{3}H]BIMA was measured using a Packard TopCount NXT and nonspecific binding determined in the presence of 50 μM unlabeled BIMA.

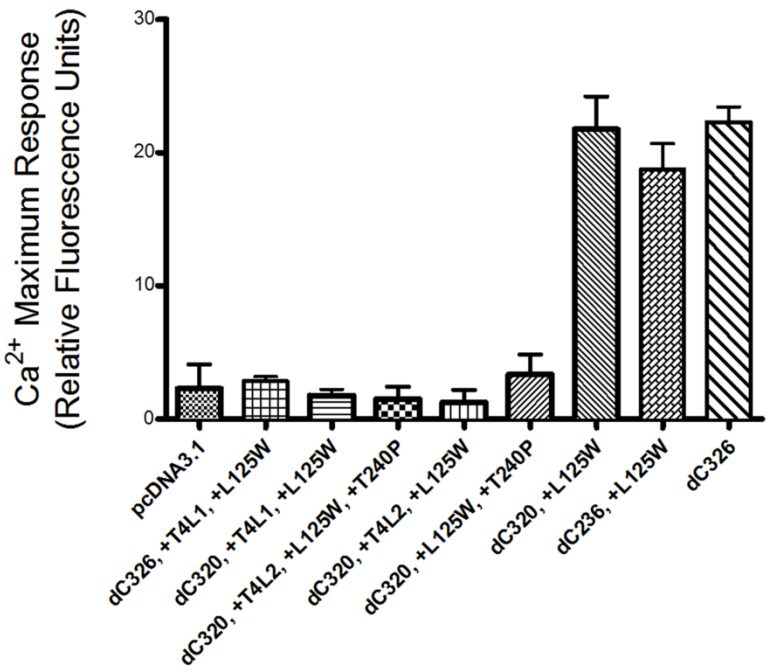
All data were analyzed by nonlinear regression analysis using GraphPad Prism. Equilibrium dissociation constants (K_d) for [^{3}H]BIMA were calculated using homologous competition binding with cold BIMA and apparent affinity **(**K_i) values for other ligands were calculated using the Cheng-Prusoff equation as $K_i = IC_{50}/(1 + [\text{ligand}] / K_d)$. All measurements were performed in triplicate and repeated at least three times. The mean values ± standard deviation for K_d of [^{3}H]BIMA and K_i of other ligands are shown in Table S3.

Mammalian Expression for ligand binding assays: HEK293 cells containing pcDNA6/TR were transformed and a stable cell line was selected for inducible expression of CXCR4, as previously described (*12*).

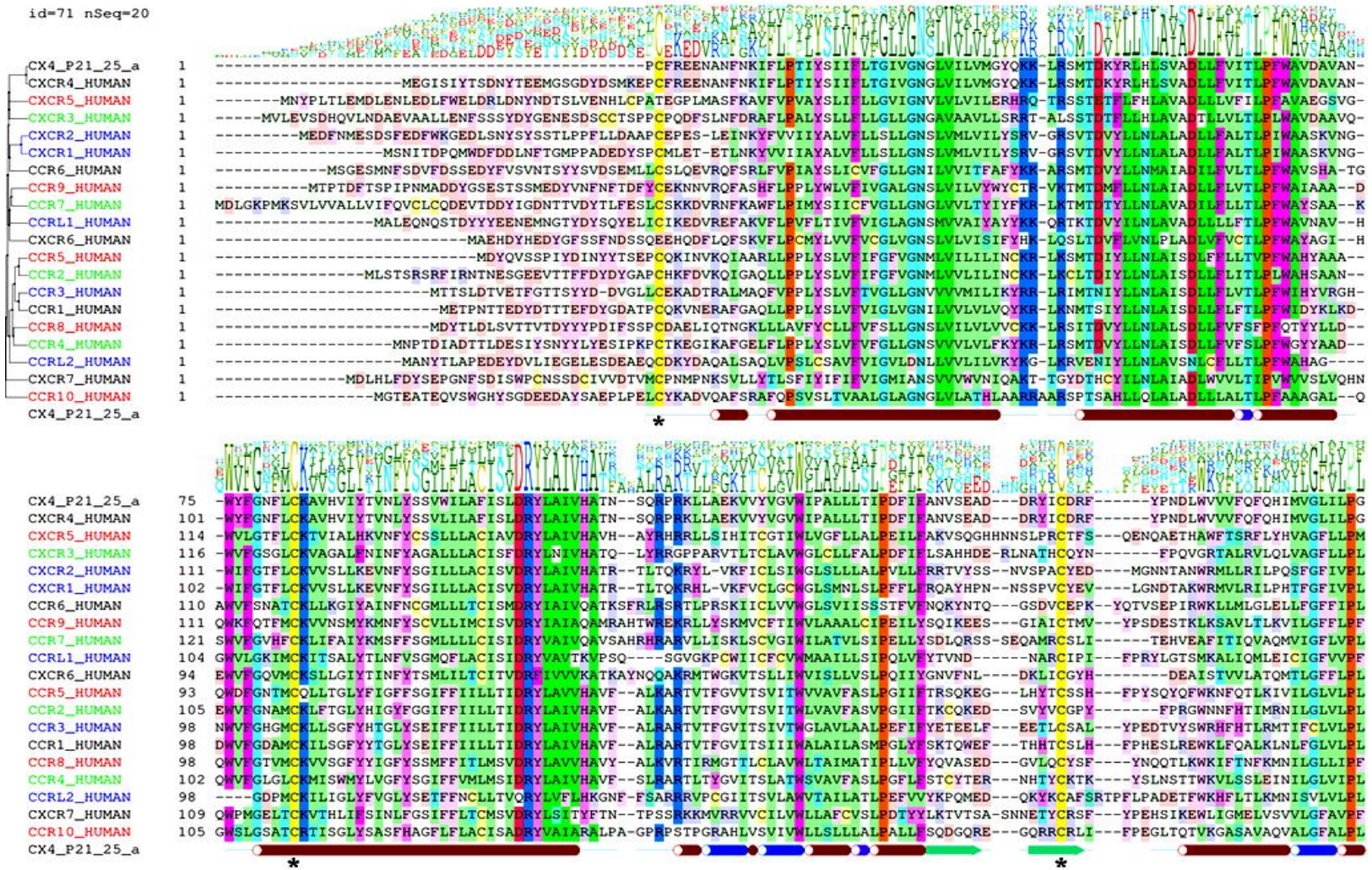
CHO expression and Calcium Flux Assays: Prior to transfection, CHO-K1 cells (ATCC) were grown in 10 cm plates at 37 °C, 5% CO_2, in DMEM/F12 (Gibco #10565) supplemented with 10% FBS. At 90% confluency, the pcDNA3.1(-) expression vectors were transiently transfected using Lipofectamine 2000 according to manufacturer instructions. After six hours, medium was replaced with fresh medium supplemented with 5 mM sodium butyrate.

