Allosteric modulation of the CXCR4:CXCL12 axis by targeting receptor nanoclustering via the TMV-TMVI domain

Eva M. García-Cuesta[1], Pablo Martínez[1], Karthik Selvaraju[2], Gabriel Ulltjärn[2], Adrián Miguel Gómez Pozo[3], Gianluca D'Agostino[1], Sofía Gardeta[1], Adriana Quijada-Freire[1], Patricia Blanco Gabella[3], Carlos Roca[3], Daniel del Hoyo[4], Rodrigo Jiménez-Saiz[1,5-7], Alfonso García-Rubia[3], Blanca Soler-Palacios[1], Pilar Lucas[1], Rosa Ayala-Bueno[1], Noelia Santander Acerete[1], Yolanda R. Carrasco[8], Carlos O. Sánchez-Sorzano[4], Ana Martínez[3,9], Nuria E Campillo[3], Lasse Jenssen[2], José Miguel Rodríguez-Frade[1], César Santiago[10], Mario Mellado[1*].

[1] Chemokine Signaling group, Department of Immunology and Oncology, Centro Nacional de Biotecnología/CSIC, Campus de Cantoblanco, 28049, Madrid, Spain.
[2] Division of Diagnostics and Specialist Medicine, Department of Health, Medical and Caring Sciences, Linköping University, 58185, Linköping, Sweden.
[3] Centro de Investigaciones Biológicas Margarita Salas (CIB-CSIC), 28040, Madrid, Spain.
[4] Biocomputing Unit, Centro Nacional de Biotecnología (CNB-CSIC), Campus de Cantoblanco, 28049 Madrid, Spain.
[5] Department of Immunology, Instituto de Investigación Sanitaria Hospital Universitario de La Princesa (IIS-Princesa), 28006, Madrid, Spain.
[6] Department of Medicine, McMaster Immunology Research Centre (MIRC), Schroeder Allergy and Immunology Research Institute, McMaster University, Hamilton, ON L8S 4L8, Canada.
[7] Faculty of Experimental Sciences, Universidad Francisco de Vitoria (UFV), 28223, Madrid, Spain.
[8] B Lymphocyte Dynamics, Department of Immunology and Oncology, Centro Nacional de Biotecnología (CNB)/CSIC, Campus de Cantoblanco, 28049, Madrid, Spain.
[9] Neurodegenerative Diseases Biomedical Research Network Center (CIBERNED), Instituto de Salud Carlos III, 28029 Madrid, Spain
[10] X-ray Crystallography Unit, Department of Macromolecules Structure, Centro Nacional de Biotecnología/CSIC, Campus de Cantoblanco, 28049, Madrid, Spain.

[*]Corresponding author: Mario Mellado, Department of Immunology and Oncology, CNB/CSIC, Darwin 3, Cantoblanco, 28049 Madrid Spain. Tel: (+34) 91/585-4852; Fax: (+34) 91/372-0493; **Email:** mmellado@cnb.csic.es

Author Contributions: J.M.R.-F., C.S. and M.M. conceptualized the project and designed all aspects of the study; A.M.G.P. carried out the virtual computational screening; P.B.G., C.R., D.H. and C.O.S.S. performed docking analysis and evaluated the molecular dynamics; A.G.-R. synthesized the organic compounds; N.E.C. and A.M. contributed to the analysis of the *in silico* results; E.M.G.-C., J.M.R.-F., P.M., G.D'A., S.G., A.Q-F., R.J.-S., B.S.-P., P.L., R.A.-B. and N.S.A. performed experiments; E.M.G.-C. and J.M.R.-F. quantified the TIRFM images; C.S. modeled the crystal packing structure and determined the mutations on CXCR4; P.M., S.G., and Y.R.C. performed the experiments on lipid bilayers; K.S., G.U. and L.J. developed the ZTX model; L.J. designed the ZTX model for the *in vivo* analysis; R.J.-S., Y.R.C., N.E.C., J.M.R.-F., C.S., and M.M. provided input into the project; J.M.R.-F., C.S., and M.M. wrote the manuscript.

Competing Interest Statement: E.M.G.-C., J.M.R.-F., N.E.C., A.M., C.S., and M.M. have a patent for compounds AGR1.131, AGR1.135 and AGR1.137. The remaining authors declare no potential conflict of interest.

Classification: Biological Science, Immunology and inflammation.

Keywords: cell migration, chemokine receptors, allosteric inhibitors

Abstract

CXCR4 is a ubiquitously expressed chemokine receptor that regulates leukocyte trafficking and arrest in both homeostatic and pathological states. It also participates in organogenesis, HIV-1 infection and tumor development. Despite the potential therapeutic benefit of CXCR4 antagonists, only one, plerixafor (AMD3100), which blocks the ligand-binding site, has reached the clinic. Recent advances in imaging and biophysical techniques have provided a richer understanding of the membrane organization and dynamics of this receptor. Activation of CXCR4 by CXCL12 reduces the number of CXCR4 monomers/dimers at the cell membrane and increases the formation of large nanoclusters, which are largely immobile and are required for correct cell orientation to chemoattractant gradients. Mechanistically, CXCR4 activation involves a structural motif defined by residues in TMV and TMVI. Using this structural motif as a template, we performed *in silico* molecular modeling followed by *in vitro* screening of a small compound library to identify negative allosteric modulators of CXCR4 that do not affect CXCL12 binding. We identified AGR1.137, a small molecule that abolishes CXCL12-mediated receptor nanoclustering and dynamics and blocks the ability of cells to sense CXCL12 gradients both *in vitro* and *in vivo* while preserving ligand binding and receptor internalization.

Significance Statement

The chemokine receptor CXCR4 and its ligand CXCL12 are key for development, hematopoiesis, neutrophil homeostasis and lymphocyte trafficking. The only commercially available CXCR4 antagonist currently approved for clinical use is plerixafor (AMD3100), a small compound that blocks the ligand-binding site. Unfortunately, its clinical use is limited by poor pharmacokinetics and adverse effects associated with long-term administration. Here, we performed *in silico* analyses of a small aromatic compound library and *in vitro* screening to identify allosteric CXCR4 modulators. These compounds abolish the ability of cells to sense chemoattractant gradients without affecting other ligand-mediated functions such as cAMP production or receptor internalization. The selected compounds were also effective *in vivo,* as demonstrated by reduced tumorigenesis and metastasis in a zebrafish tumor model. Our study describes a new approach to selectively alter some GPCR functions while maintaining other functions.

Introduction

CXC chemokine receptor 4 (CXCR4) is a homeostatic G protein-coupled receptor (GPCR) that is widely expressed in both embryonic and adult tissues (1). It is also ubiquitously expressed in the hematopoietic system, where it plays a pivotal role in leukocyte trafficking and arrest in specific niches in homeostasis and disease. CXCR4 is essential for adaptive and innate immune responses, as well as in the organization and maintenance of the bone marrow (BM) (2). Indeed, CXCR4 and its unique chemokine ligand CXC motif chemokine 12 (CXCL12) are largely responsible for the migration (3), homing (4) and survival (5) of hematopoietic stem cells in the BM.

Results from studies in conditional *Cxcr4*-knockout mice point to a relevant role for CXCR4 in several non-hematopoietic tissues, for example, in regulating central nervous system development (6), and in vasculature development in the gastrointestinal tract (7) and the kidney (8). In addition, CXCR4, together with CCR5, serve as primary co-receptors (with CD4) for HIV-1 fusion and entry into target cells (9).

CXCR4 expression is frequently elevated in many cancers, including breast (10), ovarian (11), prostate (12), melanoma (13) and neuroblastoma (14), where it participates in tumor growth, tumor cell interactions with the microenvironment (15), vasculogenesis, angiogenesis (16) and metastasis (10). Indeed, increased CXCR4 expression in metastatic lesions correlates with tumor progression and with preferential metastatic sites of the primary tumor (17). Studies in mice have suggested that CXCR4 is a good target for cancer therapy, as blockade of its signaling impairs

metastasis in several models (10, 18). Despite the essential role of the CXCR4/CXCL12 axis in physiology and pathology, the only commercially available CXCR4 antagonist approved for clinical use is plerixafor (AMD3100), which is indicated for the mobilization of stem cells from the BM to the peripheral blood in autologous transplantation (19).

Single particle tracking in total internal reflection fluorescence (SPT-TIRF) experiments and stimulated emission depletion super resolution microscopy (STED) have demonstrated that CXCR4 is organized at the cell membrane as monomers, dimers and small aggregates (groups of ≥3 receptors) termed nanoclusters, and that CXCL12 binding decreases the percentage of monomers/dimers and increases the formation of large nanoclusters (20). This mechanism is essential for CXCR4 signaling, which allows cells to orient themselves correctly in response to CXCL12 gradients (21). A triple-mutant CXCR4 (K239L/V242A/L246A; CXCR4mut), located in the N-terminal region of transmembrane (TM) VI, dimerizes but neither forms nanoclusters in response to CXCL12 nor supports CXCL12-induced directed cell migration, although it can induce some Ca^{2+} flux and is internalized after ligand binding (20). A cluster has been identified in CXCR4, defined by TMV and TMVI residues that form a hydrophobic lock connecting the signal from the orthosteric ligand-binding site to the downstream conserved domains involved in G protein coupling and signaling (22, 23).

Here, we used this signal domain of CXCR4 as a base to search for new CXCR4 antagonists. We performed an *in silico* screening of a small aromatic compound library for the identification of allosteric modulators that block directed cell migration, without interfering with ligand binding and, thus, allowing other CXCL12-mediated functions. Using molecular modeling analyses, we identified a novel class of structurally related small molecules that modulate some CXCR4 functions by binding to a novel regulatory cleft formed by the TMV and TMVI transmembrane helices. One of the selected compounds, AGR1.137, abolished CXCL12-mediated receptor nanoclustering and dynamics and the ability of the cells to specifically sense CXCL12 gradients, but did not alter ligand binding or receptor internalization. Notably, AGR1.137 had a minimal effect on ERK1/2 and AKT phosphorylation, which may reduce the side effects associated with full CXCR4 inhibition. Finally, using a zebrafish tumor model, we observed that AGR1.137 treatment of HeLa cells reduced tumorigenesis and metastasis, suggesting that AGR1.137 also impairs cell sensitivity to CXCL12 gradients by blocking CXCR4 nanoclustering *in vivo*. In conclusion, our results establish the importance of residues within the TMV–VI cleft for CXCR4 nanoclustering and CXCL12-mediated directional migration, and identify a negative allosteric modulator that exhibits both *in vitro* and *in vivo* activity.

Results

Screening for small compounds targeting CXCR4 that block CXCL12-induced CXCR4 nanoclustering

The propagation domain of CXCR4, which links ligand binding to the residues involved in G-protein association, includes residues bounded by the transmembrane helices TMV and TMVI, which enclose a cleft of 870 $Å^3$ (22) is surface-exposed for interactions with the plasma membrane, as it contains residues involved in lipid binding (24), and also includes residues implicated in receptor activation and in the transmission of conformational changes through the TM helix domains (22) (Fig. 1A). Additionally, the cleft contains the residues K239, V242, and L246, the mutation of which prevents CXCL12-mediated receptor oligomerization (20). We thus screened for allosteric modulators that fit into this cleft and the cavity formed by TMV and TMVI but do not interfere with the CXCL12 binding site. Virtual screening was performed using the Medicinal and Biological Chemistry (MBC) library (CIB-CSIC) (25) comprises more than 2,000 small heterocyclic compounds with drug-like properties. As a template for modeling CXCR4 we used the atomic structure of the receptor solved in complex with the inhibitor IT1t (PDB code: 3ODU). As this included T4 lysozyme inserted between helices TMV and TMVI to increase the stability of the

protein (Fig. 1A), a common strategy to facilitate crystallogenesis of GPCRs (26), we first generated (24, 27) a CXCR4 homology model using the SWISS_MODEL server (28). The compounds were then prepared and docked against the modeled CXCR4 structure using "Glide", and those identified were ranked by docking score, that is, by the minimal interaction energy given by Glide and visualization of the poses. The results were confirmed using a second strategy based on the CXCR4 model predicted by AlphaFold (29) and the sequence available under UniProt entry P61073.

The top 40 candidates with the best docking score in the region of interest were chosen and were selected for functional analysis (Supplementary Table 1). We first screened their ability to block CXCL12-induced cell dependent migration in Transwell chambers.

Jurkat cells that were either untreated or treated with different concentrations (1–100 μM) of the selected small compounds (30 min, 37°C, 5% CO_2) were allowed to migrate towards CXCL12. To enhance their possible inhibitory effects, the compounds were maintained in the upper chamber containing the cells throughout the chemotaxis experiment (Supplementary Fig. 1A, B).

Based on these results, we selected two compounds, AGR1.135 and AGR1.137, which showed reproducible dose-dependent inhibitory effects on CXCR4-induced cell migration (Fig. 1B) and had the highest docking score on the cleft between TMV and TMVI (-51.4 and -37.2 kcal /mol, respectively). As a control for further analysis, we used AGR1.131, which shares a common scaffold with the other compounds and interacts with TM domains of CXCR4, but did not affect CXCL12-promoted cell migration. It also theoretically targeted the same motif on CXCR4 as AGR1.135 and AGR1.137 (-39.8 kcal/mol), but showed a better docking score on a pocket between TMI and TMVII (-43.6 kcal/mol). Under these experimental conditions, none of the selected compounds were toxic against Jurkat cells, as shown by propidium iodide incorporation and cell cycle analysis (Supplementary Fig. 2). The structures of the three selected compounds was verified by NMR analysis (Supplementary information) and were characterized by the presence of a common core (4-(1-benzyl-1H-1,2,3-triazol-4-yl)phenyl) methanol, linked through a benzamide group to a lateral chain bearing amines of different length and chemical nature. AGR1.131 bears a simple benzylamine, AGR1.135 contains a complex lateral chain due to the presence of the 1-(2-nitro-4-(trifluoromethyl)phenyl)piperazine, and AGR1.137 incorporates ethyl piperidine-4-carboxylate (Fig. 1C).

AGR1.135 and AGR1.137 block CXCL12-mediated CXCR4 nanoclustering and dynamics

We next evaluated whether the selected compounds altered CXCL12-mediated CXCR4 nanoclustering and dynamics. To do this, we utilized SPT-TIRF on CXCR4-deficient Jurkat cells ($JK^{-/-}$) transiently transfected with CXCR4 fused to the AcGFP monomeric protein (CXCR4-AcGFP) and pretreated or not with 50 μM of the selected small compounds (Supplementary videos 1–8), a concentration that inhibited CXCL12-induced cell migration by 50–75% (Fig. 1B). The compounds had no effect on DMSO-stimulated cells, and we observed mainly receptor monomers and dimers (~82–98%), with a very low percentage of complexes with >3 receptors (~2–18%) (Fig. 2A). Accordingly, the basal mean spot intensity (MSI) of CXCR4 was similar in all cases (~1,500 a.u.) (Fig. 2B). Under these experimental conditions, most of the particles corresponded to mobile particles (~88%) (Fig. 2C), and none of the selected compounds affected the short time-lag diffusion coefficient (D_{1-4}) for CXCR4 trajectories, with a median value of ~0.02 $\mu m^2/s$ (Fig. 2D).

We next examined how pre-treatment with the selected compounds affected receptor dynamics upon CXCL12 stimulation. Pre-treatment with AGR1.135 and AGR1.137, but not with AGR1.131, significantly altered CXCL12-mediated receptor nanoclustering (32–69% of nanoclusters with ≥3 receptors/particle in untreated and in AGR1.131-treated cells *versus* 7–22% in cells pretreated with AGR1.135 or AGR1.137) (Fig. 2A), with an MSI of 3,219 a.u. for CXCR4 in untreated cells and 3,618 a.u. in AGR1.131-treated cells *versus* 1,611 a.u. in AGR1.135- and 1,403 a.u. in AGR1.137-treated cells (Fig. 2B). This effect was also evident when we evaluated the dynamic parameters of CXCR4. Under control conditions (untreated or AGR1.131-treated cells), CXCL12 significantly increased the percentage of immobile particles (12% in untreated cells *versus* 25% in untreated cells + CXCL12 and 18% in AGR1.131-treated cells + CXCL12) (Fig. 2C),

whereas AGR1.135 and AGR1.137 pretreatment had no effect on the percentage of immobile particles after CXCL12 stimulation (14% for both AGR1.135 and AGR1.137). Moreover, the expected reduction in CXCR4 diffusivity triggered by CXCL12 in untreated (median D_{1-4} = 0.0088 $\mu m^2/s$) or AGR1.131-treated (median D_{1-4} = 0.0099 $\mu m^2/s$) cells was abolished with AGR1.135 (D_{1-4} = 0.0169 $\mu m^2/s$) and AGR1.137 (D_{1-4} = 0.0165 $\mu m^2/s$) treatments (Fig. 2D).

Altogether, these results suggest that AGR1.135 and AGR1.137 behave as negative allosteric modulators of CXCR4 and alter the CXCL12-mediated receptor nanoclustering and dynamics.

AGR1.135 and AGR1.137 incompletely abolish CXCR4-mediated responses in Jurkat cells

CXCR4 internalizes in response to CXCL12, a process that mediates receptor desensitization (30). This process is associated with clathrin-mediated endocytosis of chemokine receptors (31), and requires receptor aggregation. Our previous results indicate that CXCR4 nanoclustering and the aggregates involved in receptor internalization are independent processes (20). Indeed, CXCL12 triggers CXCR4mut internalization but fails to promote receptor nanoclustering (20). We thus tested whether the selected antagonists affected CXCL12-mediated CXCR4 internalization. Jurkat cells treated with vehicle or the selected compounds were stimulated with CXCL12 and receptor internalization was evaluated by flow cytometry using anti-CXCR4 antibodies. Neither AGR1.135 nor AGR1.137 treatment altered the internalization of CXCR4 that was observed in untreated Jurkat cells or in AGR1.131-treated cells (Fig. 3A). These data suggest that none of the allosteric modulators block CXCL12 binding to CXCR4, supporting the *in silico* screening strategy of preserving ligand binding integrity. Indeed, using flow cytometry analysis and CXCL12-ATTO 700 (32) we confirmed that none of the selected compounds blocked ligand binding, whereas this was blocked by the ligand-binding inhibitor AMD3100 (Fig. 3B, Supplementary Fig. 3). Taken together, the data indicate that the selected compounds do not abrogate ligand binding or receptor internalization, but blocked other functions such as cell migration.

Because CXCL12 promotes Gi-protein activation, we analyzed the effects of the selected compounds on CXCL12-mediated inhibition of cAMP production, a canonical signaling pathway downstream of CXCR4 activation. In contrast to AMD3100, an orthosteric antagonist of CXCR4, none of the compounds affected CXCL12-mediated Gi protein activation (Fig. 3C). We next evaluated their effects on other signaling pathways such as ERK1/2 phosphorylation and PI3K activation (33). Jurkat cells treated with vehicle or the selected compounds were activated with CXCL12 (50 nM) for different time periods and cell lysates were analyzed by western blotting using anti-P-ERK1/2 and -P-AKT antibodies. Neither AGR1.135 nor AGR1.137 nor the control AGR1.131 treatment inhibited CXCL12-mediated activation of the two signaling pathways (Fig. 3D–F, Supplementary Fig. 4), whereas AMD3100 treatment as a positive control did so (Supplementary Fig. 4). Overall, the data indicate that the selected compounds inhibit CXCL12-induced chemotaxis in Transwell assays, but do not affect G protein activation or other ligand-mediated signaling pathways such as ERK1/2 or PI3K.

Alterations in actin cytoskeletal dynamics are linked to deficiencies in ligand-mediated receptor nanoclustering and to defects in the ability of cells to sense chemoattractant gradients. Cells treated with latrunculin A, an actin polymerization inhibitor (20), and cells expressing CXCR4WHIM^{R334X}, which fail to properly control actin dynamics (21, 34), have defective CXCL12-mediated nanoclustering and are unable to properly sense chemoattractant gradients. Using flow cytometry and phalloidin staining of Jurkat cells treated with vehicle or the selected compounds, we detected altered actin polymerization in cells treated with AGR1.135, with AGR1.137, and with AMD3100 (Fig. 4A). These data were corroborated by immunostaining and confocal analysis (Fig. 4B, C). Whereas Jurkat cells treated with vehicle or AGR1.131 were correctly polarized after CXCL12 activation, AGR1.135 and AGR1.137 treatments promoted a reduction in the number of polarized cells. As a control, AMD3100 treatment also abolished CXCL12-mediated cell

polarization (Fig. 4B, C). None of the compounds affected anti-CD3 mediated actin polymerization, ruling out a direct effect of the modulators on actin or actin-binding proteins (Supplementary Fig. 5).

CXCR4 nanoclustering has been recently associated with the migratory phenotype of T cell blasts (21, 34). To explore how the allosteric modulators affect the phenotype of migrating cells, we added T cell blasts previously treated with vehicle or with the selected compounds to a 2D lipid bilayer system with embedded ICAM-1, alone or with CXCL12. We observed that treatment with AGR1.135, AGR1.137, or AMD3100 (control), but not with AGR1.131, abolished CXCL12-induced cell migration and cell adhesion (Fig. 4D, E).

As the antagonistic effects of AGR1.135 and AGR1.137 are also compatible with a partial agonist behavior, we used Transwell chambers to evaluate the ability of the selected compounds to promote Jurkat cell migration. None of the compounds induced cell migration (Supplementary Fig. 6). Overall, our results suggest that AGR1.135- and AGR1.137-driven modulation of CXCR4 nanoclustering blocks certain receptor-associated functions, including actin dynamics, directional cell migration, integrin-mediated adhesion and migration, while leaving other functions unaffected (receptor internalization, inhibition of cAMP production and ERK1/2 activation).

AGR1.135 and AGR1.137 antagonists act by direct binding on CXCR4

We developed a fluorescence spectroscopy strategy and FRET assays to formally examine the ability of the compounds to bind CXCR4, but our efforts were unsuccessful. AGR1.135 exhibited a yellow color that interfered with these techniques, and AGR1.137 did not alter the FRET efficiency of CXCR4 dimers (Supplementary Fig. 7). Therefore, we designed an *in silico* approach including a final step of molecular dynamics (MD) analysis, which allowed us to define the CXCR4 residues involved in the binding of the compounds, and to generate point mutant receptors to determine the inhibitory effect of AGR1.135, AGR1.137 and AGR1.131 (control) on CXCL12-mediated chemotaxis. We initially performed a binding-site search using PELE software, a Monte Carlo-based technique (35), to identify the most promising CXCR4 binding sites of the selected compounds. AGR1.135 and AGR1.137 exhibited one of the most stable trajectories (less than 0.25$Å^2$ RMSD fluctuation) upon binding to the cleft formed by TMV and TMVI (Fig. 5A, B). By contrast, the most stable trajectories for AGR1.131 corresponded to the binding to a contiguous region located between TMI and TMVII (Fig. 5C). Once the binding site of the compounds was confirmed, we performed docking studies to identify the best poses as a starting point for further studies.

Results showed that AGR1.135 and AGR1.137 shared a similar binding mode (Fig. 5A, B) and might interact with CXCR4 in a molecular cavity between TMV and TMVI. Specifically, AGR1.135 formed hydrogen bonds with residues G207, Y256 and R235, although a pi-cation interaction was also possible in the latter case (Fig. 5A). Additionally, AGR1.135 could interact hydrophobically with several residues in TMV and TMVI, comprising two independent cavities of 434 and 1,381 $Å^3$, respectively (Supplementary Fig. 8A), which were not connected to the orthosteric site as indicated by SURFNET analysis (36) included in the PDBsum server (37) (Supplementary Fig. 8A). AGR1.137 could use the carboxyl group of V124 in TMIII and overlap with AGR1.135 binding in the cavity, interacting with the other 19 residues scattered between TMV and VI to create several interaction surfaces of 790 and 580 $Å^3$ (Fig. 5B; Supplementary Fig. 8B). AGR1.137 did not have the phenyl ring present in AGR1.135, likely explaining why it did not interact with residue R235 (Fig. 1C, Fig. 5B). Regarding AGR1.131, its best binding site was the pocket between helices TMI and TMVII (Fig. 5C), which is likely the reason why this compound had no effect on receptor oligomerization. These results were also confirmed using the CXCR4 model predicted by AlphaFold (29) and the sequence available under UniProt entry P61073.

