A STUDY OF VIRAL ENTRY PROTEINS AND RECEPTORS USING DEEP MUTAGENESIS

BY

KEVIN SEAN GILL

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biochemistry
in the Graduate College of the
University of Illinois Urbana-Champaign, 2023

Urbana, Illinois

Doctoral Committee:

Assistant Professor Beth M. Stadtmueller, Chair
Dr. Erik Procko, Director of Research
Professor Emeritus David Kranz
Associate Professor Christopher Brooke

ABSTRACT

Mutagenesis allows for the analysis and manipulation of protein interactions and can be especially insightful when protein complexes are challenging for structure determination. Both targeted and random mutagenesis have their benefits but also have limitations, especially as tools for testing the phenotypes of large numbers of mutations. Deep mutational scanning can overcome the limitation of finding desired mutations by testing many thousands of mutations for a specific phenotype in a single experiment. In particular, a single site-saturation mutagenesis library is composed of sequence variants where each variant has one mutation, and the library cumulatively contains each possible amino acid substitution for all regions of interest. This Big Data method can be used to discover binding sites or to find desirable mutations. I have used this method to investigate the dimerization interfaces of chemokine receptors CXCR4 and CCR5. CXCR4 has five homodimeric conformations crystallized, but CCR5 has only had the monomeric form crystallized. Using bimolecular fluorescence complementation (BiFC), I found a mutationally less tolerant region on CXCR4 that matched the known dimerization interface and found a similarly mutationally intolerant region for CCR5, which could be a potential dimerization interface. In the process, I found that many mutationally tolerant regions for self-association were disruptive, leading to non-specific aggregation. Mutations that decreased self-associations had increased binding to ligand but decreased calcium signaling activity in response to an agonist. Overall, the data suggest multiple mechanisms are involved in receptor dimerization. I then assisted in the search for mutations in HIV-1 Env trimer that promoted a closed conformational state, as the naturally conformationally flexible viral spike uses this property to evade the immune system by 'directing' antibody production towards strain-specific epitopes rather than generating broadly neutralizing antibodies. In another scan sorting for the

ability of HIV-1 Env to bind to co-receptor CD4, I found that the glycosylation site N262 is highly conserved for binding to CD4 by stabilizing a high affinity-conformation, and that mutations to that site resulted in an intrinsic binding deficit to CD4. Lastly, I assisted in investigating the binding interactions between HCMV and the receptor it uses for viral entry, PDGFRα. We found a mutation in PDFGRα that resulted in preferential binding to HCMV over its endogenous ligands PDGFs. Creating a soluble Fc-fused variant containing the mutation, we are developing an antiviral for HCMV. I found that murine PDGFRα is sufficiently homologous to human PDGFRα that the receptors and their ligands cross react, which will allow us to test in a mouse model in the future the tolerability and pharmacokinetics of the decoy receptor, with that data being relevant to humans.

ACKNOWLEDGEMENTS

I would like to start by acknowledging Erik Procko for guiding and supporting me all this time. When I first joined the lab, he helped me get my bearings and taught me to trust my judgement and speak my mind. He respected my time and trusted me to work to the best of my ability. He helped me be a better scientist. But the most important thing is that he was present, willing to meet and talk about my work when early in the lab, and after joining Cyrus Biotech he was still willing to meet on zoom frequently to give whatever guidance he could. I would like to thank and acknowledge my committee, Beth Stadtmueller, Dave Kranz, and Chris Brooke, for their guidance and insight. And I would like to acknowledge my undergraduate lab mentors Eric Vimr and Susan Steenbergen, who trained me to work in a lab and without whom I wouldn't have joined the graduate program.

As for former and fellow graduate students and post-docs, I would like to acknowledge Jihye Park, Jeremiah Heredia, and Heather Ellis as my seniors in the lab who helped me whenever they could with any questions I had and made me feel welcomed. I would like to acknowledge my other lab mates and friends Christine Devlin and Krishna Narayanan, who continue to support me now, and helped me get here today. I would like to acknowledge Savanna Skeeters and Page Daniels, two members of my cohort who have been great friends to me in the biochemistry department together. Other former and fellow graduate students I would like to acknowledge for their friendship are Kritika Mehta, Stella Ekaputri, and Sonya Bharathkar.

I would also like to acknowledge my family and friends outside of graduate school. My parents, sisters and grandmother have been supporting me this whole time, and I am forever grateful for them going out of their way to help me. I acknowledge my best friends Jason Su, Daniel

Baldwin, and Michael Chan, who listen to my ramble on about whatever topic on my mind and being able to join in on a call whenever I was the last person in lab at night.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Transmembrane proteins are a major method to relay signals from outside of the cell to induce a wide variety of cellular responses. However, the membrane proteins that decorate the cell surface are also exposed as targets for viruses. In the early replication cycle of a virus, virions can take advantage of host membrane proteins as receptors to attach and infect cells[1]. In understanding these receptors' endogenous mechanisms as well as their roles for viral entry, we might better treat viral infection while leaving undisturbed the endogenous function of these transmembrane proteins. To accomplish this level of understanding, insights may be learned from Big Data methods to analyze many different interactions in a smaller number of experiments.

DEEP MUTATIONAL SCANNING

Mutagenesis is a common means of investigating the molecular mechanism of receptor-ligand binding and protein activity. There are many approaches of applying mutagenesis, including site-specific mutagenesis like alanine scans or random mutagenesis[2,3]. These methods have been very useful in discovering important aspects of protein function. However, the desire of a Big Data method that can achieve a more comprehensive and, compared to targeted site-specific mutagenesis, a less biased analysis of protein mutations has led to deep mutational scanning. Deep mutational scanning[4,5] takes a protein of interest and makes mutations at each amino acid site of interest and mutates the site to each of the other amino acids, generating a library of variants. These protein variants are then measured for a desired phenotype, with those variant sequences that have the desired phenotype getting collected by some in vitro selection method (e.g. phage panning, fluorescence-activated sorting of expressing cells, etc). This library can be sorted/selected as many times as needed until there is a final library of selected variants that can

be compared to the original library by deep sequencing. This method allows for the testing of up to tens of thousands of variants quickly in an unbiased manner to find variants that have been enriched or depleted for the phenotype of interest.

Deep mutational scanning has been used to great effect in a number of studies, including research into interactions between ACE2 and the receptor binding domain of SARS-CoV-2[6,7], the mutational tolerance of HIV Envelope protein [8], TAPBPR and its interactions with MHC-I[9,10], and the generation of antibody variants [11]. This method can be used with a wide range of library sizes, from a single domain to an entire protein sequence, with many kinds of mutational diversity, from all possible variations at every position to a focused single site-saturation mutagenesis library.

CXCR4 AND CCR5 DIMERIZATION

G Protein Coupled Receptors (GPCRs) are integral membrane proteins that activate G proteins leading to many potential downstream effects. They make up the largest class of membrane proteins in the human genome[12] and about 35% of drug targets[13]. There are six classes of GPCRs, with each class having multiple subtypes. The largest of the classes are class A GPCRs, also known as rhodopsin-like GPCRs, which have a variety of functions such as acting as photoreceptors[14], providing pain relief[15], stimulating the sympathetic nervous system [16], and chemotaxis[17–19]. A commonality among GPCRs is that they have 7 integral transmembrane helices.

Those class A GPCRs that perform chemotaxis are known as chemokine receptors. Human chemokine receptors are named in reference to disulfide linkage motifs found in the chemokines they bind, i.e. C, CC, CXC and CX3C[20]. In response to a concentration gradient of chemokines, chemokine receptors cause downstream signaling causing the cell to move along the gradient[21]. These chemokine receptors are necessary for migration of all types of immune cells, from cell development to recruitment at sites of infection. Of the chemokine receptors, CXCR4 and CCR5 have been some of the most studied.

CXCR4 is one of the most studied chemokine receptors due to its roles in immune cell homing to bone marrow, metastasis, and as a co-receptor in HIV infection[19,22–24]. CXCR4 has the canonical ligand stromal cell-derived factor 1 (SDF-1), also known as CXCL12[25], though is also known to bind and respond to macrophage migration inhibitory factor [26]. When CXCL12 binds to CXCR4, it can lead to a variety of downstream pathways, resulting in chemotaxis, cell proliferation/differentiation, and the activation of other genes[27,28].

CCR5 is similarly well studied. Along with CCR5's role in leukocyte migration to areas of inflammation, CCR5 is most well known as a co-receptor to HIV infection[29,30]. The CCR5-Δ32 mutation is particularly famous for conferring resistance to HIV infection. CCR5 was once considered non-essential due to individuals with the mutation appearing seemingly healthy. Recent studies however have suggested that there are risks to losing functionality of CCR5[31,32]. The importance of these receptors has further fueled more research investigating their activation and interactions with other receptors.

It is generally agreed upon that GPCRs dimerize. Many GPCRs have been crystallized as dimers. Generally, class A GPCRs have been observed to form homodimers using transmembrane helix 1 or transmembrane helices 4-6. Rhodopsin forms a dimer at TM1 and H8 in nanodiscs, while also forming contacts with other rhodopsin proteins with the cytoplasmic ends of TM 5-6[33], whereas μ-opioid receptor dimerizes at TM 5-6[34]. CXCR4 was found to dimerize in five different crystal structures with different ligands bound[35]. These structures are symmetrical in nature, with contacts primarily at the extracellular end of TM 5-6, though also showing intracellular contacts with TM 3-4 when in a complex with a larger ligand like CVX15 (**Figure 1**).

There are other methods of examining GPCR dimers besides crystallization. Dopamine D2 receptor has been shown to dimerize through co-immunoprecipitation and single molecule tracking[36,37]. Co-immunoprecipitation has also been used to suggest a dimer of β_2-adrenergic receptor using two different tags[38] and M_2 muscarinic cholinergic receptor[39]. Fluorescence resonance energy transfer (FRET) and Bioluminescence energy transfer (BRET) are also useful methods for viewing dimers and higher order oligomers[40,41].

These methods are useful tools but are not able to answer all questions. While experiments like coimmunoprecipitation and FRET are useful as evidence for dimerization, they cannot reveal specific dimerization interfaces. Adding to these issues is that crystallographic methods are not always reliable for finding dimers. There is evidence to suggest that CCR5 dimerizes/oligomerizes[42,43], but to date there is no crystal structure of CCR5 in a homodimer.

Therefore, there is a need for other ways of looking at homodimer interactions, such as deep mutational scanning.

THE HIV-1 ENV TRIMER

Human immunodeficiency virus (HIV) infects CD4+ T cells, over time damaging the immune system to a point where the body can no longer fend off against other secondary infections. As medicine has developed, HIV has become treatable, and now there are over 38 million people worldwide living with HIV (World Health Organization [WHO], 2021). However, despite these advances, there is no FDA approved vaccine for HIV, and access to treatments for HIV are less accessible depending on income and location. There are two types of HIV, HIV-1 and HIV-2. HIV-1 is the more common of the two, having been spread worldwide and having a higher virulency and transmissibility.

HIV infects cells via its spike, the HIV Envelope protein (Env), the only protein on the surface of the virion[44]. Env exists as a trimer of glycoproteins each 160 kDa (gp160) that is modified through a furin cleavage site to form mature gp120 and gp41, resulting in a trimer of heterodimers[45,46]. The trimer is also heavily glycosylated, resulting in up to half of its weight coming from glycans depending on the variant or strain[47]. Starting in a closed trimer state, Env first binds to CD4, resulting in conformational changes that opens the trimer, allowing for binding to either CXCR4 or CCR5 depending on the tropism of the strain. This results in further conformational changes that allows gp41 to initiate fusion with the membrane.

Alongside glycan coverage, mutational tolerance, and low surface expression, this conformational flexibility provides a means to escape antibody neutralization, as many antibodies are generated to conformations that expose non-conserved epitopes and are thus not broadly neutralizing. However, in the closed conformation, conserved epitopes are more likely to elicit broadly neutralizing antibodies[48]. SOSIPs are constructs of HIV Env that have been solubilized by removing the transmembrane region of gp41 and have had mutations introduced that insert disulfide bonds and I559P, which destabilizes the open conformation and promotes the closed state (**Figure 2**)[49,50]. They are commonly generated as they are easy to make and purify stably. However, in situations where making full length Env is desirable, it is difficult to isolate properly folded proteins from misfolded proteins. Making targeted mutations in full length Env that minimally disrupt structure while promoting the closed state allows for another useful tool in research and could be used to promote more broadly neutralizing antibody production.

AN ANTIVIRAL FOR HCMV

Human cytomegalovirus (HCMV) is a pathogen that infects over 50% of people in the United Stated by 6 years of age and over 90% for individuals 80 or over[51]. The virus remains dormant in most individuals but can cause harm in immunocompromised patients. Another worry is that HCMV when acquired congenitally can lead to neurological complications. And due to the ease of transmissibility, it is difficult to prevent infection. As there is no FDA approved antiviral for neonates, there is a demand for a safe antiviral for use.

HCMV has two major mechanisms of entry into cells. The first is the viral trimer complex gH/gL/gO which will bind to platelet-derived growth factor receptor alpha (PDGFRα) and the

second is the pentamer complex gH/gL/UL128/UL130/UL131 which binds to Nrp2[52]. A soluble version of PDGFRα has been shown to compete off HCMV binding to endogenous PDGFRα and blocks infection of all cell types[53–55]. While an interesting starting point for a decoy receptor, using soluble PDGFRα is not a safe option. PDGFRα is a receptor tyrosine kinase that is responsible for mesenchymal cell growth, survival, and migration from embryonal development into adulthood upon activation by PDGFs. Soluble PDGFRα will also compete off PDGFs from binding to endogenous receptor, leading to potentially dire side effects. We can take advantage of certain traits of the PDGF binding site, primarily that the binding site for PDGFs likely overlaps with the binding site for HCMV trimer[54,56]. Using deep mutational scanning methods, we can look for mutants in PDGFRα that will preferentially bind to viral trimer complex over PDGFs.

This method of developing decoy receptors has an advantage over antibody treatments. As the decoy receptor will look very similar to the wild type receptor, any mutations that allow the virus to escape binding to the decoy receptor also risk escape from binding to the endogenous receptor. The use of decoy receptors can potentially be further expanded into use with other viruses whose receptors can be solubilized. Similar attempts are currently being made with other receptors, such as work to create a soluble variant of ACE2 that preferentially binds the spike protein of SARs-CoV-2[7].

In this thesis, I use deep mutational scanning to analyze single site-saturation mutagenesis libraries of chemokine receptors CXCR4 and CCR5, HIV-1 Env, and PDGFRα to learn about their interactions with protein ligands. In my main body of work, CXCR4 and CCR5 were sorted

for their ability to form self-associations using bimolecular fluorescence complementation. The mutational landscape generated from CXCR4 closely matched with known crystallographic dimers, which sets up the basis for suggesting a dimerization interface for CCR5. In the process of these experiments, I also found that the common method of overexpressing these receptors to view dimerization results in a high number of non-specific aggregation mutants when sorting for increased dimerization. Using the experience from developing the original libraries of chemokine receptors, HIV-1 Env was sorted for its ability to bind to broadly neutralizing antibody PG-16, which prefers the closed conformation, to generate a construct that might elicit an antibody response consisting of more broadly neutralizing antibodies. Lastly, we looked at interactions between PDGFRα and the trimer complex of HCMV in the absence and presence of PDGFs, so that we could generate a decoy receptor that preferentially binds to viral receptor over the endogenous PDGF ligands.

FIGURES

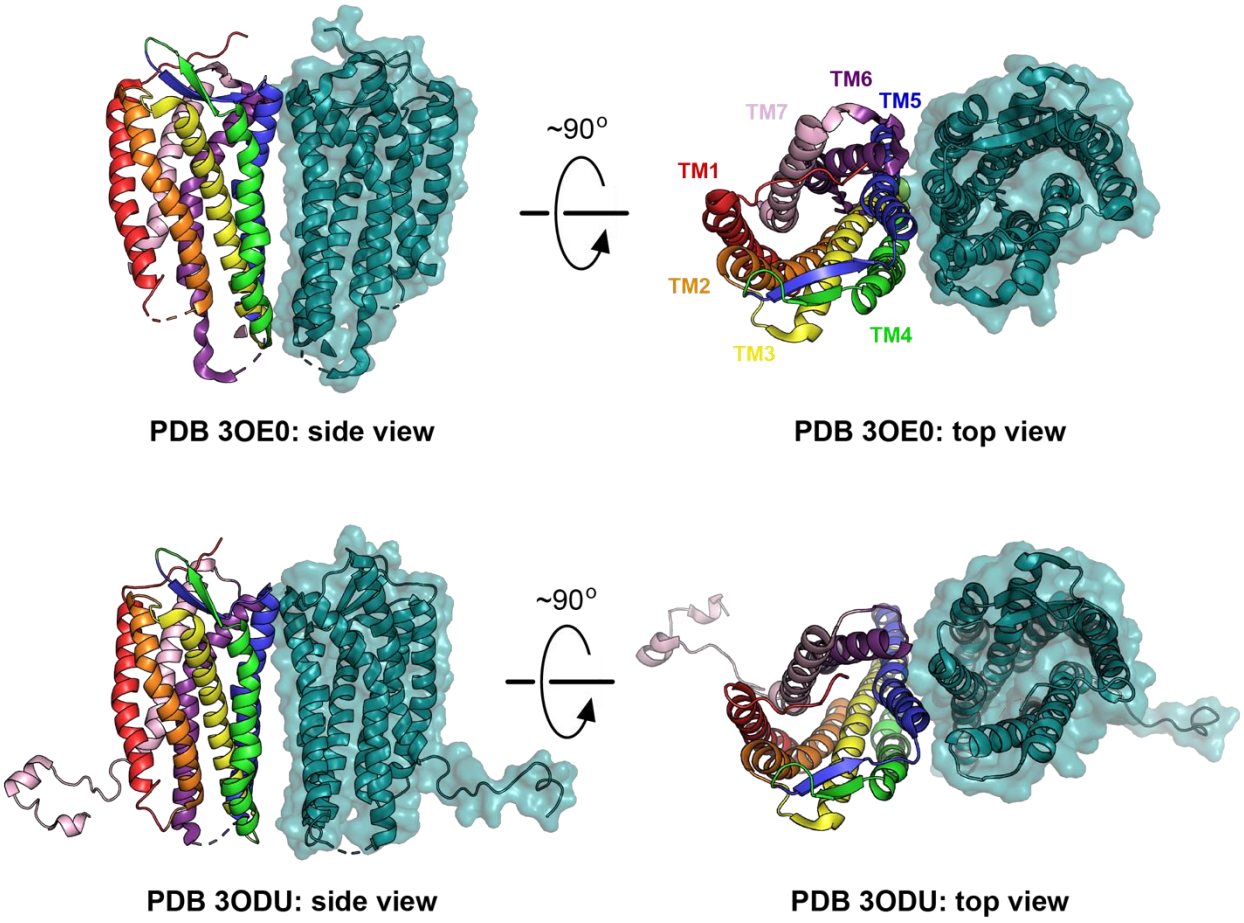

Figure 1.1. Dimeric organization of CXCR4. Two representative crystal structures of dimeric CXCR4 (PDB IDs 3OE0 and 3ODU) are shown as ribbons, with one monomer colored by transmembrane segments and the second colored teal. The transmembrane segments are colored red (N-terminus and TM1), orange (TM2), yellow (TM3), green (TM4), blue (TM5), purple (TM6), and pink (TM7 and cytosolic tail). Despite their differences, the two structures show similar dimeric organization with TM3, TM4, and TM5 at the interface.