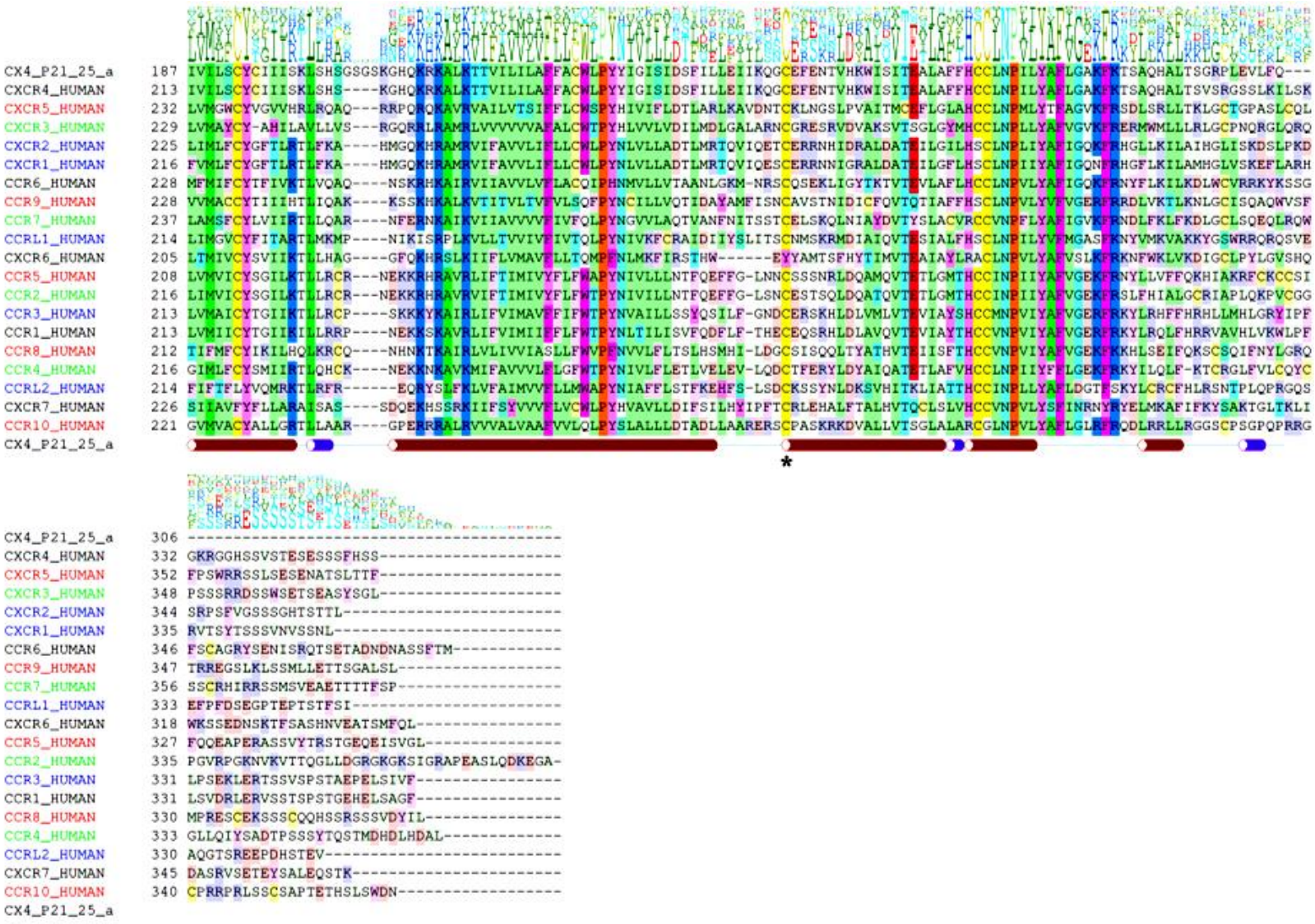
Calcium flux assays were performed 24 hours after transfection using the FLIPR Calcium 4 assay kit (Molecular Devices) and 2.0×10^5 cells per well in a 96-well plate format. The provided assay buffer was supplemented with 0.1% BSA and 2 mM probenecid. Briefly, cells were detached using PBS + 1 mM EDTA, washed twice in PBS + 0.5% BSA and resuspended in assay buffer. After counting on a Vi-CELL automated cell analyzer (Beckman Coulter), cells were normalized to 2.0×10^6 cells/ml and plated at 100 μl per well plus 100 μl dye. Plates were centrifuged for 3 minutes at 200 × g then incubated for 1 hour at 37°C. CXCL12-dependent increases in cytosolic Ca^{2+} were measured at 37 °C using a FlexStation III microplate reader (Molecular Devices). CXCL12 was prepared as previously described (*13*).