To confirm the putative binding sites for AGR1.135 and AGR1.137, we generated several point mutations in CXCR4 using the identified residues that were not critical for signal transduction processes. These included the G207I, L208K, R235L, F249L and Y256F mutants. We also included the mutations I204K, S260A as controls, as they are located in the same region (TMV-TMVI cleft) but did not interfere with compound binding or the CXCL12 binding site. All CXCR4 mutants were normally expressed at the cell surface, as demonstrated by flow cytometry (Supplementary Fig.

9A), and were fully functional in CXCL12-mediated cell chemotaxis assays (Supplementary Fig. 9B).

We next used transient transfection of $JK^{-/-}$ cells with each of the mutants or with wild-type CXCR4 treated with AGR1.131, AGR1.135 or AGR1.137 (50 μM) prior to assessing migration to CXCL12 gradients in Transwell chambers. The antagonistic effect of AGR1.135 was reversed in cells transfected with CXCR4 carrying the specific point mutations L208K or Y256F (Fig. 6A). Similarly, the antagonistic effect of AGR1.137 was reversed in cells expressing G207I, L208K, R235L, F249L or Y256F mutants (Fig. 6A). Collectively, these data indicate that the selected compounds bind directly to CXCR4, that the L208 and Y256 residues in the CXCR4 TMVI domain are critical for the allosteric modulation activities of AGR1.135 and AGR1.137, and that other residues, G207, R235 and F249, are also important for the effect of AGR1.137 (Fig. 6A-D).

AGR1.137 treatment reduces tumor volume and dissemination in a zebrafish xenograft model

Finally, to determine the *in vivo* activity of the selected compounds, we considered the role of CXCR4 in tumor growth and metastasis (10, 15) and the transparency of zebrafish larvae to develop a tumor xenograft (ZTX) model. ZTX models have proven to be a powerful complementary *in vivo* system for research in oncology and tumor biology (38, 39), particularly for early tumor invasion and dissemination.

We used HeLa cells, which are derived from a human cervical tumor that expresses functional CXCR4 (Fig. 7A, B; Supplementary Fig. 10A), and which have been used previously in ZTX models (40, 41). HeLa cells respond to CXCL12 gradients in directional cell migration assays (Fig. 7A, B) and, as in Jurkat cells, their treatment with AGR1.137 but not with AGR1.131 (control) abolished directional cell migration toward a CXCL12 gradient (Fig. 7A, B, Supplementary videos 9–12). As a control for specificity, CXCL2-driven HeLa cell migration was unaffected by AGR1.137 treatment but was blocked by the specific CXCR2 inhibitor AZD5069 (Fig. 7C, D).

DiI-labeled HeLa cells treated with vehicle, AMD3100, AGR1.135, AGR1.137 or AGR1.131 (control) were implanted subcutaneously into the perivitelline space of 2-day-old zebrafish larvae. Images of each tumor-bearing larva were taken immediately after implantation and 3 days later, and the relative change in tumor size was determined. AGR1.135 was discarded because most of the treated larvae died, indicating a toxic effect. By contrast, AGR1.137 was not toxic to embryos, as shown by visual inspection of the larvae at the end of the experiment, and no other non-lethal toxic phenotypes (e.g., pericardial edema, head and tail necrosis, head or tail malformation, brain hemorrhage or yolk sac edema) were evident. AGR1.137 10 μM had no effect but 50 μM significantly reduced the size of high-intensity red fluorescent tumors (~40%), whereas the control AGR1.131 had no effect on tumor size as compared with vehicle-treated fish, and AMD3100 reduced tumor size to a similar extent as AGR1.137 (Fig. 7E). These data demonstrate an antiproliferative effect of AGR1.137 on HeLa cells. As a control, AGR1.137 also reduced HeLa cell proliferation in *in vitro* assays mimicking the *in vivo* experimental conditions in zebrafish (addition of compounds every 24 hrs for 3 days) (Supplementary Fig. 10B). We next evaluated the ability of the compounds to block cell dissemination by determining the number of labeled HeLa cells that appeared in the main metastatic niche at the caudal hematopoietic plexus. AGR1.137 50 μM treatment of HeLa cells reduced cell dissemination by ~40% (Fig. 7F), strongly suggesting that this compound exerts anti-metastatic effects. As control AMD3100 (10 μM) treatment also reduced cell dissemination (~30%). The combined ability of AGR1.137 to reduce proliferation and dissemination *in vivo* is likely to have therapeutic value.

Discussion

The CXCL12/CXCR4 axis is involved in myriad functions, including leukocyte recruitment, embryogenesis, vascular development, hematopoiesis, heart development, nervous system organization, tumor progression, autoimmune diseases and, together with CD4, CXCR4 is a co-

receptor used by HIV-1 to infect immune cells (42). This evidence supports the clinical interest in developing specific antagonists to modulate or directly block CXCR4 functions.

Plerixafor (AMD3100), which blocks ligand binding, is the first FDA-approved CXCR4 antagonist, and is used in peripheral blood stem cell transplantation regimens (43). Unfortunately, its clinical use is limited by poor pharmacokinetics and adverse effects associated with long-term administration (44, 45). These limitations and the poor clinical success of other chemokine receptor antagonists have prompted the search for alternative strategies to block chemokine receptor functions.

As for other GPCRs, CXCR4 exists in multiple conformations on the plasma membrane and interacts with other chemokine receptors (46) and with cell surface proteins such as the T-cell receptor (47), CD4 and other receptors (20, 42). Advanced imaging-based techniques such as SPT-TIRF and super-resolution microscopy have revealed the existence of CXCR4 monomers, dimers and oligomers at the cell membrane, diffusing into the lipid bilayer. CXCL12 binding induces receptor nanoclustering and reduces the proportion of monomers and dimers, which is essential for the full activation of CXCR4 signaling and the correct orientation of the cell towards chemokine gradients (20, 21, 34). $CXCR4R^{334X}$, a truncated mutant receptor that causes WHIM syndrome, fails to form nanoclusters in the presence of CXCL12 and is unable to sense chemoattractant gradients, although it remains capable of triggering Ca^{2+} flux and other signaling pathways (21). Due to the inability of CXCR4 to associate with the actin cytoskeleton, β-arrestin-1-deficient cells do not exhibit CXCL12-mediated receptor nanoclustering (21). Similar observations are seen CXCR4-expressing cells treated with latrunculin A, an inhibitor of actin polymerization (20, 42). Collectively, these observations suggest that receptor nanoclustering is a ligand-mediated process that requires activation of a signaling pathway involving β-arrestin-1 and proper actin cytoskeletal dynamics.

CXCR4 contains a pocket formed by two TM domains, TMV and TMVI, and mutations in this pocket can affect the dynamics of the conformational changes triggered by ligand binding. The mechanism involves the initial interaction of CXCL12 with the extracellular region of CXCR4, which promotes conformational changes at a series of hydrophobic residues located mainly on TMV and TMVI that continue downward in the TM domains of the receptor, ultimately allowing G-protein interaction and activation (22). Interestingly, a complete alignment of GPCR class A chemokine receptors reveals that hydrophobic residues in this region that are not in highly conserved positions can have any side-chain length at any site depending on the receptor, likely conferring a specificity that could be exploited for the discovery and/or design of new antagonists (48) (Supplementary Fig. 11). A recent report using meta-dynamic approaches highlights the importance of the TM regions for dimerization and oligomerization of CXCR4 and CCR5 (24).

The majority of GPCR-targeted drugs are functionally active by binding to the orthosteric site of the receptor. Nonetheless, these orthosteric compounds have off-target effects and poor selectivity due to highly homologous receptor orthosteric sites and abrogation of spatial and/or temporal endogenous signaling patterns. The use of negative allosteric modulators can tune receptor-associated functions without affecting the orthosteric binding site. For example, the negative allosteric modulator 873140 blocks CCL3 binding to CCR5 but does not alter CCL5 binding (49). Maraviroc, an anti-HIV-1 molecule is an allosteric modulator of CCR5 that regulates CCR5 dimer populations, their localization to the cell membrane, and their subsequent subcellular trafficking (50). Allosteric modulators of the kappa opioid receptor have been demonstrated to have desirable antinociceptive and antipruritic effects without sedative and dissociative effects in rodent models (51).

To identify new allosteric modulators, we focused on an *in silico* screen targeting the pocket in CXCR4 created by TMV and TMVI, without affecting the residues involved in CXCL12 binding (20, 42). Subsequent functional analysis of the selected compounds identified AGR1.135, AGR1.137, and AGR1.131. The latter, although selected with the same *in silico* criteria, showed a better docking score on a cleft between TMI and TMVII and did not block CXCL12-mediated chemotaxis. In addition, the compounds had no effect on CXCR2-mediated cell migration, supporting their specificity for CXCR4. The three compounds had a common core of (4-(1-benzyl-

1H-1,2,3-triazol-4-yl)phenyl) methanol but different side chains carrying amines of different length and chemical nature, which may explain their different biological activity. AGR1.131 seems to be more flexible than the other two compounds. It has an amide group and an additional CH2 linking the phenyl ring. The absence of the piperidine or piperazine rings in AGR1.131 may be related to its lack of activity on CXCR4. AGR1.137 is expected to be less hydrophilic than AGR1.135, as it contains only one tertiary amine while AGR1.135 contains two. These tertiary amines are susceptible to protonation to form a quaternary ammonium salt. In addition, the ethyl ester group present in AGR1.137 is less polar than the nitro group in AGR1.135, a feature that could affect the solubility of the compounds as well as their interactions on the receptor.

SPT-TIRF analysis showed that AGR1.135 and AGR1.137 abolished the ability of CXCL12 to induce CXCR4 nanoclustering and altered receptor dynamics at the cell membrane. Moreover, as expected from the selection criteria, these compounds had no effect on CXCL12 binding, receptor internalization, cAMP-production, or ERK1/2 or PI3K activation. The results also suggested the existence of different CXCR4 conformational states responsible for the activation of different signaling pathways. It is well known that GPCRs are thought to reside on the plasma membrane in equilibrium between distinct states, depending on complex allosteric interactions and ligand-induced conformational changes, as well as cell-specific parameters (52, 53). The emerging view is that GPCRs are highly dynamic proteins, and ligands with varying pharmacological properties differentially modulate the balance between multiple conformations. In the absence of agonists, the receptor population is dominated by conformations closely related to those observed in inactive-state crystal structures (54). While agonist binding drives the receptor population towards conformations similar to those in active-state structures, a mixture of inactive and active conformations remains, reflecting “loose” or incomplete allosteric coupling between the orthosteric and transducer pockets (55). Surprisingly, for some GPCRs, and under some experimental conditions, a substantial fraction of unliganded receptors are already in an active-like conformation, which may be related to their level of basal or constitutive signaling (56, 57). These basal activities can be modulated by ligands of varying potency. Full agonists can induce the maximal signaling response, whereas partial agonists and inverse agonists promote submaximal signaling and decrease basal activity, respectively. In addition, some ligands are known to be biased by selectively activating specific receptor-associated pathways at the expense of others, supporting the existence of distinct receptor conformations (58). Studies on the β2-adrenergic receptor support the idea that different agonists stabilize different receptor conformations (59), raising the possibility that allosteric ligands shift the balance to favor a particular receptor conformation. Biased ligands or allosteric modulators may thus achieve their distinctive signaling profiles by modulating this distribution of receptor conformations (60). For example, some analogs of angiotensin II do not appreciably activate Gq signaling (e.g., increase IP3 and Ca^{2+}), but still induce receptor phosphorylation, internalization, and mitogen-activated protein kinase (MAPK) signaling (61). Some of these ligands activate Gi and G_{12} in bioluminescence resonance energy transfer experiments (62). A similar observation has been described for CCR5, where some chemokine analogs promote G protein subtype-specific signaling bias (63). Structural analyses of various GPCRs in the presence of different ligands vary considerably in the extent and nature of the conformational changes in the orthosteric ligand-binding site upon agonist binding (64). However, these changes modify conserved motifs in the interior of the receptor core and induce common conformational changes in the intracellular site involved in signal transduction.

We failed to detect a direct interaction between the compounds and CXCR4. Therefore, we employed an indirect strategy using docking and MD to confirm this interaction and to predict the most promising CXCR4 binding residues involved in the binding of the selected compounds. This methodology was combined with a mutation strategy to confirm the binding specificity of the allosteric modulators. AGR1.135 and AGR1.137 lost antagonism on cells expressing CXCR4L208K or CXCR4Y256F. LigPlot+ analysis (65) indicated that AGR1.135 interacts directly with the receptor through a hydrogen bond with Y256 (Supplementary Fig.12). When this residue was mutated to F, one of the anchors for the compound was lost, weakening the potential interaction in the region of the upper anchor point. It is not clear how the Y256F mutation affects the binding of AGR1.137, but

other potential contacts cannot be ruled out since this part of the compound is identical in both AGR1.135 and AGR1.137. This is particularly important for its neighboring residues in the alpha helix, F249, L208, as shown in the 3ODU structure, which appear to be directly involved in the interaction of both compounds with CXCR4. Alternatively, we cannot rule out an interaction between Y256 and other TMs or lipids stabilizing the overall structure, which could reverse the effect of the mutant at a later stage.

The inhibitory behavior of AGR1.137 was also lost in cells expressing CXCR4F249L. Both allosteric modulators showed stable trajectories bound to CXCR4, being able to occupy the cavity formed by TMV and TMVI and maintaining specific hydrogen-bond interactions and hydrophobic contacts along the interaction surface. Moreover, there were differences between the two modulators in terms of binding; whereas AGR1.135 showed a more elaborated interaction due to three hydrogen bonds that curved the position of the molecule as it fitted into the cavity formed by TMV and TMVI, AGR1.137 bound to the carboxyl group of V124 in TMIII, penetrating deeper into the receptor and then following the TMV–TMVI cleft down to the bottom of the cavity. In both cases, hydrophobic interactions with several residues of the receptor allowed the compounds to attach to the cleft and partially abolish the movement of TMVI required for complete signal transmission (22). AGR1.137 allowed for a different conformation of the TMVI helix, one that is more shifted towards an active conformation (Supplementary Fig. 13). By contrast, predictions indicated that AGR1.135 established hydrogen bonds with residues G207, Y256 and R235, complemented by hydrophobic interactions in TMV and TMVI, comprising a total surface area of 400 $Å^2$ with a length of 20 Å. AGR1.135 appeared in a more stretched conformation than AGR1.137 due to its ability to share a hydrogen bond with R235. AGR1.137 could use the carboxyl group of V124 in TMIII and overlap with AGR1.135 binding in the cavity, interacting with the other 19 residues scattered between TMV and VI to create an interaction surface of 370 $Å^2$ along 20 Å. AGR1.137 did not have the phenyl ring present in AGR1.135, resulting in a shorter moiety with greater difficulty in reaching the lower part of TMVI where R235 sits. These differences may explain the different behavior of the two compounds in cells expressing CXCR4F249L and suggest a closer interaction between the triazole group of AGR1.137 and the F249 residue on CXCR4. The main difference between AGR1.135 and AGR1.137 is the way in which the triazole group interacts with F249 and L216. For AGR1.137, the three groups are aligned in a parallel organization, which appears to be more effective. This is because there is only one hydrogen bond with V124, which significantly contributes to a more flexible interaction. Contrastingly, AGR1.135 has two anchor points on either side of the link, which, together with the interaction with F248, results in a more rigid position in the cleft. When F249 is changed to L and the two most likely rotamers resulting from the mutation are shown, rotamer B appears in close proximity to the compound, which may cause the binding to either be displaced or to adopt an alternative conformation that is easier to bind into the cleft. As mentioned above, it is likely that AGR1.135 can displace the mutant rotamer and bind more easily to the cleft due to its higher affinity. This region has also been implicated in the attachment of the orthosteric chemokine binding pocket to the lipid bilayer. F249 in CXCR4 as well as L132, V214 and L216 are involved in cholesterol binding (24). Thus, AGR1.137 may also affect cholesterol binding to CXCR4, thereby altering receptor oligomerization. The x-ray structure of CXCR4 with PDB code 3ODU (27) shows experimentally the presence of two fatty acid molecules in contact with both TMV and TMVI. Although further experiments are needed to clarify the mechanism, our results suggest that cholesterol and/or other lipids might also play an important role in CXCR4 oligomerization and function, as seen for other GPCRs (66). We have recently shown that the lipid composition of the cell membrane modulates CXCR4 oligomerization (67). The chemokine field is very complex, and most receptors can form dimers (homo- and heterodimers) as well as oligomers (68). It is therefore possible that the compounds might also affect the function of other chemokine receptors that heterodimerize with CXCR4. Nonetheless, it should be noted that the residues involved in CXCR4 oligomerization might differ from those involved in dimeric structures (24). These observations highlight the flexibility of the cavity and its potential for modulation. They also suggest that this cavity could be used to guide the development of a new generation of antagonists that, without affecting CXCL12 binding, allosterically modulate some GPCR-mediated functions without altering others.

CXCR4 overexpression contributes to tumor growth, invasion, angiogenesis, metastasis, relapse, and therapeutic resistance (69). As might be expected, CXCR4 antagonism has been shown to disrupt tumor-stromal interactions, sensitize cancer cells to cytotoxic drugs, and reduce tumor growth and metastatic burden (69). We used a ZTX model with HeLa cells expressing functional CXCR4 to evaluate the *in vivo* relevance of the selected antagonists. This is a well-studied model to assess tumor progression and metastasis (40). Our results demonstrate that AGR1.137 treatment of HeLa cells induce a dose-dependent reduction tumor size by 40% and limits cell dissemination in the absence of toxicity, suggesting that this compound has both anti-proliferative and anti-metastatic properties. While the effect of AGR1.137 on cell dissemination may be a consequence of its effect on tumorigenesis, it also abolished directed HeLa cell migration *in vitro*, supporting the inhibition of cell metastasis. The data support that AGR1.137 reduces tumor size and metastasis to a similar extent as AMD3100, but the allosteric modulator is highly hydrophobic and therefore it will be very important to improve solubility, increase affinity and optimize delivery, to facilitate its use, particularly *in vivo*. AMD3100, which is FDA-approved for clinical use, is an orthosteric inhibitor of CXCR4 that blocks the entire signaling cascade triggered by CXCL12. Indeed, we observed that AMD3100 treatment blocked CXCL12 binding, cAMP inhibition, calcium flux, cell adhesion and cell migration, whereas the effect of AGR1.137 was limited to CXCL12-mediated directed cell migration. Although AMD3100 was well tolerated by healthy volunteers in a single-dose study, it also promotes some mild and reversible events such as increases in white blood cell counts and variations in urinary calcium just outside the normal range (70). For the treatment of viral infections, continuous daily dosing of AMD3100 was impractical due to severe side effects including cardiac arrhythmias (71). It is therefore critical to control the timing of administration of AMD3100 in the context of therapeutic use. In addition, side effects after long-term administration are a potential problem. Shorter-term usage and lower doses would be fundamental keys to success in clinical use (72). The use of new drugs, such as the negative allosteric modulator AGR1.137, which block cell migration but do not affect other CXCL12-mediated signaling pathways, would, at least in theory, be more specific and produce fewer side effects.

Our data support the notion that CXCR4 is a flexible protein that can adopt a spectrum of conformations depending on several factors such as the presence of a bound ligand, the lipid composition of the cell membrane (67) and the presence of other interacting proteins. As with many GPCRs, the stabilization of distinct CXCR4 states is a key element to modulate its function (73, 74). Our data show that by targeting the pocket between TMV and TMVI in CXCR4, AGR1.137 can block (both *in vitro* and *in vivo*) CXCL12-mediated receptor nanoclustering and cell sensitivity to chemoattractant gradients without altering ligand binding, thus preserving other signaling events. Ultimately, these results demonstrate the value of stabilizing specific CXCR4 conformations and considering targets other than the ligand-binding site for the design of negative allosteric modulators to block a specific set of CXCR4 functions.

Materials and Methods

Cells and reagents

HeLa cells (CCL-2) and HEK293T cells (CRL-11268) were obtained from the American Type Culture Collection (Rockville, MD). Jurkat human leukemia $CD4^+$ cells were kindly provided by Dr. J. Alcamí (Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain). Where indicated, Jurkat cells lacking endogenous *CXCR4* ($JK^{-/-}$) (21) were transiently transfected with plasmids expressing wild-type or mutant CXCR4-AcGFP receptors (20 μg), as described (21). CXCR4 mutants were generated by PCR using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) with full-length CXCR4-AcGFP serving as a template, and specific primers (Supplementary Table 2).

Human peripheral blood mononuclear cells isolated from buffy coats were activated *in vitro* for 1 week with 20 U/ml of IL-2 (Teceleukin; Roche, Nutley, NJ) and 5 μg/ml phytohemagglutinin (Roche) to generate T cell blasts (34).

The following antibodies were used: monoclonal mouse anti-human CXCR4 (clone 44717) and phycoerythrin-conjugated human CXCR4 (clone 12G5; both from R&D Systems, Minneapolis, MN); purified mouse anti-human CD3 (clone HIT3a, BD Biosciences, Franklin Lakes, NJ); goat F(ab')2 anti-mouse IgG-PE (Southern Biotech, Birmingham, AL); and anti-phospho-AKT (Ser473), anti-AKT (#9272), anti-phospho-ERK1,2 (#9191) and anti-ERK (#9102) (all from Cell Signaling Technology, Danvers, MA). Human CXCL12, CXCL2 and CCL17 were obtained from PeproTech (Rocky Hill, NJ). Human CXCR4 was cloned into the pAcGFPm-N1 plasmid (Clontech Laboratories, Palo Alto, CA), as described (20).

Unless otherwise indicated, cells were pre-treated with 50 μM of the selected small compounds, 10 μM AMD3100 (Merck, Darmstadt, Germany), 1 μM AZD5069 (MedChemExpress, Monmouth Junction, NJ) or vehicle (DMSO) as control (30 min, 37°C, 5% CO_2).

Binding experiments were performed using CXCL12-Atto-700 (32), kindly donated by Prof. Marcus Thelen (Institute for Research in Biomedicine, Università della Svizzera italiana, Bellinzona, Switzerland).

Compounds database

All the compounds used in the present study were synthesized in the Centro de Investigaciones Biológicas Margarita Salas (CIB-CSIC, Madrid) following previously described procedures (75). Synthetic and analytical data are shown as Supplementary information. All the compounds have a purity ≥95% determined by high-performance liquid chromatography. Compounds are collected in the MBC library (CIB-CSIC) (25), which contains more than 2,000 drug-like compounds.

Computational studies of CXCR4 modulators

Protein modeling

The CXCR4 model was built on the SWISS-MODEL server (28) using the human CXCR4 sequence and the crystallographic structure of a CXCR4 dimer in complex with the small molecule IT1t as template (PDB code: 3ODU (27)). At the time we performed these analyses, several crystallographic structures of CXCR4 in complex with different molecules and peptides were deposited in the PDB. None of them were fully resolved because the N- and C-terminal ends of CXCR4 are very flexible loops. In addition, the CXCR4 constructs contained T4 lysozyme inserted between helices TMV and TMVI to increase the stability of the protein; this was a common strategy to facilitate crystallogenesis of GPCRs (26). Therefore, a CXCR4 homology model was generated using the SWISS-MODEL server (28). This program reconstructed the loop between TMV and TMVI, a domain of particular importance in this study and not present in any of the crystal structures available in the PDB.

The model was further optimized by adding hydrogens, ionizing the structure at pH 7.2, and adjusting size chain positions using the Maestro Protein Preparation Wizard tool included in the Schrödinger software package (76).

Virtual Screening

Virtual screening of ligands was performed from the compounds included in the MBC library using the "Glide" module as a docking tool on the CXCR4 model using standard-position (SP) and extra-position (XP) scoring functions. The center of the catalytic pocket was selected as the centroid of the grid. In the grid generation, a scaling factor of 1.0 in van der Waals radius scaling and a partial charge cutoff of 0.25 were used. A rescoring of the SP poses of each compound was then performed with the XP scoring function of Glide. The XP mode in Glide was used in the virtual

screening, the ligand sampling was flexible, epik state penalties were added, and an energy window of 2.5 kcal/mol was used for ring sampling. In the energy minimization step, the distance dependent dielectric constant was 4.0 with a maximum number of minimization steps of 100,000. In the clustering, poses were considered as duplicates and discarded if both RMS deviation is less than 0.5 Å and maximum atomic displacement is less than 1.3 Å.