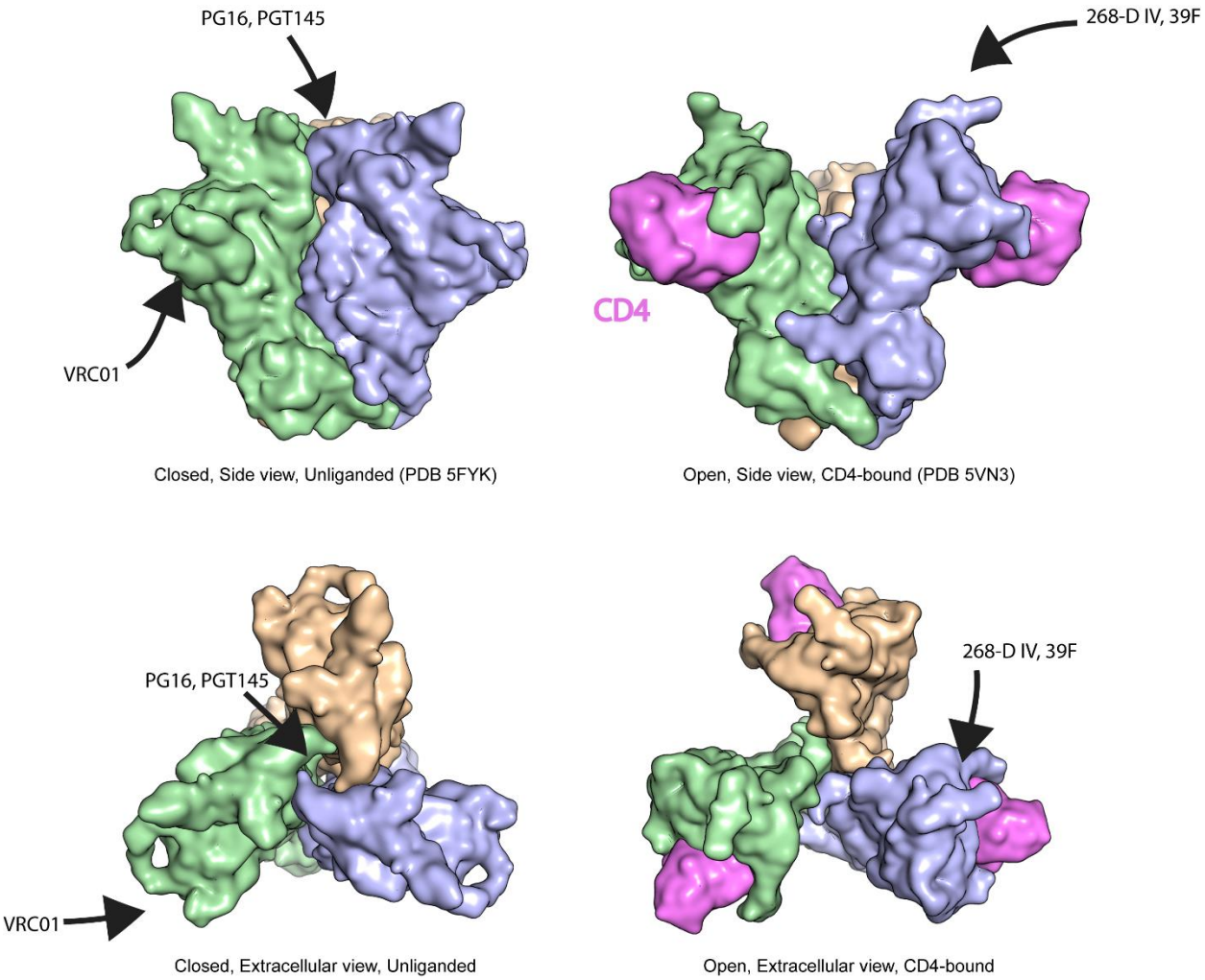

Figure 1.2 Antibody Binding Sites of HIV Env-1. Two representative structures of prefusion JR-FL SOSIP.664 and open conformation B41 SOSIP.664 in complex with soluble CD4 (D1-D2) (PDB IDs 5FYK and 5VN3) are shown with individual gp140 monomers colored in green, blue, and beige, and sCD4 colored in pink. Arrows point to binding sites of broadly neutralizing antibodies VRC01, PG16, and PGT145, as well as anti-V3 antibodies 268-D IV and 39F.

CHAPTER 2: MULTIPLE MECHANISMS OF SELF-ASSOCIATION OF CHEMOKINE RECEPTORS CXCR4 AND CCR5 DEMONSTRATED BY DEEP MUTAGENESIS

ABSTRACT

Chemokine receptors are members of the rhodopsin-like class A GPCRs whose signaling through G proteins drives the directional movement of cells in response to a chemokine gradient. Chemokine receptors CXCR4 and CCR5 have been extensively studied due to their roles in white blood cell development and inflammation and their status as coreceptors for HIV-1 infection, among other functions. Both receptors form dimers and oligomers but the function/s of self-associations are unclear. While CXCR4 has been crystallized in a dimeric arrangement, available atomic resolution structures of CCR5 are monomeric. To investigate the dimerization interfaces of these chemokine receptors, we used a bimolecular fluorescence complementation (BiFC)-based screen and deep mutational scanning to find mutations that modify receptor self-association. Many disruptive mutations promoted self-associations nonspecifically, suggesting they aggregated in the membrane. A mutationally intolerant region was found on CXCR4 that matched the crystallographic dimer interface, supporting this dimeric arrangement in living cells. A mutationally intolerant region was also observed on the surface of CCR5 by transmembrane helices 3 and 4. Mutations from the deep mutational scan that reduce BiFC were validated and were localized in the transmembrane domains as well as the C-terminal cytoplasmic tails where they reduced lipid microdomain localization. The reduced self-association mutants of CXCR4 had increased binding to the ligand CXCL12 but diminished calcium signaling. There was no change in syncytia formation with cells expressing HIV-1 Env. The data lead us to conclude that multiple mechanisms are involved in receptor dimerization.

INTRODUCTION

G-protein coupled receptors (GPCRs) are integral membrane proteins that activate heterotrimeric G protein signaling pathways, leading to diverse downstream signals. They are involved in sight, hormonal regulation, white blood cell trafficking, and a wide range of other physiological functions. While it is accepted that many GPCRs dimerize, the functional consequences of dimerization on ligand interactions and downstream signaling remain incompletely understood, with structural experiments of class A rhodopsin-like GPCRs suggesting they bind to ligand as monomers[57], and coimmunoprecipitation and FRET experiments suggesting they are active as dimers[18,38,58]. Ligand-receptor interactions may have additional complexities where the ligands may act as monomers or dimers with different effects[59] and there may be negative cooperativity between subunits in a GPCR dimer such that effectively only one of the receptors is ligand-bound[60]. To determine whether dimerization is important for GPCR activity, researchers have tried different methods to identify dimerization interfaces for class A GPCRs with limited success. While class C GPCRs such as $GABA_B$ receptors, metabotropic glutamate receptors and sweet and umami taste receptors are obligate dimers[61,62], the dimerization of class A GPCRs is less understood, although it has been observed in multiple crystal structures including those of rhodopsin[33] and μ-opioid receptor[34]. Different transmembrane helices (TM) have been observed at dimer interfaces, including TM1 and TM4-6.

We have previously used conformation-dependent antibodies against the class A GPCRs CXCR4 and CCR5 to develop a method for deep mutational scanning (DMS) of transmembrane proteins expressed on human cells[63]. CXCR4 and CCR5 are chemokine receptors, which in response to a

chemokine gradient cause cell migration, often of white blood cells, but also of other cell types and some cancers[19,23]. These two chemokine receptors are also well studied for their role as coreceptors for HIV-1 infection[24,29,30]. There is extensive evidence for CXCR4 dimerization and oligomerization[28,64–66] and the protein crystallizes as a dimer[35]. Wu et al. showed a dimerization interface using TM3-5 and the extracellular tip of TM6 that was consistent across 5 different crystal forms. Structures of CCR5 have been determined with and without ligands as a monomer[67–69] but not yet as a homodimer. Models for CCR5 dimerization have been made through comparisons to other class A GPCR dimeric structures and molecular dynamics simulations[70,71]. These methods have suggested multiple different dimerization interfaces that collectively cover almost all sides of the protein. Given GPCRs are known to oligomerize and hetero-oligomerize with other GPCRs[72,73], finding physiologically relevant dimerization interfaces has been challenging.

By examining the mutational landscapes of CXCR4 and CCR5 for binding to antibodies 12G5 and 2D7, respectively, we found conserved sites on the surface of the transmembrane domains that, with respect to CXCR4, matched previous crystallographic dimers[35]. To further elucidate how these chemokine receptors dimerize, we used bimolecular fluorescence complementation (BiFC) with deep mutational scanning to search for membrane-exposed surface patches that were intolerant of mutations for high BiFC signal. We found regions on the chemokine receptors that are hypothesized sites for dimerization, which on CXCR4 again matched closely with the crystallographic dimer conformation. We also found many structurally-disruptive mutations that increased BiFC signal, including nonspecifically with an unrelated GPCR, suggesting overexpression of mutant chemokine receptors may lead to aggregation. Finally, mutations were

found in the cytoplasmic tails that reduced BiFC between receptors and also reduced colocalization with a marker for lipid microdomains. We conclude that there are multiple mechanisms that mediate self-association of chemokine receptor polypeptides.

MATERIALS AND METHODS

Cell Culture

The X4-KO Expi293F and the MA-VN expressing X4-KO Expi293F cell lines are previously described[63,74]. Cells (Thermo Fisher) were cultured in Expi293 Expression Media (Thermo Fisher) at 37 °C, 8% CO_2, and 125 rpm. Cells were transfected at a density of 2×10^6 cells/ml with 500 ng plasmid DNA per ml of culture using Expifectamine (Thermo Fisher) according to the manufacturer's directions. To generate stable lines for BiFC, X4-KO Expi293F cells were transfected with linearized pCEP4-FLAG-CXCR4-VN or pCEP4-FLAG-CCR5-VN and selected with hygromycin B (200 μg/ml). The population of CXCR4-VN or CCR5-VN positive cells was enriched by FACS, in which cells were washed in Dulbecco's phosphate buffered saline (PBS) supplemented with 0.2% bovine serum albumin (BSA), stained with 1:200 chicken anti-FLAG-FITC (Immunology Consultants Laboratory) in PBS-BSA, washed and resuspended for sorting on a BA FACS Aria II. The collected cells were expanded by culturing in Expi293 Expression Medium supplemented with penicillin-streptomycin.

Plasmids

Plasmids for myc-tagged CCR5 and CXCR4, used in signaling and syncytia assays, are previously described[63] and available on Addgene (# 98948 and # 98946). Plasmids for myc-CXCR4-VC (Addgene # 98967), myc-CCR5-VC (# 98966), myc-GRM3-VC (# 98968), FLAG-

CXCR4-VN (# 98964), FLAG-CCR5-VN (# 98963), and FLAG-GRM3-VN (# 98965) are previously described[63]. Briefly, the N-terminal half (VN) of the yellow fluorescent protein Venus (a.a. 1-154; mutant I152L) or the C-terminal half (VC) of Venus (a.a. 155-238) were fused to the C-termini of tagged receptors and cloned into the NheI-XhoI sites of pCEP4 (Invitrogen). Plasmids for HIV-1 MN gp160 and BaL gp160 (# 100919) are previously described[63]. pCMV3-CD4 was from Sino Biological (# HG10400-UT).

Library Generation

Single site-saturation mutagenesis (SSM) libraries of human myc-tagged CXCR4 and CCR5 were generated previously[63]. Here, the libraries were modified by using PCR-based fragment assembly to fuse the 3' ends of mutated chemokine receptor sequences to VC (a.a. 155-238 of Venus). The PCR products were digested and ligated into the NheI-XhoI sites of pCEP4 and electroporated into *E. coli* DH5α cells. The number of transformants was > 100x the theoretical sequence diversity. The myc-CXCR4-VC and myc-CCR5-VC plasmid libraries were transfected into stable FLAG-CXCR4-VN or FLAG-CCR5-VN Expi293F lines under conditions where cells typically express no more than one mutant VC-fused chemokine receptor gene per cell: 1 ml culture at a density of 2×10^6 cells was transfected with 1 ng library DNA diluted with 1.5 µg pCEP4ΔCMV carrier DNA using Expifectamine (Thermo Fisher). The media was replaced 2 h post-transfection and cells were harvested for sorting 24 h post-transfection.

Sorting Cells for Surface Expression and High BiFC Signal

Cells transfected with the libraries were washed with cold PBS-BSA and incubated with 1/300 anti-myc Alexa 647 (clone 9B11; Cell Signaling Technology) plus 1/300 anti-Flag Cy3 (clone

M2; Sigma-Aldrich) in PBS-BSA for 30 minutes before being washed twice and resuspended in cold PBS-BSA supplemented with 1/100 fetal bovine serum (FBS). Cells were sorted on a BD FACS Aria II at the Roy J. Carver Biotechnology Center. Cells were first gated by forward-side scatter and side scatter for the main population of cells. For sorting based on surface expression, cells positive for the myc tag (Alexa 647 fluorescence) were collected. For sorting based on high BiFC signal, within the expression gate the top 5% of cells for YFP/Venus fluorescence relative to surface receptor expression were collected. Cells were sorted for no longer than 4 h to maintain high viability and were collected in tubes coated overnight in FBS. Collected cells were centrifuged and the pellets stored at -80 °C.

Illumina Sequencing and Analysis

RNA was extracted from frozen cells from FACS using the GeneJET RNA Purification Kit (Thermo Fisher) and reverse transcribed with AccuScript primed with the EBV reverse sequencing primer for first strand cDNA synthesis. The receptor cDNA was PCR amplified as three overlapping fragments to achieve full coverage of the gene. In a second round of PCR, experiment-specific barcodes and adaptors for annealing to the Illumina flow cell were added. Samples were sequenced (2 x 250 nt) using Illumina HiSeq 2500 at the UIUC Roy J. Carver Biotechnology Center. Data were analyzed using Enrich (Fowler et al., 2011). Log_2 enrichment ratios of mutants were adjusted by subtracting the enrichment of the wild type sequence. Data and analysis commands are deposited with NCBI's Gene Expression Omnibus under series accession number GSE125426.

Validation of Receptor Mutants with Changes in BiFC

pCEP4 plasmids containing wild type VN and VC fused receptors were used as templates for targeted mutagenesis by overlap extension PCR. Mutated inserts in all plasmids were confirmed by Sanger sequencing. Plasmids were transiently transfected as described above into the relevant cell line and analyzed by flow cytometry for receptor expression and BiFC signal 24 h post-transfection. Cells were washed with PBS-BSA, incubated in 1:100 anti-myc Alexa 647 and 1:100 anti-FLAG Cy3 in PBS-BSA for 20 minutes, washed twice, and resuspended in PBS-BSA for analysis on a BD LSR II. Data were collected using instrument software and analyzed with FCS Express (De Novo Software). Cells were gated by forward-side scatter for the main population, and then gated by Alexa 647/Cy3 fluorescence to control for a consistent level of surface expressed receptor across the samples. Mean YFP/Venus fluorescence was recorded.

Co-Localization with Matrix

Using a MA-VN cell line (Heredia et al., 2019), cells were transiently transfected as described above with wild-type or mutant myc-tagged receptors fused to VC and cloned into pCEP4. Cells were processed 24 h post-transfection. Cells were washed PBS-BSA and incubated in a 1:200 dilution of anti-myc Alexa 647 (clone 9B11; Cell Signaling Technology). Cells were washed and analyzed on a BD LSRII flow cytometer.

CXCL12 Binding Assay

Transfected X4-KO Expi293F cells were harvested 24 h post-transfection, washed with PBS-BSA and incubated in 1:200 anti-myc-Alexa 647 (clone 9B11; Cell Signaling Technology) and

10 µM CXCL12-sfGFP for 30 minutes. Cells were washed and analyzed on a BD LSR II. The preparation of CXCL12-sfGFP is described elsewhere (Heredia et al., 2018).

Calcium Mobilization

X4-KO Expi293F cells were transiently transfected as described above with myc-tagged CXCR4 or CCR5 plasmids. Cells were prepared at room temperature 24 h post-transfection. Cells were washed with assay buffer (PBS containing 0.2% BSA and 1 mM $CaCl_2$) and resuspended in assay buffer with 2 µM Fluo-4-AM (Life Technologies). Fluo-4-AM was prepared as a 1 mM stock in DMSO. Cells were incubated with 1:250 anti-myc-Alexa-647 (clone 9B11; Cell Signaling Technology) for 30 minutes with frequent mixing, and washed and resuspended in assay buffer. Cells were analyzed on a BD Accuri C6 Cytometer. Cells were gated by forward-side scatter for the main population and the Fluo-4 fluorescence of the Alexa 647 positive population was monitored over time. To cells at baseline, CXCL12 (PeproTech) or CCL5 (PeproTech) were added at final concentrations of 1, 10 and 100 ng/mL. Fluorescence returned to baseline within 120 s, at which point ionomycin was added to a final concentration of 4 µM. The spike in Ca2+-dependent Fluo-4 fluorescence in response to the chemokine was normalized to the maximal response induced by ionomycin.

Syncytia Formation Assay

X4-KO Expi293F cells (2×10^6 cells in 1 ml) were transfected with 50 ng pCMV3-CD4 and 450 ng pCEP4-myc-CXCR4 or pCEP4-myc-CCR5. A separate set of cells (2×10^6 cells in 1 ml) were transfected with 500 ng pCEP4-gp160 from the MN or BaL HIV-1 strains. Empty vector was used for control transfections. After 5 h incubation at 37 °C, 8% CO_2, 125 rpm, 0.2×10^6

receptor-expressing cells and gp160-expressing cells were mixed and added to 0.6 ml Expi293 Expression Medium. The cells were added to wells of a 12-well tray that had been incubated for 30 minutes with 0.01% poly-L-lysine (Sigma) to coat the plastic surface. Cells were incubated without agitation for 20 h. Wells were washed with warm PBS and attached cells removed with 0.25% trypsin-2.21 mM EDTA for 15 minutes at 37 °C. Detached cells were washed with cold PBS-BSA and analyzed on a BD LSR II flow cytometer to detect syncytia based on high forward-side scattering. The positive gate was set at less than 1% of negative control cells.

Single Molecule Imaging of CXCR4

A halo tag sequence was inserted into the myc-CXCR4 wild type and low BiFC mutant sequences, in between the myc tag and the chemokine receptor sequence. Coverslips measuring 24 mm × 40 mm (VWR Cat. No. 48393230) were cleaned for single molecule experiments as described previously [75]. The washed coverslips were stored in molecular grade water, air-dried, and flamed for a few seconds. Coverslips were then immersed in a clean petri dish with 50 µg/mL poly-L-Lysine (PLL; Cat. No. P1274; Sigma-Aldrich) and incubated overnight. Coverslips were rinsed twice in molecular grade water and air dried. A large PDMS chamber was carefully assembled over a dried coverslip and placed in a 60 mm dish.