Supplementary Figures:

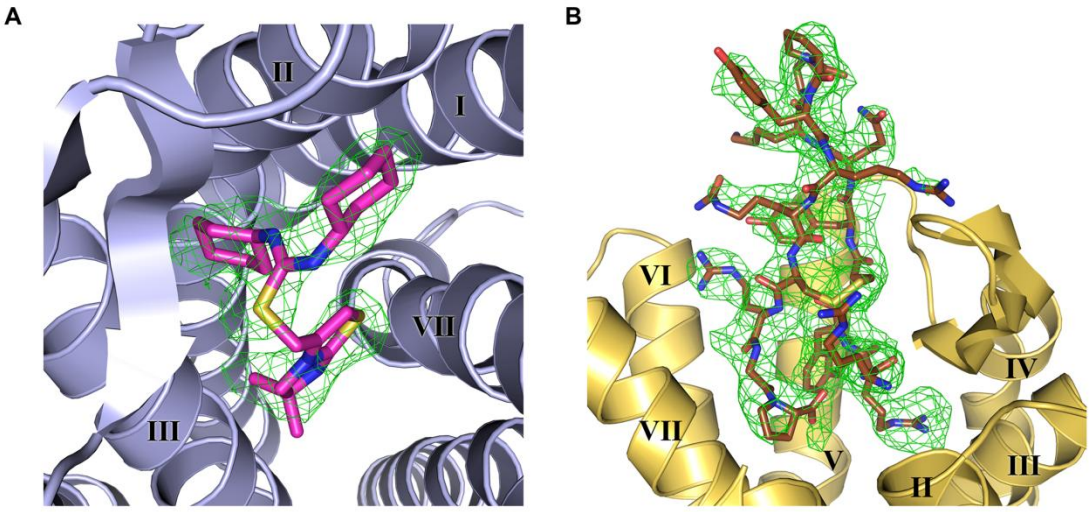


Supplementary Figure S1. Calcium flux assay in CHO cells transiently transfected with CXCR4 receptor constructs or empty vector pcDNA3.1, showing the effects of T4L, L125W, T240P and various C-terminal truncations on signaling. T4L1 refers to the T4L junction in constructs CXCR4-1 (dC326, +T4L1, +L125W) and CXCR4-2 (dC320, +T4L1, +L125W), and T4L2 is in CXCR4-3 (dC320, +T4L2, +L125W, +T240P). Details are shown in supplementary molecular biology section. C-terminal truncations, dC326 and dC320, refer to truncating C-terminal residues 326-352 and 320-352, respectively. Real-time changes in relative fluorescence units were measured for 150 seconds with addition of 200 nM CXCL12 at 20 seconds. Bars represent the maximum fluorescence change observed +/- standard deviation. Experiments were repeated at least two times in triplicate.