Simulations were calculated with an Asus 1151 h170 LVX-GTX-980Ti workstation, with an Intel Core i7-6500 K Processor (12 M Cache, 3.40 GHz) and 16 GB DDR4 2133 MHz RAM, equipped with a Nvidia GeForce GTX 980Ti available for GPU (Graphics Processing Unit) computations.

The computational sequential workflow applied in this project included: i) Protein model construction, ii) Virtual screening (Glide), iii) PELE, iv) Docking (autoDock and Glide) and v) MD (AMBER).

PELE

A Protein Energy Landscape Exploration (35) study was carried out for the control molecule IT1t, which is a small druglike isothiourea derivative that has been crystallized in complex with CXCR4 (PDB code: 3ODU). PELE is a method to explore the energy landscape of a protein that can perform unconstrained ligand exploration to search its binding site. PELE combines a Monte Carlo stochastic approach with protein structure prediction algorithms. The script "Unconstrained ligand exploration and binding search" was used to calculate 20 PELE trajectories for each ligand of the PELE program, yielding 20 PELE trajectories for each ligand. From the best five trajectories, two of them presented clearly better binding energies, and corresponded to almost the same predicted pose of the molecule. Although the predicted binding mode was not exactly the same as that in the crystal structure, the approximation was very good, allowing the approach validation. Although PELE is a suitable technique to find potential binding sites, the predicted poses must be subsequently refined using docking programs.

Docking studies

All compounds were adjusted to a pH of 7.2 $\pm$ 0.2 using the Maestro LigPrep tool (77). The processed geometry of these ligands was optimized using Gaussian09 software (78), which calculates RESP (Restrained ElectroStatic Potential) charges and other molecular properties. The method used for calculations was Hartree-Fock/6-311++G(2d,2p), using the tight option to ensure adequate convergence. Subsequently, all ligands were parameterized using Antechamber (79). In this work, docking calculations were performed using Autodock v4.2 (80).

All docking assays were performed using a box that contained the intracellular half of the protein. The CXCL12 binding pocket was excluded from the search based on the experimental results showing that none of these molecules prevent CXCL12 binding (unpublished results). The grid was formed by 126×126×126 points with a spacing of 0.310 Å (~39×39×39 Å3). Ligand posing was performed using a stochastic search with Lamarckian Genetic Algorithms (GAs). GAs use a population of poses and combine them to generate offspring that are evaluated with the scoring function to select the best ones to create the next generation. The GA parameters were 200 runs, a population size of 150 and 2,500,000 evaluations The results were analyzed clustering the best 200 poses with a Root Mean Square Deviation (RMSD) of 2.0 Å.

Molecular Dynamics

The most promising docking poses from the docking studies were further studied using MD simulations to determine their stability and interactions with the receptor. The CXCR4-IT1t complex was used as a control to validate the protocol. Simulations were performed using AMBER14 (81) with ff14SB (82) and lipid14 (83) force fields in the NPT thermodynamic ensemble (constant pressure and temperature). All bonds involving hydrogen atoms were constrained with the SHAKE algorithm (84). To stabilize CXCR4 and more accurately reproduce the real environment in the MD

simulation, the system was embedded in a lipid bilayer using the Membrane Builder tool (85) from the CHARMM-GUI server. The membrane was composed of 175 molecules of the fatty acid 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) in each leaflet. The protein-membrane complex was solvated with TIP3 water molecules. Chloride ions were added up to a concentration of 0.15 M in water, and sodium ions were added to neutralize the system. Minimization was performed using 3500 Steepest Descent steps and 4500 Conjugate Gradient steps three times, firstly considering only hydrogens, then considering only water molecules and ions, and finally minimizing all atoms. Equilibration raises system temperature from 0 to 300 K at a constant volume, fixing everything but ions and water molecules. After thermalization, several density equilibration phases were performed. In the production phase 50 ns MD simulations without position restraints were calculated using a time step of 2 fs. Trajectories of the most interesting poses were extended to 150 ns. "Stable trajectories" have been defined to refer to those trajectories of the ligand-protein complexes whose RMSD does not fluctuate more than 0.25 Å. To estimate the affinity of the ligands for CXCR4, the binding energy of each representative structure obtained with the principal components analysis was calculated using the PRODIGY-LIGAND server (86).

Virtual screening validation

We also used a strategy based on the model of CXCR4 predicted by AlphaFold (87) and the sequence available under UniProt entry P61073 to validate our studies. We prepared the ligands using the OpenBabel (88), with a gasteiger charge assignment, and generated 10 conformers for each input ligand using OpenBabel. We then prepared the target structure with Openmm, removing all water and possible heteroatoms, and adding all missing atoms. We next predicted the target binding pockets with fPocket (89), p2rank (90), and AutoDock autosite (91). We chose only those pockets between TMV and TMVI. We merged the results of the three programs into so-called consensus pockets, as two pockets are said to be sufficiently similar if at least 75% of their surfaces are shared (92). From the consensus pockets, one pocket was significantly larger than the others and was, therefore, selected. We then docked the ligand conformers in this pocket using AutoDock GPU (93), LeDock (94), and Vina (95). The number of dockings varied from 210 to 287 poses. We scored each pose with the Vina score using ODDT (96). Then, we clustered the different solutions into groups whose maximum RMSD was 1Å. This resulted in 40 clusters; the representative of each cluster was the one with maximum Vina score, and confirmed that the selected compounds bound this pocket. When required, we calculated the binding affinity using Schrodinger's MM-GBSA procedure (97) in two ways: first, assuming that the ligand and target are fixed; second, with an energy minimization of all the atoms within a distance of 3 Å from the ligand.

Cleft volume calculations

Once we detected that several compounds bound CXCR4 in the target region, the cleavage properties were calculated by subtracting the compound structure. The resulting PDB was analyzed using the PDBsum server (98). Volume calculations were obtained using the SURFNET server to analyze surface clefts (36). The theoretical interaction surface between the selected compounds and CXCR4 and the atomic distances between the protein residues and the compounds was calculated using the PISA server (99).

Transwell migration assay

Cells (3×10^5) in 0.1 ml of RPMI medium containing 10 mM HEPES and 0.1% BSA were placed in the upper wells of 5-μm pore size transmigration chambers (Transwell, Costar, Corning, NY). CXCL12 (12.5 nM) or CCL17 (50 nM) in 0.6 ml of the same medium was added to the lower well. Cell migration was evaluated as described (120 min, 37ºC, 5% CO_2) (21).

Cell cycle analysis

Cells (5×10^5 cells/well) were collected from microplates and washed twice in PBS and then resuspended in 50 μl of detergent (DNA-Prep Reagent Kit; Beckman Coulter, Brea, CA) containing 10 ng/ml propidium iodide (DNA-Prep Reagent Kit; 30 min, 37°C). Cell cycle phases were analyzed by flow cytometry on a Beckman Coulter FC500 flow cytometer and results were expressed as the percentage of stained cells. H_2O_2 (10%) and staurosporine (10 μM) treatments were used as positive controls of cell death.

Single molecule TIRF imaging and analysis

Transfected cells expressing 8,500–22,000 receptors/cell (<4.5 particles/μm^2) were plated on glass-bottomed microwell dishes (MatTek Corp., Ashland, MA) coated with fibronectin (Sigma-Aldrich, St. Louis, MO: 20 μg/ml, 1 hr, 37°C) as a substrate for cell adhesion and to facilitate image aquisition, as TIRF determinations are restricted to the cell membrane in contact with the coverslip. These are conditions corresponding to the unstimulated cells. To observe the effect of the ligand, we coated dishes with CXCL12 (100 nM, 1 hr, 37°C); stimulated cells were incubated (20 min, 37°C, 5% CO_2) before image acquisition. Experiments were performed at 37ºC with 5% CO_2 using a total internal reflection fluorescence microscope (Leica AM TIRF inverted microscope; Leica Microsystems, Wetzlar, Germany). Image sequences of individual particles (500 frames) were then acquired at 49% laser power (488-nm diode laser) with a frame rate of 10 Hz (100 ms/frame). The penetration depth of the evanescent field was 90 nm. Particles were detected and tracked using the U-Track2 algorithm (100) implemented in MATLAB, as described (101). Mean spot intensity (MSI), number of mobile and immobile particles and diffusion coefficients (D_{1-4}) were calculated from the analysis of thousands of single trajectories over multiple cells (statistics provided in the respective figure captions) using described routines (101). The receptor number along individual trajectories was determined as reported (20), using the intensity of the monomeric protein CD86-AcGFP as a reference (102). Briefly, for each particle detected we measured the intensity of pixels around the particle (in a 3×3 pixel matrix whose central position is given by the coordinates x,y obtained by U-Track2), in each frame along its trajectory. We also estimated the particle background (K0) for each frame. K0 for each frame was calculated by locating the cell in the video and analyzing the intensities of the cell background in each frame. K0 was chosen as the gray value at a given quantile of this distribution (80%). The intensity value for each particle is then given by the difference between particle intensity and the background in each frame. To minimize photon fluctuations within a given frame, we considered the particle intensity as the average value (background subtracted) obtained over the first 20 frames. To ensure that no photobleaching events occurred within this range of frames that would affect the quantitation of the data, we measured the photobleaching times (s) from individual CD86-AcGFP (electroporated in Jurkat CD4+ cells, (103) particles that exclusively showed a single photobleaching step. Fitting of the distribution to a single exponential decay renders a $\tau 0$ value of ~5 s, which corresponds to 50 frames (100 ms/frame). The total number of receptors/particles was finally estimated by dividing the average particle intensity by the particle intensity arising from individual AcGFP molecules. To unequivocally identify the intensity emitted by an individual AcGFP, we used as a calibrator Jurkat $CD4^+$ cells electroporated with the monomeric CD86-AcGFP. We performed SPT experiments on CD86-AcGFP and analyzed the data in similar way as described above. Distribution of monomeric particle intensities was analyzed by Gaussian fitting, rendering a mean value of 980 ± 89 a.u. This value was then used as the monomer reference to estimate CXCR4-AcGFP particle size.

Internalization and flow cytometry studies

Cells (5×10^5 cells/well) were activated with CXCL12 (50 nM) at the indicated time points at 37ºC. After incubation, cells were washed twice with staining buffer and receptor internalization was determined by flow cytometry using an anti-CXCR4 mAb (clone 44717, 30 min, 4°C), followed by secondary staining with PE-coupled goat anti-mouse IgG (30 min, 4ºC). Results are expressed as a percentage of the mean fluorescence intensity of treated cells relative to that of untreated

cells. When required, cells were pretreated with the small compounds (50 μM, 30 min, 37ºC) or AMD3100 (1 μM, 30 min, 37ºC) before activation with CXCL12.

Western blotting

Cells (3×10^6) were activated with CXCL12 (50 nM) at the time points indicated and then lysed in RIPA detergent buffer supplemented with 1 mM PMSF, 10 μg/ml aprotinin, 10 μg/ml leupeptin and 10 μM sodium orthovanadate, for 30 min at 4ºC. Extracts were analyzed by western blotting using specific antibodies. Densitometric evaluation of blots was performed using ImageJ (NIH, Bethesda, MD). When required, cells were pretreated with the small compounds (50 μM, 30 min, 37ºC) or AMD3100 (1 μM, 30 min, 37ºC) before activation with CXCL12.

cAMP determination

cAMP levels were determined using the cAMP-Glo™ Max Assay (Promega, Madison, WI, in cells (5×10^5 cells/well) untreated or pre-treated with CXCL12 (50 nM, 5 min, 37°C) followed by the addition of forskolin (10 μM, 10 min, 37°C). When required, cells were pretreated with the small compounds (50 μM, 30 min, 37ºC) or AMD3100 (1 μM, 30 min, 37ºC) before activation with CXCL12.

CXCL12 binding

Cells (5×10^5 cells/well) were incubated with CXCL12 ATTO-700 (32), (30 min, 37°C) and maintained at 4°C before analyzing the bound fluorescence in a Beckman Coulter Cytoflex flow cytometer. Results are expressed as mean fluorescence intensity. When required, cells were pretreated with the small compounds (50 μM, 30 min, 37ºC) or AMD3100 (1 μM, 30 min, 37ºC) before activation with CXCL12.

Fluorescence Resonance Transfer Analysis by Sensitized Emission

HEK-293T cells transiently transfected at a fixed CXCR4-YFP:CXCR4-CFP ratio (15 μg and 9 μg, respectively) were treated or not with the compounds (50 μM, 30 min, 37ºC) or their vehicle (DMSO), and FRET efficiency was evaluated ($n = 3$, mean ± SEM, $***p \leq 0.001$). Emission light was quantified using the Wallac Envision 2104 Multilabel Reader (Perkin Elmer, Foster City, CA) equipped with a high-energy xenon flash lamp (donor: receptor fused to C-CFP, 8-nm bandwidth excitation filter at 405 nm; acceptor:receptor fused to YFP, 10 nm bandwidth excitation filter at 510 nm).

Actin polymerization

Cells (5×10^5 cells/well) were incubated with CXCL12 (50 nM, 37°C) or anti-CD3 (1 μg/ml, 37°C) at the indicated time periods and then fixed with 2% paraformaldehyde and transferred to ice for 10 min. Fixed cells were permeabilized with 0.01% saponin (10 min, 4ºC) and labeled with Phalloidin-TRITC (Merck, 30 min, 4ºC). After washing, bound fluorescence was analyzed by flow cytometry on a Beckman Coulter FC500 flow cytometer. When required, cells were pretreated with the small compounds (50 μM, 30 min, 37ºC) or AMD3100 (1 μM, 30 min, 37ºC) before activation with CXCL12.

Immunofluorescence analyses

Cells on fibronectin (20 μg/ml, Sigma)-coated glass slides were stimulated or not with 100 nM CXCL12 (5 min at 37 °C), fixed with 4% paraformaldehyde (10 min), permeabilized with 0.25% saponin (10 min), and stained with phalloidin-TRITC (Sigma-Merck; 30 min), all at room temperature (RT). Preparations were analyzed using a Leica TCS SP8 confocal multispectral microscope. When required, cells were pretreated with the small compounds (50 μM, 30 min, 37ºC) or AMD3100 (1 μM, 30 min, 37ºC) before activation with CXCL12.

Cell adhesion/migration on planar lipid bilayers

Planar lipid bilayers were prepared as reported (104). Briefly, unlabeled GPI-linked intercellular adhesion molecule 1 (ICAM-1) liposomes were mixed with 1,2-dioleoyl-phosphatidylcoline. Membranes were assembled in FCS2 chambers (Bioptechs, Butler, PA), blocked with PBS containing 2% FCS for 1 hr at RT, and coated with CXCL12 (200 nM, 30 min, RT). Cells (3×10^6 cells/ml) in PBS containing 0.5% FCS, 0.5 g/l D-glucose, 2 mM $MgCl_2$ and 0.5 mM $CaCl_2$ were then injected into the pre-warmed chamber (37ºC). Confocal fluorescence, differential interference contrast (DIC) and interference reflection microscopy (IRM) images were acquired on a Zeiss Axiovert LSM 510-META inverted microscope with a 40× oil-immersion objective. Imaris 7.0 software (Bitplane, Zurich, Switzerland) and ImageJ 1.49v were used for qualitative and quantitative analysis of cell dynamics parameters, fluorescence and IRM signals. The fluorescence signal of the planar bilayer in each case was established as the background fluorescence intensity. The frequency of adhesion (IRM^+ cells) per image field was estimated as [nº of cells showing IRM contact/total nº of cells (estimated by DIC)] × 100; similarly, we calculated the frequency of migration (cells showing and IRM contact and moving over time). When required, cells were pretreated with the small compounds (50 μM, 30 min, 37ºC) or AMD3100 (1 μM, 30 min, 37ºC) before activation with CXCL12.

Directional cell migration

Pre-treated cells were diluted to 10×10^6 cell/ml in RPMI medium containing 1% BSA and 20 mM HEPES (chemotaxis medium) and then seeded into the channel of a chemotaxis slide (μ-Slide Chemotaxis System, 80326; Ibidi, Munich, Germany) (1 hr, 37ºC, 5% CO_2). The reservoirs were then filled with chemotaxis medium containing 50 μM AGR1.137 or AGR1.131, or vehicle (DMSO) as a control condition, and 100 nM CXCL12 or CXCL2 was added to the right reservoir. Phase-contrast images were recorded over 20 hrs with a time lapse of 15 min using a Microfluor inverted microscope (Leica) with a 10× objective and equipped with an incubation system set to 5% CO_2 and 37°C. Single-cell tracking was evaluated by selecting the center of mass in each frame using the manual tracking plug-in tool in ImageJ. Spider plots, representing the trajectories of the tracked cells, forward migration index (FMI), and straightness values, were obtained using the Chemotaxis and Migration Tool (ImageJ).

Drug efficacy assay in zebrafish

HeLa cells were labeled with 8 μg/ml Fast-DiI™ oil (Cat #D3899; ThermoFisher Scientific, Waltham, MA) in RPMI medium supplemented with 2% FBS for 30 min at 37°C. Cells were then washed and filtered, and cell viability was determined using trypan blue-exclusion. Transgenic zebrafish Tg(fli1a:EGFP)y1 (105) were bred naturally and maintained in E3 embryo medium (deionized water containing 0.5 mM NaCl, 0.4 mM $CaCl_2$, 0.7 mM MgSO4 and 0.2 mM KCl, pH 7.2) supplemented with 0.2 mM PTU (E3/PTU) at 28.5°C. DiI-labeled HeLa cells were implanted into the dorsal perivitelline space of 2-day-old embryos, as described (38). Tumors were imaged within 2 hrs of implantation and tumor-baring embryos were treated with either vehicle (DMSO), AMD-3100 (1 and 10 μM) or with AGR1.131 or AGR1.137 (10 and 50 μM) for 3 days, followed by re-imaging. Changes in tumor size were evaluated as tumor area at day 3 divided by tumor area at day 0, and metastasis was evaluated as the number of cells disseminated to the caudal hematopoietic plexus.

Statistical analyses

All data were analyzed with GraphPad Prism software version 9 (GraphPad Inc., San Diego, CA). Cell migration in Transwell assays and in planar lipid bilayers and directional cell migration assays and cell polarization under the various conditions were analyzed to determine significant differences between means using one-way analysis of variance (ANOVA) followed by

Tukey's multiple comparison test. A two-tailed Mann-Whitney non-parametric test was used to analyze the diffusion coefficient (D_{1-4}) of single particles. We used contingency tables to compare two or more groups of categorical variables, such as the percentages of mobile or immobile particles, and these were compared using a Chi-square test with a two-tailed p-value. Statistical differences were reported as n.s. = not significant $p > 0.05$, *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$ and ****$p \leq 0.0001$.

Acknowledgments

This work was supported by grants from the Spanish Ministry of Science and Innovation (PID2020-114980RB-I00) Agencia Estatal de Investigación/Fondo Europeo de Desarrollo Regional (AEI/FEDER), Unión Europea, and by a grant from the Regional Government of Madrid P2022/BMD-7274. CS is supported by MCIN /AEI /10.13039/501100011033 / FEDER, UE) (PID2022-140651NB-I00). NEC is supported by a grant of the Spanish Ministry of Economy, Industry and Competitivity (RTI2018-096100B-100). RJS is supported by FSE/FEDER through the Instituto de Salud Carlos III (ISCIII; CP20/00043). EG-C was supported by the program Apoyos Centros de Excelencia S.O. of the Spanish Ministry of Science and Innovation (SEV-2017-0712). PM and SG are included in the doctoral program of the Department of Molecular Biosciences, Universidad Autónoma de Madrid, and are supported by the Fondo de Personal Investigador (FPI) program of the Spanish Ministry for Science and Innovation (BES2015-071302 and PRE2018-083201 respectively). RA-B is supported by the Garantía Juvenil program of the Regional Government of Madrid, Spain (CAM20_CNB_AI_07).

We also acknowledge the technical help of the Advance Light Microscopy Unit at the CNB/CSIC. Compounds AGR1.31, AGR1.35 and AGR1.37 are included in the patent PCT/ES2022/070379.

References

1. A. Zlotnik, O. Yoshie, The chemokine superfamily revisited. *Immunity* **36**, 705–716 (2012).

2. Y. Nie, Y. C. Han, Y. R. Zou, CXCR4 is required for the quiescence of primitive hematopoietic cells. *J Exp Med* **205**, 777–783 (2008).

3. D. E. Wright, E. P. Bowman, A. J. Wagers, E. C. Butcher, I. L. Weissman, Hematopoietic stem cells are uniquely selective in their migratory response to chemokines. *J Exp Med* **195**, 1145–1154 (2002).

4. T. Lapidot, O. Kollet, The essential roles of the chemokine SDF-1 and its receptor CXCR4 in human stem cell homing and repopulation of transplanted immune-deficient NOD/SCID and NOD/SCID/B2m(null) mice. *Leukemia* **16**, 1992–2003 (2002).

5. H. E. Broxmeyer, *et al.*, Stromal cell-derived factor-1/CXCL12 directly enhances survival/antiapoptosis of myeloid progenitor cells through CXCR4 and G(alpha)i proteins and enhances engraftment of competitive, repopulating stem cells. *J Leukoc Biol* **73**, 630–638 (2003).

6. T. Cash-Padgett, A. Sawa, H. Jaaro-Peled, Increased stereotypy in conditional Cxcr4 knockout mice. *Neurosci Res* **105**, 75–79 (2016).

7. K. Tachibana, *et al.*, The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract. *Nature* **393**, 591–594 (1998).

8. Y. Takabatake, *et al.*, The CXCL12 (SDF-1)/CXCR4 axis is essential for the development of renal vasculature. *J Am Soc Nephrol* **20**, 1714–1723 (2009).

9. E. A. Berger, P. M. Murphy, J. M. Farber, Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol* **17**, 657–700 (1999).

10. A. Müller, *et al.*, Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50–56 (2001).

11. J. M. Hall, K. S. Korach, Stromal cell-derived factor 1, a novel target of estrogen receptor action, mediates the mitogenic effects of estradiol in ovarian and breast cancer cells. *Mol Endocrinol* **17**, 792–803 (2003).

12. R. S. Taichman, *et al.*, Use of the stromal cell-derived factor-1/CXCR4 pathway in prostate cancer metastasis to bone. *Cancer Res* **62**, 1832–1837 (2002).

13. S. Scala, *et al.*, Human melanoma metastases express functional CXCR4. *Clin Cancer Res* **12**, 2427–2433 (2006).

14. H. Geminder, *et al.*, A possible role for CXCR4 and its ligand, the CXC chemokine stromal cell-derived factor-1, in the development of bone marrow metastases in neuroblastoma. *J Immunol* **167**, 4747–4757 (2001).

15. A. Orimo, *et al.*, Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**, 335–348 (2005).

16. Z. Liang, *et al.*, CXCR4/CXCL12 axis promotes VEGF-mediated tumor angiogenesis through Akt signaling pathway. *Biochem Biophys Res Commun* **359**, 716–722 (2007).

17. O. L. Bohn, *et al.*, Biomarker profile in breast carcinomas presenting with bone metastasis. *Int J Clin Exp Pathol* **3**, 139 (2010).

18. S. Y. Kim, *et al.*, Inhibition of the CXCR4/CXCL12 chemokine pathway reduces the development of murine pulmonary metastases. *Clin Exp Metastasis* **25**, 201–211 (2008).

19. E. De Clercq, Mozobil® (Plerixafor, AMD3100), 10 years after its approval by the US Food and Drug Administration. *Antivir Chem Chemother* **27** (2019).

20. L. Martínez-Muñoz, *et al.*, Separating Actin-Dependent Chemokine Receptor Nanoclustering from Dimerization Indicates a Role for Clustering in CXCR4 Signaling and Function. *Mol Cell* **70**, 106-119.e10 (2018).

21. E. M. García-Cuesta, *et al.*, Altered CXCR4 dynamics at the cell membrane impairs directed cell migration in WHIM syndrome patients. *Proc Natl Acad Sci U S A* **119** (2022).

22. M. P. Wescott, *et al.*, Signal transmission through the CXC chemokine receptor 4 (CXCR4) transmembrane helices. *Proc Natl Acad Sci U S A* **113**, 9928–9933 (2016).