X4-KO Expi293F cells were transfected with wildtype or mutant myc-halo-CXCR4. 2 h post-transfection, cells were diluted 1:10 in Expi293 Expression Media and placed on the assembled PDMS coverslip chamber. 24 h post-transfection, cells were gently washed on the coverslip with PBS containing calcium and magnesium (Cat. No. 21-030CM; Corning) and incubated in 2 nM JF549-Halo-ligand (Cat. No. GA1110; Promega) for 15 minutes at 37 °C. Cells were washed

three more times and incubated in PBS for another 15 minutes before four more washes and covered in PBS.

Single molecule tracking was performed at room temperature in a custom-built TIRF microscope (TIRFM). A 100X oil immersion objective (100X, N.A. 1.49, oil immersion) was assembled on an inverted microscope. A 561 nm laser was used to excite the labelled molecules. Power of the laser was controlled using neutral density filters. Individual labeled molecules were seen as diffraction-limited spots on the cell surface in immediate contact with the coverslip. The fluorescence from the single spots was collected by the same objective, passing an emitter and captured by an Electron Multiplying Charge Coupled Device (EMCCD) camera. A total of 2400 frames/trajectory were acquired for each field of view with an integration time of 50 ms. The collected data was exported to FIJI and single spots were tracked using the plugin TrackMate [76]. Individual spots were selected as 3X3 pixel bright features. Threshold was selected such that all the single spots were selected in a frame. For track generation, the linking distance was fixed to 2 pixels, merging was not allowed, Gap closing distance was set as 2 pixels, and the Max frame gap was 2 frames. Data files were exported to MATLAB 2016b (MATLAB Release 2016b, The MathWorks, Inc.). For Divide and Conquer Moment Scaling spectrum transient diffusion analysis (DC-MSS), we used the MATLAB based software[77]. All the tracks were segregated into super diffusion, free, confined, and immobile motions.

RESULTS

Mutations to membrane-exposed surfaces of CXCR4 and CCR5 are deleterious for binding conformation-dependent antibodies

We revisited published deep mutational scans of the chemokine receptors CXCR4 and CCR5 to search for evidence of receptor oligomers[63]. These scans determined how mutations in the receptors impacted recognition by conformation-dependent monoclonal antibodies: 12G5 for CXCR4 and 2D7 for CCR5. While these antibodies bind to extracellular loops of the receptors, we hypothesized that mutations in transmembrane regions that disrupt oligomeric organization may alter how the extracellular epitope is presented. Mapping conservation scores from the mutational scan of CXCR4 for binding to 12G5 onto the CXCR4 crystal structure[35] revealed that the dimer interface is indeed weakly but discernably more conserved than other membrane-exposed surfaces (**Figure 1A**). We thus turned our attention to CCR5, for which a dimeric structure has not been determined at atomic resolution. Mapping conservation scores from the mutational scan of CCR5 for binding to 2D7 to the CCR5 monomer structure[67], we found that a surface primarily formed by TM4 was more conserved than other membrane-exposed surfaces and might therefore form the dimer interface (**Figure 1B**). We thus introduced mutations into this region of CCR5 that were depleted in the scan for 2D7 binding (**Figure 1C**). Two control mutations located elsewhere on CCR5 and expected to have no effect were also evaluated. We found as expected that most of the mutations did indeed decrease 2D7 affinity (**Figure 1D**). To understand whether the mutations decreased receptor associations as hypothesized, we co-expressed CCR5 mutants fused to N-terminal (VN) or C-terminal (VC) segments of split Venus (a variant of yellow fluorescent protein or YFP). If the receptors closely associate, the VN and VC polypeptides will fold together to produce fluorescent Venus. This method is known as

bimolecular fluorescence complementation (BiFC)[78]. However, to our surprise, mutations that reduced CCR5 affinity for 2D7 caused increased BiFC signal (**Figure 1D**). This suggested that the mutations to the hypothesized dimer interface of CCR5 were generally disruptive of structure, reducing 2D7 recognition and causing putative aggregation in the membrane, measured as increased BiFC. These initial observations inspired us to use BiFC as the basis for deep mutational scans of CXCR4 and CCR5 to see whether we could identify mutations that decreased receptor associations.

Deep mutational scans of CXCR4 and CCR5 based on BiFC

Deep mutational scanning couples an in vitro screen or selection of a diverse library of sequence variants with next generation sequencing. BiFC is well suited to illuminating how mutations in chemokine receptors influence receptor self-association, as BiFC facilitates the separation of cells using fluorescence-activated cell sorting (FACS). Using the split Venus BiFC method [78], single site saturation mutagenesis (SSM) libraries of CXCR4 and CCR5 [63] were modified by fusing the C-terminal half of split Venus to the C-termini, creating CXCR4-VC and CCR5-VC libraries. These were expressed in an Expi293F cell line that was modified to (i) first remove endogenous CXCR4 using CRISPR-Cas9 (CXCR4-knockout or X4-KO [63]) and (ii) second to stably integrate a gene encoding either CXCR4 or CCR5 fused at their C-termini to the N-terminal half of split Venus, creating CXCR4-VN and CCR5-VN stable lines without endogenous CXCR4 expression. The VC and VN fused receptors had different N-terminal (and thus extracellular) epitope tags for their detection. The stable lines were transiently transfected with the matching VC-fused receptor libraries under conditions where the cells typically express no more than one gene per cell, providing a tight link from genotype to phenotype. The cell

libraries were sorted by fluorescence-activated cell sorting (FACS), in which collection gates were applied to cells that expressed CXCR4-VC or CCR5-VC mutants at the plasma membrane. In a second sorting experiment, cells were gated not only for surface CXCR4-VC or CCR5-VC expression but also for high levels of Venus fluorescence (**Figure 2**), indicative of an association between a VC-fused mutant with a VN-fused wild-type receptor. RNA transcripts from the sorted libraries were analyzed by Illumina sequencing and compared to the naïve plasmid libraries to calculate an enrichment ratio for each mutation. The enrichment or depletion of single mutations in the CXCR4-VC and CCR-VC genes are presented as heat maps that represent mutational landscapes for receptor associations **(Figures 3 and 4).**

Sorting experiments were independently replicated. Enrichment ratios weakly agree between the replicate experiments and the agreement is higher when the libraries were sorted for expression only (**Figures 3A and 4A**) versus sorting on expression and BiFC signal (**Figures 3B and 4B**). Low frequency mutations in the naïve library had weaker agreement between the replicate experiments, suggesting low frequency mutations were not deeply sampled during sorting and thus yielded higher variation between replicates. Overall, we consider the 'noise' in the data to be similar to earlier deep mutational scans of chemokine receptors[63] but considerably higher than in recent mutational scans of the sweet taste receptor, SARS-CoV-2 receptor ACE2, and serotonin transporter[7,79,80], which all use equivalent methods. The high noise in this data set may represent difficulties in achieving consistent sort conditions on BiFC signal. Large variations in mutation frequencies in the naïve library is also expected to have contributed to uneven sampling of mutations during sorting of the transfected cell culture. The enrichment ratios for individual mutations should thus be considered as estimates or predictions until verified by targeted

mutagenesis. The $\log_2$ enrichment ratios for all substitutions at a given residue position may also be averaged to calculate a conservation score. Conservation scores show higher agreement between replicate sorting experiments and are considered a more reliable indicator of whether a particular residue is tolerant of mutations (**Figures 3C-D and 4C-D**).

There are two key features in the mutational landscapes. First, for CCR5 but less prominently for CXCR4, mutations of transmembrane residues to polar amino acids tend to reduce expression at the cell surface (**Figures 3E and 4E**). Such mutations are expected to adversely impact folding of the transmembrane domain. This is consistent with previous deep mutational scans of both receptors[63]. Second, there are a large number of mutations that are highly enriched for BiFC and these mutations are often found within the transmembrane regions (**Figures 3E and 4E**), where many are expected to be highly destabilizing for tertiary structure (for example, introduction of charged and polar residues in transmembrane helices). It is likely that destabilizing mutations enhance BiFC through non-specific aggregation of mutant proteins. This is consistent with our preliminary data showing that mutations in CCR5 that decreased binding to a conformation-dependent monoclonal antibody (and thus were damaging folded structure) were associated with elevated BiFC. These data indicate the challenges in using mutagenesis to identify dimerization sites in chemokine receptors; a mutation within a dimer interface may disrupt the native dimer but simultaneously increase aggregation and association of misfolded proteins.

To support these hypotheses, we chose 12 mutations for both CXCR4 and CCR5 that were enriched for high BiFC signal. These mutations substitute amino acids for side chains with very

different size and chemical properties. In some cases they introduce prolines, which often disrupt secondary structure. We validated that the mutations within CXCR4-VC and CCR5-VC increased BiFC signal in the respective CXCR4-VN and CCR5-VN cell lines, above the BiFC signal of wild type receptors (**Figure 5A,C**). As a control, the metabotropic glutamate receptor mGluR3, which has no known association with chemokine receptors, produced low BiFC signals in these cell lines. Cells were gated based on surface levels of the different receptors to partially control for potential differences in expression, although we do not exclude the possibility that increased BiFC signal arises from elevated levels of aggregating protein trapped in intracellular compartments. Furthermore, we generated a stable mGluR3-VN expressing cell line that provides a high BiFC signal when transfected with mGluR3-VC. mGluR3 is a class C GPCR that forms disulfide bonded dimers. The BiFC signal is low when the mGluR3-VN line is transfected with wild type CXCR4-VC or CCR5-VC. However, the mutant CXCR4 and CCR5 proteins had very high BiFC signal with mGluR3 (**Figure 5B,D**), again suggesting they are non-specifically associating with other proteins in the membrane due to destabilized structure.

Mutations in CXCR4 and CCR5 that reduce receptor associations as measured by BiFC

To understand whether the data provides insights into specific dimer interactions surfaces, the conservation scores from the BiFC selection experiments were mapped to the crystal structures of CXCR4 and CCR5 (**Figures 4F and 5F**). CXCR4 has been crystalized as a dimer, with the same dimeric arrangement being resolved in multiple crystal forms with different molecular packing, providing high confidence that the structure represents a relevant complex[35]. Residues of CXCR4 known to be buried in the dimeric assembly were weakly more conserved in the BiFC-based selection, with lower mutational tolerance compared to other exposed surfaces of the

protein (**Figure 4F**). These more conserved regions are principally on TM3, TM4, and TM5. We note that these regions approximately correspond to the transmembrane surface that was more highly conserved in the published deep mutational scan of CXCR4 for binding to a conformation-dependent antibody (**Figure 1A**). Overall, we conclude that the BiFC-based selection and deep mutational scan succeeded in identifying the known dimer interface of CXCR4. We thus turned our attention to CCR5 where the dimer interface has not been determined by structural methods at atomic resolution. Conservation scores from the BiFC-based deep mutational scan of CCR5 mapped to the protein's monomeric structure indicate that the transmembrane surface formed by TM3-TM4 has reduced mutational tolerance compared to other surfaces (**Figure 5F**). The TM3-TM4 surface of CCR5 was also more conserved in the published deep mutational scan of CCR5 for binding to a conformation-dependent antibody (**Figure 1**). Accordingly, we hypothesize that this surface is buried in the CCR5 dimer.

We further inspected the mutational landscapes for mutations that were highly depleted in the BiFC-based selection and found a small number in the C-terminal cytoplasmic tails of the receptors. GPCRs are known to cluster in lipid microdomains[81,82], and their clustering would be anticipated to increase BiFC signal. CXCR4 has a number of basic residues at its C-terminus and CCR5 has palmitoylation sites[83], both of which are often associated with lipid microdomain localization.

We selected mutations of CXCR4 and CCR5 that were highly depleted following the BiFC-based selection but were close to neutral in the selection for surface expression; low BiFC signal was therefore not predicted to be due to decreased expression. We validated 5-6 mutations in

CXCR4-VN and CCR5-VC as decreasing BiFC in the wild-type CXCR4-VC and CCR5-VC stable cell lines (**Figure 6B,E**). We considered that decreased BiFC between a mutant and wild-type receptor might occur if the mutant or wild-type proteins preferentially self-associated, thus excluding VN- and VC-fused proteins from a dimer complex. We therefore introduced the mutations into VN and VC receptors, which were co-transfected **(Figure 6C,F).** In this arrangement, both receptor chains carry the same mutation. This led to the confirmation of 4 mutants in CXCR4 and 2 in CCR5 as decreasing self-association of receptor chains (**Figure 6H**). In CXCR4, the mutations are P42W, which induces a kink in TM1 and likely has structural consequences; I204W, which is located on TM5 in the crystallographic dimer interface and is thus predicted to disrupt CXCR4 dimers (**Figure 6A**); and R322M and K327W, which change properties in the basic C-terminal tail of CXCR4 and are hypothesized to alter receptor clustering into lipid microdomains. In CCR5, the mutations are V157I, which is exposed to the membrane on TM4 (**Figure 6D**) and is predicted to lie within a dimer interface; and C323I, which modifies a palmitoylation site in the C-terminal tail and is predicted to alter clustering into lipid microdomains.

Effects of CXCR4 and CCR5 mutations

To explore the mechanisms by which these mutations reduce receptor associations, we tested their colocalization with the Matrix protein (MA) of HIV-1, which localizes to cholesterol-rich lipid microdomains[84]. Expi293F cells that stably express MA fused at the C-terminus to VN were transfected with VC-fused CXCR4 and CCR5 mutants (**Figure 6G**). We found that mutations in the C-terminal tails of CXCR4 (R322M and K327W) and CCR5 (C323I) substantially reduced BiFC interactions with MA, consistent with these mutations reducing

localization into lipid microdomains. By comparison, mutations at the putative dimer interfaces (I204W for CXCR4 and V157I for CCR5) had no impact on clustering with MA. Mutation P42W in CXCR4, which is anticipated to have substantial structural effects by targeting a kink-forming proline in TM1, also reduced association with the MA marker for lipid microdomains, indicating that the C-terminal tails are not solely responsible for membrane microdomain localization.

The dynamics of the CXCR4 mutants in the plasma membrane were analyzed by single molecule imaging. Halo-tagged CXCR4 proteins were labeled under conditions that permitted tracking of single molecules by fluorescence microscopy. In addition to testing the aforementioned mutants of CXCR4, we included as a control D97W, a mutation located at the end of TM2 and oriented towards the ligand-binding cavity, which was shown to increase BiFC signal (**Figure 5A**). With the exception of CXCR4-D97W, wild-type and mutant CXCR4 proteins were similar, with most molecules mobile within a confined space and with diffusion coefficients < 0.005 $\mu m^2/s$, although a small population of molecules were more mobile (**Figure 7**). Only CXCR4-K327W single molecules had substantially higher mobility based on increased diffusion coefficients. Molecules of the high-BiFC control, CXCR4-D97W, were nearly all immobile (approximately 80%) with low diffusion coefficients, consistent with our speculation that mutations which increase BiFC between receptors promote non-specific aggregation in the membrane.

The effects of the mutations on ligand binding and calcium mobilization were measured. Using flow cytometry, binding of sfGFP-tagged CXCL12 to the low-BiFC CXCR4 mutants (P42W, I204W, R322M, and K327W) significantly increased compared to wild-type CXCR4 (**Figure**

8A), despite no change or slightly reduced receptor expression at the plasma membrane (**Figure 8B**). CXCL12 binding to CXCR4-P42W and I204W was particularly prominent, increasing 2.5-fold and 5-fold over wild-type, with smaller increases for the two CXCR4 C-terminal mutants. However, increased ligand binding was not associated with higher signaling (**Figure 8C**). Rather, Ca^{2+} mobilization trended lower in the CXCR4 mutants. The CXCR4-P42W mutant was inactive, consistent with P42 having an important structural role in inducing a kink in TM1. The CXCR4-I204W and R322M mutants had decreased signaling, while CXCR4-K327W was similar to wild-type. Hence while the mutations favor high ligand-binding states of CXCR4, those states are decoupled from increased Ca^{2+} signaling. The CCR5-V157I and C232I mutants were also trending towards decreased Ca^{2+} mobilization relative to wild-type CCR5 when stimulated with CCL5 (**Figure 8D**). The data suggest that receptor associations and/or localized clustering, either via oligomeric assemblies or through localization to lipid microdomains, facilitate signaling downstream of ligand binding.

CXCR4 and CCR5 are coreceptors with CD4 for HIV-1 entry into a host cell. The viral spike protein Env engages CD4 on the host cell, facilitating conformational changes that expose binding sites for CXCR4 (X4-tropic viruses) or CCR5 (R5-tropic viruses)[85]. We tested whether the low-BiFC receptor mutations altered syncytia formation when Env-expressing cells were cultured with receptor-expressing cells. Cells expressing CD4 together with wild-type or mutant CXCR4 had no significant differences in syncytia formation when incubated with cells expressing Env from the X4-tropic MN strain (**Figure 9A**), and likewise cells expressing CD4 together with wild-type or mutant CCR5 had no significant differences in syncytia formation when incubated with cells expressing Env from the R5-tropic BaL strain[30] (**Figure 9B**).

DISCUSSION

Extensive research efforts have found that chemokine receptors CXCR4 and CCR5 homo- and heterodimerize [35,43,60,66,72,86], a feature that is common to many other class A GPCRs [18,33,34]. However, the organization of chemokine receptor dimers and their functional importance remain incompletely understood, in part because identifying mutations that alter dimer strength has been difficult. Here, we set out to define dimer interfaces and find mutations to modulate receptor associations through a BiFC-based deep mutational scan. The use of deep mutagenesis to identify dimerization interfaces was previously applied to the sweet taste receptor and informed the modeling of that receptor's dimeric architecture, in close agreement with a cryo-EM structure of a related class C GPCR[80]. However, the mutational landscapes we report for CXCR4 and CCR5 unexpectedly revealed many hundreds of mutations that increased receptor associations based on increased BiFC signal. We show that these mutations not only increased BiFC between chemokine receptor chains but also with an unrelated class C GPCR, suggesting the mutations cause non-specific associations and aggregation. This is supported by the observation that the mutations frequently altered chemical properties of amino acids within transmembrane helices and may thus be deleterious for proper folding. Single molecule imaging of CXCR4 with a high-BiFC mutation showed the mutant receptors were mostly immobile in the membrane. The ease with which CCR5 and CXCR4 could be driven towards aggregation means over-expression studies of receptors must be approached with caution, lest non-specific aggregation is falsely interpreted as biologically relevant dimer formation. Despite the high prevalence of BiFC-enhancing mutations in the scans, there were surfaces on CXCR4 and CCR5 where mutations

tended to be depleted following BiFC-based sorting. We consider these relatively more conserved surfaces in the sorting experiments to be potential interfaces for dimer formation.

CXCR4 has been crystallized as a symmetric dimer with inter-subunit contacts formed primarily by TM3-5[35]. The TM3-5 surface exposed to the membrane is less tolerant of mutations in the BiFC-based deep mutational scan and our data therefore support this dimeric arrangement of CXCR4 in living cells. Substitution I204W in TM5 caused a partial decrease in CXCR4 homodimerization based on BiFC. Cells expressing CXCR4-I204W had 5-fold enhanced binding to CXCL12, consistent with CXCL12/CXCR4 forming a 1:1 complex. However, signaling by the mutant receptor was diminished, although the data do not answer whether decreased signaling was due to reduced dimerization or other structural perturbations caused by the mutation that decouple G protein activation from ligand binding.