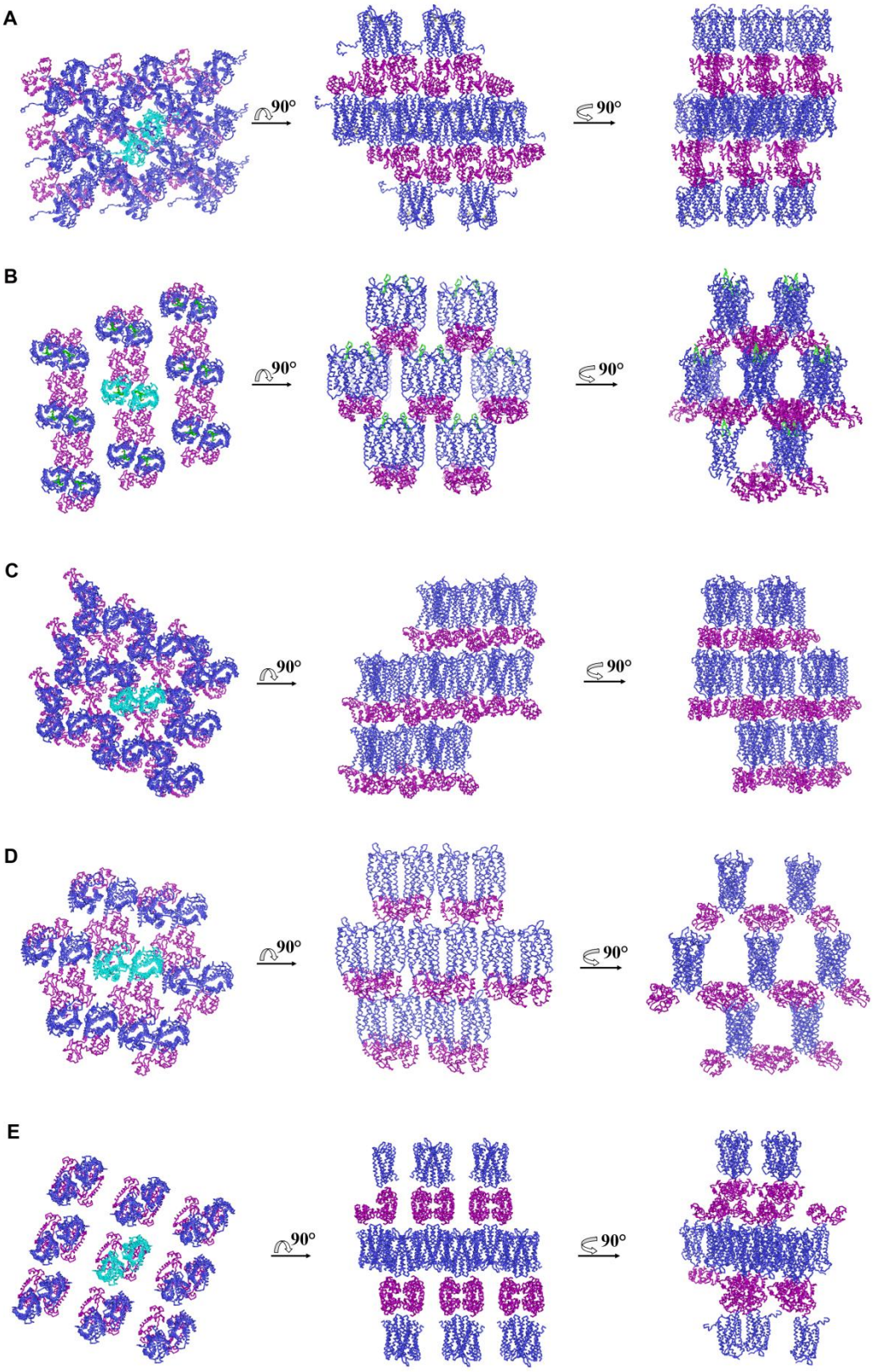

id=71 nSeq=20
CX4_P21_25_a
CXCR4_HUMAN
CXCR5_HUMAN
CXCR3_HUMAN
CXCR2_HUMAN
CXCR1_HUMAN
CCR6_HUMAN
CCR9_HUMAN
CCR7_HUMAN
CCRL1_HUMAN
CXCR6_HUMAN
CCR5_HUMAN
CCR2_HUMAN
CCR3_HUMAN
CCR1_HUMAN
CCR8_HUMAN
CCR4_HUMAN
CCRL2_HUMAN
CXCR7_HUMAN
CCR10_HUMAN