23. Q. Zhou, *et al.*, Common activation mechanism of class A GPCRs. *Elife* **8** (2019).

24. D. Di Marino, P. Conflitti, S. Motta, V. Limongelli, Structural basis of dimerization of chemokine receptors CCR5 and CXCR4. *Nat Commun* **14**, 1–16 (2023).

25. V. Sebastián-Pérez, *et al.*, Medicinal and Biological Chemistry (MBC) Library: An Efficient Source of New Hits. *J Chem Inf Model* **57**, 2143–2151 (2017).

26. Y. Zou, W. I. Weis, B. K. Kobilka, N-Terminal T4 Lysozyme Fusion Facilitates Crystallization of a G Protein Coupled Receptor. *PLoS One* **7**, e46039 (2012).

27. B. Wu, *et al.*, Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066–1071 (2010).

28. A. Waterhouse, *et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296–W303 (2018).

29. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

30. B. Haribabu, *et al.*, Regulation of human chemokine receptors CXCR4. Role of phosphorylation in desensitization and internalization. *J Biol Chem* **272**, 28726–28731 (1997).

31. S. Venkatesan, J. J. Rose, R. Lodge, P. M. Murphy, J. F. Foley, Distinct mechanisms of agonist-induced endocytosis for human chemokine receptors CCR5 and CXCR4. *Mol Biol Cell* **14**, 3305–3324 (2003).

32. R. Ameti, *et al.*, Characterization of a chimeric chemokine as a specific ligand for ACKR3. *J Leukoc Biol* **104**, 391–400 (2018).

33. R. K. Ganju, *et al.*, The alpha-chemokine, stromal cell-derived factor-1alpha, binds to the transmembrane G-protein-coupled CXCR-4 receptor and activates multiple signal transduction pathways. *J Biol Chem* **273**, 23169–23175 (1998).

34. S. R. Gardeta, *et al.*, Sphingomyelin Depletion Inhibits CXCR4 Dynamics and CXCL12-Mediated Directed Cell Migration in Human T Cells. *Front Immunol* **13** (2022).

35. K. W. Borrelli, A. Vitalis, R. Alcantara, V. Guallar, PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J Chem Theory Comput* **1**, 1304–1311 (2005).

36. R. A. Laskowski, SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* **13**, 323–330 (1995).

37. R. A. Laskowski, *et al.*, PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends Biochem Sci* **22**, 488–490 (1997).

38. P. Rouhi, *et al.*, Hypoxia-induced metastasis model in embryonic zebrafish. *Nat Protoc* **5**, 1911–1918 (2010).

39. J. Xiao, E. Glasgow, S. Agarwal, Zebrafish Xenografts for Drug Discovery and Personalized Medicine. *Trends Cancer* **6**, 569–579 (2020).

40. H. K. Brown, K. Schiavone, S. Tazzyman, D. Heymann, T. J. A. Chico, Zebrafish xenograft models of cancer and metastasis for drug discovery. *Expert Opin Drug Discov* **12**, 379–389 (2017).

41. Y. You, *et al.*, Silibinin Induces G2/M Cell Cycle Arrest by Activating Drp1-Dependent Mitochondrial Fission in Cervical Cancer. *Front Pharmacol* **11** (2020).

42. T. Pozzobon, G. Goldoni, A. Viola, B. Molon, CXCR4 signaling in health and disease. *Immunol Lett* **177**, 6–15 (2016).

43. S. Fruehauf, W. J. Zeller, G. Calandra, Novel developments in stem cell mobilization: Focus on CXCR4. *Novel Developments in Stem Cell Mobilization: Focus on CXCR4* 1–496 (2012). https://doi.org/10.1007/978-1-4614-1960-0/COVER.

44. H. Y. Choi, C. S. Yong, B. K. Yoo, Plerixafor for stem cell mobilization in patients with non-Hodgkin's lymphoma and multiple myeloma. *Ann Pharmacother* **44**, 117–126 (2010).

45. C. W. Hendrix, *et al.*, Safety, pharmacokinetics, and antiviral activity of AMD3100, a selective CXCR4 receptor inhibitor, in HIV-1 infection. *J Acquir Immune Defic Syndr* **37**, 1253–1262 (2004).

46. L. Martínez-Muñoz, B. L. Holgado, C. Martínez-A, J. M. Rodríguez-Frade, M. Mellado, Chemokine receptor oligomerization: a further step toward chemokine function. *Immunol Lett* **145**, 23–29 (2012).

47. A. Kumar, *et al.*, CXCR4 physically associates with the T cell receptor to signal in T cells. *Immunity* **25**, 213–224 (2006).

48. N. Ma, A. K. Nivedha, N. Vaidehi, Allosteric communication regulates ligand-specific GPCR activity. *FEBS J* **288**, 2502–2512 (2021).

49. C. Watson, S. Jenkinson, W. Kazmierski, T. Kenakin, The CCR5 receptor-based mechanism of action of 873140, a potent allosteric noncompetitive HIV entry inhibitor. *Mol Pharmacol* **67**, 1268–1282 (2005).

50. J. Jin, *et al.*, CCR5 adopts three homodimeric conformations that control cell surface delivery. *Sci Signal* **11** (2018).

51. T. F. Brust, *et al.*, Biased agonists of the kappa opioid receptor suppress pain and itch without causing sedation or dysphoria. *Sci Signal* **9** (2016).

52. G. Vauquelin, I. Van Liefde, G protein-coupled receptors: a count of 1001 conformations. *Fundam Clin Pharmacol* **19**, 45–56 (2005).

53. J. Wess, S. J. Han, S. K. Kim, K. A. Jacobson, J. H. Li, Conformational changes involved in G-protein-coupled-receptor activation. *Trends Pharmacol Sci* **29**, 616–625 (2008).

54. A. Manglik, *et al.*, Structural Insights into the Dynamic Process of β2-Adrenergic Receptor Signaling. *Cell* **161**, 1101–1111 (2015).

55. R. O. Dror, *et al.*, Activation mechanism of the β 2-adrenergic receptor. *Proc Natl Acad Sci U S A* **108**, 18684–18689 (2011).

56. D. P. Staus, L. M. Wingler, D. Pichugin, R. S. Prosser, R. J. Lefkowitz, Detergent- And phospholipid-based reconstitution systems have differential effects on constitutive activity of G-protein– coupled receptors. *Journal of Biological Chemistry* **294**, 13218–13223 (2019).

57. L. Ye, N. Van Eps, M. Zimmer, O. P. Ernst, R. Scott Prosser, Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature* **533**, 265–268 (2016).

58. T. Kenakin, New concepts in pharmacological efficacy at 7TM receptors: IUPHAR review 2. *Br J Pharmacol* **168**, 554–575 (2013).

59. X. Yao, *et al.*, Coupling ligand structure to specific conformational switches in the beta2-adrenoceptor. *Nat Chem Biol* **2**, 417–422 (2006).

60. L. M. Wingler, R. J. Lefkowitz, Conformational Basis of G Protein-Coupled Receptor Signaling Versatility. *Trends Cell Biol* **30**, 736–747 (2020).

61. H. Wei, *et al.*, Independent beta-arrestin 2 and G protein-mediated pathways for angiotensin II activation of extracellular signal-regulated kinases 1 and 2. *Proc Natl Acad Sci U S A* **100**, 10782–10787 (2003).

62. Y. Namkung, *et al.*, Functional selectivity profiling of the angiotensin II type 1 receptor using pathway-wide BRET signaling sensors. *Sci Signal* **11** (2018).

63. E. Lorenzen, *et al.*, G protein subtype–specific signaling bias in a series of CCR5 chemokine analogs. *Sci Signal* **11** (2018).

64. A. J. Venkatakrishnan, *et al.*, Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* **536**, 484–487 (2016).

65. R. A. Laskowski, M. B. Swindells, LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778–2786 (2011).

66. J. Jakubík, E. E. El-Fakahany, Allosteric Modulation of GPCRs of Class A by Cholesterol. *Int J Mol Sci* **22**, 1–18 (2021).

67. S. R. Gardeta, *et al.*, Sphingomyelin Depletion Inhibits CXCR4 Dynamics and CXCL12-Mediated Directed Cell Migration in Human T Cells. *Front Immunol* **13** (2022).

68. L. Martínez-Muñoz, *et al.*, Receptor oligomerization: A pivotal mechanism for regulating chemokine function. *Pharmacol Ther* [Preprint] (2011). Available at: https://pubmed.ncbi.nlm.nih.gov/21600920/ [Accessed 6 May 2024].

69. S. Chatterjee, B. Behnam Azad, S. Nimmagadda, The intricate role of CXCR4 in cancer. *Adv Cancer Res* **124**, 31–82 (2014).

70. C. W. Hendrix, *et al.*, Pharmacokinetics and safety of AMD-3100, a novel antagonist of the CXCR-4 chemokine receptor, in human volunteers. *Antimicrob Agents Chemother* **44**, 1667–1673 (2000).

71. E. De Clercq, AMD3100/CXCR4 Inhibitor. *Front Immunol* **6** (2015).

72. T. Liu, X. Li, S. You, S. S. Bhuyan, L. Dong, Effectiveness of AMD3100 in treatment of leukemia and solid tumors: from original discovery to use in current clinical practice. *Exp Hematol Oncol* **5** (2016).

73. K. Bourque, *et al.*, Distinct Conformational Dynamics of Three G Protein-Coupled Receptors Measured Using FlAsH-BRET Biosensors. *Front Endocrinol (Lausanne)* **8** (2017).

74. R. Nygaard, *et al.*, The dynamic process of β(2)-adrenergic receptor activation. *Cell* **152**, 532–542 (2013).

75. V. Sebastián-Pérez, *et al.*, Medicinal and Biological Chemistry (MBC) Library: An Efficient Source of New Hits. *J Chem Inf Model* **57**, 2143–2151 (2017).

76. G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* **27**, 221–234 (2013).

77. Schrödinger Resease 2016-3: LigPrep. Schrödinger, LLC, New York, NY, 2016 (2016).

78. M. J. Frisch, *et al.*, Gaussian 09. Revision A. 02. [Preprint] (2016).

79. J. Wang, W. Wang, P. A. Kollman, D. A. Case, Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* **25**, 247–260 (2006).

80. G. M. Morris, *et al.*, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30**, 2785–2791 (2009).

81. D. A. Case, *et al.*, Amber 14. (2014).

82. J. A. Maier, *et al.*, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713 (2015).

83. C. J. Dickson, *et al.*, Lipid14: The Amber Lipid Force Field. *J Chem Theory Comput* **10**, 865–879 (2014).

84. R. A. Lippert, *et al.*, A common, avoidable source of error in molecular dynamics integrators. *J Chem Phys* **126** (2007).

85. S. Jo, J. B. Lim, J. B. Klauda, W. Im, CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys J* **97**, 50–58 (2009).

86. A. Vangone, A. Bonvin, PRODIGY: A Contact-based Predictor of Binding Affinity in Protein-protein Complexes. *Bio Protoc* **7** (2017).

87. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

88. N. M. O'Boyle, *et al.*, Open Babel: An open chemical toolbox. *J Cheminform* **3** (2011).

89. V. Le Guilloux, P. Schmidtke, P. Tuffery, Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10** (2009).

90. R. Krivák, D. Hoksza, P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J Cheminform* **10** (2018).

91. P. A. Ravindranath, M. F. Sanner, AutoSite: an automated approach for pseudo-ligands prediction-from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics* **32**, 3142–3149 (2016).

92. D. Del Hoyo, *et al.*, Scipion-Chem: An Open Platform for Virtual Drug Screening. *J Chem Inf Model* **63**, 7873–7885 (2023).

93. D. Santos-Martins, *et al.*, Accelerating AutoDock4 with GPUs and Gradient-Based Local Search. *J Chem Theory Comput* **17**, 1060 (2021).

94. W.-T. Chu, *et al.*, Using LeDock as a docking tool for computational drug design. *IOP Conf Ser Earth Environ Sci* **218**, 012143 (2019).

95. J. Eberhardt, D. Santos-Martins, A. F. Tillack, S. Forli, AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model* **61**, 3891–3898 (2021).

96. M. Wójcikowski, P. Zielenkiewicz, P. Siedlecki, Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field. *J Cheminform* **7**, 26 (2015).

97. P. A. Greenidge, C. Kramer, J. C. Mozziconacci, R. M. Wolf, MM/GBSA binding energy prediction on the PDBbind data set: successes, failures, and directions for further improvement. *J Chem Inf Model* **53**, 201–209 (2013).

98. R. A. Laskowski, J. Jabłońska, L. Pravda, R. S. Vařeková, J. M. Thornton, PDBsum: Structural summaries of PDB entries. *Protein Sci* **27**, 129–134 (2018).

99. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774–797 (2007).

100. K. Jaqaman, *et al.*, Robust single-particle tracking in live-cell time-lapse sequences. *Nat Methods* **5**, 695–702 (2008).

101. C. O. S. Sorzano, *et al.*, Image processing protocol for the analysis of the diffusion and cluster size of membrane receptors by fluorescence microscopy. *Journal of Visualized Experiments* **2019** (2019).

102. S. Dorsch, K. N. Klotz, S. Engelhardt, M. J. Lohse, M. Bünemann, Analysis of receptor oligomerization by FRAP microscopy. *Nat Methods* **6**, 225–230 (2009).

103. S. Dorsch, K. N. Klotz, S. Engelhardt, M. J. Lohse, M. Bünemann, Analysis of receptor oligomerization by FRAP microscopy. *Nat Methods* **6**, 225–230 (2009).

104. Y. R. Carrasco, S. J. Fleire, T. Cameron, M. L. Dustin, F. D. Batista, LFA-1/ICAM-1 interaction lowers the threshold of B cell activation by facilitating B cell adhesion and synapse formation. *Immunity* **20**, 589–599 (2004).

105. N. D. Lawson, B. M. Weinstein, In vivo imaging of embryonic vascular development using transgenic zebrafish. *Dev Biol* **248**, 307–318 (2002).

Figures

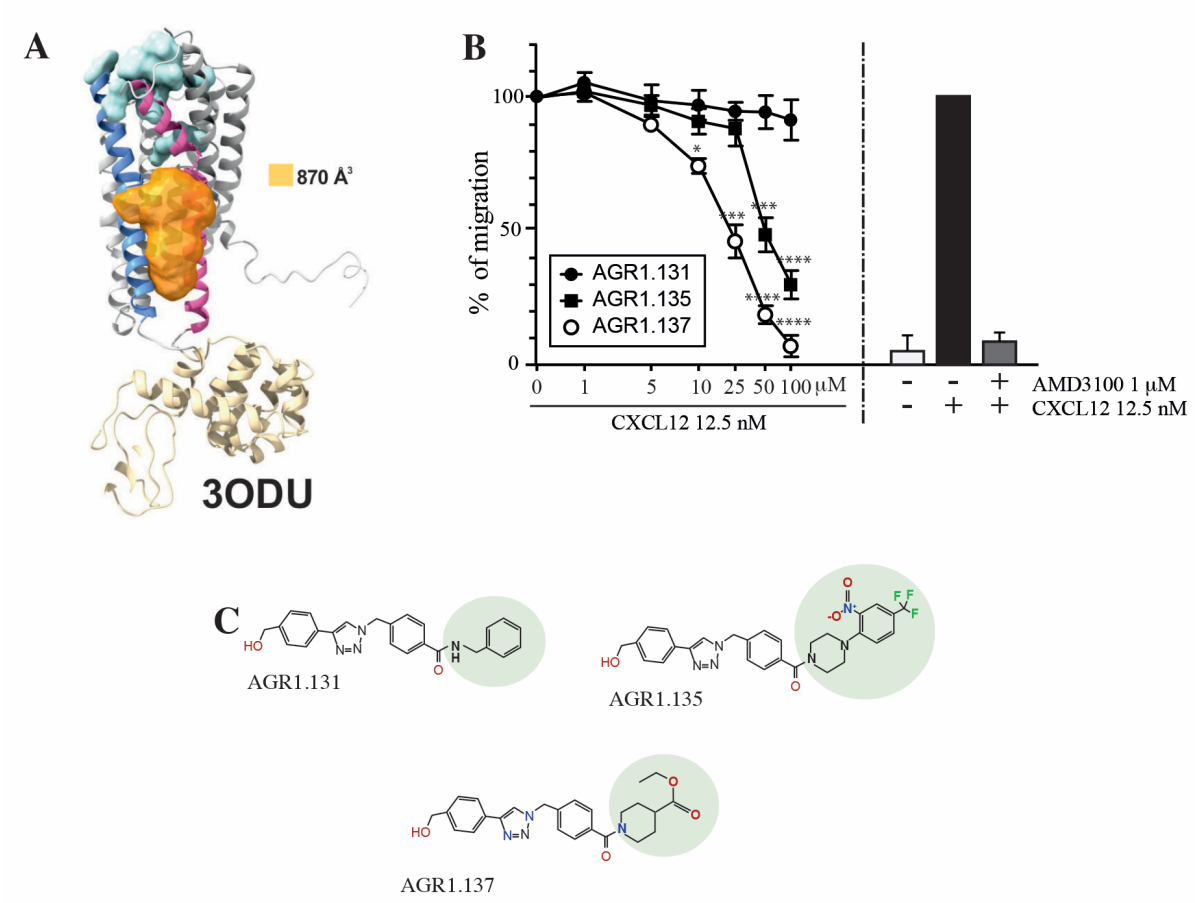


Figure 1. Screening for small compound antagonists acting at the oligomerization site of CXCR4. A) Cartoon and surface representation of CXCR4 and the cleft identified between TMV and TMVI. The protein structure is shown in gray, with TMV and TMVI colored in blue and pink, respectively. In green, residues involved in CXCL12 binding. The cavity identified by SurfNet is shown in orange. Volume for the cavity is measured and shown in Å^3. T4 lysozyme inserted between TMV and TMVI in the crystallized version of CXCR4 (PDB: 3ODU) is also shown in yellow. **B)** Dose-response curve of the selected antagonists in Jurkat cell migration experiments in response to 12.5 nM CXCL12. Data are shown as percentage of migrating cells (mean ± SD; n = 5; * ≤ 0.05, *** p ≤ 0.001, **** p ≤ 0.0001). **C**) Chemical structure of the selected compounds (AGR1.131, AGR1.135 and AGR1.137). The differences between the lateral chains of the three compounds are shaded in green.

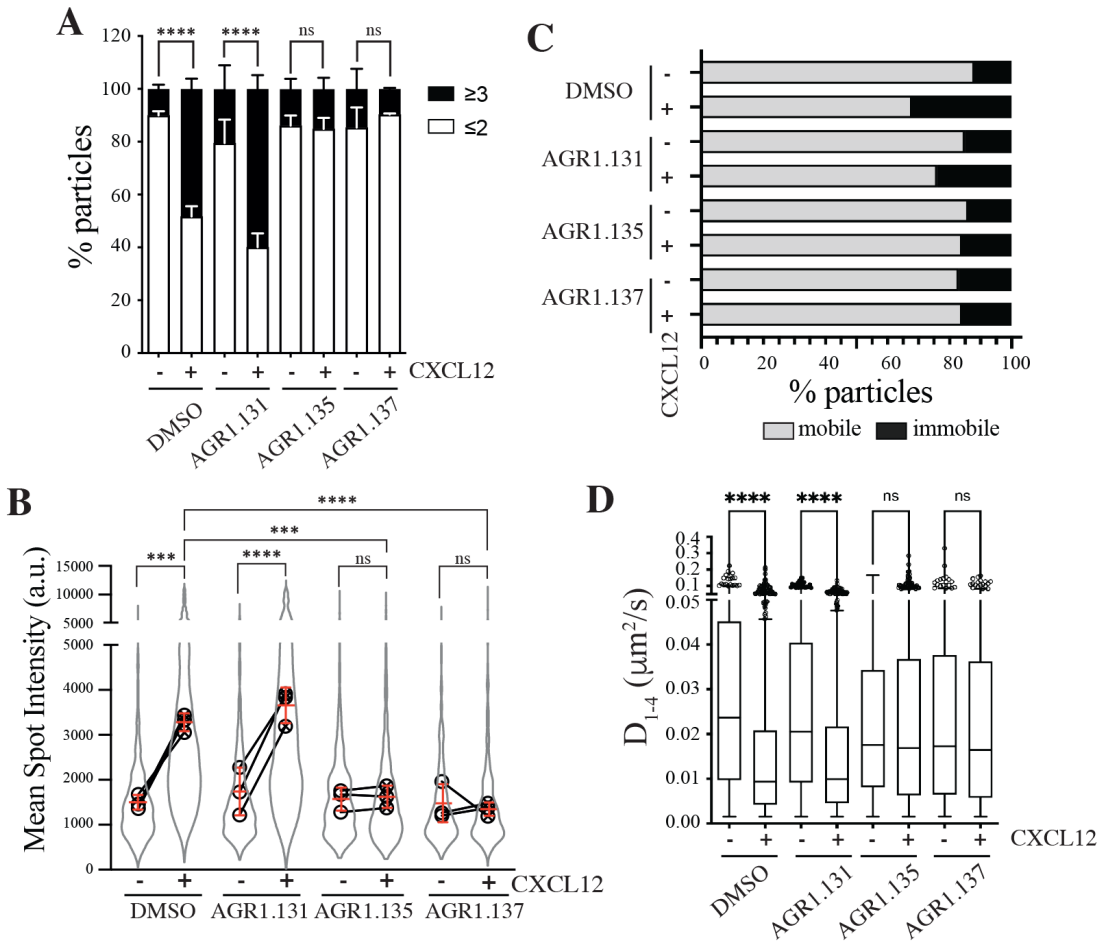


Figure 2. AGR1.135 and AGR1.137 alter CXCL12-mediated CXCR4 dynamics and nanoclustering. Single-particle tracking analysis of JK$^{-/-}$ cells transiently transfected with CXCR4 (JK$^{-/-}$-X4) treated with DMSO (control), AGR1.131, AGR1.135 or AGR1.137 on fibronectin (FN)- or FN+CXCL12-coated coverslips (DMSO: 581 particles in 59 cells on FN; 1365 in 63 cells on FN+CXCL12; AGR1.131: 1019 particles in 71 cells on FN; 1291 in 69 cells on FN+CXCL12; AGR1.135: 862 particles in 70 cells on FN; 1003 in 77 cells on FN+CXCL12; AGR1.137: 477 particles in 66 cells on FN; 566 in 64 cells on FN+CXCL12) n = 3. **A)** Frequency of CXCR4-AcGFP particles containing monomers plus dimers (≤2) or nanoclusters (≥3), ± SEM, calculated from mean spot intensity values of each particle as compared with the value of monomeric CD86-AcGFP (n = 3; n.s., not significant; ****p ≤ 0.0001). **B)** Intensity distribution (arbitrary units, a.u.) from individual CXCR4-AcGFP trajectories on unstimulated and CXCL12-stimulated JK$^{-/-}$-X4 cells pretreated or not with the indicated compounds or vehicle (DMSO). Graph shows the distribution of all trajectories (gray outline), with the mean value of each experiment (black circles) and the mean of all experiments ± SD (red lines) (n = 3; n.s., not significant; ***p ≤ 0.001, ****p ≤ 0.0001). **C)** Percentage of mobile and immobile CXCR4-AcGFP particles at the membrane of cells treated as indicated. **D)** Diffusion coefficients (D_{1-4}) of mobile single particle trajectories at the membrane of cells treated as indicated, represented by box-and-whiskers plots (Tukey method) with median (black line) indicated. (n.s., not significant, ****p ≤ 0.0001).