The structure of a CCR5 dimer has not been determined experimentally at atomic resolution, although chemical cross-linking has supported close contacts between TM5 helices across the interface[70]. There is no surface on CCR5 that unequivocally stands out as the putative dimer interface in our BiFC-based deep mutational scan. Our data is most supportive of TM4 forming the CCR5 dimer interface, as the membrane-exposed surface of this helix was relatively more conserved than other transmembrane regions, albeit weakly. It is difficult to compare our findings to the published cross-linking study of CCR5 that emphasized TM5, as cross-linking mutations were not explored in TM4. Substitution V157I, located on TM4 and pointing outwards towards the lipid phase of the membrane, was confirmed as decreasing BiFC between CCR5 receptors.

We also found mutations in the cytoplasmic tails of CXCR4 and CCR5 that decreased BiFC. The cytoplasmic tails are known to contain features that promote lipid microdomain localization[83,87–89] and decreased BiFC is readily explained by reduced clustering of receptors in membrane microdomains. In agreement, mutations in the C-terminal tails also reduced co-localization with Matrix, a marker for cholesterol-rich microdomains in cellular membranes.

Overall, deep mutational scanning of the chemokine receptors CXCR4 and CCR5 through the use of BiFC demonstrated that these receptors may associate through multiple mechanisms: (1) non-specific associations or aggregation, (2) dimerization, and (3) clustering into lipid microdomains. These findings have implications for studies into chemokine receptor associations, as results that are often interpreted as receptor dimerization or oligomerization may instead arise due to other mechanisms.

FIGURES

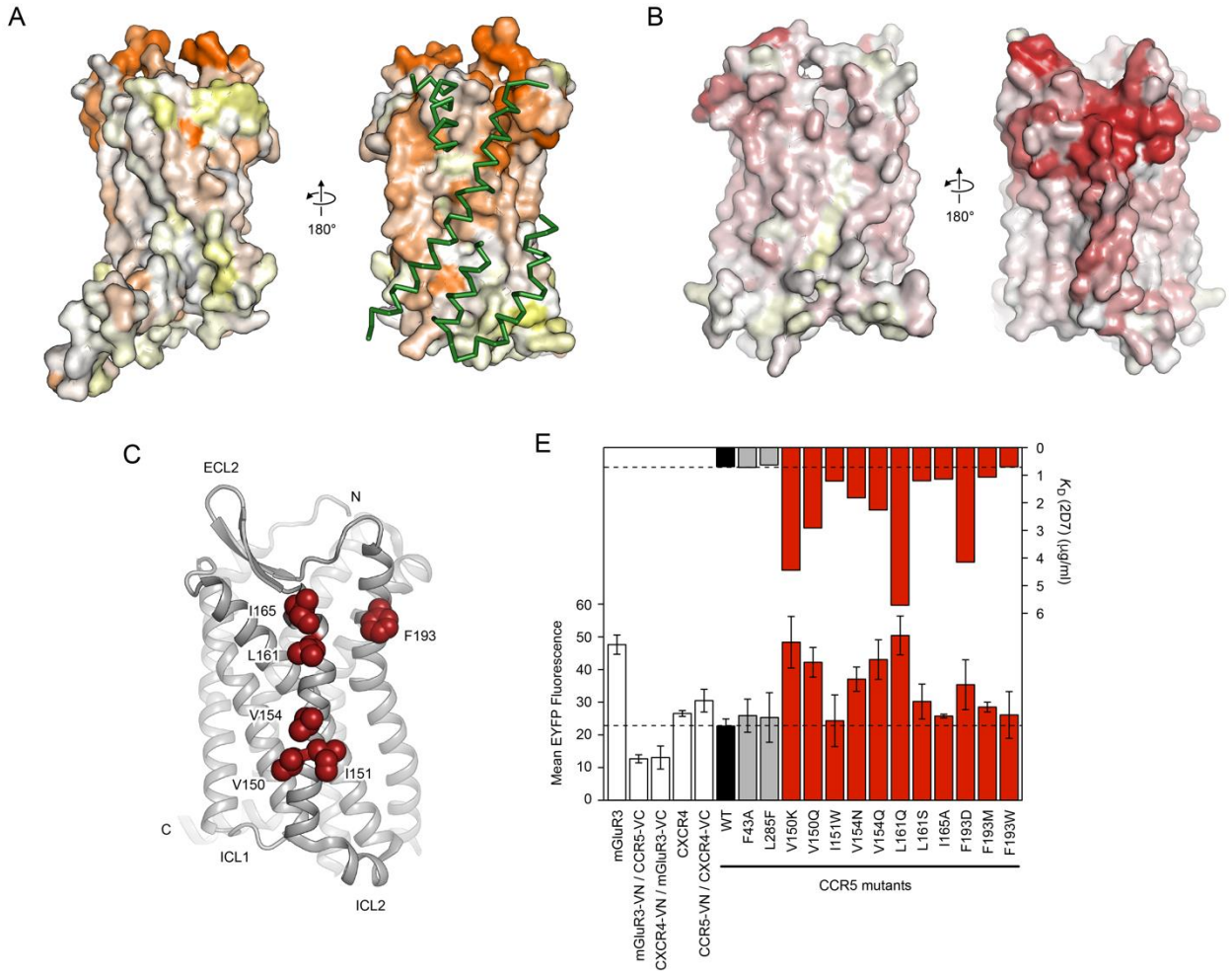

Figure 2.1. Deep mutational scans of chemokine receptors based on recognition by conformation-dependent antibodies highlights conserved surfaces facing the membrane. (A) Using the published deep mutational scan of CXCR4 for binding to monoclonal 12G5, experimental conservation scores are mapped to the CXCR4 surface (PDB 3ODU). Mutational tolerance of residues is shown from low in orange to high in white and yellow. At right, a more conserved surface exposed to the membrane corresponds to the interface for a second CXCR4 subunit in the crystal structure (regions in close contact are shown as green ribbons). **(B)** Conservation scores from the published deep mutational scan of CCR5 for binding to monoclonal 2D7 are mapped to the CCR5 crystal structure (PDB 4MBS), from conserved residues in red to mutationally tolerant residues in white and yellow. At right, one side of CCR5 exposed to the membrane is more conserved. **(C)** Cartoon of the CCR5 structure oriented with the conserved surface facing the reader. Residues targeted for mutagenesis are shown as red spheres. **(D)** *(Upper plot)* CCR5 mutants were expressed on Expi293F cells and the apparent K_D of 2D7 was measured by flow…

Figure 2.1 (cont.) cytometry. Wild-type CCR5 and two control mutations not expected to impact 2D7 affinity are shown in black and grey, respectively. Mutations in the conserved membrane-exposed surface are shown in red. *(Lower plot)* Mutations were introduced into CCR5-VC and CCR5-VN constructs and BiFC measured in transfected Expi293F cells. Various controls are in white (see description in main text). Data are mean ± SD, n=3 independent replicates.

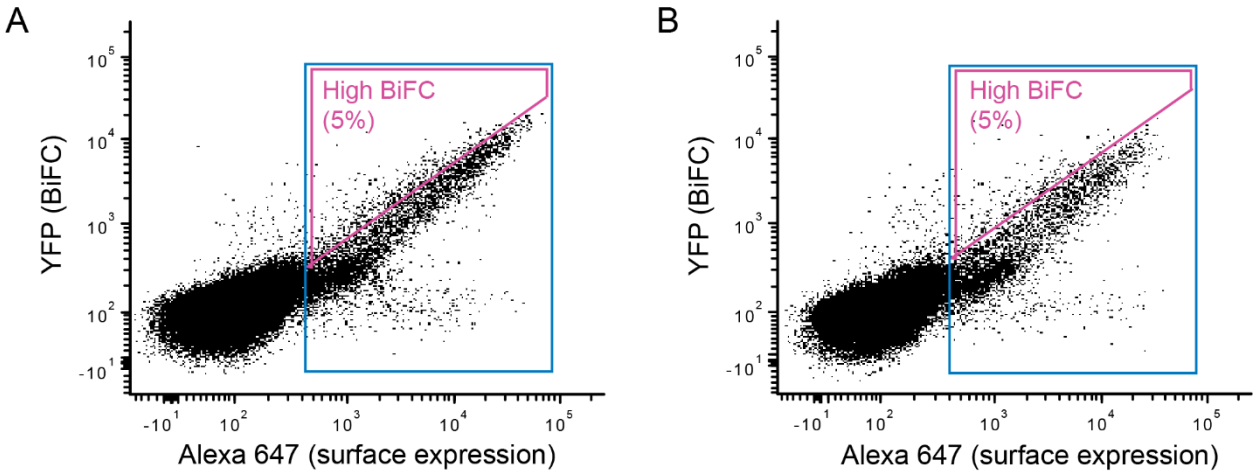

Figure 2.2. FACS collection gates for sorting human cells expressing SSM libraries of chemokine receptors CXCR4 and CCR5. (A, B) X4-KO Expi293F cells stably expressing (A) FLAG-CXCR4-VN or (B) FLAG-CCR5-VN were transfected with SSM libraries of (A) myc-CXCR4-VC or (B) myc-CCR5-VC. After gating the main population of viable cells based on forward-side scatter, cells were gated for surface expression of the VC-fused chemokine receptor mutants by detection of the myc tag (expression gate in blue). Within the expression gate, the top 5% of cells for Venus/YFP fluorescence were gated (high BiFC gate in magenta). Cells in the expression gate or in the high BiFC gate were collected in independent sorting experiments.

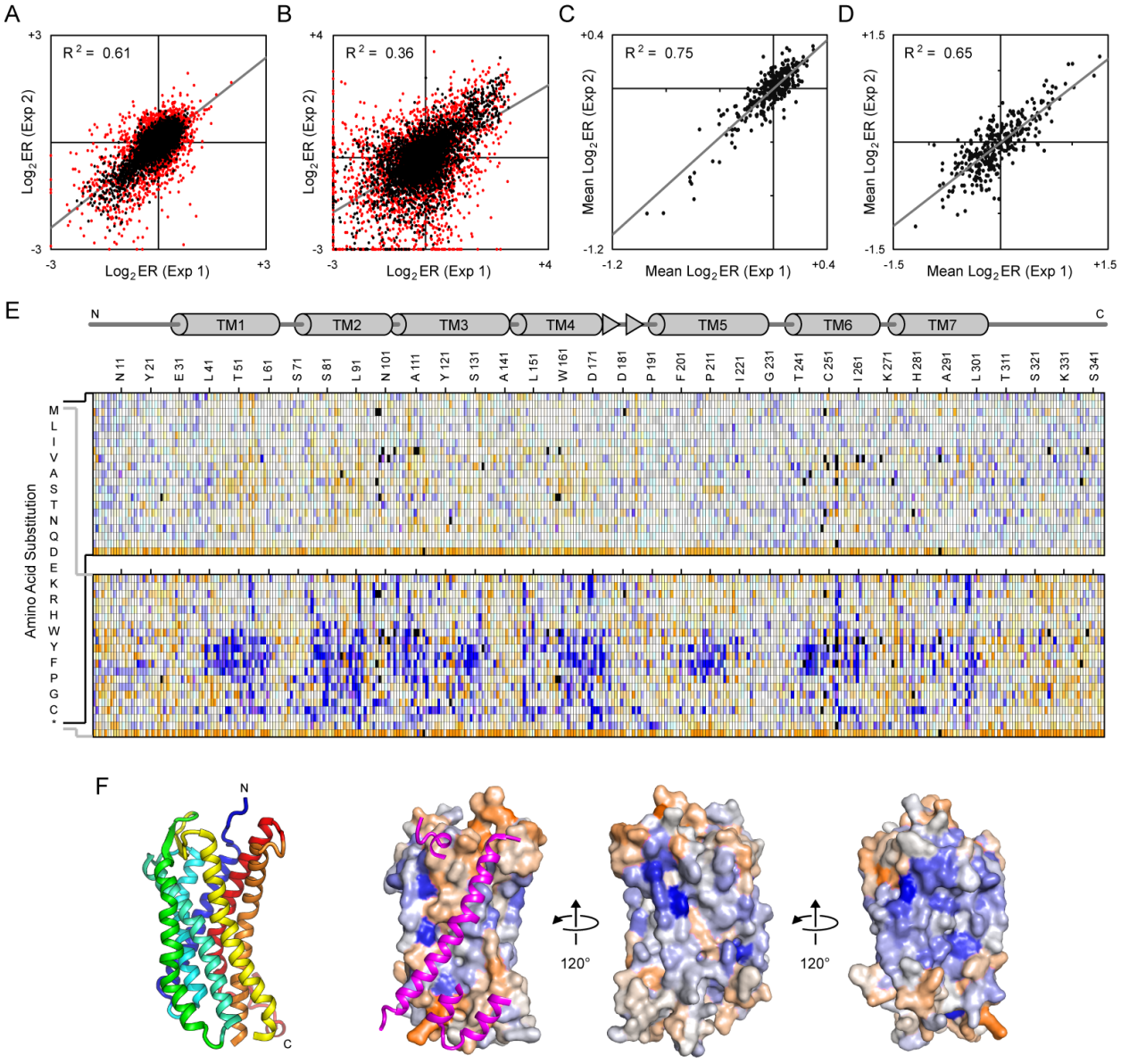

Figure 2.3. Deep mutational scan and mutational landscape of CXCR4 based on receptor associations. (A-B) A single site saturation mutagenesis (SSM) library of CXCR4-VC was expressed in CXCR4-VN Expi293F cells and sorted by FACS for CXCR4-VC surface expression and BiFC. Log2 enrichment ratios for mutations were calculated following Illumina sequencing of the naïve and sorted libraries. Plots show agreement between mutation log2 enrichment ratios after sorting for **(A)** expression or **(B)** expression and BiFC. R-squared values correspond to mutations with frequency ≥ 5×10^{-5} in the naïve library (shown in black). Rare mutations with frequency < 5×10^{-5} in the naïve library are red. Mutations with frequency < 5×10^{-6} were considered absent from the library. **(C-D)** Log_2 enrichment…

Figure 2.3 (cont.) ratios for all amino acid substitutions at a given residue position were averaged to determine a mean conservation score. Plots show agreement of residue conservation scores between two independent sorting experiments for **(C)** expression or **(D)** expression and BiFC. (**E**) Mutational landscapes of CXCR4-VC sorted for surface expression (top) and BiFC (bottom). Log_2 enrichment ratios are plotted from depleted/deleterious (≤ -3, orange) to enriched (≥ +3, dark blue). Mutations missing in the naïve library are in black. The CXCR4 sequence is on the horizontal axis and amino acid substitutions are on the vertical axis. A schematic of CXCR4 secondary structure is shown at top, with cylinders representing α-helices and arrows representing β-strands. (**F**) Ribbon structure of CXCR4 (PDB 3ODU) is shown at left, viewed from the plane of the membrane and colored blue to red from N- to C-terminus. The helices forming the dimer interface are facing out. Adjacent on the right, the same orientation of CXCR4 is shown as a surface colored by conservation scores from the BiFC sorting experiment, with conserved residues in orange and residues that are hot spots for enriched mutations in blue. A second CXCR4 molecule is shown as magenta ribbons; for clarity, only regions of the second CXCR4 that are at the dimer interface are shown. To the right, the structure is rotated.

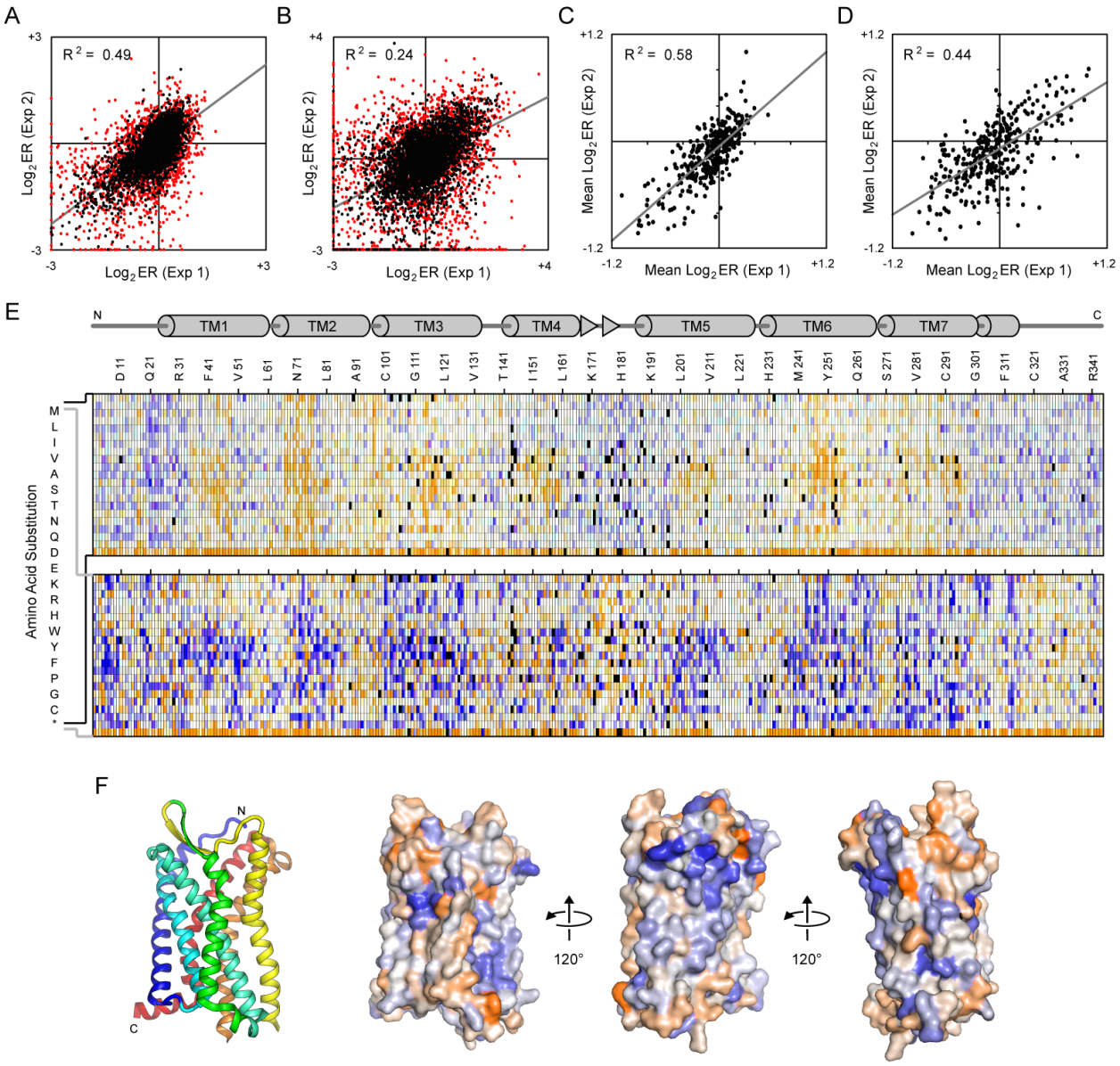

Figure 2.4. Deep mutational scan and mutational landscape of CCR5 based on receptor associations. (A-B) A SSM library of CCR5-VC was expressed in CCR5-VN Expi293F cells and sorted by FACS for CCR5-VC surface expression and BiFC. Agreement between $\log_2$ enrichment ratios for all mutations is shown when the libraries were sorted for **(A)** expression or **(B)** expression and BiFC. R-squared values correspond to mutations with frequency $\geq 5\times10^{-5}$ in the naïve library (black). Rare mutations with frequency $< 5\times10^{-5}$ in the naïve library are red. Mutations with frequency $< 5\times10^{-6}$ were considered absent from the library. **(C-D)** Agreement between residue conservation scores from two independent sorting experiments for **(C)** expression or **(D)** expression and BiFC. (**E**) Mutational…

Figure 2.4 (cont.) landscapes of CCR5-VC sorted for surface expression (top) and BiFC (bottom). Log_2 enrichment ratios are plotted from depleted/deleterious (≤ -3, orange) to enriched (≥ +3, dark blue). Missing mutations are in black. The CCR5 sequence is on the horizontal axis and amino acid substitutions are on the vertical axis. A schematic of CCR5 secondary structure is shown at top, with cylinders representing α-helices and arrows representing β-strands. (**F**) Ribbon structure of CCR5 (PDB 4MBS) is shown at left, viewed from the plane of the membrane and colored blue to red from N- to C-terminus. The proposed dimer interface of CCR5 is facing out. Adjacent on the right, the same orientation of CCR5 is shown as a surface colored by conservation scores from the BiFC sorting experiment, with conserved residues in orange and residues that are hot spots for enriched mutations in blue. To the right, the structure is rotated.