CX4_P21_25_a 187 IVILSCYCIIISKLSHSGSGSKGHQKRKALKTTVILILAFFACWLPYYIGISIDSFILLEIIKQGCEFENTVHKWISITEALAFFHCCLNPILYAFLGAKFKTSAQHALTSGRPLEVLFQ---
CXCR4_HUMAN 213 IVILSCYCIIISKLSHS----KGHQKRKALKTTVILILAFFACWLPYYIGISIDSFILLEIIKQGCEFENTVHKWISITEALAFFHCCLNPILYAFLGAKFKTSAQHALTSVSRGSSLKILSK
CXCR5_HUMAN 232 LVMGWCYVGVVHRLRQAQ---RRPQRQKAVRVAILVTSIFFLCWSPYHIVIFLDTLARLKAVDNTCKLNGSLPVAITMCEFLGLAHCCLNPMLYTFAGVKFRSDLSRLLTKLGCTGPASLCQL
CXCR3_HUMAN 229 LVMAYCY-AHILAVLLVS---RGQRRLRAMRLVVVVVVAFALCWTPYHLVVLVDILMDLGALARNCGRESRVDVAKSVTSGLGYMHCCLNPLLYAFVGVKFRERMWMLLLRLGCPNQRGLQRQ
CXCR2_HUMAN 225 LIMLFCYGFTLRTLFKA----HMGQKHRAMRVIFAVVLIFLLCWLPYNLVLLADTLMRTQVIQETCERRNHIDRALDATEILGILHSCLNPLIYAFIGQKFRHGLLKILAIHGLISKDSLPKD
CXCR1_HUMAN 216 FVMLFCYGFTLRTLFKA----HMGQKHRAMRVIFAVVLIFLLCWLPYNLVLLADTLMRTQVIQESCERRNNIGRALDATEILGFLHSCLNPIIYAFIGQNFRHGFLKILAMHGLVSKEFLARH
CCR6_HUMAN 228 MFMIFCYTFIVKTLVQAQ----NSKRHKAIRVIIAVVLVFLACQIPHNMVLLVTAANLGKM-NRSCQSEKLIGYTKTVTEVLAFLHCCLNPVLYAFIGQKFRNYFLKILKDLWCVRRKYKSSG
CCR9_HUMAN 228 VVMACCYTIIIHTLIQAK----KSSKHKALKVTITVLTVFVLSQFPYNCILLVQTIDAYAMFISNCAVSTNIDICFQVTQTIAFFHSCLNPVLYVFVGERFRRDLVKTLKNLGCISQAQWVSF
CCR7_HUMAN 237 LAMSFCYLVIIRTLLQAR----NFERNKAIKVIIAVVVVFIVFQLPYNGVVLAQTVANFNITSSTCELSKQLNIAYDVTYSLACVRCCVNPFLYAFIGVKFRNDLFKLFKDLGCLSQEQLRQW
CCRL1_HUMAN 214 LIMGVCYFITARTLMKMP----NIKISSFLKVLLTVVIVFIVTQLPYNIVKFCRAIDIIYSLITSCNMSKRMDIAIQVTESIALFHSCLNPILYVFMGASFKNYVMKVAKKYGSWRRQRQSVE
CXCR6_HUMAN 205 LTMIVCYSVIIKTLLHAG----GFQKHRSLKIIFLVMAVFLLTQMPFNLMKFIRSTHW------EYYAMTSFHYTIMVTLAIAYLRACLNPVLYAFVSLKFRKNFWKLVKDIGCLPYLGVSHQ
CCR5_HUMAN 208 LVMVICYSGILKTLLRCR---NEKKRHRAVRLIFTIMIVYFLFWAPYNIVLLNTFQEFFG-LNNCSSSNRLDQAMQVTETLGMTHCCINPIIYAFVGEKFRNYLLVFFQKHIAKRFCKCCSI
CCR2_HUMAN 216 LIMVICYSGILKTLLRCR---NEKKRHRAVRVIFTIMIVYFLFWTPYNIVILLNTFQEFFG-LSNCESTSQLDQATQVTETLGMTHCCINPIIYAFVGEKFRSLFHIALGCRIAPLQKPVCGG
CCR3_HUMAN 213 LVMAICYTGIIKTLLRCP----SKKKYKAIRLIFVIMAVFFIFWTPYNVAILLSSYQSILF-GNDCERSKHLDLVMLVTEVIAYSHCCMNPVIYAFVGERFRKYLRHFFHRHLLMHLGRYIPF
CCR1_HUMAN 213 LVMIICYTGIIKILLRRP----NEKKSKAVRLIFVIMIIFFLFWTPYNLTILISVFQDFLF-THECEQSRHLDLAVQVTEVIAYTHCCVNPVIYAFVGERFRKYLRQLFHRRVAVHLVKWLPF
CCR8_HUMAN 212 TIFMFCYIKILHQLKRCQ----NHNKTKAIRLVLIVVIASLLFWVPFNVVLFLTSLHSMHI-LDGCSISQQLTYATHVTEIISFTHCCVNPVIYAFVGEKFKKHLSEIFQKSCSQIFNYLGRQ
CCR4_HUMAN 216 GIMLFCYSMIIRTLQHCK----NEKKNKAVKMIFAVVVLFLGFWTPYNIVLFLETLVELEV-LQDCTFERYLDYAIQATETLAFVHCCLNPIIYFFLGEKFRKYILQLF-KTCRGLFVLCQYC
CCRL2_HUMAN 214 FIFTFLYVQMRKTLRFR------EQRYSIFKLVFAIMVVFLLMWAPYNIAFFLSTFKEHFS-LSDCKSSYNLDKSVHITKLIATTHCCINPLLYAFLDGTFSKYLCRCFHLRSNTPLQPRGQS
CXCR7_HUMAN 226 SIIAVFYFLLARAISAS----SDQEKHSSRKIIFSYVVVFLVCWLPYHVAVLLDIFSILHYIPFTCRLEHALFTALHVTQCLSLVHCCVNPVLYSFINRNYRYELMKAFIFKYSAKTGLTKLI
CCR10_HUMAN 221 GVMVACYALLGRTLLAAR----GPERRRALRVVVALVAAFVVLQLPYSLALLLDTADLLAARERSCPASKRKDVALLVTSGLALARCGLNPVLYAFLGLRFRQDLRRLLRGGSCPSGPQPRRG
CX4_P21_25_a
CX4_P21_25_a 306 ----------
CXCR4_HUMAN 332 GKRGGHSSVSTESESSSFHSS----------
CXCR5_HUMAN 352 FPSWRRSSLSESENATSLTTF----------
CXCR3_HUMAN 348 PSSSRRDSSWSETSEASYSGL----------
CXCR2_HUMAN 344 SRPSFVGSSSGHTSTTL----------
CXCR1_HUMAN 335 RVTSYTSSSVNVSSNL----------
CCR6_HUMAN 346 FSCAGRYSENISRQTSETADNDNASSFTM----------
CCR9_HUMAN 347 TRREGSLKLSSMLLETTSGALSL----------
CCR7_HUMAN 356 SSCRHIRRSSMSVEAETTTTFSP----------
CCRL1_HUMAN 333 EFPFDSEGPTEPTSTFSI----------
CXCR6_HUMAN 318 WKSSEDNSKTFSASHNVEATSMFQL----------
CCR5_HUMAN 327 FQQEAPERASSVYTRSTGEQEISVGL----------
CCR2_HUMAN 335 PGVRPGKNVKVTTQGLLDGRGKGKSIGRAPEASLQDKEGA-
CCR3_HUMAN 331 LPSEKLERTSSVSPSTAEPELSIVF----------
CCR1_HUMAN 331 LSVDRLERVSSTSPSTGEHELSAGF----------
CCR8_HUMAN 330 MPRESCEKSSSCQQHSSRSSSVDYIL----------
CCR4_HUMAN 333 GLLQIYSADTPSSSYTQSTMDHDLHDAL----------
CCRL2_HUMAN 330 AQGTSREEPDHSTEV----------
CXCR7_HUMAN 345 DASRVSETEYSALEQSTK----------
CCR10_HUMAN 340 CPRRPRLSSCSAPTETHSLSWDN----------
CX4_P21_25_a

Supplementary Figure S2. Sequence alignment between human chemokine receptors. Conserved cysteine residues are highlighted in yellow, and the four cysteines forming disulfide bonds in CXCR4 (Cys28-Cys274 and Cys109-Cys186) are marked with stars. Other colors represent properties of the conserved residues: blue, positively charged; red, negatively charged; cyan, polar; magenta, aromatic; green, hydrophobic; orange, proline. Intensity of the color is function of residue conservation. First row of the alignment shows CXCR4-2 sequence.



Supplementary Figure S3. Electron density of (**A**) IT1t in CXCR4-2/IT1t, and (**B**) CVX15 in CXCR4-3/CVX15. IT1t and CVX15 are shown in stick representation. Electron density is contoured at 2.5σ for IT1t and 2.0σ for CVX15 from an F_{obs}-F_{calc} omit map calculated without the contribution of ligands.