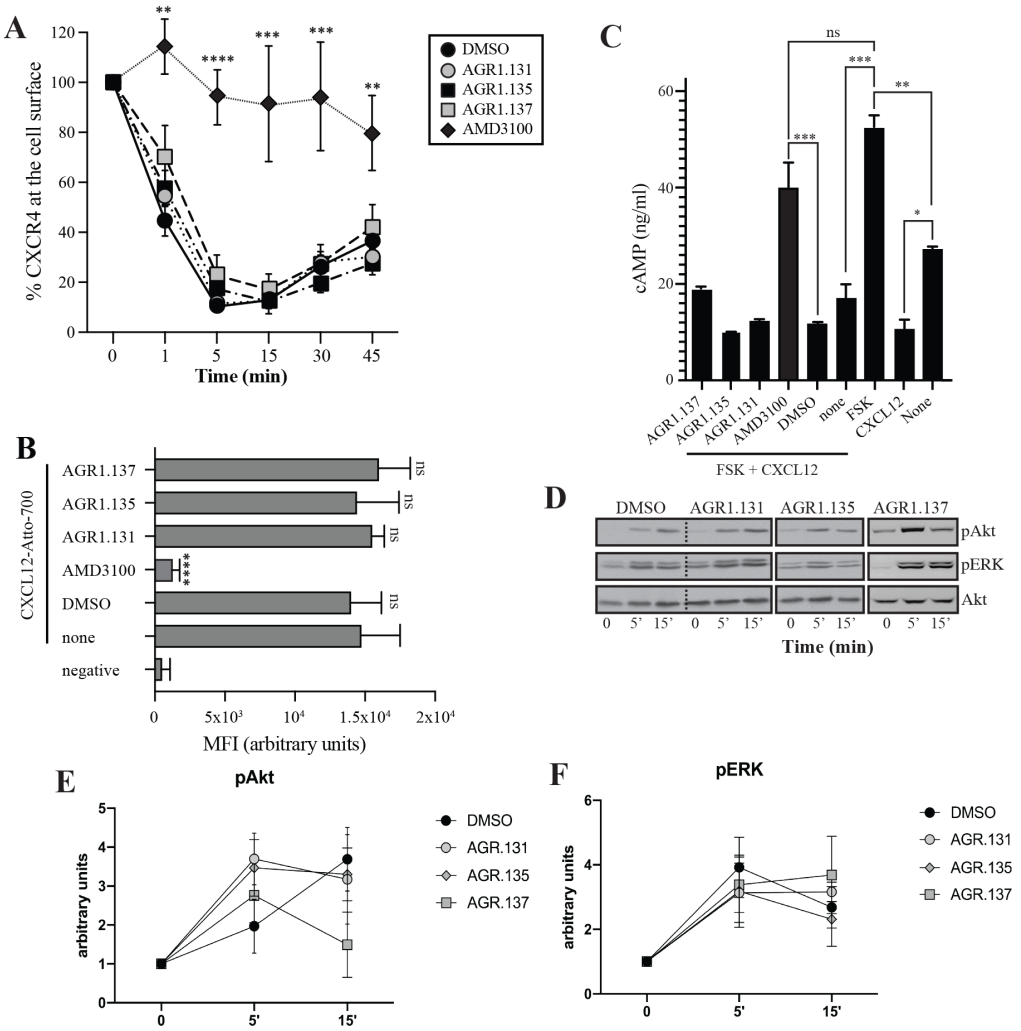
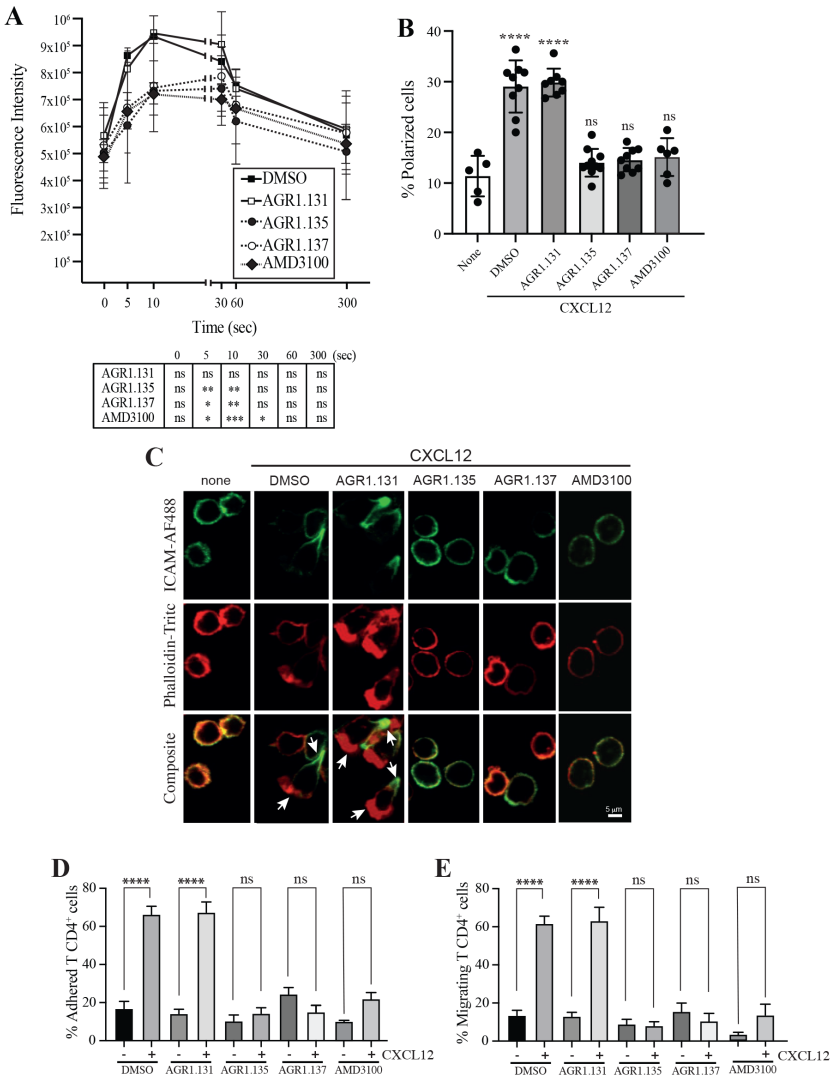


Figure 3. AGR1.135 and AGR1.137 treatments do not block CXCL12-mediated signaling pathways. A) Cell surface expression of CXCR4 in Jurkat cells after stimulation with CXCL12 (12.5 nM) at different time points and analyzed by flow cytometry using an anti-CXCR4 antibody in nonpermeabilized cells. Results show mean ± SEM of the percentage of CXCR4 expression at the cell surface (*n* = 4; n.s. not significant; **p ≤ 0.01, ***p ≤ 0.001, **** p ≤ 0.0001). **B)** CXCL12-Atto-700 binding on untreated or Jurkat cells pre-treated with DMSO (vehicle), AGR1.131, AGR1.135, AGR1.137 or with AMD3100 as a control, followed by flow cytometry analysis. Results are expressed as mean fluorescence intensity (MFI) values (arbitrary units). Negative corresponds to basal cell fluorescence in the absence of CXCL12-Atto-700 (mean ± SD, *n* = 3 n.s. not significant; ****p ≤ 0.0001). **C)** Cells untreated or pretreated with the small compounds (50 µM, 30 min 37ºC) or AMD3100 (1 µM, 30 min 37ºC) were stimulated with CXCL12 (50 nM, 5 min, 37ºC) followed by forskolin (10 µM, 10 min, 37ºC). Cells were then lysed and cAMP levels were determined (mean ± SD; *n* = 3; n.s. not significant; *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001). **D)** Western blot analysis of phospho (p)Akt and pERK in Jurkat cells pre-treated with DMSO, AGR1.131, AGR1.135 or AGR1.137, in response to CXCL12. As a loading control, membranes were re-blotted with an anti-Akt antibody. Representative experiments are shown (*n* = 4). **E, F**) Densitometry evaluation of the specific bands in D) (mean ± SD; *n* = 4).



	0	5	10	30	60	300 (sec)
AGR1.131	ns	ns	ns	ns	ns	ns
AGR1.135	ns	**	**	ns	ns	ns
AGR1.137	ns	*	**	ns	ns	ns
AMD3100	ns	*	***	*	ns	ns

Figure 4. AGR1.135 and AGR1.137 treatment alter CXCL12-mediated actin polymerization. A) Actin polymerization in response to CXCL12 as determined by F-actin (phalloidin-TRITC) staining in Jurkat cells treated with DMSO (vehicle) or the indicated modulators. Statistical significance of the different time points in comparison with the control (DMSO) is shown in the table (mean ± SD; *n* = 3; n.s. not significant; *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001). **B)** Percentage of polarized T cell blasts adhered to fibronectin and treated or not with CXCL12 in the presence of the indicated antagonists, as analyzed by immunostaining with anti-ICAM3-Alexa fluor 488 and phalloidin-TRITC. More than 500 cells were analyzed in each condition. Data are presented as percentage of polarized cells (mean ± SD; *n* = 3; n.s. not significant; ****p < 0.0001). **C)** Representative images of T cell blasts adhered to fibronectin and treated or not with CXCL12 in the presence of the indicated antagonists, as analyzed by immunostaining with anti-ICAM3-Alexa fluor 488 and phalloidin-TRITC as in B). **D)** $CD4^+$ T cells pretreated with AGR1.131, AGR1.135 or AGR1.137 were perfused in flow chambers coated with ICAM-1-containing lipid bilayers, alone or CXCL12-coated, and analyzed for cell contacts with the substrate. Data are presented as percentage of adhered cells (mean ± SD; *n* = 3; n.s. not significant; ****p ≤ 0.0001). **E)** Cells in D) were analyzed for cell migration. Data are presented as percentage of migrating cells (mean ± SD; *n* = 3; n.s. not significant; ****p ≤ 0.0001).

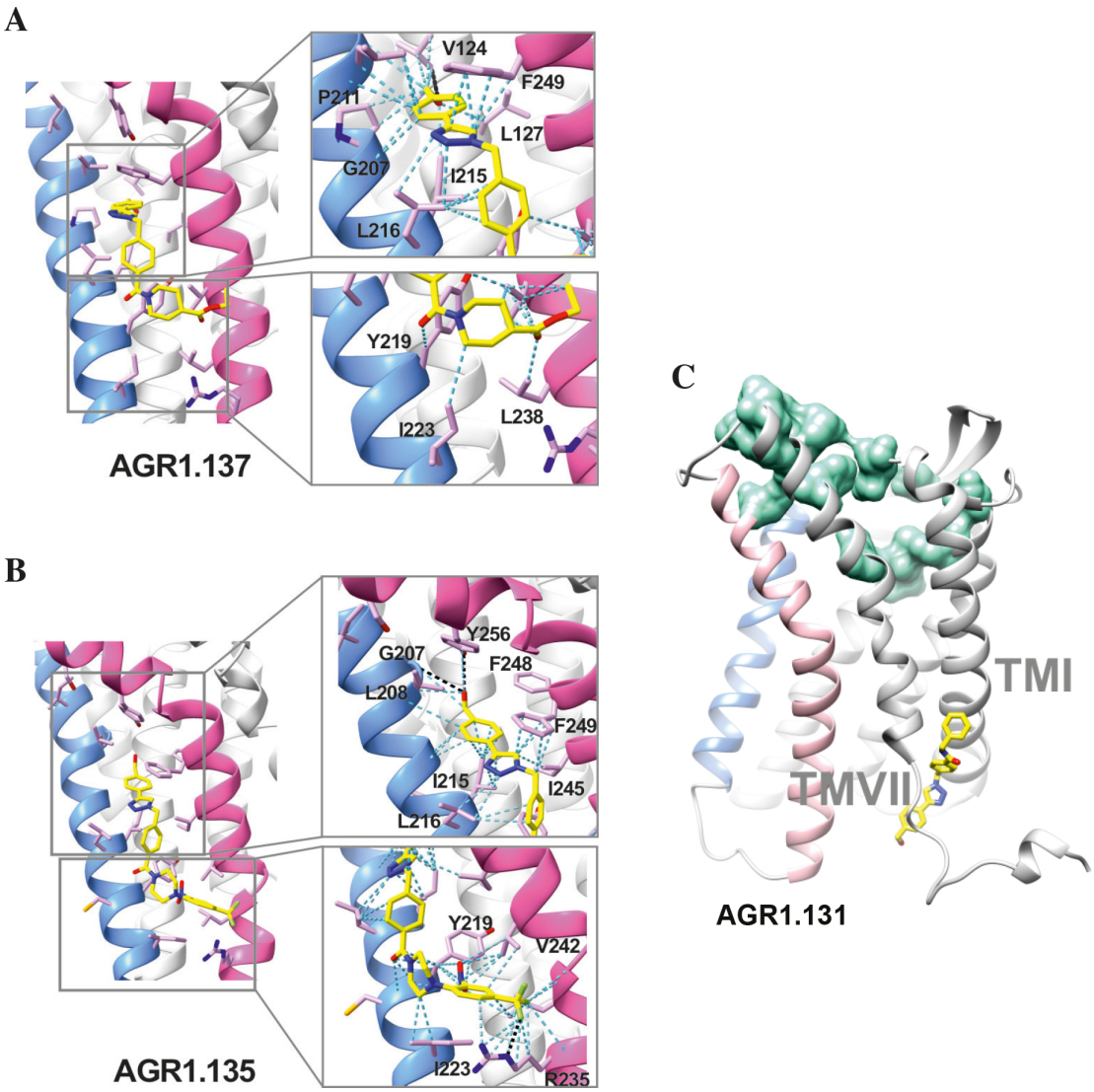


Figure 5. Ribbon and sticks representation of the CXCR4 modulators bound to the receptor. In the ribbon and sticks representations, the TMV helix is colored in blue and TMVI in pink. Figures show the binding of ligands AGR1.135 (**A)** and AGR1.137 (**B)** Zoom images showing the residues involved in the interaction of both compounds are also shown (right panels). **C)** Binding of AGR1.131 represented as sticks with carbon atoms in yellow, oxygen in red, nitrogen in blue and fluorine in green. CXCR4 residues involved in CXCL12 engagement and initial signal transmission are represented as green spheres.

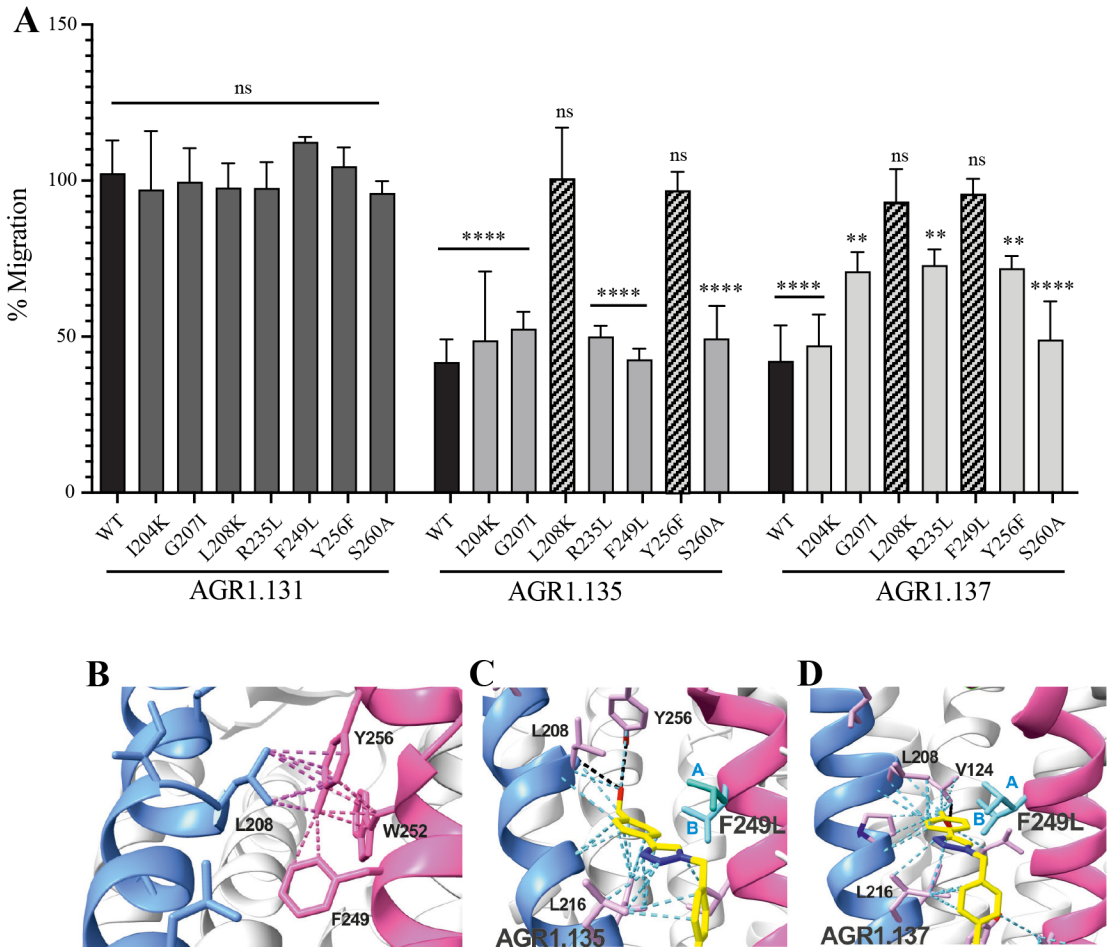


Figure 6. The antagonistic behavior of AGR1.135 and AGR1.137 depends on specific residues of CXCR4. A) CXCL12-induced migration of AGR1.131- AGR1.135- or AGR1.137-pretreated $JK^{-/-}$ cells transiently transfected with CXCR4wt or the different mutants described. Data are shown as the mean percentage ± SD of input cells that migrate (n = 4; n.s. not significant, **p ≤ 0.01, ****p ≤ 0.0001). **B)** Cartoon representation of Y256 and its intramolecular interactions in the CXCR4 X-ray solved structure 3ODU. **C, D)** Cartoon representation of the interaction of CXCR4 F249L mutant with AGR1.135 and AGR1.137, respectively. The two most probable conformations of leucine rotamers are represented in cyan. Van der Waals interactions are depicted in cyan dashed lines and hydrogen bonds in black dashed lines.

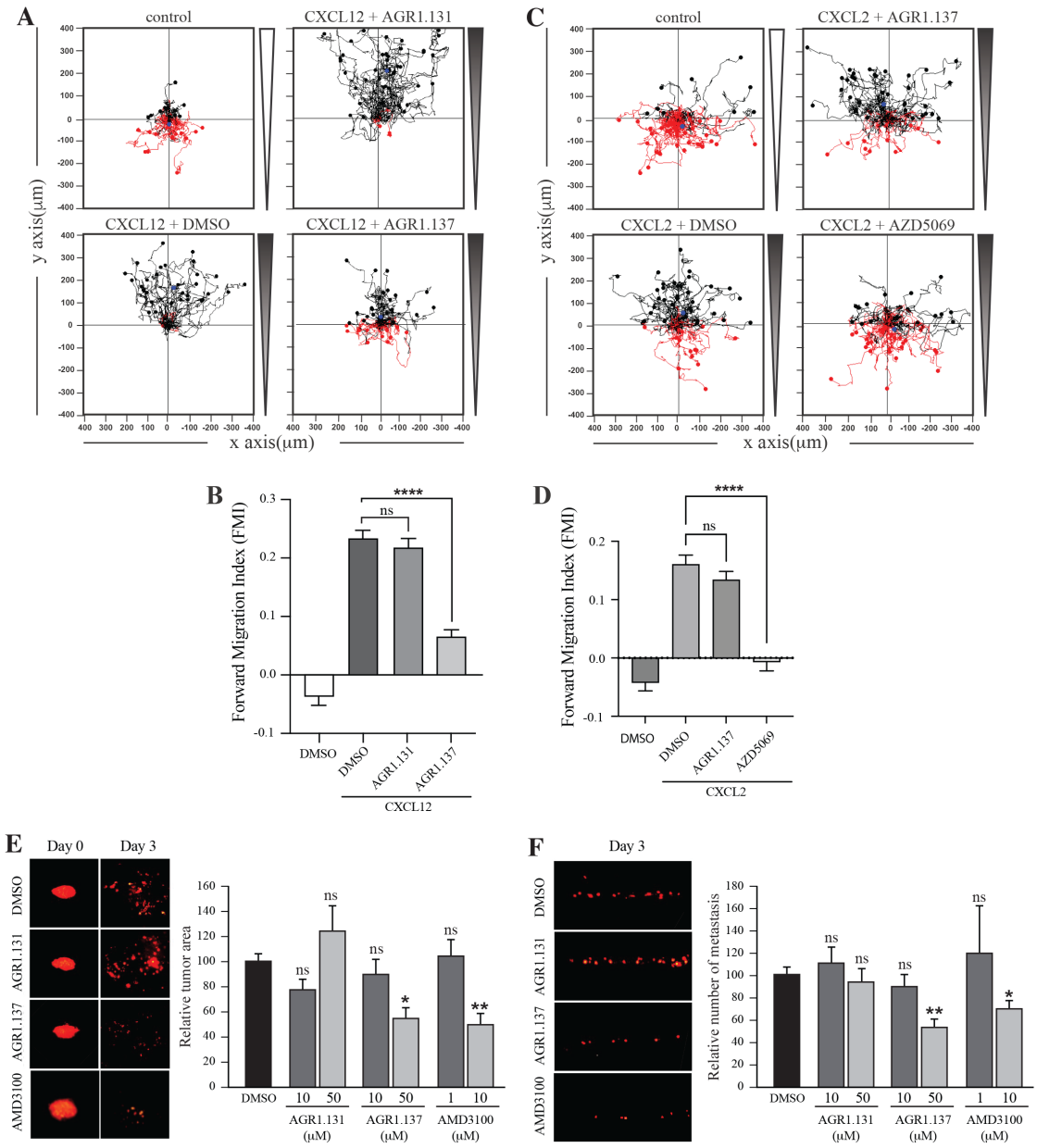


Figure 7. AGR1.137 reduces tumorigenesis and metastasis in a zebrafish model. (A–D) Migration of HeLa cells treated with vehicle (DMSO) or with the selected small compounds or inhibitor as indicated, on µ-chambers in response to a CXCL12 **(A, B)** or CXCL2 **(C, D)** gradient (*n* = 2, in duplicate, with at least 50 cells tracked in each condition). Figures A) and C) show representative spider plots with the trajectories of tracked cells migrating along the gradient (black) or moving in the opposite direction (red). Black and red dots in the plots represent the final position of each single tracked cell. **(B, D)** Quantification of the Forward Migration Index of experiments performed in A) and C) (mean ± SD; *n* = 3; n.s. not significant, ****p ≤ 0.0001). **E)** Representative fluorescent images of DiI-labeled HeLa cells in zebrafish larvae treated with vehicle (DMSO), AGR1.131 50 µM, AGR1.137 50 µM or AMD3100 10 µM at 0 or 3 days post-implantation and

treatment. Quantitation of the relative tumor size at day 3 compared with that of day 0 normalized to the relative tumor size in the DMSO control group, is shown for each experimental group (mean ± SD; $n = 20$; n.s. not significant, **$p \leq 0.01$). **F)** Representative fluorescent images of the caudal hematopoietic plexus of larvae from the same groups as shown in E) at 3 days postimplantation. Quantitation of the relative amount of metastasized cells in each group relative to DMSO-controls is shown (mean ± SD; $n = 20$, n.s. not significant, * $p \leq 0.05$, *** $p \leq 0.001$).

Supplementary Information for
Allosteric modulation of the CXCR4:CXCL12 axis by targeting receptor nanoclustering via the TMV-TMVI domain
Compound synthesis and characterization

1. General Experimental Details

2. Synthesis and Characterization of Triazol derivatives

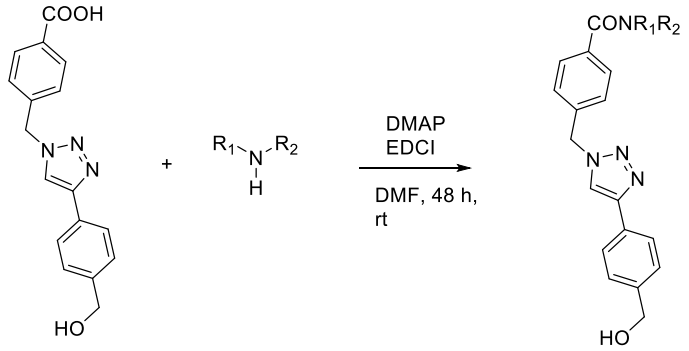
3. NMR Spectra and HRMS Analysis

1. General Experimental Details

All solvents were purchased from Sigma-Aldrich (anhydrous solvents) and commercially available reagents were used as received. All reactions were followed by TLC analysis (TLC plates GF254, Merck) or LC–MS (liquid chromatography mass spectrometry). Melting points were determined in a Büchi Melting Point M-560 apparatus. NMR spectra were recorded at ambient temperature unless otherwise stated using standard pulse methods on one of the following spectrometers and signal frequencies: Bruker AV-300 (1H = 300 MHz, 13C = 75 MHz). Chemical shifts are reported in ppm and are referenced to tetramethylsilane (TMS) or to the following solvent peaks: CDCl3 (1H = 7.27 ppm, 13C = 77.00 ppm), DMSO-d6 (1H = 2.50 ppm, 13C = 39.51 ppm). Coupling constants are quoted to the nearest 0.1 Hz, and multiplicities are given by the following abbreviations and combinations thereof: s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), and br (broad). Column chromatography was performed on prepacked silica gel columns using a Biotage SP4 or Isolera One system. High resolution mass spectra (HRMS-ESI) were recorded on a Micromass Q-Tof Ultima hybrid quadrupole time-of-flight mass spectrometer, with analytes separated on an Agilent 1100 liquid chromatograph equipped with a Phenomenex Luna C18(2) reversed phase column (100 mm × 2.1 mm, 3 μm packing diameter).