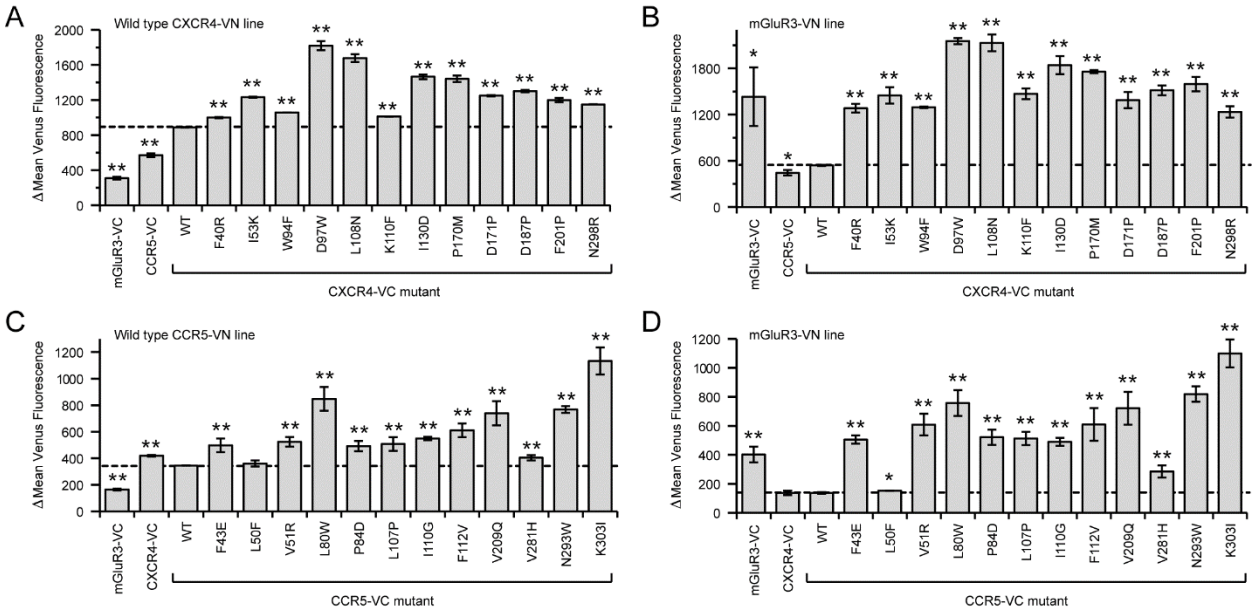

Figure 2.5. Validation of mutations in chemokine receptors that increase BiFC. (A,B) Mutations of CXCR4 predicted from the deep mutational scan to increase BiFC with wild-type CXCR4 were validated. The mutations were introduced into CXCR4-VC and transfected into X4-KO Expi293F cell lines stably expressing either (**A**) CXCR4-VN or (**B**) mGluR3-VN. **(C, D)** Validation of CCR5 mutants enriched in the deep mutational scan for high BiFC signal. The mutations in the CCR5-VC construct were tested in cell lines stably expressing either (**C**) CCR5-VN or (**D**) mGluR3-VN. BiFC signal was measured by flow cytometry. Data are mean ± SD, n=3 independent replicates. *, $p < 0.05$. **, $p < 0.01$, Student's t test.

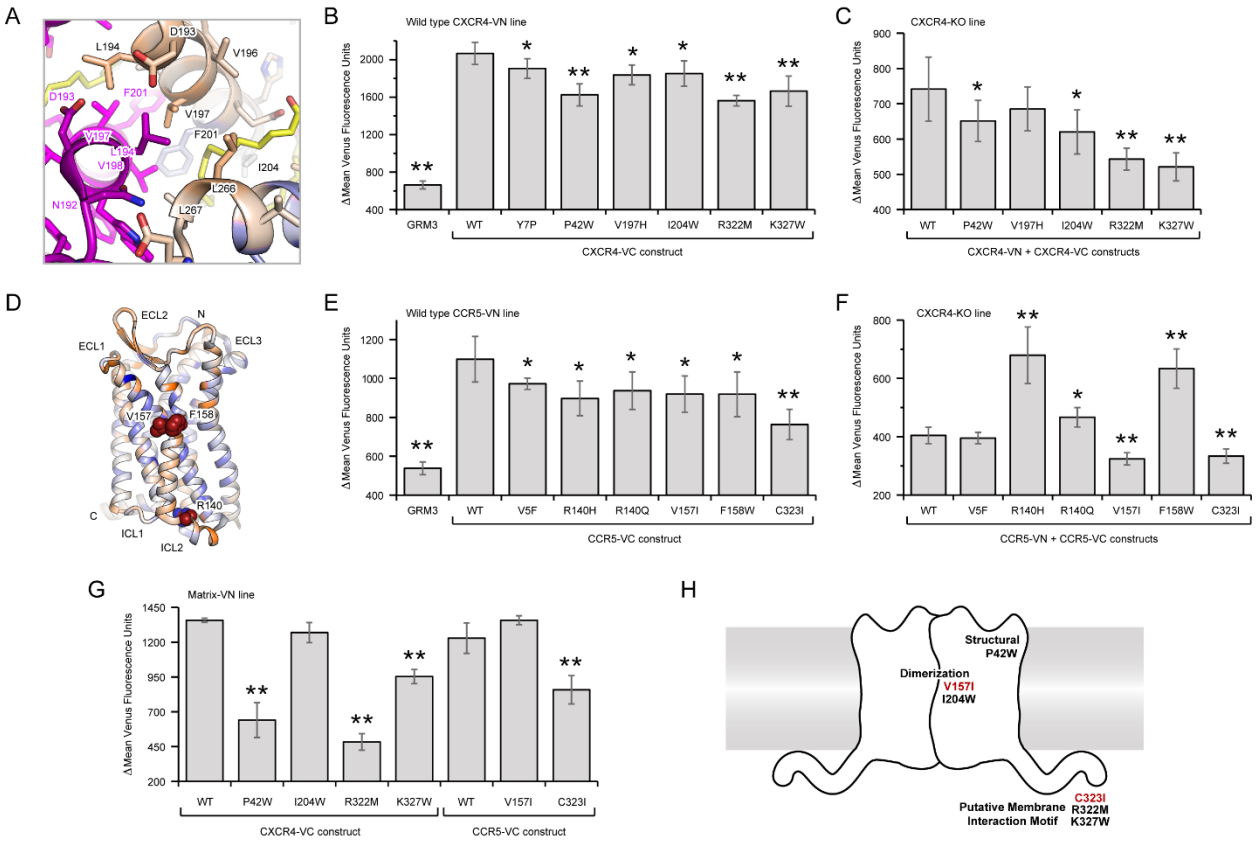

Figure 2.6. Mutations in CXCR4 and CCR5 that decrease receptor associations based on BiFC signal. (A) Dimerization interface of CXCR4 (PDB 3ODU) showing one monomer in magenta and the other monomer colored by the averaged conservation scores from the BiFC-expression deep mutational scan (low mutational tolerance in orange to high mutational tolerance in blue). V197 and I204 are packed against a lipid (yellow sticks) near the interface. **(B, C)** Mutations in CXCR4-VC that were depleted in the BiFC-based deep mutational scan were validated in (**B**) a X4-KO Expi293F cell line stably expressing wild-type CXCR4-VN or (**C**) in X4-KO Expi293F cells in which mutant CXCR4-VC and CXCR4-VN constructs were transiently transfected. (**D**) Ribbon representation of CCR5 (PDB 4MBS) with conservation scores from the BiFC-expression mutational scan mapped to the structure. Residues of low mutational tolerance are orange; residues of high mutational tolerance are blue. A membrane-exposed surface patch that is less tolerant of mutations and is the proposed dimerization interface is oriented towards the reader, with mutations depleted in the mutational scan in red spheres. **(E, F)** Mutations in CCR5-VC that were depleted in the BiFC-based scan were transfected in (**E**) X4-KO Expi293F cells stably expressing wild-type CCR5-VN and BiFC measured by flow cytometry. The mutations were also tested in (**F**) X4-KO Expi293F cells in which both mutant CCR5-VC and CCR5-VN constructs were

Figure 2.6 (cont.) cotransfected. (**G**) Mutations in CXCR4-VC and CCR5-VC that reduced BiFC in panels C and F were transfected into a stable MA-VN Expi293F line as a marker for lipid microdomains and BiFC was measured. (**H**) Cartoon of a chemokine receptor dimer annotated with mutations that diminish BiFC between homodimers of CXCR4 (black labels) and CCR5 (red labels). In all data panels, plotted are means ± SD, n=4 biological independent replicates. *, $p < 0.05$; **, $p < 0.01$, Student's t test.

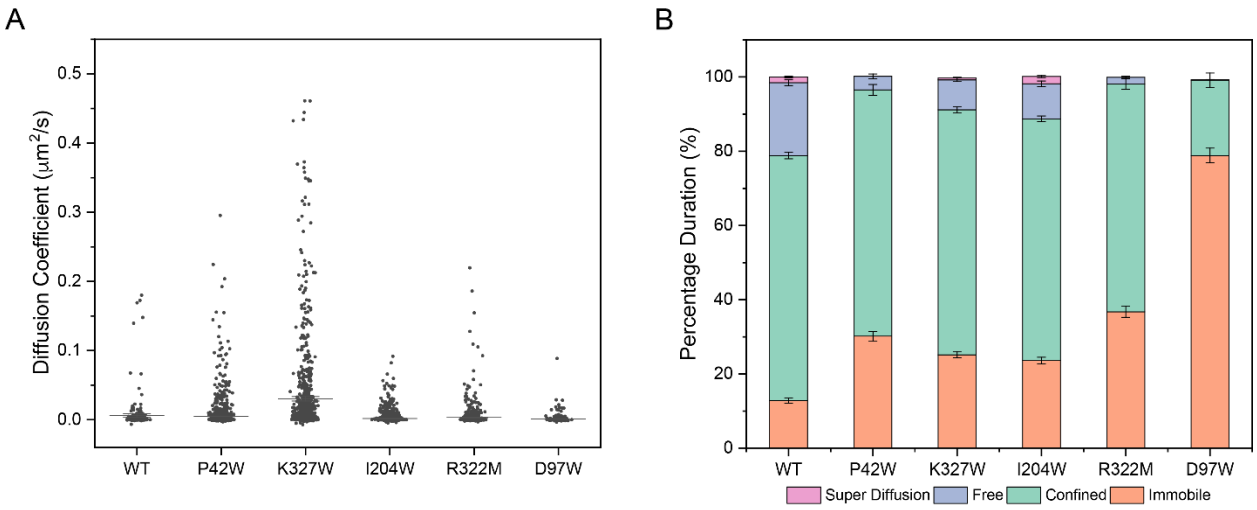

Figure 2.7. Single molecule imaging of CXCR4. Data generated by Kritika Mehta with samples prepared by Vishnu Krishnamurthy and Kevin Gill. (A) Single Molecule Imaging analysis of CXCR4 and its mutants. (**A**) Distribution of diffusion coefficients derived from DC-MSS analysis[77] for myc-Halo-CXCR4, comparing reduced self-association mutants (P42W, K327W, I204W and R322M) to wild type and one increased self-association mutant (D97W). (**B**) DC-MSS analysis for the tracked trajectories of myc-Halo-CXCR4, dissecting transient motion in the tracks.

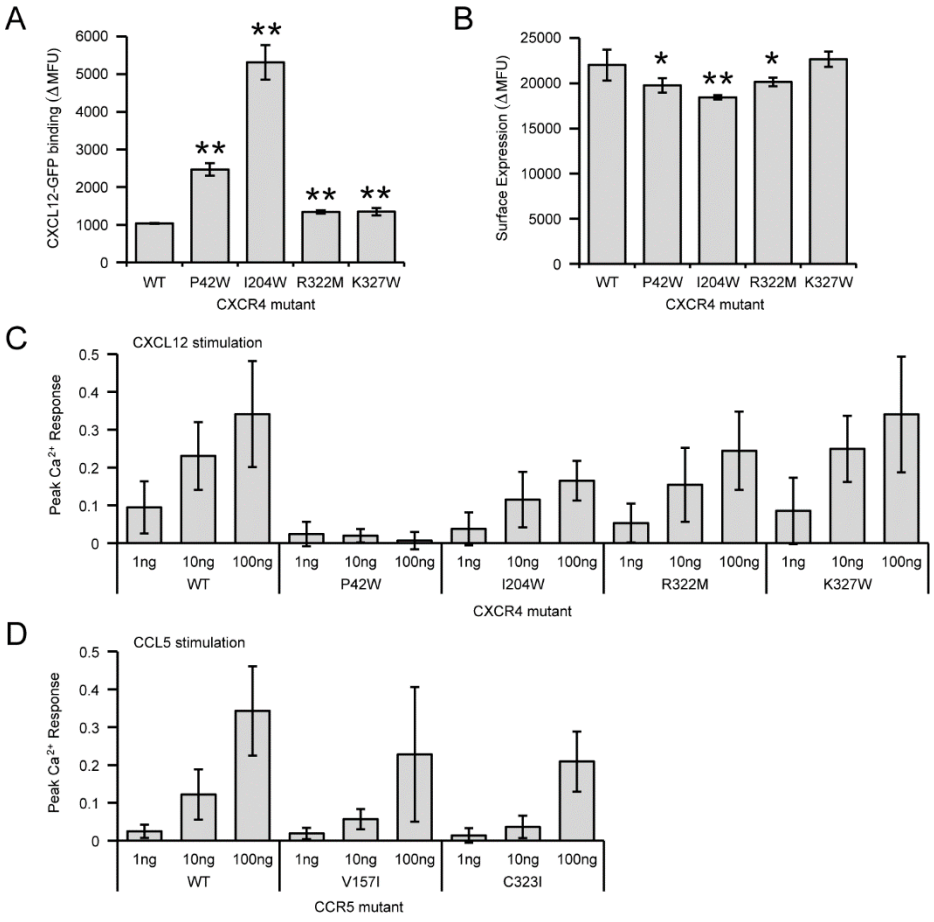

Figure 2.8. Ligand binding and signaling of chemokine receptor mutants with reduced self-association. (A) X4-KO Expi293F cells expressing CXCR4 mutants were incubated with 10 μM CXCL12-sfGFP and bound ligand was detected by flow cytometry. (**B**) Surface expression of CXCR4 mutants measured by flow cytometry. In panels A and B, data are mean ± SD, n=6 biological independent replicates. *, $p < 0.05$; **, $p < 0.01$, Student's t test. **(C)** Calcium mobilization was recorded in Fluo-4-AM loaded X4-KO Expi293F cells transfected with the indicated CXCR4 mutants. Calcium mobilization at 1, 10, and 100 ng/ml CXCL12 concentrations are reported relative to maximum Ca^{2+} response to 4 μM ionomycin. Data are mean ± SD, n=8 biological independent replicates. **(D)** As in panel C, with cells now expressing CCR5 proteins and stimulated with CCL5. Data are mean ± SD, n=6 biological independent replicates.

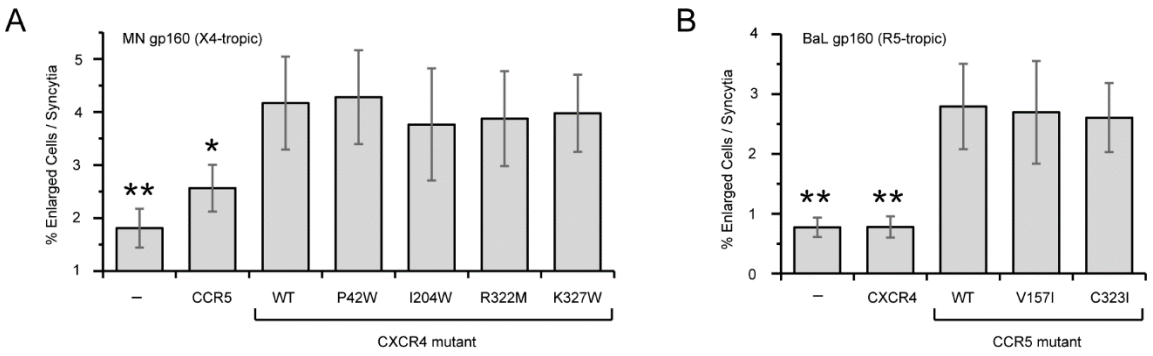

Figure 2.9. HIV-1 Env-mediated syncytia formation is not affected by mutations in chemokine receptors that reduce self-associations. Formation of syncytia was measured by flow cytometry after co-culturing X4-KO Expi293F cells expressing either (**A**) MN gp160 or (**B**) BaL gp160 with cells expressing CD4 and the indicated CXCR4 or CCR5 mutants, respectively. Data are mean ± SD, n = 6 independent experiments. *, $p < 0.05$; **, $p < 0.01$, Student's t test.

CHAPTER 3: CONFORMATIONAL ENGINEERING OF HIV-1 ENV BASED ON MUTATIONAL TOLERANCE IN THE CD4 AND PG16 BOUND STATES[1]

ABSTRACT

HIV-1 infection is initiated by viral Env engaging the host receptor CD4, triggering Env to transition from a 'closed' to 'open' conformation during the early events of virus-cell membrane fusion. To understand how Env sequence accommodates this conformational change, mutational landscapes decoupled from virus replication were determined for Env from BaL (clade B) and DU422 (clade C) isolates interacting with CD4 or antibody PG16 that preferentially recognizes closed trimers. Sequence features uniquely important to each bound state were identified, including glycosylation and binding sites. Notably, the Env apical domain and trimerization interface are under selective pressure for PG16 binding. Based on this key observation, mutations were found that increase presentation of quaternary epitopes associated with properly conformed trimers when Env is expressed at the plasma membrane. Many mutations reduce electrostatic repulsion at the Env apex, and increase PG16 recognition of Env sequences from clades A and B. Other mutations increase hydrophobic packing at the gp120 inner-outer domain interface, and were broadly applicable for engineering Env from diverse strains spanning tiers 1, 2, and 3 across clades A, B, C, and BC recombinants. Core mutations predicted to introduce steric strain in the open state show markedly reduced CD4 interactions. These findings and methods may assist vaccine design.

[1] The research described in this chapter was a combined effort of myself with Jeremiah D. Heredia, Jihye Park, Hannah Choi, and Erik Procko. Data I was responsible for collecting are shown in Figure 3.