A
90°
90°
B
90°
90°
C
90°
90°
D
90°
90°
E
90°
90°

Supplementary Figure S4. Crystal-packing in five crystal forms, showing conserved CXCR4 dimers. CXCR4 is colored blue, and T4 lysozyme is shown in purple. One of the dimers is colored cyan. (**A**) CXCR4-2/IT1t (space group $P2_1$). (**B**) CXCR4-3/CVX15 (space group *C2*). (**C**) CXCR4-2/IT1t (space group *P1*). (**D**) CXCR4-3/IT1t (space group *P1*). (**E**) CXCR4-1/IT1t (space group *I222*).

Supplementary Tables:

Supplementary Table S1. Data collection (APS GM/CA 23ID-B/D, *10-μm* beam) and refinement statistics. Highest resolution shell is shown in parentheses.

Structure	**CXCR4-2 /IT1t$_1$ (A)**	**CXCR4-3 /CVX15 (B)**	**CXCR4-2 /IT1t$_2$ (C)**	**CXCR4-3 /IT1t (D)**	**CXCR4-1 /IT1t (E)**
Construct definition	+T4L1†, dC320‡, +L125W	+T4L2†, dC320‡, +L125W, +T240P	+T4L1†, dC320‡, +L125W	+T4L2†, dC320‡, +L125W, +T240P	+T4L1†, dC326‡, +L125W
PDB ID	3ODU	3OE0	3OE8	3OE9	3OE6
		Data collection			
Number of crystals	2	14	11	9	3
Space group	*P2$_1$*	*C2*	*P1*	*P1*	*I222*
Cell dimensions					
a, b, c (Å)	64.5, 83.7, 120.0	82.1, 144.9, 74.0	69.4, 76.6, 91.7	72.8, 72.9, 84.5	71.1, 78.7, 240.6
α, β, γ (°)	90.0, 102.2, 90.0	90.0, 104.5, 90.0	96.0, 97.8, 97.4	64.6, 73.8, 61.3	90.0, 90.0, 90.0
Number of reflections measured	94,558	106,704	93,481	76,109	36,485
Number of unique reflections	41,569	17,656	28,801	24,127	10,233
Resolution (Å)	50.0-2.50 (2.59-2.50)	50.0-2.90 (3.00-2.90)	50.0-3.10 (3.15-3.10)	50.0-3.10 (3.27-3.10)	50.0-3.30 (3.42-3.30)
R_{sym}*	0.12 (0.49)	0.12 (0.63)	0.15 (0.51)	0.12 (0.51)	0.14 (0.67)
Mean *I/σ(I)*	7.2 (1.8)	9.5 (1.5)	12.7 (1.8)	11.2 (1.5)	10.2 (1.5)
Completeness (%)	95.8 (89.0)	94.9 (73.6)	87.3 (59.3)	96.9 (87.7)	94.4 (78.2)
Redundancy	2.3 (1.9)	6.0 (3.9)	3.2 (1.9)	3.2 (2.5)	3.6 (2.2)
		Refinement			
Resolution (Å)	20.0-2.50	20.0-2.90	20.0-3.10	20.0-3.10	6.0-3.20
Number of reflections (test set)	41,455 (2,079)	16,707 (898)	28,647 (1,441)	24,209 (1,238)	8,310 (406)

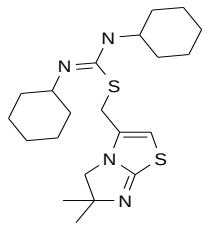
R_{work} / R_{free}	0.239 / 0.286	0.209 / 0.267	0.247 / 0.308	0.251 / 0.283	0.230 / 0.308
Number of atoms	7,803	3,644	10,397	6,804	3,444
Protein	7,425	3,488	10,316	6,750	3,368
Lipids, ligand, and other	243	150	81	54	76
Water	135	6	0	0	0
Overall *B* values (Å^2)	43	68	98	101	68
CXCR4	38	66	91	99	55
T4 lysozyme	54	72	108	104	93
Ligand	42	67	96	106	69
Lipid	50	—	—	—	59
RMSD					
Bond lengths (Å)	0.012	0.011	0.009	0.009	0.012
Bond angles (°)	1.25	1.28	1.07	1.12	1.43
Ramachandran plot statistics (%) (excluding Gly, Pro)					
Most favored regions	90.3	89.6	89.3	87.0	89.6
Additionally allowed regions	9.1	10.0	9.5	12.2	9.8
Generously allowed regions	0.6	0.7	0.7	0.5	0.5
Disallowed regions	0.0	0.0	0.4	0.3	0.0

*$R_{sym} = \Sigma_{hkl} |I(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl} \langle I(hkl)\rangle$, where $\langle I(hkl)\rangle$ is the mean of the symmetry-equivalent reflections of $I(hkl)$.

†T4L1 & T4L2: different T4L junction sites. Details are shown in supplementary molecular biology section.

‡dC320 and dC326: C-terminal truncations. C-terminal residues 320-352 and 326-352 were truncated in different constructs.

Supplementary Table S2. CXCR4 antagonists in ligand optimization.

Antagonist	Structure or sequence	Molecular weight, Da	CXCR4 binding, IC_{50}, nM	T_m*, °C
IT1t		404	8.0†	57
CVX15	^{1}Arg-^{2}Arg-^{3}Nal-^{4}Cys-^{5}Tyr-^{6}Gln-^{7}Lys-^{8}dPro-^{9}Pro-^{10}Tyr-^{11}Arg-^{12}Cit-^{13}Cys-^{14}Arg-^{15}Gly-^{16}dPro (^{4}Cys—^{13}Cys disulfide bond)	2115	0.6‡	57

* Melting temperature measured using CXCR4-2 construct.
†Radioligand competition assay using [^{125}I]CXCL12 and CEM cell membranes (*14*).
‡Radioligand competition assay using [^{125}I]CXCL12 and Jurkat cells (whole cell RLBA).