2. Synthesis and Characterization of Triazol derivatives

General Procedure (GP):



Over a solution of 4-((4-(4-(hydroxymethyl)phenyl)-1H-1,2,3-triazol-1-yl)methyl) benzoic acid (123 mg, 0.4 mmol) in 4 ml of DMF, DMAP (97 mg, 0.8 mmol) and EDC.HCl (92 mg, 0.48 mmol) were added and stirred for 15 min at room temperature. Subsequently, the corresponding amine (0.4 mmol) was added and stirred at room temperature for 48 hrs. Next, H_2O (10 ml) was added and extracted with DCM (25 ml × 3). The organic phase was dried over $MgSO_4$, filtered and evaporated under reduced pressure. The crude sample obtained was purified by column chromatography, eluting with a DCM-MeOH solvent mixture (from 99:1 to 90:10).

_N_-Benzyl-4-((4-(4-(hydroxymethyl)phenyl)-1H-1,2,3-triazol-1-yl)methyl)benzamide (AGR1.131)

The reaction was performed according to GP using benzylamine (44 μl, 0.4 mmol), and the title product was obtained after purification by silica gel column chromatography as a white solid (35% yield). Pf: 193 – 195 ºC.

^{1}H NMR (300 MHz, DMSO-d_6) δ 9.09 (t, J = 6.0 Hz, 1H), 8.65 (s, 1H), 7.98 – 7.89 (m, 2H), 7.86 – 7.78 (m, 2H), 7.46 (d, J = 8.2 Hz, 2H), 7.40 (d, J = 8.1 Hz, 2H), 7.36 – 7.20 (m, 5H), 5.73 (s, 2H), 5.25 (t, J = 5.7 Hz, 1H), 4.53 (d, J = 5.7 Hz, 2H), 4.49 (d, J = 6.0 Hz, 2H). ^{13}C NMR (75 MHz, DMSO) δ 165.69, 146.69, 142.29, 139.56, 139.01, 134.08, 128.99, 128.25, 127.83, 127.74, 127.12, 126.88, 126.71, 124.90, 121.42, 62.61, 52.58, 42.56.
HRMS (ESI$^+$) m/z Calc. for $C_{24}H_{23}N_4O_2$ $[M+H]^+$: 399.1816. Found: 399.1809.

(4-((4-(4-(Hydroxymethyl)phenyl)-1H-1,2,3-triazol-1-yl)methyl)phenyl)(4-(2-nitro-4-(trifluoromethyl)phenyl)piperazin-1-yl)methanone (AGR1.135)

The reaction was performed according to GP using (2-nitro-4-(trifluoromethyl)phenyl)piperazine (110 mg, 0.4 mmol), and the title product was obtained after purification by silica gel column chromatography as a yellow solid (37% yield). Pf: 184 – 186 ºC.

^{1}H NMR (300 MHz, DMSO-d_6) δ 8.65 (s, 1H), 8.19 (d, J = 2.3 Hz, 1H), 7.89 (dd, J = 8.9, 2.3 Hz, 1H), 7.81 (d, J = 8.0 Hz, 2H), 7.53 – 7.35 (m, 7H), 5.71 (s, 2H), 5.23 (t, J = 5.7 Hz, 1H), 4.52 (d, J = 5.4 Hz, 2H), 3.73 (s, 2H), 3.48 (s, 2H), 3.21 (s, 4H). ^{13}C NMR (75 MHz, DMSO) δ 169.12, 147.70, 147.08, 142.67, 140.04, 137.84, 135.74, 130.59, 130.54, 129.38,

128.27, 128.00, 127.25, 125.65, 125.29, 124.12, 124.07, 122.06, 121.92, 121.82, 120.57, 120.12, 62.99, 62.87, 52.97, 50.33, 47.01.
HRMS (ESI$^+$) Calc. for $C_{28}H_{26}F_3N_6O_4$ $[M+H]^+$: 567.1962. Found: 567.1959.

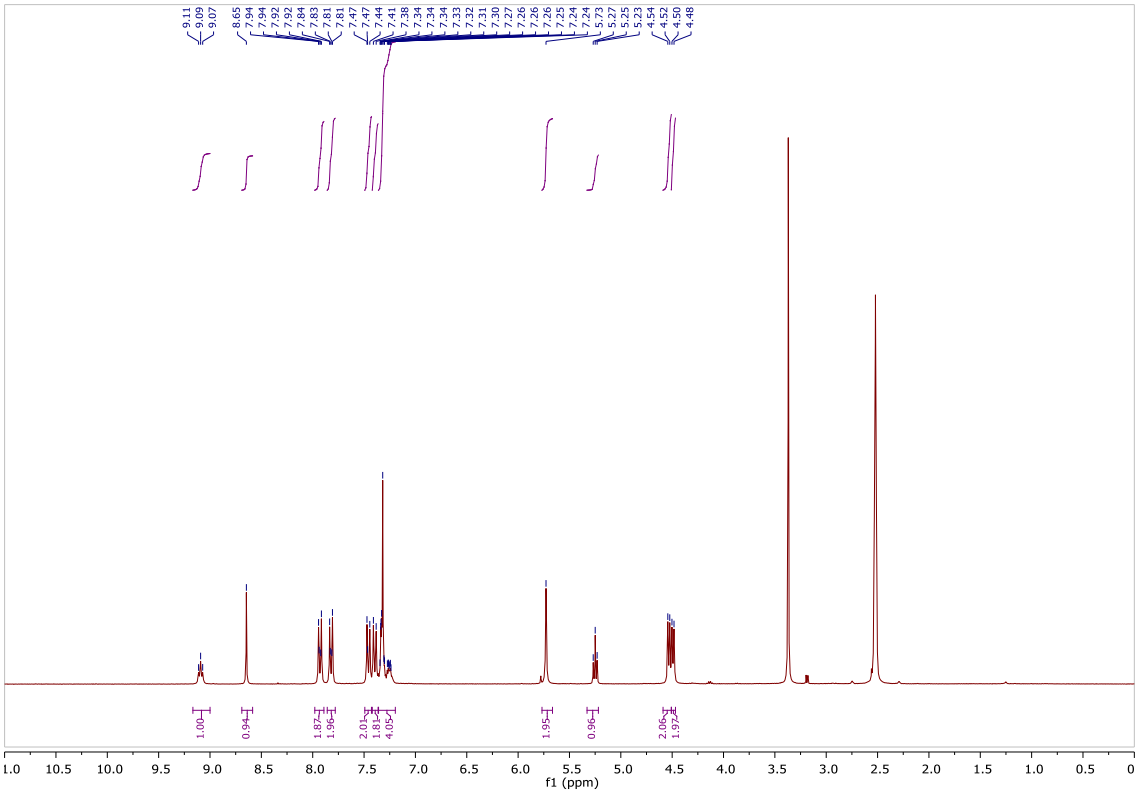
Ethyl 1-(4-((4-(4-(hydroxymethyl)phenyl)-1H-1,2,3-triazol-1-yl)methyl)benzoyl) piperidine-4-carboxylate (AGR1.137)

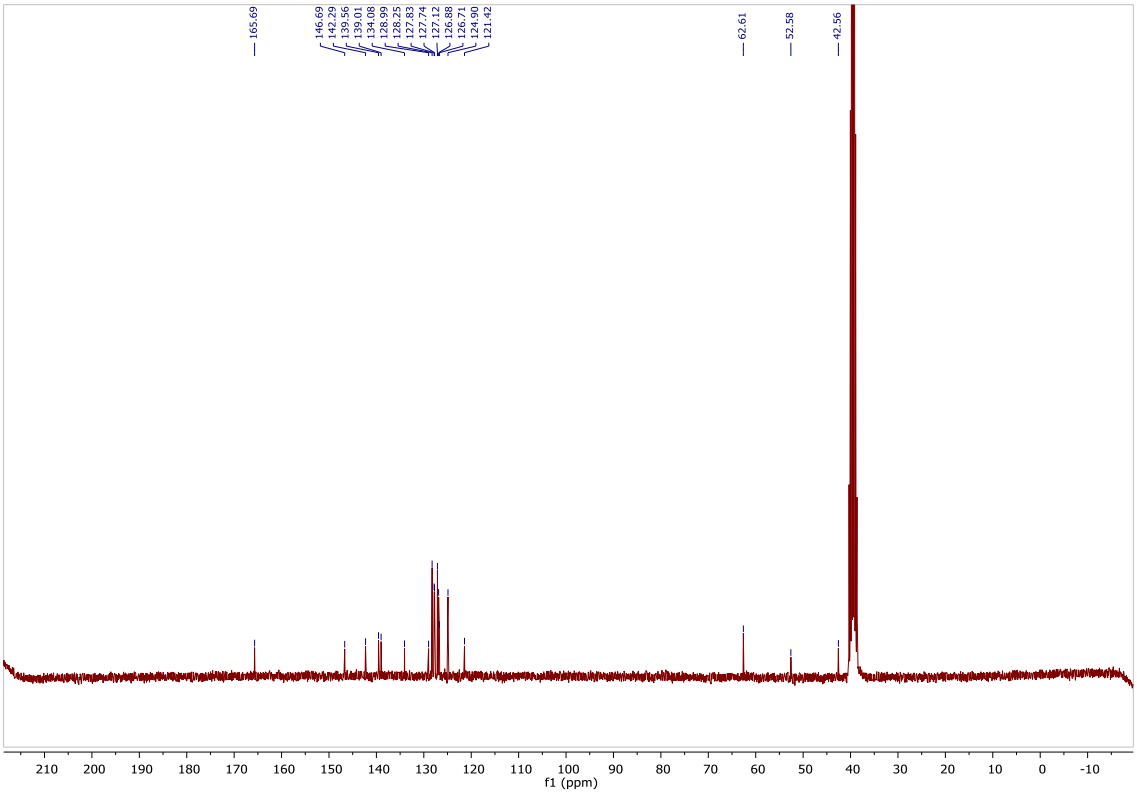
The reaction was performed according to GP using ethyl piperidinee-4-carboxylate (63 mg, 0.4 mmol) and the title product was obtained after purification by silica gel column chromatography as a white solid (30% yield). Pf: 119 – 121 ºC.

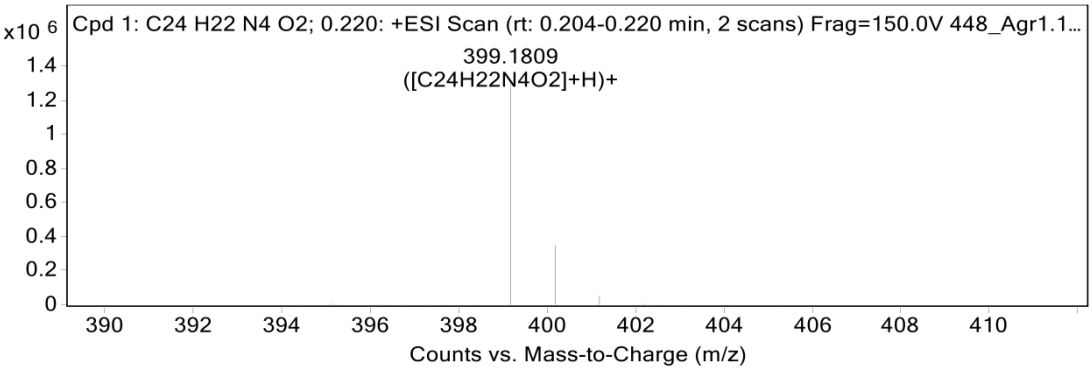
^{1}H NMR (300 MHz, Methanol-d_4) δ 8.40 (s, 1H), 7.89 – 7.74 (m, 2H), 7.56 – 7.31 (m, 6H), 5.73 (s, 2H), 4.66 (s, 2H), 4.48 (d, J = 13.2 Hz, 1H), 4.16 (q, J = 7.1 Hz, 2H), 3.69 (d, J = 13.0 Hz, 1H), 3.24 – 2.96 (m, 2H), 2.78 – 2.52 (m, 1H), 2.15 – 1.52 (m, 4H), 1.27 (t, J = 7.1 Hz, 3H). ^{13}C NMR (75 MHz, MeOD) δ 175.82, 171.87, 149.21, 143.15, 138.81, 137.25, 130.56, 129.43, 128.59, 128.56, 126.71, 122.39, 64.88, 61.79, 54.58, 42.65, 41.96, 29.67, 28.96, 14.51.
HRMS (ESI$^+$) Calc. For $C_{25}H_{29}N_4O_4$ $[M+H]^+$: 449.2183. Found: 449.2160.

3. NMR Spectra and HRMS Analysis

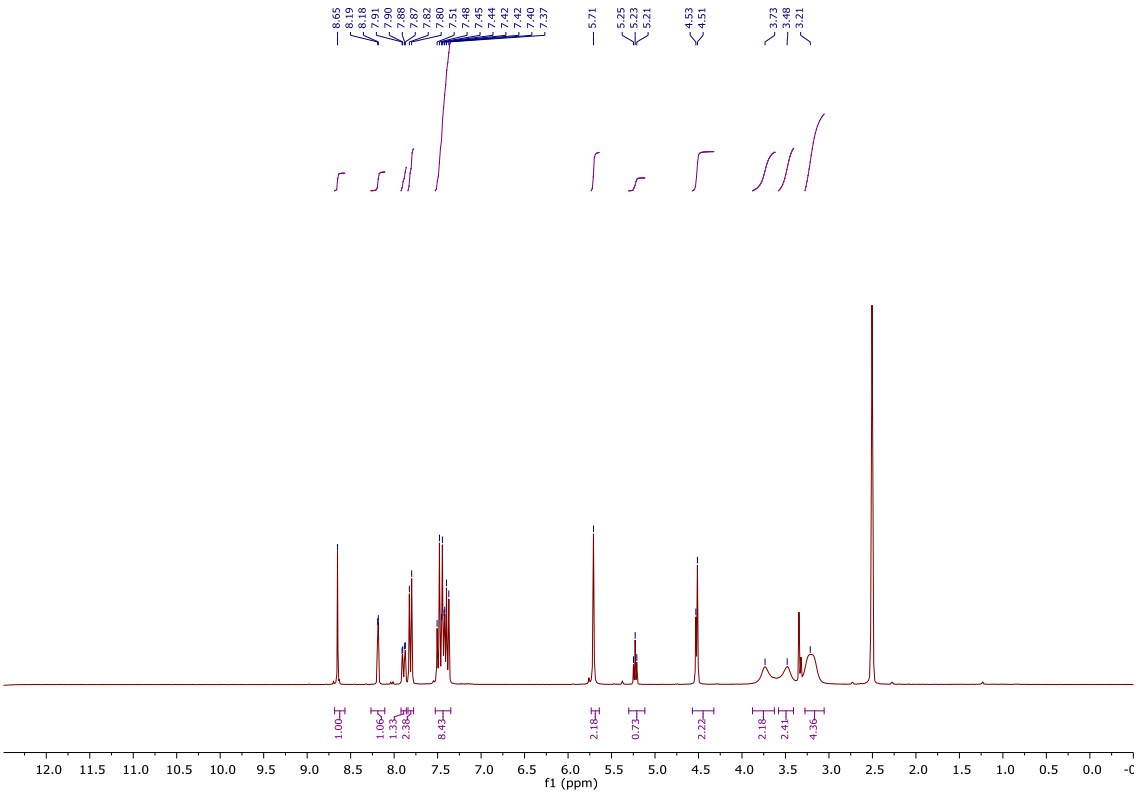
(AGR1.131)

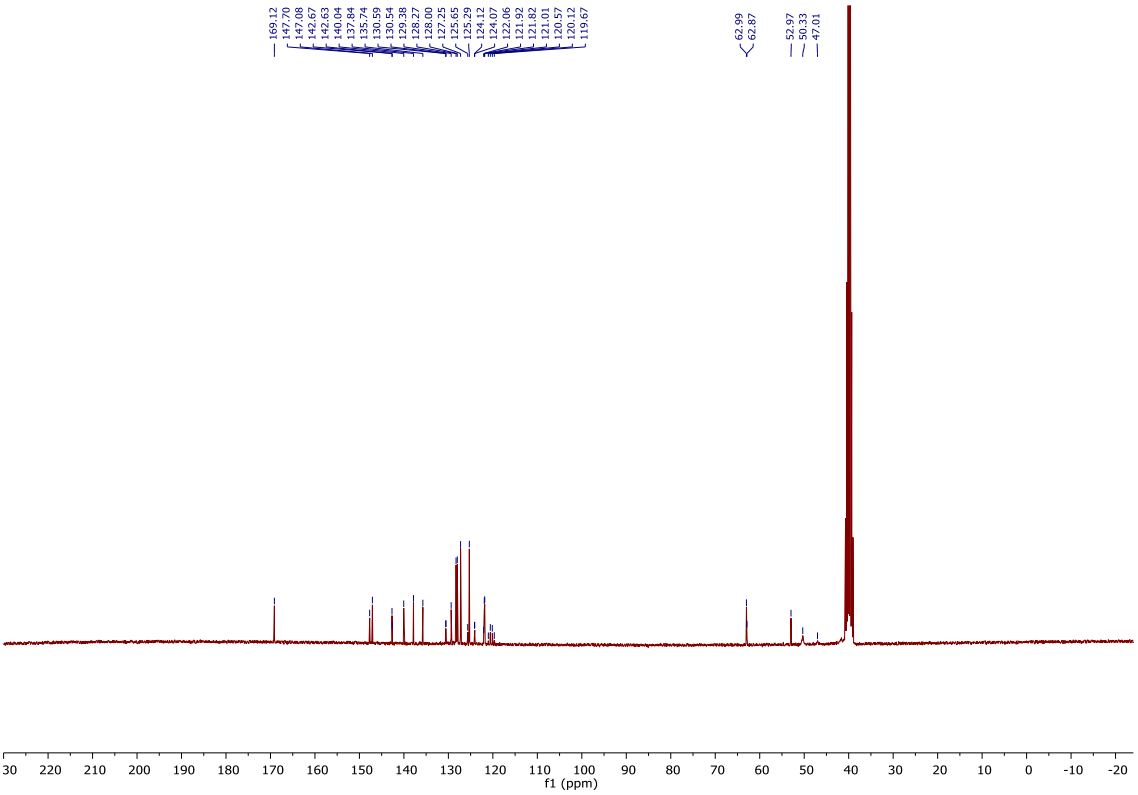


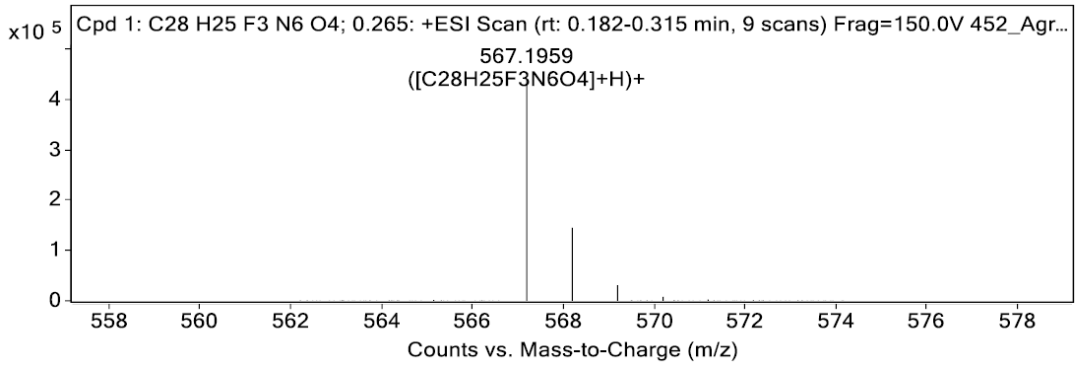

165.69
146.69
142.29
139.56
139.01
134.08
128.99
128.25
127.83
127.74
127.13
126.88
126.71
124.90
121.42
62.61
52.58
42.56
f1 (ppm)


Cpd 1: C24 H22 N4 O2; 0.220: +ESI Scan (rt: 0.204-0.220 min, 2 scans) Frag=150.0V 448_Agr1.1...
x10 6
399.1809
([C24H22N4O2]+H)+
Counts vs. Mass-to-Charge (m/z)

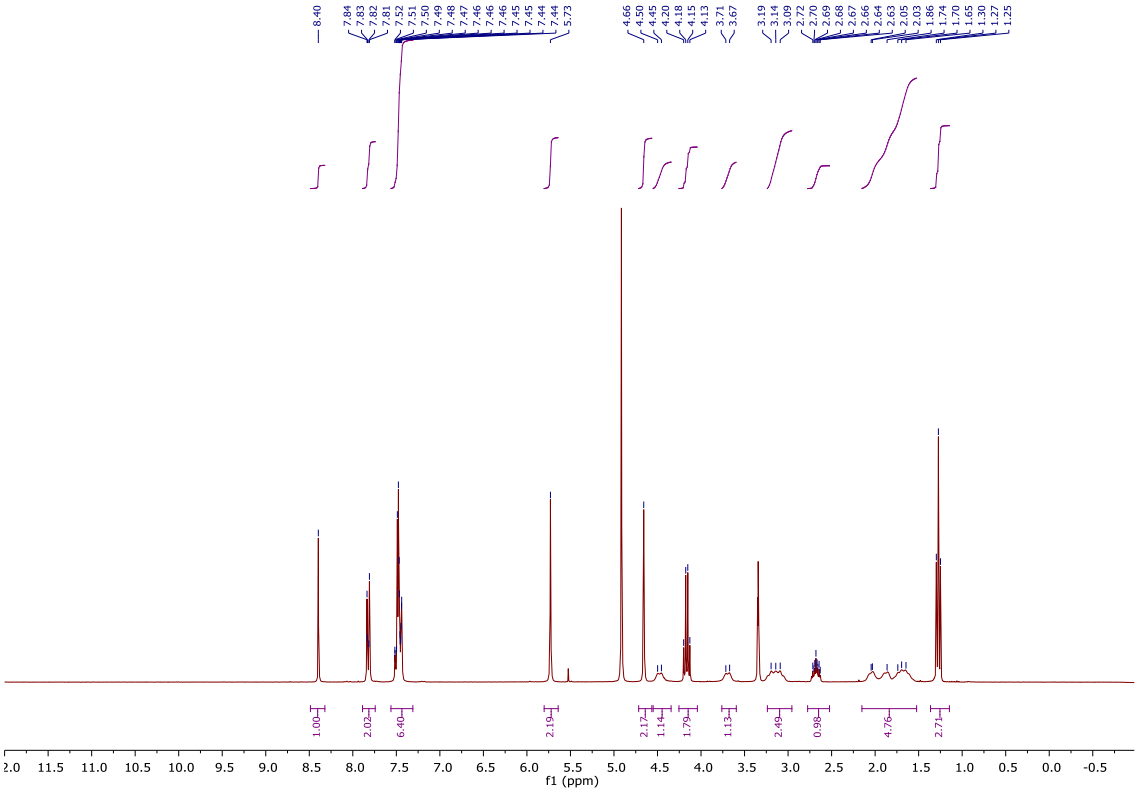
(AGR1.135)