INTRODUCTION

HIV-1 engages target cells through interactions between the viral glycoprotein Env and host receptors. HIV-1 Env is formed by a homotrimeric complex of gp160 subunits that are cleaved by host proteases during maturation into extracellular gp120 and membrane-tethered gp41, which remain non-covalently associated in a 'closed' conformation that can dynamically sample alternative conformational states[90]. During infection, the gp120 subunit binds the primary host receptor CD4, inducing an 'open' conformation of Env that exposes binding sites for a secondary co-receptor[91–93]. This co-receptor is one of either two chemokine receptors, CCR5 or CXCR4, and once bound, further conformational changes release fusogenic regions of gp41 that mediate membrane fusion and viral entry into the host cell.

Env is the only viral protein on the outside of a HIV-1 virion accessible to the humoral immune system, and therefore has been extensively studied for vaccine development [94]. Conformational diversity that causes exposure of non-neutralizing or strain-specific immunodominant epitopes[95–99], Env sequence diversity, and epitope shielding by extensive glycosylation, all act to limit potency and breadth of the host response. Many broadly neutralizing antibodies (bNAbs) recognize Env in the closed state, but the design and purification of Env immunogens that correctly fold into native-like, pre-fusion closed trimers is challenging due to intrinsic conformational flexibility[48,100]. A widely used construct for purifying the extracellular region of Env as a soluble trimer, called SOSIP, incorporates a disulfide between gp41 and gp120 subunits, and an I559P substitution to destabilize the post-fusion state[49,50,101]. SOSIP constructs from various strains have been heavily engineered with additional mutations for favorable antigenic profiles, trimer stability, purification properties, structure, and reduced exposure of

strain-specific V3 and CD4i epitopes[102–104]. Closed SOSIP trimers can be isolated with high purity using conformation specific antibodies, but for full-length or close to full-length Env that is membrane anchored, appropriately folded protein cannot be easily separated from misfolded forms. This means that nucleic acid vaccines or virus-like particles (VLPs) can present a heterogeneous mixture of full-length Env in monomeric, oligomeric, closed, open, and partially open conformers. While VLPs can be treated with protease to remove sensitive misfolded forms[105], there would ideally exist sequence variants of Env that are primarily expressed as closed trimers. Understanding how Env sequence dictates conformation therefore not only addresses important biophysical questions on how protein dynamics are governed by primary structure but may also assist the design of sophisticated Env immunogens.

Deep mutational scanning couples directed evolution of diverse sequence populations with next generation sequencing to track the phenotypic fitness of thousands of mutations simultaneously[4]. The method has been used for vaccine design to screen for mutations within extracellular Env constructs that enhance direct interactions with an antibody and its germline precursors, reduce exposure of V3 epitopes, and improve SOSIP thermostability[103,106,107], however none of these data sets have been analyzed to address the central question behind this study; how does Env sequence relate to conformational change? Tissue culture propagation experiments of viruses expressing Env variants have also been extensively followed by next generation sequencing, and have shown that a surprisingly small number amongst the thousands of possible mutations in Env are positively selected for escape from antibody neutralization[108,109]. However, virus propagation demands that Env fold, traffic to the cell surface, incorporate into a budding virion,

bind target receptors, and catalyze membrane fusion, and these mutational landscapes therefore closely resemble allowed diversity observed in natural sequences[8].

Here, we determine Env sequence preferences independent of infection and virus propagation for interactions with three protein ligands that act as conformational probes: CD4, which induces the open Env conformation and binds monomeric gp120 with highest affinity; bNAb VRC01, which binds tightly to both monomeric gp120 and mature Env without inducing the open conformation; and bNAb PG16, which preferentially binds closed trimeric Env[90,110–116]. The mutational scans provide detailed landscapes for how sequence relates to Env conformation, in particular highlighting how electrostatic repulsion at the apical tips and underpacking at the gp120 inner-outer domain interface contribute to conformational change. We identify mutations that increase presentation of the PG16 quaternary epitope in full-length Env sequences from representative tier 1B, 2 and 3 strains in clades A, B, C, and BC recombinants, while simultaneously enhancing presentation of epitopes recognized by other broadly neutralizing antibodies, and reducing CD4 binding. Finally, we demonstrate how our methodology can be applied to investigate interactions between two proteins at the membrane (Env and MA) that are both expressed by the same cell. How these results relate to the design of Env immunogens that are anchored in the plasma membrane is discussed.

RESULTS

Deep mutational scanning of HIV-1 Env from the BaL isolate

Codon-optimized Env of the BaL isolate (a tier 1B virus from clade B[117]) bound soluble CD4 (domains D1-D2), PG16 and VRC01 by flow cytometry when expressed on human Expi293F cells. Previously, deep mutational scanning of HIV-1 receptors CCR5 and CXCR4 found only qualitative agreement between replicate experiments[63], likely in part due to under sampling the diverse mutant libraries. To increase sampling of mutations in the much larger Env protein, three separate single site-saturation mutagenesis (SSM) libraries were constructed focused on the Env N-terminus, center, and C-terminus. Together, the SSM libraries cover 16,332 of the possible 16,520 single amino acid substitutions.

To maintain a tight link between genotype and phenotype, the Env libraries were transfected into Expi293F cells after diluting with a large excess of carrier DNA, such that typically no more than one sequence variant is expressed per cell[63]. Expression was increased by addition of an artificial intron in the 5' untranslated region, and cotransfection with carrier DNA hypothesized to promote extra-chromosomal replication of the episomal Env plasmids[80].

Expi293F cells expressing the BaL Env SSM libraries were bound to soluble CD4, VRC01, or PG16 near the apparent dissociation constants, and were screened by fluorescence-activated cell sorting (FACS) for highest binding signal. The enrichment or depletion of Env mutants was determined by comparing the frequencies in the naïve plasmid libraries with the transcripts in the evolved populations (Figure 1A-1C). Enrichment ratios for each amino acid substitution qualitatively agreed between replicate experiments, with higher agreement for more abundant

mutations that were better sampled (Figure 1D-1F). To further explore data quality, we validated by targeted mutagenesis that 20 representative mutations depleted in both replicate selections for PG16 binding did indeed have reduced binding signal (mutations are listed in Material and Methods). However, when 43 highly enriched substitutions were individually tested, only 21 were validated to cause a slight to moderate increase in PG16 binding signal (see Materials and Methods). Identifying rare gain-of-function mutations is therefore challenging, as they can be difficult to find amongst the noise of many neutral or deleterious substitutions. By comparison, conservation scores (mean of the $\log_2$ enrichment ratios for all substitutions at a specific position) were closely correlated between replicate experiments and are used to define functional sites of mutational intolerance (Figure 1G-1I).

Env sequence-activity landscapes for interacting with CD4, VRC01 and PG16

The BaL Env sequence-activity landscapes are similar whether screened for CD4, VRC01 or PG16 binding (Figure 1A-1C); this is because features of the landscapes that impact protein folding and surface expression will be shared. The highest conservation is in the gp120 inner domain and in regions maintaining non-covalent association between gp41 and gp120 subunits. Polar substitutions within the hydrophobic transmembrane (TM) helix are appropriately depleted, as are extracellular premature stop codons. However, within the cytosolic tail there are regions where stop codons are tolerated or enriched, most notably between residues 731-759, a region that has previously been targeted with nonsense mutations for elevated Env surface expression[118–120]. Interestingly, stop codons are depleted at cytosolic sites approximately corresponding to the lentivirus lytic peptide-2 (LLP-2) and LLP-3 to LLP-1 regions, even though

most amino acid substitutions are tolerated in these regions suggesting a lack of compact folded structure.

Unique features in the respective mutational landscapes are apparent when the conservation scores are plotted on a model of trimeric BaL Env (Figure 2A-2C). To highlight regions explicitly conserved for interacting with different ligands, difference plots were mapped to the structure, in which conservation scores for interacting with one ligand are subtracted from those for interacting with a second (Figure 2D-2G). The most notable differences are localized to the structurally characterized VRC01 and PG16 epitopes, the known CD4 binding site, and trimer interfaces. The data are in close agreement with prior structural and mutagenesis studies of the respective interactions[114,121–124], and highlight how PG16 and especially VRC01 primarily make sequence-independent contacts that tolerate many Env amino acid substitutions.

Glycosylation sites preferentially conserved for CD4 or PG16 binding

Env is heavily glycosylated on asparagine residues at canonical NX(S/T) motifs. Of the 28 predicted Env_{BaL} N-glycosylation sites, only a small subset are moderately-to-highly conserved in the mutational landscapes (Figure 3A). The four N-glycosylation sites in gp41 are similarly conserved for CD4, VRC01 and PG16 interactions, suggesting these are necessary for appropriate Env folding and surface expression. However, within gp120, glycosylation of N262 stands out as being more highly conserved for CD4 binding. N262 mutations were previously identified as reducing CD4 interactions due to diminished Env packaging into virions[125,126]. Our deep mutational scan suggests N262 site mutants also have an intrinsic CD4 binding deficit. N262Q and S264A mutants were found to have substantially reduced CD4 interactions compared

to wild type and Env S264T, which could not be explained by a large loss of protein expression (Figures 3B and 3C). The N262 glycan fits in a deep cleft bridging the outer and inner domains[127] that is remodeled during the closed to open conformational transition[124]. Other residues lining this cleft are highly conserved for CD4 interaction (Figure 3D).

Crystal structures of PG16 and related bNAb PG9 bound to scaffolded V1-V2 demonstrate extensive contacts to the glycan on Env-N160, with a smaller contact surface to the adjacent glycan on Env-N156 (alternatively N173 in other HIV-1 strains)[121,128]. This is supported by mutagenesis [111]. Accordingly, the N160 glycosylation motif is uniquely conserved in our mutational landscape for PG16 binding (Figure 3A).

Env residues at the trimer interface are under selection for PG16 binding

PG16 binds the junction between two gp120 subunits at the Env apex, possibly forming bridging contacts to glycans from each subunit and explaining the antibody's preference for trimeric quaternary structure[112]. When the BaL Env library is evolved for PG16 interactions, trimer interface residues are under selection and show moderate sequence conservation (Figures 2F, 2G and 2I), further supporting the claim that PG16 binds with highest affinity to closed trimers. We therefore hypothesized that Env may be stabilized in closed, trimeric conformations by mutations that increase PG16 binding. Extensive efforts have been invested in stabilizing closed SOSIP trimers, frequently using recognition by trimer-specific bNAbs such as PGT145 as indicators of successful engineering[129]. However, PGT145-class antibodies insert a long HCDR3 loop into the Env apical cavity to make direct contacts at the apical trimer interface[130]. We reasoned that by using PG16, which binds to the outer surface of trimeric Env, we might identify mutations within

the apical cavity that stabilize the closed conformation, despite loss of atomic contacts to PGT145, and thereby complement previous engineering studies.

Nearly a hundred mutations in the Env sequence-activity landscape for PG16 binding had $\log_2$ enrichment ratios greater than 1.5 in both selection replicates, though we note that because most substitutions in the N-terminal SSM library were severely depleted, neutral substitutions in V1-V2 could be misleadingly enriched. We initially focused on mutations at subunit interfaces and validated 18 as causing a slight to moderate increase in PG16 binding, which are clustered at three sites (Figure 4A; a fourth site at subunit interfaces identified from the mutational scan of $gp140_{DU422}$ will be described later). We refer to these as Quaternary Epitope Selected (QES) mutations. The first and most prominent site is located at the trimerization interface near the apex, and includes Q114A, K117V/Y, P124D, T163D, R166E/F/L, V200E/T, R315A/Q, and R432T. Nearly all of these mutations reduce positive charge within the apical cavity. Furthermore, nearly all substitutions of Env-K117, R166, and R432 are predicted to enhance PG16 binding in the mutational landscape. This unambiguously highlights that neutralization of the electropositive apical trimer interface stabilizes a conformation with increased PG16 binding signal, and electrostatic repulsion between apical tips likely primes the subunits for conformational changes upon receptor binding. This same hypothesis was recently proposed based on the structure of BG505 SOSIP bound to bNAb PGT145[130].

The second interfacial site of mutations for enhanced PG16 binding is a centrally located contact between Env protomers, and between gp41 and gp120 subunits (Figure 4A). F223Y in one Env protomer may add a hydrogen bond contact to R557 or N553 from a neighbor. R557Q reduces

the desolvation penalty of burying a charged group at the interface, while conversely T49D or L581D can add favorable electrostatic interactions.

Mutation I595M is found at the third interfacial site. I595 of one gp41 subunit occupies a hydrophobic pocket on an adjacent gp41; the methionine substitution may better pack in this pocket (Figure 4A).

We combined mutations to engineer BaL Env for enhanced PG16 binding. High PG16 binding was generally achieved by combining mutations from separate sites, suggesting mutations within a single site have epistatic interactions. Two sequences, termed BaL-QES.i01 and BaL-QES.i02 ("i" for "interfaces", where most of these mutations are localized), were found to increase PG16 binding at saturation, with only slightly tighter apparent affinities (Figure 5A). This is consistent with expression of diverse conformations on the cell surface that do not substantially interconvert on the timescale of the experiment, and the QES mutations increase the pool of Env occupying a closed trimeric state competent for PG16 interaction. However, cells expressing these engineered variants still bound CD4 (Figure 5B) and antibodies targeting V3 epitopes (Figure 6A) at similar levels to wild type Env. We reasoned that mutations predicted to positively stabilize the closed state due to improved interface complementarity are insufficient to prevent induction or sampling of the open conformation, and additional mutations are required that actively destabilize the open state by steric hindrance. Due to the form of the Lennard-Jones potential, steric clashes can have an outsized effect on the energy gap between conformational states.

A QES mutation within the core of Env can destabilize the CD4-bound state

Based on the deep mutational scan, we tested and validated three mutations within the Env core for enhanced PG16 binding (Figure 4B). I181L is just below the surface of the apical tip where PG16 binds, and perhaps stabilizes local packing. V254T and V255M are centrally positioned in the linker connecting the inner and outer domains of gp120. V254T adds a hydrogen bond to the backbone carbonyl of L261 in strand β9 of the outer domain, while V255M increases hydrophobic packing to aromatic residues in the inner domain (Figure 4B). The cavity occupied by V255M collapses in the open conformation with insufficient space for large hydrophobics. V255M is therefore unique amongst the QES mutations described thus far in that modeling predicts it also destabilizes the open state via introducing steric strain. It is notable that the Env V255M mutation was also identified in an *in vitro* replication screen for HIV-1 variants resistant to CD4 binding site inhibitors[131]; resistance was also associated with reduced entry efficiency and gp120-V255M had substantially decreased affinity for CD4 [131].

Core mutations were combined with the previous QES mutants at the trimer interface to generate BaL-QES.i01.c01 ("c" for "core"), which, when expressed on the cell surface, displays enhanced binding to PG16 and decreased binding to CD4 (Figures 5A and 5B). BaL-QES.i01.c01 also has reduced exposure of V3 region epitopes (Figure 6B), consistent with destabilization of the open conformation.

QES mutations are transferable to other Env strains and enhance binding to some bNAbs

Most of the QES mutation sites are conserved across HIV-1 strains and clades, though in some cases other strains already carry the substituted amino acid, such as a threonine, tyrosine or glutamine at positions 200, 223 and 315, respectively. We experimentally tested whether a subset of the mutations in BaL Env that enhance PG16 binding are effective in HIV-1 strains from clades A and C. Q769.d22 and Q842.d12 are tier 2 strains from clade A, and 25711 and DU422 are tier 1B and 2 strains from clade C, respectively[117]. QES mutations at the trimer interface were effective in the clade A but not clade C sequences, while at least one core mutation could be transferred to each of the strains. As observed for BaL Env, K_D changes are small, whereas PG16 binding at saturation is substantially increased (Figure 5A), consistent with putative stabilization of the closed conformation increasing the fraction or total amount of folded Env on the cell surface recognized by PG16. Only variants harboring V255M had reduced CD4 binding (Figure 5B), again emphasizing that substitutions modeled as positively stabilizing the closed state are insufficient to prevent dynamic sampling or induction of the CD4-bound open state, which instead requires explicit destabilization.

While engineering of Env was based on maximizing binding signal to a single conformation-dependent antibody, the QES variants also have increased presentation of tertiary epitopes recognized by bNAbs PGT121 and PGT128 (Figures 5C and 5D), which contact the N332 glycan supersite on the outer surface of gp120[132–135]. The N332 glycan epitope is recognized by a variety of bNAb clones that originated from various germline precursors in different individuals, implying that this structural region may form a supersite of vulnerability for immune targeting. PGT145, which makes salt-bridges between acidic residues on a CDR loop and basic amino

acids in the apical cavity[130], has either increased or decreased binding to the different QES constructs (Figure 5E), reflecting antagonism between stabilization of the closed conformation (which will enhance PGT145 binding) versus amino acid changes that disrupt direct antibody contacts (specifically, decreased PGT145 binding is perfectly correlated with the inclusion of mutation P124D in the apical cavity). Since gains in PG16, PGT121 and PGT128 binding could in part be due to increases in Env surface expression, we fused c-myc tags to the extracellular N-termini followed by a 14-residue glycine/serine-rich linker connecting to the first residue of gp120, and measured Env surface expression independent of any folded structure. Cells expressing myc-tagged or untagged Env proteins bound PG16 similarly, with QES variants showing robust increases in PG16 binding (Figure 5F). There was no apparent increase in total surface expression (Figure 5F), consistent with a shift in the conformational ensemble expressed on the cell surface towards closed trimeric states recognized by bNAbs.

A deep mutational scan of DU422 gp140 for binding to CD4 and PG16

Ideally, we would identify a suite of mutations for applying broadly to Env from any HIV-1 strain to increase surface presentation of closed trimers that expose conserved epitopes. However, the QES mutations described thus far, based on the deep mutational scan of BaL Env, were only partially transferable to other strains. In particular, mutations that neutralize electropositive charge in the apical cavity failed to increase presentation of the PG16-recognized conformer in clade C Env sequences, despite clade C Env sharing similar electrostatic properties at the apex. To address this shortcoming, we mutationally scanned clade C DU422 Env for binding to CD4 and PG16 and identified new substitutions for an expanded set of QES mutations against which Env from any strain can be screened.

When transfected with a large excess of carrier DNA to limit the copy number of coding sequences acquired per cell, DU422 gp160 surface expression in Expi293F cells was not detected. Expression remained undetectable even when the coding sequence was downstream of an intron, or when $gp160_{DU422}$ was cotransfected with carrier DNA designed to promote episomal plasmid replication. Since the cytoplasmic tail of gp160 features multiple motifs that diminish surface expression [136], we expressed only extracellular $gp140_{DU422}$ (a.a. N31-N677) anchored to the membrane via a flexible gly/ser-rich linker and canonical transmembrane helix. Soluble gp140 has increased conformational heterogeneity and imperfectly mimics the ectodomains of full-length Env[48,137,138], and therefore we have been careful to subsequently validate interesting mutations in the full protein. DU422 gp140 was expressed on the cell surface at high levels based on flow cytometric analysis of PG16 and soluble CD4 binding.

Three SSM libraries spanning $gp140_{DU422}$ residues N31-N279, N280-Q577, and T578-N677 together encoded all possible 12,820 single amino acid substitutions. As described for the mutational scan of $gp160_{BaL}$, Expi293F cells were transfected with the libraries, selected by FACS for sCD4 and PG16 binding, and mutation frequencies in RNA transcripts were compared to the naïve plasmid libraries (Figures 7A and 7B). Sorting experiments were independently replicated, and the $\log_2$ enrichment ratios for each mutation weakly agree; neutral mutations in the correlation plots are clustered near the origin, while deleterious mutations fall in the negative quadrant (Figures 7C and 7D). Residue conservation scores are better correlated between replicates (Figures 7E and 7F).