Supplementary Table S3. The equilibrium constant (K_d) values for BIMA and apparent affinity (K_i) values determined for various CXCR4 constructs using the filter plate and SPA methods. All measurements were performed in triplicate and repeated at least three times. The mean values ± standard deviation for K_d of [^{3}H]BIMA and K_i of other ligands are shown.

Constructs	BIMA (K_d, nM)		IT1t (K_i, nM)		CVX15 (K_i, nM)	
	Filter plate	SPA	Filter plate	SPA	Filter plate	SPA
CXCR4-WT (Sf9)	27.5 ± 1.2	3.5 ± 1.5	11.2 ± 2.7	8.3 ± 1.0	46.7 ± 8.7	1.1 ± 1.0
CXCR4-1 (Sf9)	25.6 ± 1.5	7.2 ± 1.2	14.0 ± 1.7	8.2 ± 1.0	268 ± 53.0	162.1 ± 1.1
CXCR4-2 (Sf9)	36.9 ± 1.2	7.4 ± 1.3	25.6 ± 4.8	7.3 ± 1.4	201 ± 36.1	189.2 ± 1.4
CXCR4-3 (Sf9)	24.4 ± 1.2	5.2 ± 1.2	19.8 ± 5.0	6.6 ± 1.4	55.3 ± 7.0	2.5 ± 1.0
CXCR4-WT (HEK293T)	- - -	3.7 ± 1.4	- - -	22 ± 1.5	- - -	2.3 ± 0.9

Note: K_d and K_i values for CXCR4 constructs were determined using both filter plate and SPA methods for cross-validation of the data. Due to nonspecific interaction between CVX15, Sf9 membranes, and the filter plate, extensive optimization of assay buffer and membrane treatment was required to obtain reproducible K_d and K_i values in the radioligand binding assay using the filter plate. We then used another assay method (SPA) to ensure that the relative binding affinities between the different CXCR4 constructs are not skewed by the assay methods and conditions.

Supplementary Table S4. Five crystal forms obtained by LCP crystallization.

Crystal form	Construct	Compound	Precipitant solution	Average Crystal size	Average Crystal growth time	Space group	Molecule(s) per AU*	Estimated solvent content
A	CXCR4-2 (+T4L1†,dC320‡, +L125W)	IT1t	100 mM sodium citrate, pH 5.5, 20% (v/v) PEG400, 300 mM sodium malonate, 5 mM Taurine and 2 mM IT1t.	~ 90 µm × 40 µm × <5 µm	10 days	$P2_1$	2	57%
B	CXCR4-3 (+T4L2†, dC320‡, +L125W,+T240P)	CVX15	100 mM Tris, pH 7.0, 25% (v/v) PEG400, 300 mM potassium sodium tartrate and 1 mM CVX15.	~ 40 µm × 30 µm × 10 µm	10 days	*C2*	1	68%
C	CXCR4-2 (+T4L1†,dC320‡, +L125W)	IT1t	100 mM MES, pH 6.0, 26% (v/v) PEG400, 300 mM Sodium malonate, 5 mM strontium chloride and 2 mM IT1t.	~ 20 µm × 10 µm × 10 µm	3 days	*P1*	3	57%
D	CXCR4-3 (+T4L2†, dC320‡, +L125W,+T240P)	IT1t	100 mM MES, pH 6.0, 27-35% (v/v) PEG400, 270-330 mM sodium malonate, 5 mM hexamine cobalt chloride and 2 mM IT1t.	~ 60 µm × 40 µm × 15 µm	10 days	*P1*	2	61%
E	CXCR4-1 (+T4L1†,dC326‡, +L125W)	IT1t	100 mM sodium citrate, pH 5.0-5.5, 20-26% (v/v) PEG400, 280-320 mM sodium malonate, 5 mM nickel chloride and 2 mM IT1t.	~ 60 µm × 10 µm × <5 µm	7 days	*I222*	1	59%

*AU: asymmetric unit.

†T4L1 & T4L2: different T4L junction sites. Details are shown in supplementary molecular biology section.
‡dC320 and dC326: C-terminal truncations. C-terminal residues 320-352 and 326-352 were truncated in different constructs.

Supplementary Table S5. Direct contacts between CXCR4 and ligands.

Hydrogen bond and salt bridge contacts between CXCR4 and IT1t		
CXCR4	**IT1t**	**Distance (Å)**
Asp97$^{2.63}$ (OD1)	N4	2.7
Glu288$^{7.39}$ (OE1)	N1	2.8
Hydrophobic interactions between CXCR4 and IT1t		
CXCR4	**IT1t**	**Distance (Å)**
Trp94$^{2.60}$ (CZ2)	C18	3.6
Trp94$^{2.60}$ (CZ3)	S1	3.8
Trp102 (CZ3)	C20	3.9
Val112$^{3.28}$ (CG2)	C19	3.8
Tyr116$^{3.32}$ (CE2)	C3	3.6
Arg183 (CZ)	C12	3.8
Ile185 (CD1)	C12	3.7
Ile185 (CD1)	C11	3.5
Cys186 (CB)	C21	3.7
Cys186 (O)	S2	3.0
Asp187 (CG)	S2	3.7
Hydrogen bond and salt bridge contacts between CXCR4 and CVX15		
CXCR4	**CVX15**	**Distance (Å)**
Asp187 (OD1)	Arg1 (NH1)	3.1
Asp187 (OD2)	Arg1 (N)	3.2
Arg188 (N)	Arg1 (O)	3.1
His113$^{3.29}$ (ND1)	Arg2 (NH1)	2.9
Thr117$^{3.33}$ (OG1)	Arg2 (NH2)	2.9
Asp171$^{4.60}$ (OD2)	Arg2 (NH2)	3.0
Arg188 (NH2)	Arg2 (O)	3.2
Arg188 (NE)	Arg2 (O)	2.9