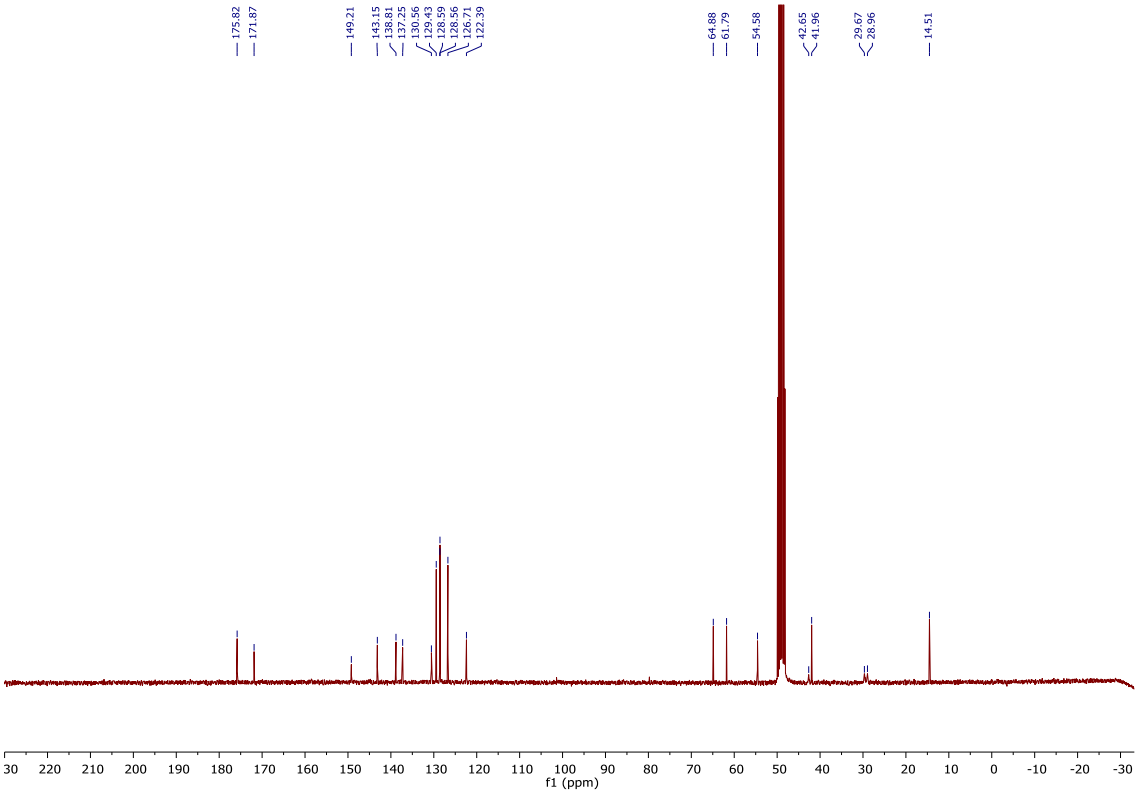

8.65
8.19
8.18
7.91
7.90
7.88
7.87
7.82
7.80
7.51
7.48
7.45
7.44
7.42
7.40
7.37
5.71
5.25
5.23
5.21
4.53
4.51
3.73
3.48
3.21
1.00
1.06
1.33
2.38
8.43
2.18
0.73
2.22
2.18
2.41
4.36
f1 (ppm)

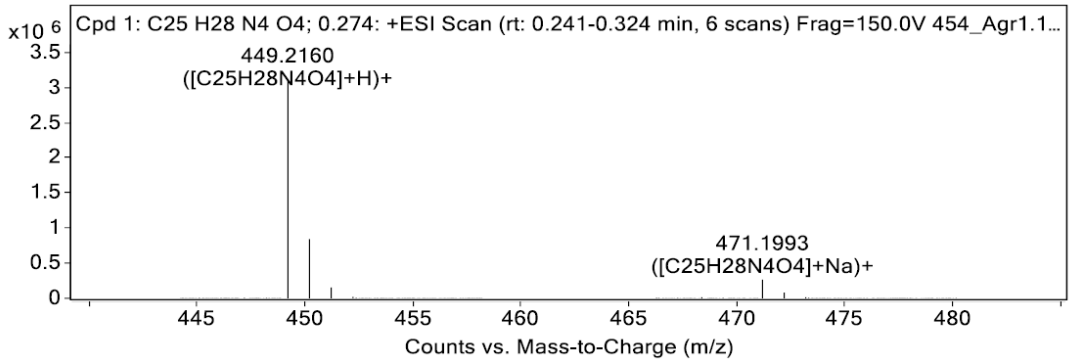

169.12
147.70
147.08
142.67
142.63
140.04
137.84
135.74
130.59
130.54
129.38
128.27
128.00
127.25
125.65
125.29
124.12
124.07
122.06
121.92
121.82
121.01
120.57
120.12
119.67
62.99
62.87
52.97
50.33
47.01
f1 (ppm)

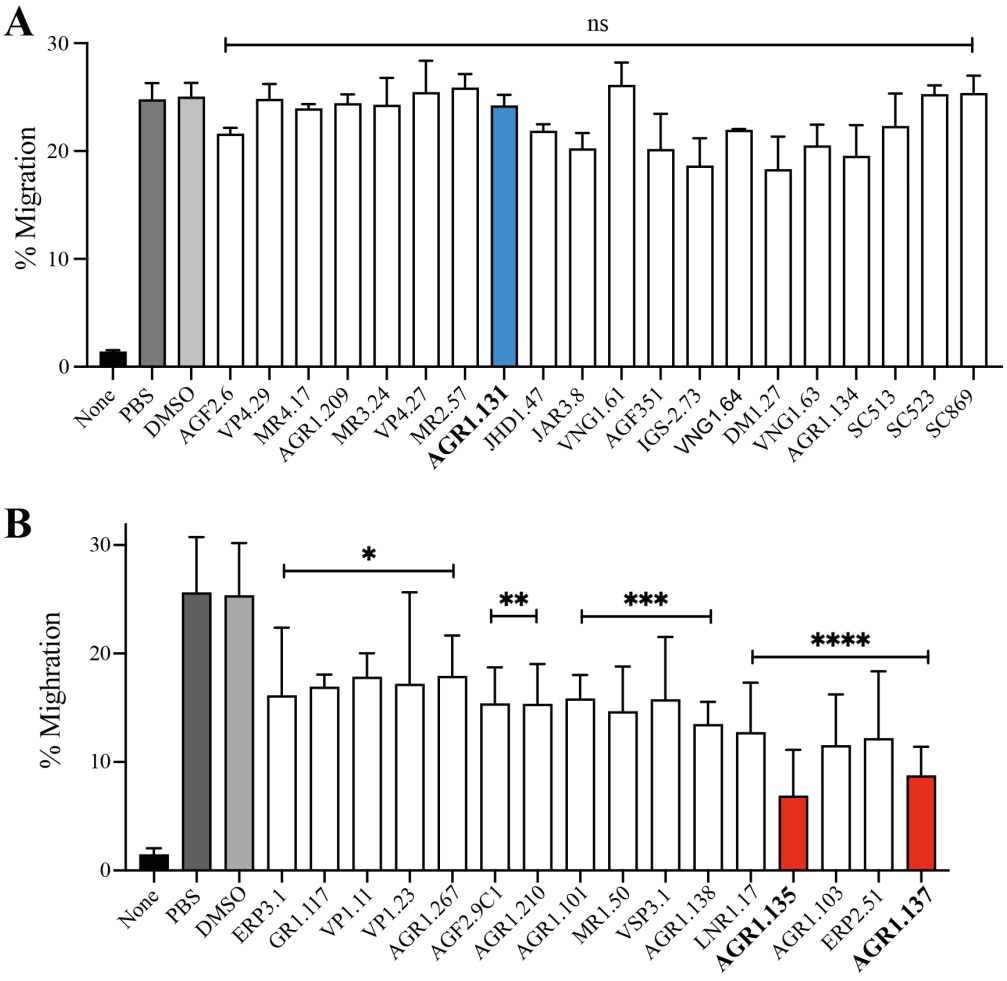

Cpd 1: C28 H25 F3 N6 O4; 0.265: +ESI Scan (rt: 0.182-0.315 min, 9 scans) Frag=150.0V 452_Agr...
x10 5
567.1959
([C28H25F3N6O4]+H)+
Counts vs. Mass-to-Charge (m/z)

(AGR1.137)

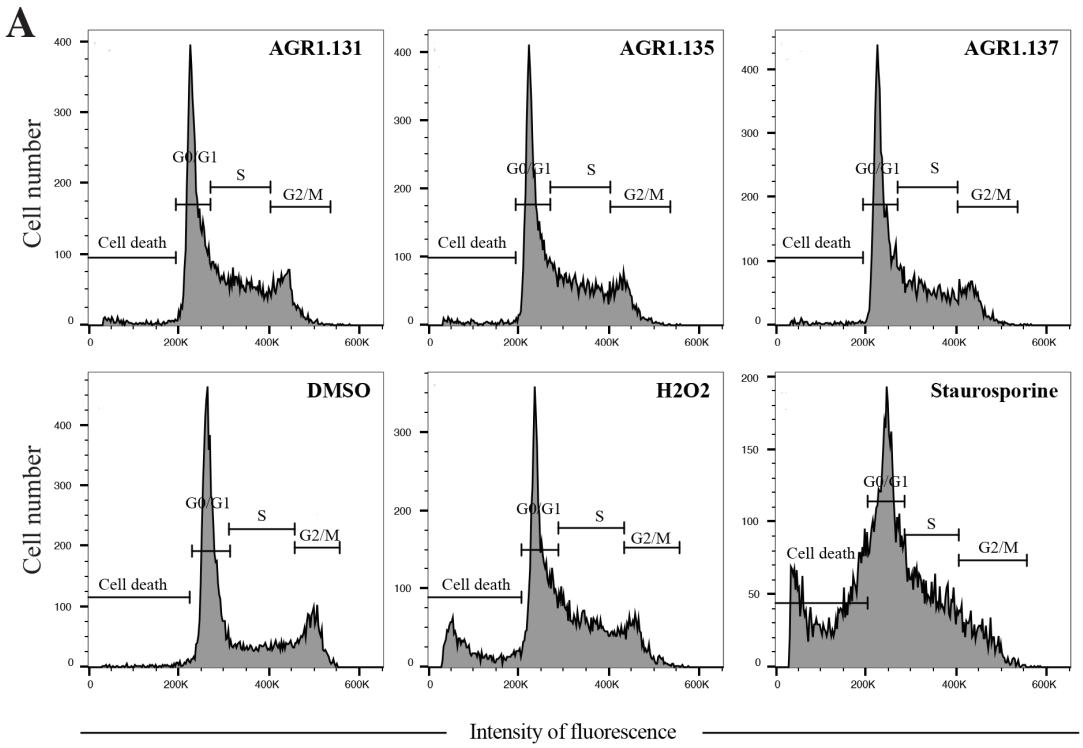

f1 (ppm)


175.82
171.87
149.21
143.15
138.81
137.25
130.56
129.43
128.59
128.56
126.71
122.39
64.88
61.79
54.58
42.65
41.96
29.67
28.96
14.51
f1 (ppm)


Cpd 1: C25 H28 N4 O4; 0.274: +ESI Scan (rt: 0.241-0.324 min, 6 scans) Frag=150.0V 454_Agr1.1...
x10 6
449.2160
([C25H28N4O4]+H)+
471.1993
([C25H28N4O4]+Na)+
Counts vs. Mass-to-Charge (m/z)



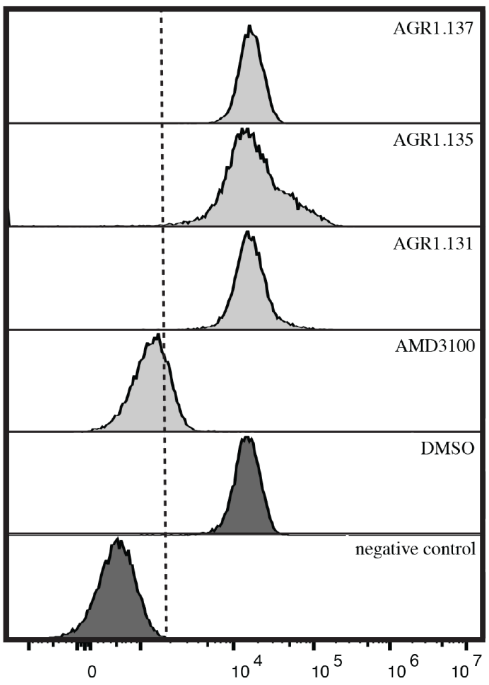
Supplementary Figure 1. Small compound-mediated inhibition of CXCL-12-induced cell migration. A, B). Migration of Jurkat cells, untreated or treated with the selected small compounds (50 µM) or DMSO (vehicle) in the previous in silico screening, in response to 12.5 nM CXCL12. Data shown as percentage of migrating cells (mean ± SD; n = 5; *p ≤0.05, **p ≤0.01, ***p ≤0.001, ****p ≤0.0001, ns : not significant).



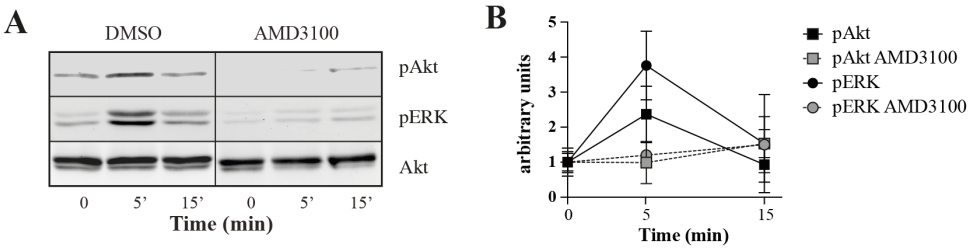
B

	Cell death	G0/G1	S	G2/M
AGR1.131	**3.3**	**46.7**	**33.6**	**16.3**
AGR1.135	**3.2**	**47.9**	**34.0**	**14.7**
AGR1.137	**2.68**	**50.5**	**32.3**	**14.4**
DMSO	**3.0**	**57.2**	**21.4**	**18.6**
H2O2	**14.8**	**41.8**	**31.9**	**11.3**
Staurosporine	**28.7**	**37.0**	**24.4**	**9.8**

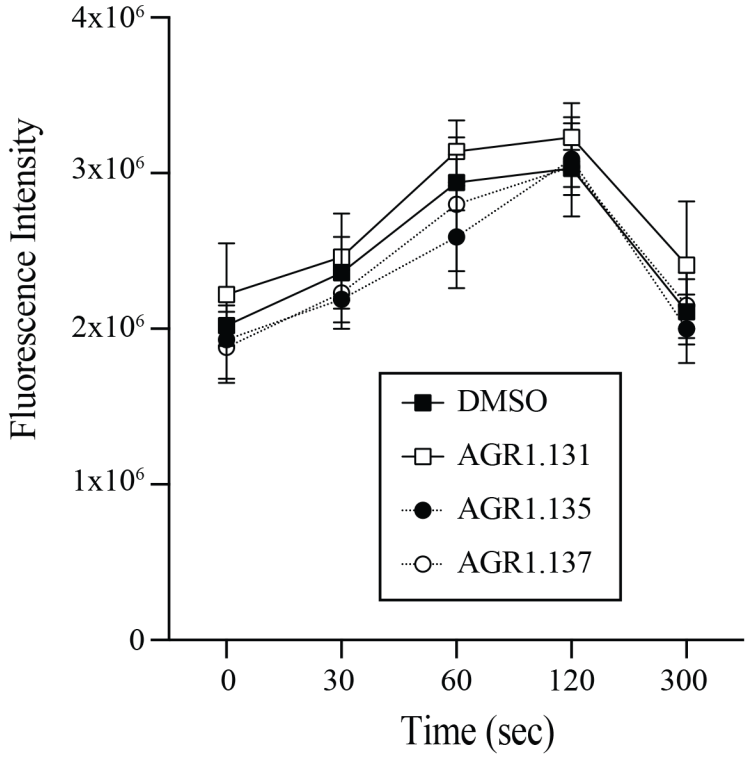
Supplementary Figure 2. Effect of small compounds on Jurkat cell cycle. A). Cell cycle analysis using propidium iodide incorporation and flow cytometry of Jurkat cells treated with AGR1.131, AGR1.135 and AGR1.137 (2 hours, 37ºC) or with DMSO (vehicle), H_2O_2 (10%) and staurosporine (10 μM) as controls. The percentage of cells in each cycle stage is shown. **B)**. Summary of the percentage of treated cells in each cell cycle phase.



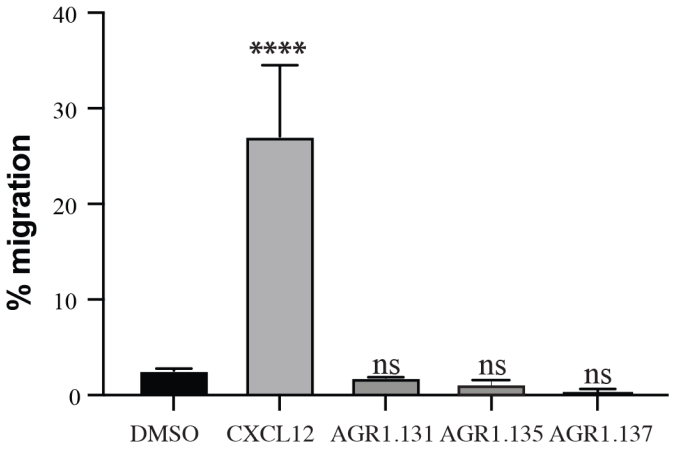
Supplementary Figure 3. Flow cytometry analysis of the binding of CXCL12-ATTO700 to JK cells. Cells were pre-incubated for 30 min at 37ºC with the indicated antagonists, followed by incubation with CXCL12- ATTO700 (30 min 37ºC). CXCL12 binding was determined in flow cytometry. A representative experiment of 4 performed is shown.



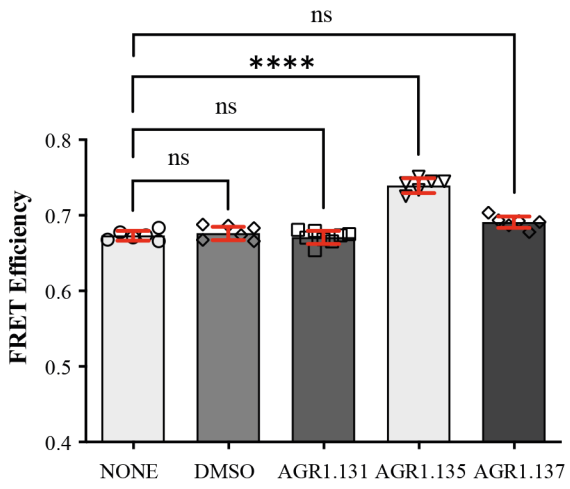
Supplementary Figure 4. Western blot analysis of AMD3100-treated JK cells. A) Cells were pre-incubated for 30 min at 37ºC with AMD3100 or DMSO (vehicle) as control, stimulated with CXCL12 and analyzed by western blot using specific antibodies. **B)** Densitometry analysis of the western blot images in A) using Image J software.



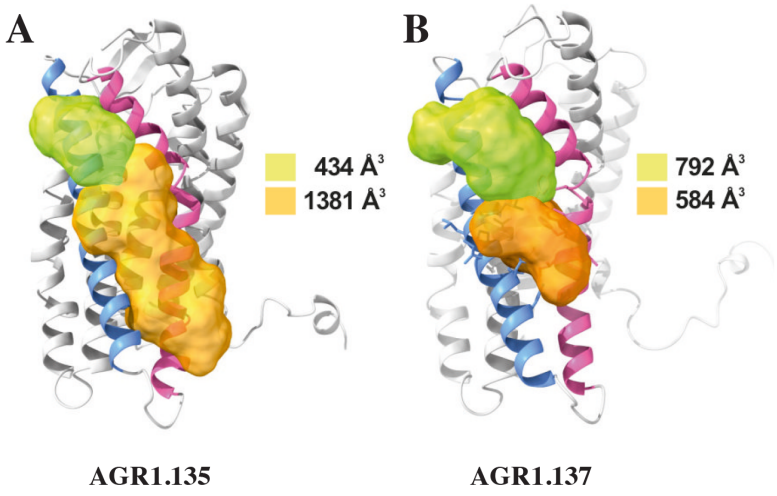
Supplementary Figure 5. AGR1.135 and AGR1.137 do not alter anti-CD3-induced actin polimerization. F-actin (Phalloidin-TRITC) staining of Jurkat cells treated with the indicated compounds or DMSO (vehicle), adhered to fibronectin and stimulated with anti-CD3 antibody, as indicated (mean ± SD; n = 3).



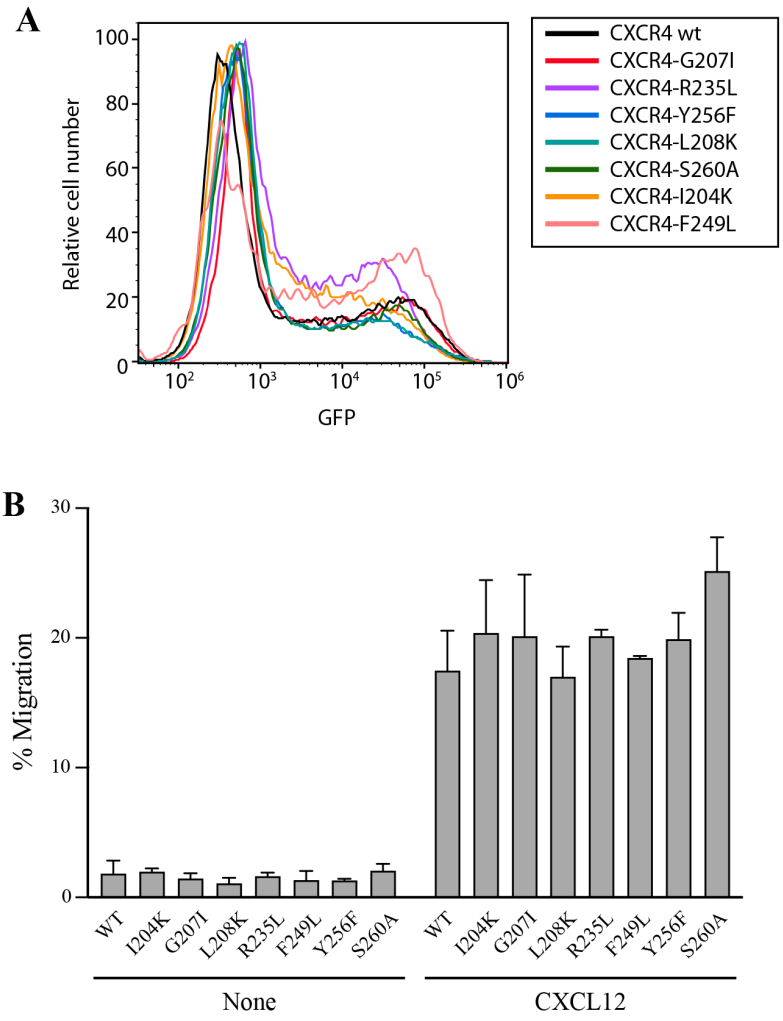
Supplementary Figure 6. AGR1.131, AGR1.135 and AGR1.137 are not CXCR4 agonists. Migration of Jurkat cells in response to 12.5 nM CXCL12, 50 µM AGR1.131, 50 µM AGR1.135 or 50 µM AGR1.137. Data shown as percentage of migrating cells (mean ± SD; n = 5; n.s. not significant, **** $p \leq 0.0001$).