Cavity-filling mutations at the gp120 inner-outer domain interface can increase presentation of the PG16-recognized closed trimer

Based on predictions from the deep mutational scan, we screened and validated five new QES mutations in full-length DU422 gp160 that show increased PG16 binding when expressed in Expi293F cells (Figure 4). One of the mutations, L544Y, localizes to a fourth site at the trimer interface and fills a cavity (Figure 4A). The other four mutations (V208M, T283P, F382W, and Y484W) are at the junction between the gp120 inner and outer domains (Figure 4B), a region where we previously identified the QES mutations V254T and V255M in BaL Env. The interface between the gp120 inner and outer domains is underpacked and undergoes substantial structural rearrangement upon CD4 ligation. As a generalization, these core mutations at the gp120 inner-outer domain interface increase hydrophobic packing. Using rational design, cavity-filling mutations have been previously described in BG505 SOSIP that stabilize closed trimers, albeit the mutations were across the entire protein rather than focused at the gp120 inner-outer domain interface[129]. Our deep mutational scan was applied to an unbiased library, and it is noteworthy that the data converged on cavity-filling substitutions for enhanced presentation of the closed conformation, with two of the QES mutations we identify, L544Y and V208M, being shared with the prior study[129]. Ultimately, we were able to use these new mutations to engineer DU422 Env proteins with favorable binding to bNAbs and diminished binding to strain-specific antibodies (Figure 8).

DISCUSSION

Alignments of protein sequences observed in nature are very effective at revealing conserved residues in primary structure for correct folding and function. However, a natural sequence like HIV-1 Env is shaped by multiple activities; Env must fold, traffic to the cell surface, be incorporated into a budding virion, bind target receptors, mediate membrane fusion, and escape antibody neutralization. Here, *in vitro* selection was instead focused on the specific activities of folding to a trimeric closed conformation and binding to the CD4 receptor. The sequence-activity landscapes provide insight into Env mutational tolerance for acquiring closed and CD4-bound conformations, information which is not at all obvious from a multiple sequence alignment. This information provided mechanistic insights and was leveraged for engineering full-length Env for enhanced presentation of vulnerable epitopes when expressed on the cell surface. By shifting the expressed conformational ensemble towards closed trimers recognized by broadly neutralizing antibodies, QES variants may prove useful for vaccines that incorporate Env in ways that cannot be easily purified in the desired conformation (e.g. VLP or nucleic acid vaccines). More importantly, this body of work highlights a mutational scanning methodology that does not require virus passaging and can be broadly applied to any complex protein expressed in human cells.

MATERIALS AND METHODS

Tissue culture.

All experiments were in the CXCR4-knockout Expi293F cell line [63]; these cells do not express CD4, CCR5, or CXCR4 receptors. Cells were cultured in Expi293 Expression Medium (Life Technologies), 8% CO_2, 37 °C, at 125 rpm.

Expression plasmids.

Synthetic, human codon-optimized Env sequences were synthesized as gBlocks (Integrated DNA Technologies) from HIV-1 isolates Q769.d22 (GenBank Accession No. AAM66234.1), Q842.d12 (AAM66242.1), BG505 (ABA61516.1; carrying the SOSIP, T332N and 6xArg proteolysis site mutations), 191084 B7-19 (ADI62025.1), BaL (AAA44191.1), H031 (EF210728), B41 (AEO83374.1), AD8 (AAB64170.1), 25711 (ABL67448.1), ZM197M (ABD49673.1), DU422 (ABD83641.1), 001428-2 (ABL67442.1), SHIV327C-Hu A10 (AIZ78012.1), CH111 (EF117258) and CH115 (EF117263). The genes encoded mature polypeptides fused to an N-terminal CD5 leader (sequence MPMGSLQPLATLYLLGMLVASVLA), and were cloned into the NheI-XhoI sites of pCEP4. Mutations were made by overlap extension PCR.

When transfecting cells under conditions that yielded a single coding variant per cell, a pCEP4 derivative vector (pCEP4-intron) containing a strong 5' chimeric intron was used for enhanced expression. This was created by cloning the intron from plasmid pRL-SV40 (Promega) into the KpnI-NheI sites of pCEP4.

Plasmid library construction.

Single site-saturation mutagenesis (SSM) libraries were generated by overlap extension PCR[139]. For the $gp160_{BaL}$ library, three separate SSM libraries were constructed focused on the Env_{BaL} N-terminus (a.a. 31-265; Library A), center (a.a. 266-529; Library B), and C-terminus (a.a. 530-856; Library C). The PCR products were cloned by restriction enzyme digestion and ligation into the NheI-BglII (Library A), BamHI-NotI (Library B), and PstI-XhoI (Library C) sites of BaLgp160 inserted into the NheI-XhoI sites of pcDNA3.1(+) (Invitrogen), with the vector PstI and BglII sites removed by QuikChange (Agilent) mutagenesis. Ligations were transformed into NEB 5-α or 10-β electrocompetent *E. coli* (New England Biolabs), and plasmid DNA for each library was prepared using GeneJET Maxiprep Kit (Thermo Scientific). Following library construction in the pcDNA3.1(+) vector, the full-length diversified BaLgp160 library inserts was subcloned into the NheI-XhoI sites of pCEP4-intron. At all cloning steps, the number of transformants was at least an order of magnitude greater than the possible library diversity. Combined, the three BaL gp160 SSM libraries covered 16,332 out of 16,520 possible single amino acid mutations, based on a minimum frequency of 5.7×10^{-6} (corresponding to approximately 10 reads) in the deep sequenced plasmid libraries.

For the $gp140_{DU422}$ library, the gp140 ectodomain (a.a. N31-N677, HXB2 reference numbering) was fused using PCR-based assembly to a C-terminal gly/ser-rich linker, 6his tag, and the transmembrane helix of HLA class I α chain for surface display. The synthetic gene was cloned into the NheI-XhoI sites of pCEP4. Three separate SSM libraries were constructed focused on the N-terminus (a.a. 31-279; NT library), center (a.a. 280-577; central library), and C-terminus (a.a. 578-677; CT library). Mutagenized PCR segments were ligated into the NheI-Pfl23II (NT

library), SbfI-HindIII (central library), or Pfl23II-Xho1 (CT library) sites of pCEP4-gp140$_{DU422}$. Ligations were electroporated and plasmid DNA purified as described above. All possible 12,820 single amino acid substitutions were present in the three SSM libraries.

To investigate Env association with MA, 001428 Env with a CD5 leader peptide was genetically fused at residue L753 to the C-terminal half of split Venus (VC: a.a. D155-K238) via a GGSG linker using PCR-based assembly. A 001428(753)-VC SSM library was generated by overlap extension PCR [139] and covered residues N677-L753. The library was cloned into the NheI-XhoI sites of pCEP4-intron and prepared as described above. All possible 1,540 single amino acid mutations were present.

Library transfections.

DNA libraries were transfected into CXCR4-knockout Expi293F cells. For transfecting the 001428(753)-VC library, the cells also stably expressed a MA-VN fusion. The MA domain of Gag (a.a. 1-125) from plasmid pGag-EGFP (Cat. No. 11468 from the NIH AIDS Reagent Program, Division of AIDS, NIAID) [140] was fused to the N-terminal half (VN: a.a. V1-A154) of split Venus (mutant I152L), a yellow fluorescent protein variant with improved signal-to-noise for the detection of protein interactions by BiFC[141]. The fusion construct was cloned back in to pGag-EGFP using KpnI-NotI, the plasmid was linearized with AflIII, and transfected into CXCR4-knockout Expi293F cells. Stable transfectants were selected with 100 µg/ml G418, and positive cells were enriched by FACS after transient transfection with pcDNA3.1(+) encoding VC.

The libraries were transfected such that on average no more than one coding sequence was acquired per cell[63]. Cultures at a density of 2×10^6 cells/ml were transfected using Expifectamine with 1 ng/ml library DNA and 1.5 μg/ml pCEP4-ΔCMV[80] as carrier DNA. Two hours post-transfection, the medium was replaced.

Sorting cell libraries for binding to sCD4, VRC01 and PG16.

Recombinant sCD4-183 (provided by Progenics), VRC01[116], and PG16[111] were obtained from the NIH AIDS Reagent Program, Division of AIDS, NIAID (Cat. No. 7356, 12033 and 12150).

For CD4 binding, cells were harvested 24-26 hours following transfection with library DNA, washed with ice cold phosphate buffered saline supplemented with 0.2% bovine serum albumin (PBS-BSA), and incubated on ice for 40 minutes with 200 nM ($gp160_{BaL}$ libraries) or 10 nM ($gp140_{DU422}$ libraries) sCD4-183. Cells were washed twice, incubated on ice for 30 minutes with fluorescein isothiocyanate (FITC)-conjugated anti-CD4 (clone M-T441, LifeSpan BioSciences, 1/200 dilution), washed twice, and resuspended in PBS-BSA.

For antibody binding, cells were washed with cold PBS-BSA 24-26 hours post-transfection and incubated on ice for 40 minutes with 5 nM VRC01 ($gp160_{BaL}$ libraries), 2 nM PG16 ($gp160_{BaL}$ libraries), or 3 nM PG16 ($gp140_{DU422}$ libraries). Cells were washed twice, incubated for 30 minutes with allophycocyanin (APC)-conjugated anti-human IgG Fc antibody (BioLegend, clone HP6017, 1/300 dilution), washed twice more, and resuspended in PBS-BSA.

Labeled cells were sorted on a BD FACS Aria II at the Roy J. Carver Biotechnology Center. Single cells were gated based on FSC/SSC, and dead cells that were positive for propidium iodide (added to a final concentration of 1 μg/ml) were excluded. Autofluorescent cells in the APC channel (after staining with FITC-conjugated anti-CD4) or FITC channel (after staining with APC-conjugated anti-IgG Fc) were also excluded. Cells were collected with the highest fluorescent signals (the top 0.3-1.0 % depending on the library and ligand, chosen so as to collect cells with similar binding signals across the respective libraries). Sort conditions are described in greater detail in the sequence depositions with NCBI's Gene Expression Omnibus (79). Sorted cell pellets were frozen at -80 °C. To maintain cell viability and mRNA quality during the experiment, samples were sorted for a maximum of 4 hours into tubes that had been coated overnight with fetal bovine serum (FBS). To collect greater numbers of cells than one 4-hour sort provided, libraries were prepared again and frozen cell pellets from multiple days' experiments were pooled during RNA extraction. Each replicate typically required 8 hours of sorting per library.

BiFC-based library sorting.

Cells stably expressing MA-VN were transfected with the 001428(753)-VC plasmid library, and were stained 24 hours post-transfection with 2 nM PG16 and APC-anti-IgG Fc as described above. During sorting, single cells were gated by FSC/SSC properties, and propidium iodide-positive dead cells and Pacific Blue-positive autofluorescent cells were excluded. APC-positive cells were gated, and the upper and lower 15 % for Venus fluorescence (Figure 10B) were collected into tubes that had been coated overnight with FBS.

Deep sequencing.

Total RNA was extracted from sorted cells using a GeneJET RNA Purification Kit (Thermo Scientific), and first strand cDNA was synthesized with high-fidelity AccuScript reverse transcriptase (Agilent Technologies) primed with oligonucleotides that annealed downstream of the diversified regions. To generate fragments for deep sequencing, the cDNA was PCR amplified in two rounds. In the first round, primer overhangs added complementary sequences to the Illumina sequencing primers. In the second round, primer overhangs added barcodes and adaptor sequences for annealing to the Illumina flow cell. Thermocycling was kept to a minimum to reduce the introduction of PCR biases and errors. Primer sequences are provided in the GEO depositions. Each of the $gp160_{BaL}$ libraries was amplified as three overlapping fragments to achieve full sequencing coverage. For $gp140_{DU422}$, the NT and central libraries were amplified as two overlapping fragments, while the CT library was amplified as a single fragment. The 001428(753)-VC library was amplified as a single fragment focused on the diversified region. DNA was sequenced at the UIUC Roy J. Carver Biotechnology Center on an Illumina MiSeq v3 (2 × 300nt kit) or HiSeq 2500 (2 × 250nt kit), and data were analyzed with Enrich [142]. Log_2 enrichment ratios of mutants were normalized by subtracting the enrichment of the wild type sequence.

Flow cytometric analysis of Env mutants binding to antibodies and sCD4.

Expi293F CXCR4-knockout cells were transfected with Expifectamine (Life Technologies) using 500 ng plasmid DNA per mL of cells at a density of 2×10^6 / ml, unless stated otherwise. Transfection Enhancers (Life Technologies) were added after 18 h, and cells were analyzed ~40 h post-transfection. Cells were washed with PBS-BSA and incubated on ice for 40 minutes with

the indicated antibodies or sCD4. Cells were washed twice, incubated for 30 minutes with secondary antibody (1/200 FITC-anti-CD4 clone M-T441 from LifeSpan BioSciences, or 1/300 APC-anti-IgG clone HP6017 from BioLegend), washed twice, and analyzed on a BD LSR II flow cytometer. For titrating different antibodies or sCD4, a 1:3 serial dilution of the ligand was prepared in a 96-well round-bottomed plate, and plates were incubated with the antibody at 4 °C on a rocker. 39F[143,144] was provided by the NIH AIDS Reagent Program from Dr. James E. Robinson. 268-D IV[145], 2442[146], and 3074[147,148] were provided by Dr. Susan Zolla-Pazner through the NIH AIDS Reagent Program. PGT121[149], PGT128[149], and PGT145[149] were provided by IAVI through the NIH AIDS Reagent Program. 35O22[150] was provided by the NIH AIDS Reagent Program from Drs. Jinghe Huang and Mark Connors.

Assessment of data quality by targeted mutagenesis.

20 representative mutations were chosen that were depleted following FACS-based selection for PG16 binding in both replicate experiments. All 20 mutants were found to have reduced PG16 (2 nM) binding signal by flow cytometry when transfected in to Expi293F cells. The mutations were: V36R, F93G, D133W, F159V, Y191I, S199C, I201M, V208G, F223L, T248E, I251S, T319K, C331I, W338A, I491S, Q540V, W614A, K617Q, E634A, and I686R.

43 mutations biased towards subunit interfaces and the furin cleavage site were tested by targeted mutagenesis for increased PG16 (0.5 and 2 nM) binding when expressed by Expi293F cells. All the mutations were enriched in the replicate selections for PG16 binding. 21 of the mutations were validated as increasing PG16 binding signals by small to moderate degrees: T49D, Q114A, K117V, K117Y, P124D, T163D, R166E, R166F, R166L, V200E, V200T, F223Y, R315A,

R315Q, R432T, G514P, G516Q, R557Q, L581D, I595M, and L663N. The 22 mutations that did not increase PG16 binding were: L34Y, W35Y, T51Q, Y61K, Y61Q, E106S, I161L, N164P, I165H, I165L, I165Q, T244I, V430E, K500Q, A501E, E509Q, H564A, L568Y, Q658F, E662Q, K665N, and S700Q.

Structural modeling.

Homology modeling of BaL Env was based on crystal structures of sequences from other HIV-1 strains. VRC01-bound $gp140_{BaL}$ in the closed conformation was modeled by threading the BaL strain sequence onto PDB 5FYK [151] with glycans removed and rebuilding missing loops and minimizing side chain conformations in FoldIt[152]. CD4-bound $gp120_{BaL}$ was generated by threading the BaL sequence onto PDB 1GC1[122] and minimizing side chain and backbone conformations with FoldIt. For modeling open-state $gp140_{BaL}$ bound to CD4, the sequence of BaL Env was threaded onto PDB 5VN3[124], which was then minimized with C3 symmetry using ROSETTA[153]. The model of the PG16-bound apical epitope was generated by threading the BaL sequence onto PDB 4DQO[121]. Coordinates for the glycan on N160 were kept, and a single N-acetyl-D-glucosamine sugar was added to N156. The structure in Figure 4 was generated by superimposing the model of the PG16-bound epitope to the model of closed Env. Homology modeling of DU422 Env was based on PDBs 5FYK and 5VN3 as described above. Images were rendered with the PyMOL Molecular Graphics System, Schrödinger, LLC.

Total cell surface expression and PG16 binding of c-myc tagged Env.

To test for changes in total cell surface expression, the native signal peptide of Env was replaced with the influenza hemagglutinin signal peptide (MKTIIALSYIFCLVFA) followed by a c-myc

epitope tag and 14-residue flexible linker (GSPGGASSGSGSGG) prior to the first residue of gp120. Expi293F CXCR4-knockout cells were transfected using Expifectamine with 500 ng plasmid DNA encoding myc-tagged Env, and Transfection Enhancers were added 20 h post-transfection. Cells were tested 27 h post-transfection for total Env surface expression and PG16 binding. For measuring Env expression, cells were washed with cold PBS-BSA, stained for 40 minutes with anti-myc Alexa 647 (clone 9B11, 1/200 dilution; Cell Signaling Technology), and washed twice. Cells tested for PG16 binding were washed with cold PBS-BSA, incubated on ice for one hour with 2 nM PG16, washed twice, incubated on ice for 40 minutes with APC-conjugated anti-human IgG Fc (clone HP6017, 1/300 dilution; BioLegend), and washed twice. Cells were analyzed on a BD LSR II flow cytometer.

BG505 SOSIP.664 purification.

BG505 SOSIP.664 (T332N) was subcloned into pCEP4 (NheI-XhoI sites) with a C-terminal gly/ser-rich linker and 8his tag. Plasmids were transfected into Expi293F cultures using a modified protocol from [154]; per mL of culture, 1 μg DNA was mixed with 5 μg linearized polyethyleneimine (MW 25,000; Polysciences) in 100 μl of OptiMEM (Gibco). The mixture was incubated for 20 minutes at room temperature, then added to the culture at a density of 2×10^6 / ml. Expifectamine Transfection Enhancers were added 18 h post-transfection. Cells were centrifuged 4 days post-transfection, and secreted protein was purified from the culture supernatant.

Protein was purified at 4 °C. The supernatant was dialyzed against 20 mM Tris pH 8.0 / 225 mM NaCl for 6-8 h, followed by dialysis overnight against 20 mM Tris pH 8.0 / 20 mM imidazole /

300 mM NaCl. Equilibrated NiNTA (50 % slurry, 500 µl per 40 ml culture, Thermo Scientific) was incubated with the sample for 1 h on a rocker, collected in a gravity column, and washed with 20 ml of purification buffer (20mM Tris pH 8.0 / 300 mM NaCl) containing 20 mM imidazole. Protein was eluted using a step gradient of purification buffer containing 50, 100 and 250 mM imidazole (1 ml per fraction). Fractions containing protein were pooled, concentrated using a 30 kD MWCO centrifugal device (Sartorius), and separated by size exclusion chromatography using a Superose 6 increase 10/300 GL column on an ÄKTA pure system (GE Healthcare) with PBS as the running buffer.

Data Availability.

Analyzed and raw deep sequencing data are deposited in NCBI's Gene Expression Omnibus (79) under series accession numbers GSE102276 (BaL gp160), GSE117328 (DU422 gp140), and GSE126136 (001428 Env interactions with MA). The deposits include commands for running Enrich scripts to replicate data analysis, and additional details on primer sequences and sort conditions.

Plasmid Availability.

Plasmids have been deposited with Addgene under ID numbers 100918-100933, 111837-111847, and 123212-123283.