Arg188 (O)	Cys4 (N)	3.0
Tyr190 (N)	Cys4 (O)	2.9
Asp193 (OD2)	Lys7 (NZ)	2.9
Asp262$^{6.58}$ (OD2)	Arg14 (NH1)	3.2
Asp262$^{6.58}$ (OD1)	Arg14 (O)	2.9

Intramolecular hydrogen bonds of CVX15		
CVX15		**Distance (Å)**
Arg2 (N)	dPro16 (OXT)	2.6
Nal3 (O)	Arg14 (N)	2.9
Tyr5 (OH)	Arg14 (NE)	3.5
Tyr5 (N)	Cit12 (O)	2.7
Tyr5 (O)	Cit12 (N)	2.8
Lys7 (N)	Tyr10 (O)	2.7
Lys7 (O)	Tyr10 (N)	2.6

Supplementary Table S6. CXCR4 homodimer interactions.

Hydrogen bonds in CXCR4-2/IT1t dimer		
Molecule A atom	**Molecule B atom**	**Distance (Å)**
Asn192 (ND2)	Leu266$^{6.62}$ (O)	3.0
Asn192 (ND2)	Glu268 (OE2)	2.6
Trp195$^{5.34}$ (NE1)	Leu267$^{6.63}$ (O)	2.8
Leu266$^{6.62}$ (O)	Asn192 (ND2)	3.0
Leu267$^{6.63}$ (O)	Trp195$^{5.34}$ (NE1)	2.8
Glu268 (OE2)	Asn192 (ND2)	2.6
Hydrophobic interactions in CXCR4-2/IT1t dimer		
Molecule A atom	**Molecule B atom**	**Distance (Å)**
Leu194$^{5.33}$ (CD2)	Leu194$^{5.33}$ (CD2)	3.5
Leu194$^{5.33}$ (CD2)	Val197$^{5.36}$ (CG2)	3.6
Val197$^{5.36}$ (CG2)	Leu194$^{5.33}$ (CD2)	3.3
Val197$^{5.36}$ (CG1)	Val198$^{5.37}$ (CG2)	3.6
Val198$^{5.37}$ (CG2)	Val197$^{5.36}$ (CG1)	3.6
Val198$^{5.37}$ (CG1)	Phe201$^{5.40}$ (CE1)	4.0
Phe201$^{5.40}$ (CE1)	Val198$^{5.37}$ (CG1)	3.7
Phe201$^{5.40}$ (CD1)	Phe201$^{5.40}$ (CD1)	4.0
Met205$^{5.44}$ (CG)	Met205$^{5.44}$ (CE)	4.3
Leu210$^{5.49}$ (CD1)	Leu210$^{5.49}$ (CD2)	3.5
Leu210$^{5.49}$ (CD2)	Leu210$^{5.49}$ (CD1)	3.5
Trp195$^{5.34}$ (CE2)	Leu267$^{6.63}$ (CB)	3.6
Leu267$^{6.63}$ (CB)	Trp195$^{5.34}$ (CE2)	3.6
Additional hydrophobic interactions in CXCR4-3/CVX15 dimer		
Molecule A atom	**Molecule B (crystal symmetry related molecule) atom**	**Distance (Å)**
Leu136$^{3.52}$ (CD2)	His140 (CE1)	3.6

His140 (CE1)	Leu136$^{3.52}$ (CD2)	3.6
His140 (CE1)	His140 (CG)	3.6

References

1. P. J. Reeves, J. M. Kim, H. G. Khorana, *Proc Natl Acad Sci U S A* **99**, 13413 (Oct 15, 2002).
2. M. A. Hanson *et al.*, *Protein Expr Purif* **56**, 85 (Nov, 2007).
3. V. Cherezov, A. Peddi, L. Muthusubramaniam, Y. F. Zheng, M. Caffrey, *Acta Crystallogr D Biol Crystallogr* **60**, 1795 (Oct, 2004).
4. V. Cherezov, J. Liu, M. Griffith, M. A. Hanson, R. C. Stevens, *Cryst Growth Des* **8**, 4307 (2008).
5. V. Cherezov *et al.*, *J R Soc Interface* **6 Suppl 5**, S587 (Oct 6, 2009).
6. V. Cherezov *et al.*, *Science* **318**, 1258 (Nov 23, 2007).
7. M. A. Hanson *et al.*, *Structure* **16**, 897 (Jun, 2008).
8. Z. Otwinowski, W. Minor, in *Methods in Enzymology* C. W. J. Carter, R. M. Sweet, Eds. (Academic Press, New York, 1997), vol. 276, pp. 307-326.
9. W. Kabsch, *J. Appl. Cryst.* **26,** (1993).
10. G. N. Mrushudov, E. J. Dodson, *Acta Crystallogr D Biol Crystallogr* **53**, 240 (1997).
11. R. W. Cheary, Coelho, A. A., *Program XFIT, deposited in CCP14 Powder Diffraction Library, Engineering and Physical Sciences Research council, Daresbury Laboratory, Warrington, England.*, (1996).
12. S. J. Allen, S. Ribeiro, R. Horuk, T. M. Handel, *Protein Expr Purif* **66**, 73 (Jul, 2009).
13. M. O'Hayre, C. L. Salanga, P. C. Dorrestein, T. M. Handel, *Methods Enzymol* **460**, 331 (2009).
14. G. Thoma *et al.*, *J Med Chem* **51**, 7915 (Dec 25, 2008).