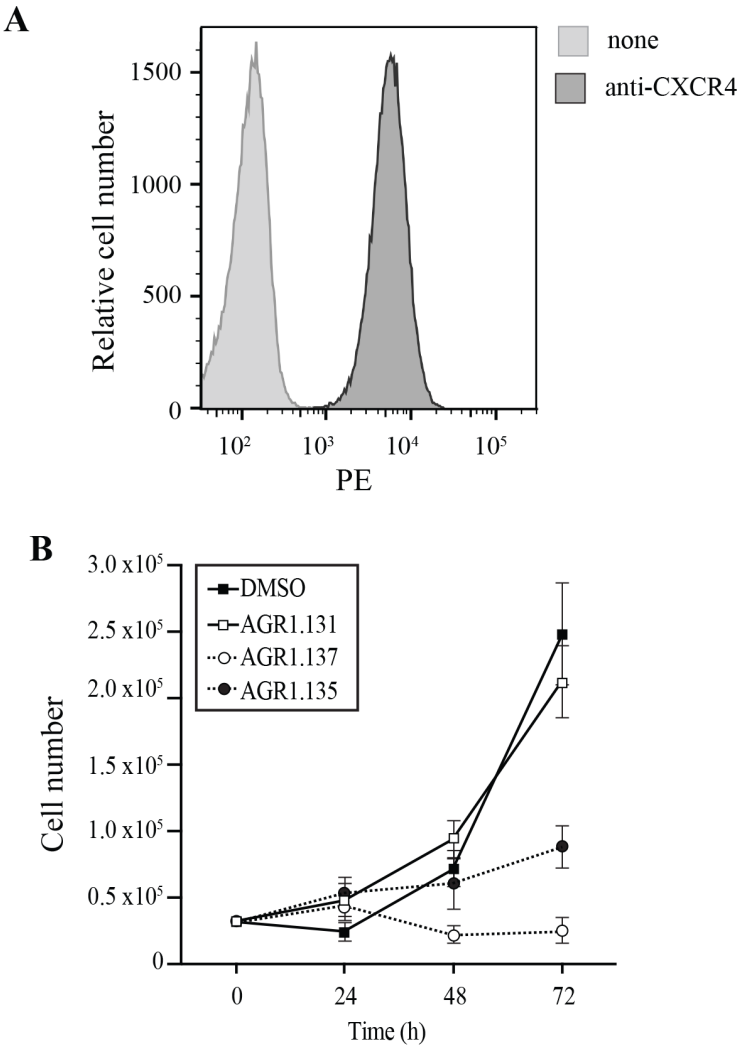
Supplementary Figure 7. FRET analysis of CXCR4 in the presence of AGR1.131, AGR1.135 and AGR1.137. FRET efficiency in HEK-293 cells transiently transfected with CXCR4-YFP/ CXCR4-CFP (ratio 15:9), in the presence of 50 µM AGR1.131, 50 µM AGR1.135, 50 µM AGR1.137 or vehicle (DMSO). Data shows FRET efficiency (a.u.) (mean ± SD; n = 6; n.s. not significant, **** $p \leq 0.0001$).



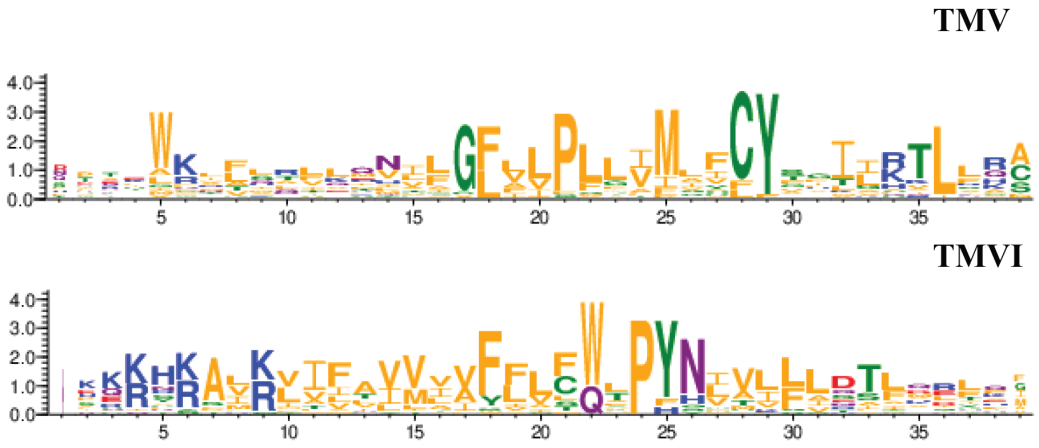
Supplementary Figure 8. Cartoon and surface representation of CXCR4 and the clefts identified between TMV and TMVI. The protein structure is shown in gray, with TMV and TMVI colored in blue and pink, respectively. Cavities associated to AGR1.135 (**A**) and AGR1.137 (**B**) binding were identified by SurfNet software are shown in orange and green. Volumes for each cavity are measured ($Å^3$).



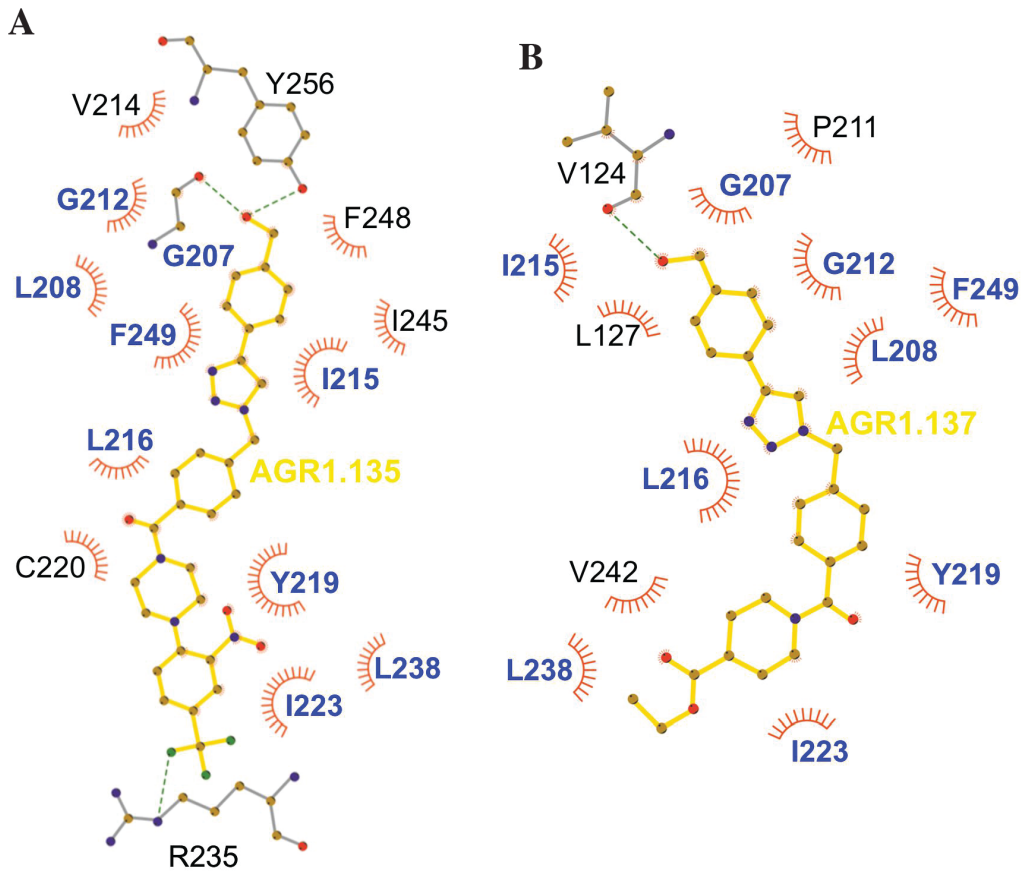
Supplementary Figure 9. Expression and function of CXCR4 and CXCR4 mutants coupled to AcGFP in transfected JK$^{-/-}$ cells. A) Flow cytometry analysis of cells were transiently transfected with the indicated CXCR4 mutants or wt CXCR4, coupled to AcGFP and their expression was determined 24 hours after transfection by GFP detection in flow cytometry. A representative experiment of 4 performed is shown. **B)** CXCL12-induced migration of CXCR4-defficient Jurkat cells (JK$^{-/-}$) transiently transfected with CXCR4wt or its mutants, CXCR4I204K, CXCR4G207I, CXCR4L208K, CXCR4R235L, CXCR4F249L, CXCR4Y256F and CXCR4S260A. Data are shown as the mean ± SD percentage of input cells that migrate (n = 3).



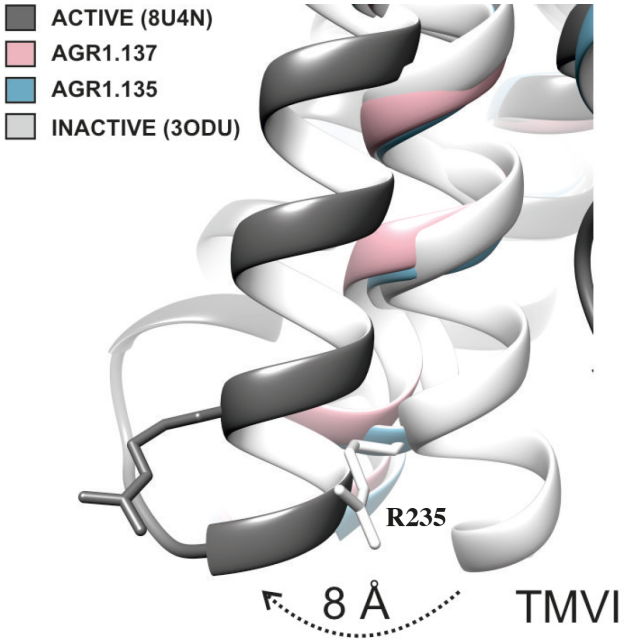
Supplementary Figure 10. Characterization of HeLa cells. A) CXCR4 expression on the surface of HeLa cells, analyzed by flow cytometry using a specific anti-CXCR4 antibody. Data show a representative experiment of 6 performed. **B**) Effect of the indicated compounds on HeLa cells growth. Cells ($0.5x10^6$) were plated on 48 well plates in the presence of 100 μM of the compounds, which were added freshly every 24 hours. Cells were counted at every 24 hours and maintained for 72 h. The mean ± SD (n=3) is shown.



Supplementary Figure 11. Residue conservation among all GPCR class A chemokine receptors for TMV and TMVI. Graphs were produced in the WebLogo server after aligning all receptors sequence in the sequence alignment tool included in GPCRdb. Font size for each residue represented in one letter code stands for the degree of conservation among all chemokine receptors. Residues are colored by their chemistry being green for polar, orange for hydrophobic, red for acidic, blue for basic and purple for neutral.

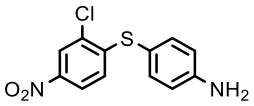
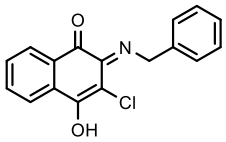
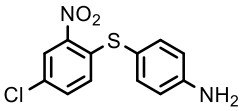
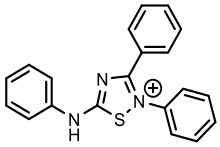
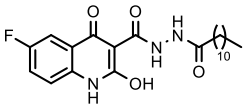
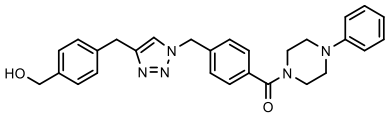
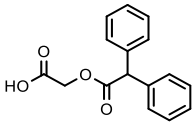
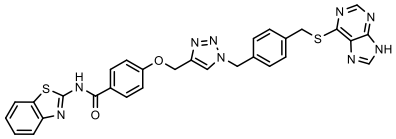
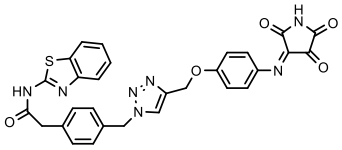
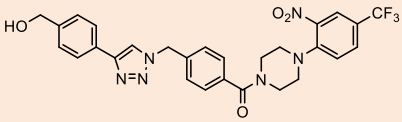
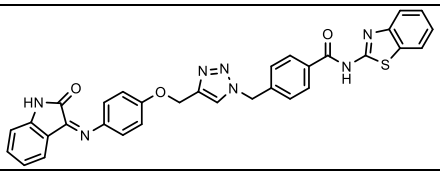
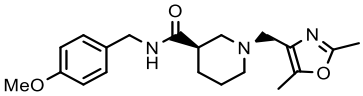
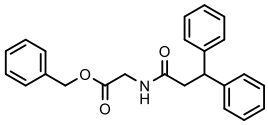
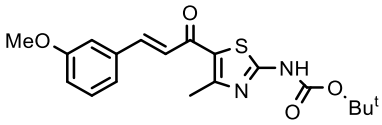
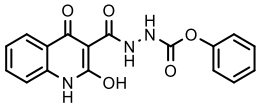
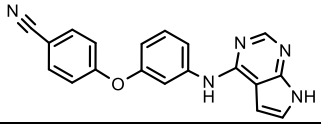
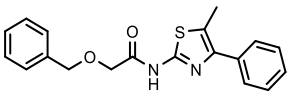
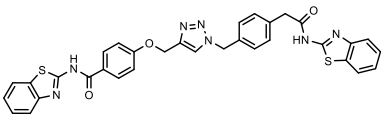
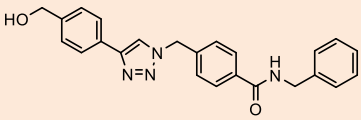
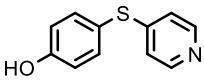


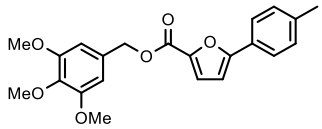
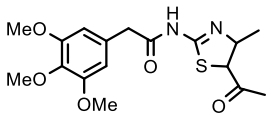
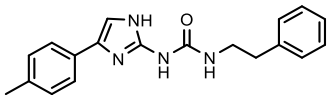
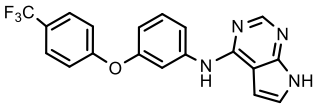
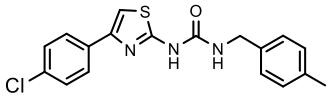
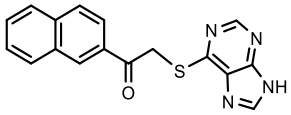
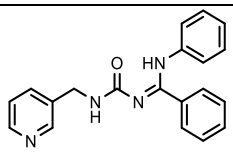
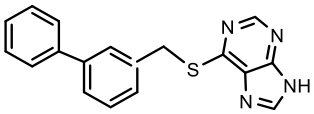
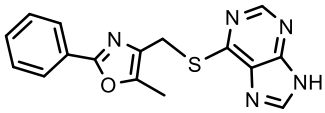
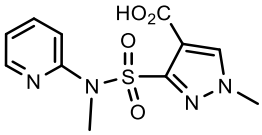
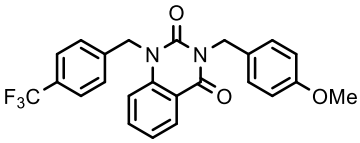
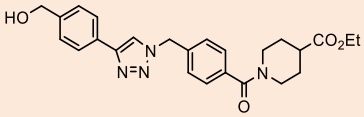
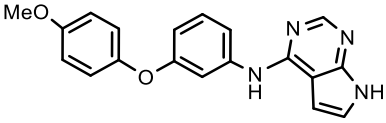
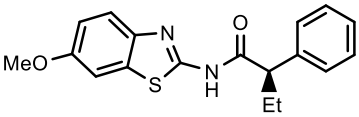
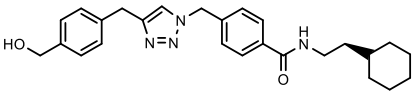
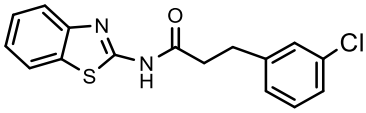
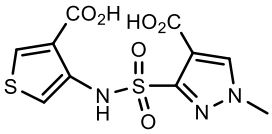
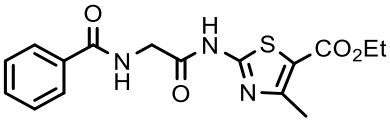
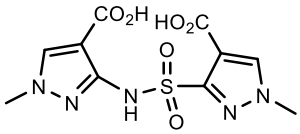
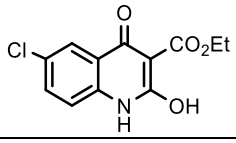
Supplementary Figure 12. Ligplot representation of antagonists interactions with CXCR4. Ligplot representations of AGR1.135 (**A**) and AGR1.137 (**B**), with hydrogen bonds represented in green dotted lines and non-bonded contacts in red spoke arcs. Residues labeled in blue font are present in the interface of both AGR1.135 and AGR1.137. CXCR4 residues forming a hydrogen bond with the ligands are shown in gray.



Supplementary Figure 13. TMVI transition from inactive to active position. Ribbon representation of TMVI of CXCR4 structures superimposed for inactive (PDB code 3ODU), intermediate active (AGR1.135, AGR1.137) and active (PDB 8U4N) upon interaction with no, negative allosteric modulators and CXCL12-induced Gi-protein interaction, respectively. Arginine 235 is depicted in stick representation as a reference for the movement of the protein.

Table S1. Compounds with minimal interaction energy in the area of interest

Ref.	Structure	Ref.	Structure
#1 AGF2.9C 1 % Inh: 38.01		#15 LNR1.17 % Inh: 48.71	
#2 AGF2.6 % Inh: 12.88		#16 VP1.23 % Inh: 30.69	
#3 VP4.29 % Inh: 0		#17 AGR1.13 8 % Inh: 45.68	
#4 MR4.17 % Inh: 3.43		#18 AGR1.10 3 % Inh: 53.55	
#5 AGR1.20 9 % Inh: 1.49		#19 AGR1.13 5 % Inh: 65.90	
#6 AGR1.10 1 % Inh: 36.19		#20 JHD1.47 % Inh: 11.79	
#7 MR3.24 % Inh: 2.14		#21 JAR3.8 % Inh: 21.03	
#8 VP4.27 % Inh: 0		#22 VNG1.61 % Inh: 0	
#9 MR2.57 % Inh: 0		#23 AGR1.21 0 % Inh: 40.15	
#10 AGR1.13 1 % Inh: 2.34		#24 AGF351 % Inh: 21.23	

#11 MR1.50 % Inh: 40.91		#25 IGS-2.73 % Inh: 27.36	
#12 AGR1.117 % Inh: 31.74		#26 VNG1.64 % Inh: 14.31	
#13 VSP3.1 % Inh: 36.43		#27 DM1.27 % Inh: 28.54	
#14 VP1.11 % Inh: 28.07		#28 ERP3.1 % Inh: 37.06	
Ref.	Structure	Ref.	Structure
#29 ERP2.51 % Inh: 52.50		#35 SC089 % Inh: 0.98	
#30 AGR1.267 % Inh: 30.06		#36 AGR1.137 % Inh: 89.82	
#31 VNG1.63 % Inh: 19.98		#37 IGS-4.32 % Inh: 0	
#32 AGR1.134 % Inh: 23.65		#38 IGS-2.3 % Inh: 0	
#33 SC513 % Inh: 12.94		#39 MR2.2 % Inh: 0	
#34 SC523 % Inh: 1.37		#40 JAR229 % Inh:	

Structure of the compounds employed in the study. The compounds studied in detail (AGR1.131, AGR1.135 and AGR1.137) are highlighted in green boxes. The percentage of inhibition of cell migration in response to CXCL12 is also shown.

Supplementary Table S2. Oligonucleotides used to generate the CXCR4 mutants

Mutation	Forward	Reverse
CXCR4 G207I	CAC ATC ATG GTT ATT CTT ATC CTG CCT	AGG CAG GAT AAG AAT AAC CAT GAT GTG
CXCR4 R235L	GGC CAC CAG AAG TTA AAG GCC CTC AAG	CTT GAG GGC CTT TAA CTT CTG GTG GCC
CXCR4 Y256F	TGG CTG CCT TAC TTT ATT GGG ATC AGC	GCT GAT CCC AAT AAA GTA AGG CAG CCA
CXCR4 S260A	TACTACATTGGGATCGCCATCGACTCCTTCATC	GATGAAGGAGTCGATGGCGATCCCAATGTAGTA
CXCR4 L208K	*CACATCATGGTTGGCAAGATCCTGCCTGGTATT*	*AATACCAGGCAGGATCTTGCCAACCATGATGTG*
CXCR4 I204K	TTCCAGTTTCAGCACAAGATGGTTGGCCTTATC	GATAAGGCCAACCATCTTGTGCTGAAACTGGAA
CXCR4 G212F	GCCCTTATCCTGCCTTTCATTGTCATCCTGTCC	GGACAGGATGACAATGAAAGGCAGGATAAGGCC
CXCR4 F249L	GTCATCCTCATCCTGGCTTTCTGCCTGTTGGCT GCCT	AGGCAGCCAACAGGCCAGGAAAGCCAGGATGAGGA TGAC

Movie S1 (separate file). Related to Figure 2. Representative movie of CXCR4-AcGFP on live JKX4$^{-/-}$ cells treated with DMSO and captured by SPT-TIRF, showing the diffusion of CXCR4 particles (monomers, dimers, and nanoclusters) at steady state (FN). The video was acquired and displayed at 10.2 frames/s.

Movie S2 (separate file). Related to Figure 2. Representative movie of CXCR4-AcGFP on live JKX4$^{-/-}$ cells treated with DMSO and captured by SPT-TIRF, showing the diffusion of CXCR4 particles (monomers, dimers, and nanoclusters) in response to CXCL12. The video was acquired and displayed at 10.2 frames/s.

Movie S3 (separate file). Related to Figure 2. Representative movie of CXCR4-AcGFP on live JKX4$^{-/-}$ cells treated with AGR1.131 and captured by SPT-TIRF, showing the diffusion of CXCR4 particles (monomers, dimers, and nanoclusters) at steady state (FN). The video was acquired and displayed at 10.2 frames/s.

Movie S4 (separate file). Related to Figure 2. Representative movie of CXCR4-AcGFP on live JKX4$^{-/-}$ cells treated with AGR1.131 and captured by SPT-TIRF, showing the diffusion of CXCR4 particles (monomers, dimers, and nanoclusters) in response to CXCL12. The video was acquired and displayed at 10.2 frames/s.

Movie S5 (separate file). Related to Figure 2. Representative movie of CXCR4-AcGFP on live JKX4$^{-/-}$ cells treated with AGR1.135 and captured by SPT-TIRF, showing the diffusion of CXCR4 particles (monomers, dimers, and nanoclusters) at steady state (FN). The video was acquired and displayed at 10.2 frames/s.

Movie S6 (separate file). Related to Figure 2. Representative movie of CXCR4-AcGFP on live JKX4$^{-/-}$ cells treated with AGR1.135 and captured by SPT-TIRF, showing the diffusion of CXCR4 particles (monomers, dimers, and nanoclusters) in response to CXCL12. The video was acquired and displayed at 10.2 frames/s.

Movie S7 (separate file). Related to Figure 2. Representative movie of CXCR4-AcGFP on live JKX4$^{-/-}$ cells treated with AGR1.137 and captured by SPT-TIRF, showing the diffusion of CXCR4 particles (monomers, dimers, and nanoclusters) at steady state (FN). The video was acquired and displayed at 10.2 frames/s.

Movie S8 (separate file). Related to Figure 2. Representative movie of CXCR4-AcGFP on live JKX4$^{-/-}$ cells treated with AGR1.137 and captured by SPT-TIRF, showing the diffusion of CXCR4 particles (monomers, dimers, and nanoclusters) in response to CXCL12. The video was acquired and displayed at 10.2 frames/s.

Movie S9 (separate file). Related to Figure 7A. Representative video of DMSO-treated HeLa cell migration on fibronectin-coated μ-Slide Chemotaxis chambers in the absence of CXCL12 gradient. Images over time (15 frames/s) are shown. Overlaid trajectories of cells shown in movie were detected and tracked using Fiji software.

Movie S10(separate file). Related to Figure 7A. Representative video of DMSO-treated HeLa cell migration on fibronectin-coated μ-Slide Chemotaxis chambers following a CXCL12 gradient on top. Images over time (15 frames/s) are shown. Overlaid trajectories of cells shown in movie were detected and tracked using Fiji software.

Movie S11 (separate file). Related to Figure 7A. Representative video of AGR1.131-treated HeLa cell migration on fibronectin-coated μ-Slide Chemotaxis chambers following a CXCL12 gradient on top. Images over time (15 frames/s) are shown. Overlaid trajectories of cells shown in movie were detected and tracked using Fiji software.

Movie S12 (separate file). Related to Figure 7A. Representative video of AGR1.137-treated HeLa cell migration on fibronectin-coated μ-Slide Chemotaxis chambers following a CXCL12

gradient on top. Images over time (15 frames/s) are shown. Overlaid trajectories of cells shown in movie were detected and tracked using Fiji software.