ACKNOWLEDGEMENTS

Barbara Pilas, Barbara Balhan, and Angela Kouris at the UIUC Roy J. Carver Biotechnology Center assisted with flow cytometry, and Alvaro Hernandez and Chris Wright helped with deep sequencing. This work was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award R01AI129719. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The University of Illinois has filed a patent with E.P. and J.D.H. as co-inventors covering aspects of this work.

FIGURES

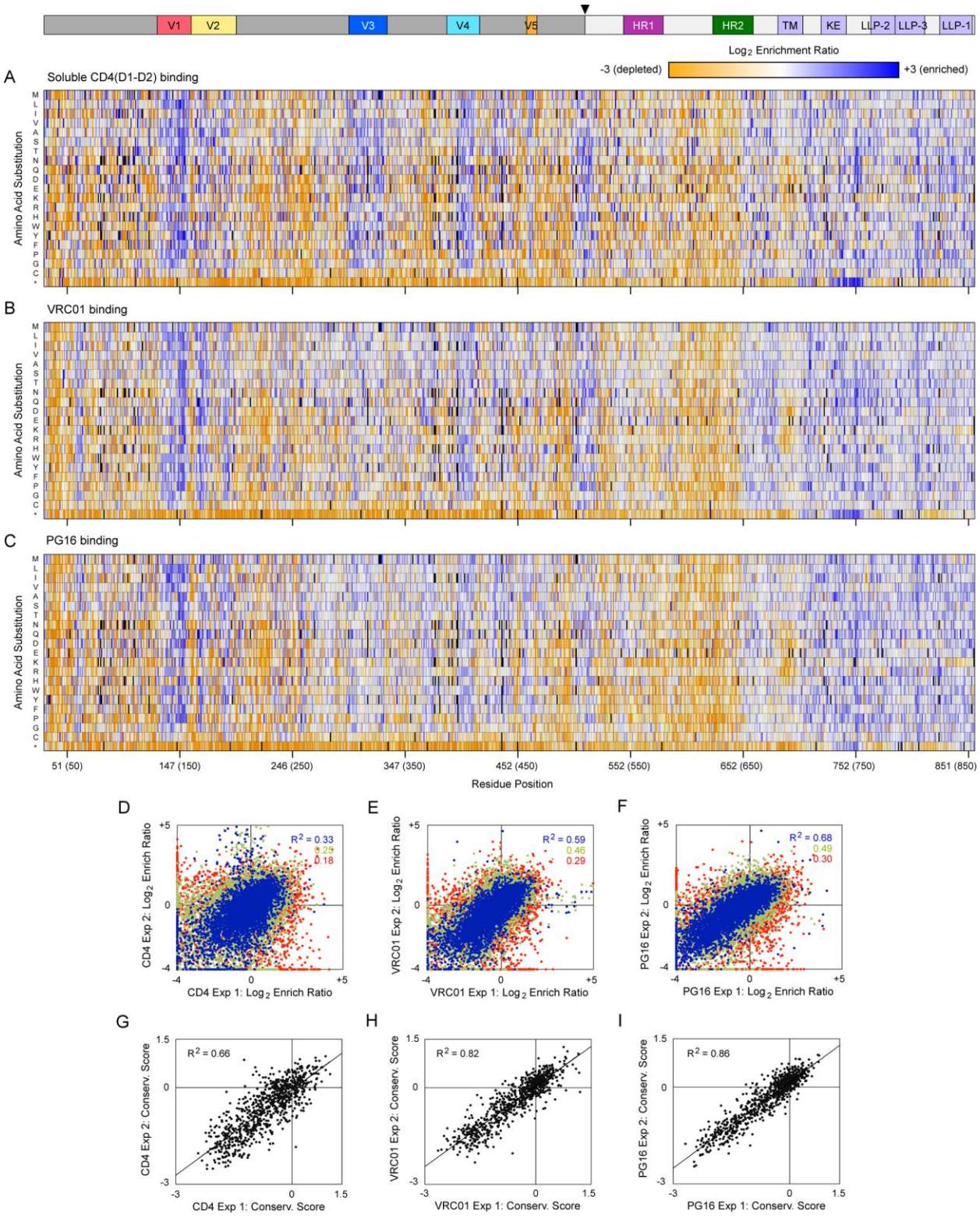

Figure 3.1. Mutational landscapes of BaL Env interacting with protein ligands. Data generated by Jeremiah Heredia. (**A-C**) Merged data from *in vitro* evolution of three SSM libraries that together…

Figure 3.1 (cont.) fully span the mature Env_{BaL} protein. The libraries were evolved by FACS for high binding signals to (A) 200 nM CD4(D1-D2), (**B**) 5 nM VRC01, and (**C**) 2 nM PG16. The Env sequence is on the horizontal axis (HXB2 numbering, BaL numbering in parentheses), and single amino acid substitutions are on the vertical axis. *, stop codon. Log_2 enrichment ratios are plotted from ≤ -3 (depleted, orange) to 0 (neutral, white) to ≥ +3 (enriched, blue). Mutations missing in the libraries (frequencies < 5×10^{-6}) are black. Primary structure of gp120 (dark grey) and gp41 (light grey) is indicated above, with an arrowhead at the proteolysis site. Average of two independent selection experiments. (**D-F**) Correlation plots of mutation log_2 enrichment ratios from independently replicated selections for high binding signals to (**D**) soluble CD4, (**E**) VRC01, and (**F**) PG16. Abundant mutations (frequencies > 2×10^{-4} in the naïve library) are blue, mutations with moderate representation (frequencies between 5×10^{-5} and 2×10^{-4}) are green, and rare mutations (frequencies between 5×10^{-6} and 5×10^{-5}) are red. (G-I) Conservation scores were calculated by averaging the log_2 enrichment ratios for all substitutions at each residue position. Conservation scores for libraries sorted for binding (**G**) soluble CD4, (H) VRC01, and (I) PG16 show agreement between replicate experiments.

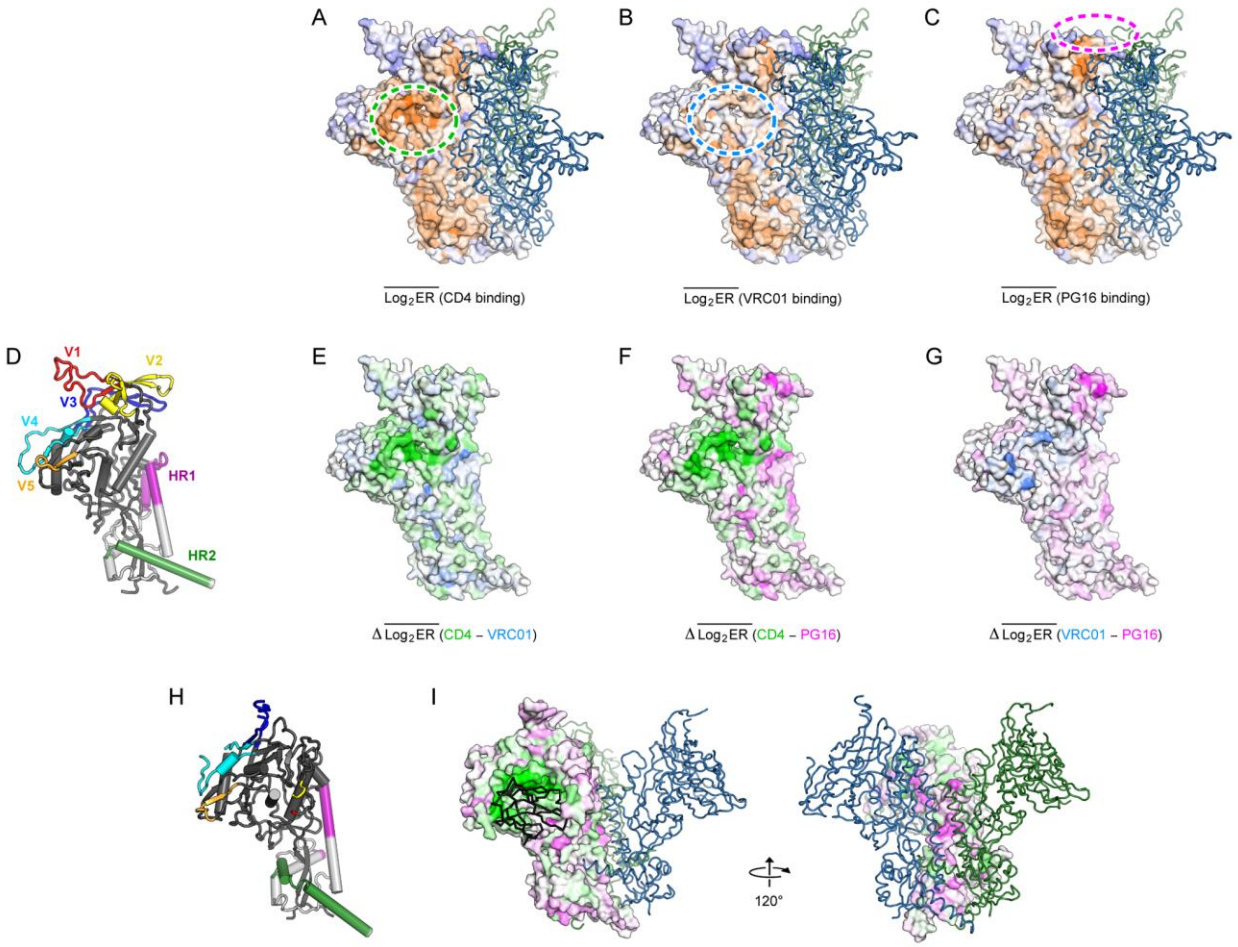

Figure 3.2. Mapping conserved sites for ligand binding to Env structure. Data generated by Jeremiah Heredia. (**A-C**) Conservation scores from selecting Env libraries for binding (A) soluble CD4, (**B**) VRC01, or (**C**) PG16 are mapped to the surface of an Env protomer, from ≤ -2 (conserved, orange) to ≥ +2 (under selection for change, dark blue). The second and third protomers in the trimeric spike are shown as dark green and blue ribbons. The binding sites for CD4, VRC01, and PG16 are shown with green, cyan, and magenta dashed lines, respectively. The model of BaL Env in the closed state was generated by sequence threading to PDB 5FYK, followed by loop building, and side chain and backbone minimization. (**D**) Cartoon representation of a single protomer (gp120 is dark grey, gp41 is pale grey), oriented as in (A-C). (**E**) Differences between conservation scores for binding soluble CD4 and VRC01 are plotted from -2 (more conserved for CD4 binding, green) to +2 (more conserved for VRC01 binding, cyan) on the surface of an Env protomer oriented as above. (**F**) CD4-PG16 difference conservation plot colored from -2 (more conserved for CD4 binding, green) to +2 (more conserved for PG16 binding, magenta). (**G**) VRC01-PG16 difference conservation plot colored from -2 (more conserved for VRC01 binding, cyan) to +2 (more conserved for PG16 binding, magenta). (**H**) Model of Env_{BaL} in the open CD4-bound conformation (based on PDB 5VN3 (42)), showing a single protomer. (**I**) As in (F), but now…

Figure 3.2 (cont.) plotting the CD4-PG16 difference conservation scores to Env_{BaL} in the open CD4-bound conformation, oriented as in (**H**). A single CD4 domain D1 is shown as a black ribbon.

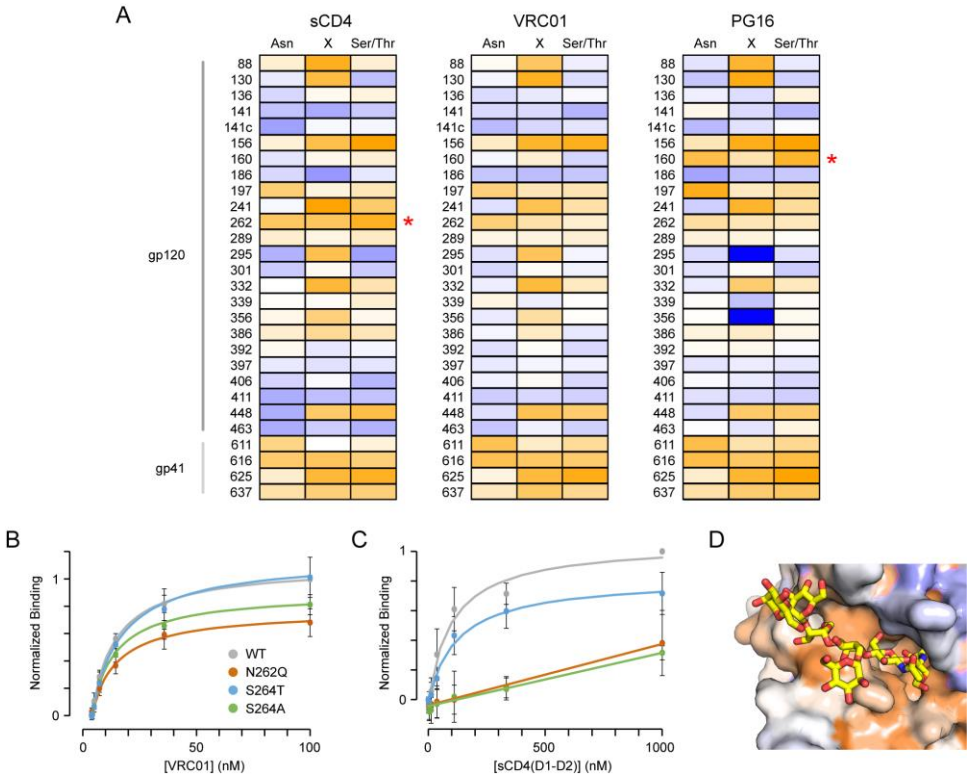

Figure 3.3. Conservation of N-glycosylation sites. Data generated by Kevin Gill. (**A**) Env libraries were evolved for binding soluble CD4, VRC01, and PG16 as indicated. Conservation scores of residues within N-glycosylation motifs are plotted from -2 (conserved, orange) to 0 (variable, white) to +2 (under selection for change, blue). Site numbering is based on Asn in the motif. Red asterisks indicate glycosylation sites preferentially conserved for binding only one of the three ligands. (**B**) Flow cytometry analysis of VRC01 binding to Expi293F cells expressing wild type (WT, grey), N262Q (dark orange), S264T (blue), and S264A (green) Env variants. (n = 4, mean ± SD) (**C**) Binding of soluble CD4 to Env variants. (n = 4, mean ± SD) (**D**) The N262 glycan (yellow sticks) occupies a cleft in the cryo-EM structure of CD4-bound Env (PDB 5VN3). The protein surface is colored by conservation score for CD4 binding, from ≤ -2 (conserved, orange) to ≥ +2 (under selection for change, dark blue).

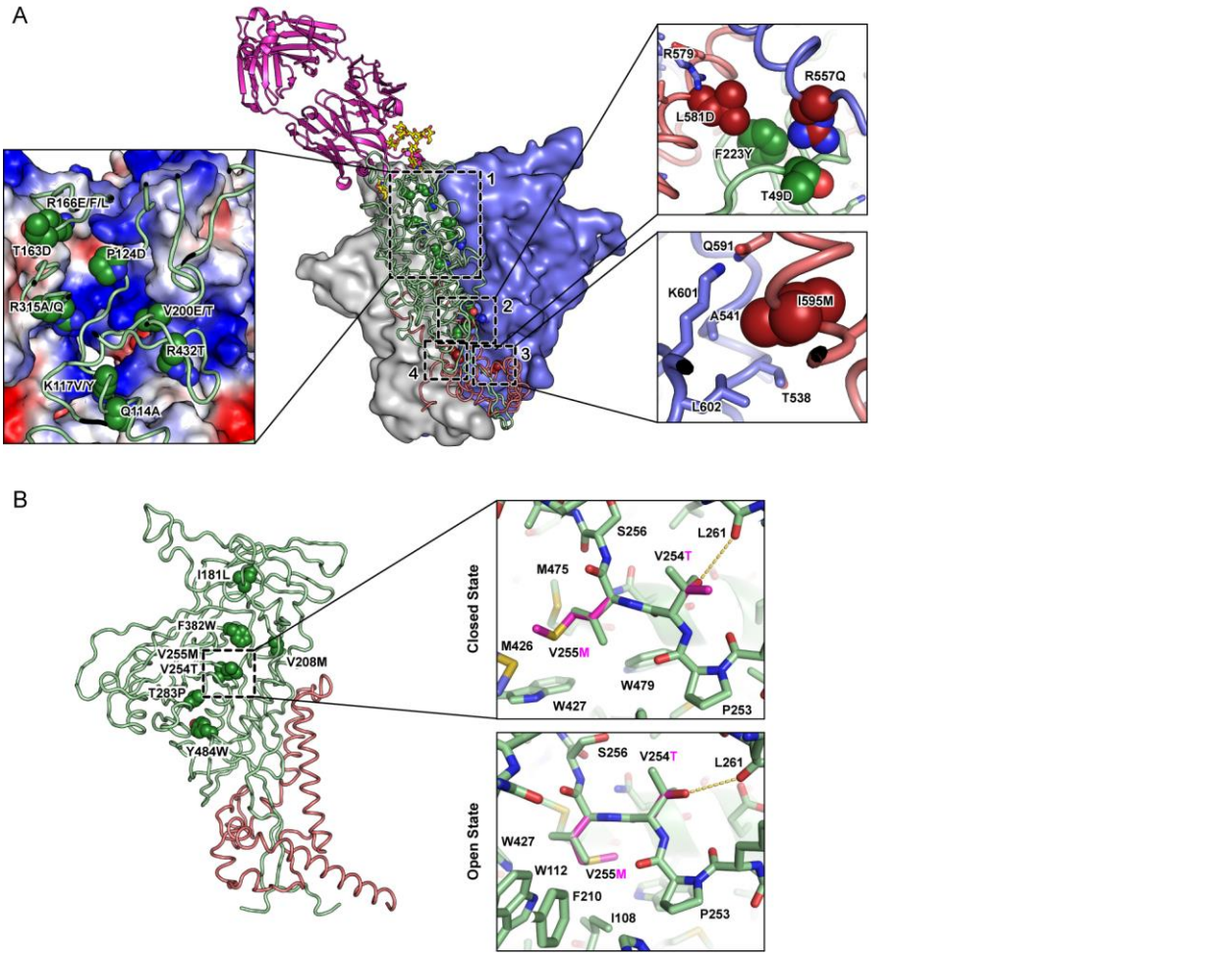

Figure 3.4. Neutralization of the electropositive apical cavity and increased packing at the inner-outer domain interface increase presentation of Env in a PG16-recognized conformation. Data generated by Jeremiah Heredia and Kevin Gill assisting with validation. From the mutational scan of $gp160_{BaL}$ and $gp140_{DU422}$, substitutions were identified and validated that enhance PG16 binding. (**A**) Mutations at subunit interfaces clustered to four sites, shown on a structural model of closed Env_{BaL}. PG16 is magenta, interacting glycans are yellow, two Env protomers are shown as grey and blue surfaces, and the third Env protomer is shown as a green (gp120) and brick red (gp41) ribbon. In the magnified inset of site 1, the electrostatic potential on the surface of two Env protomers is plotted from positive (blue) to negative (red). Previously described mutation L544Y is found at site 4. (**B**) Mutations in the core generally increased hydrophobic packing along the inner-outer domain interface of gp120. In the magnified insets, V255M (magenta) fills a void in the closed conformation, whereas modeling predicts V255M has steric clashes in the open conformation, with only a single methionine rotamer being accommodated.

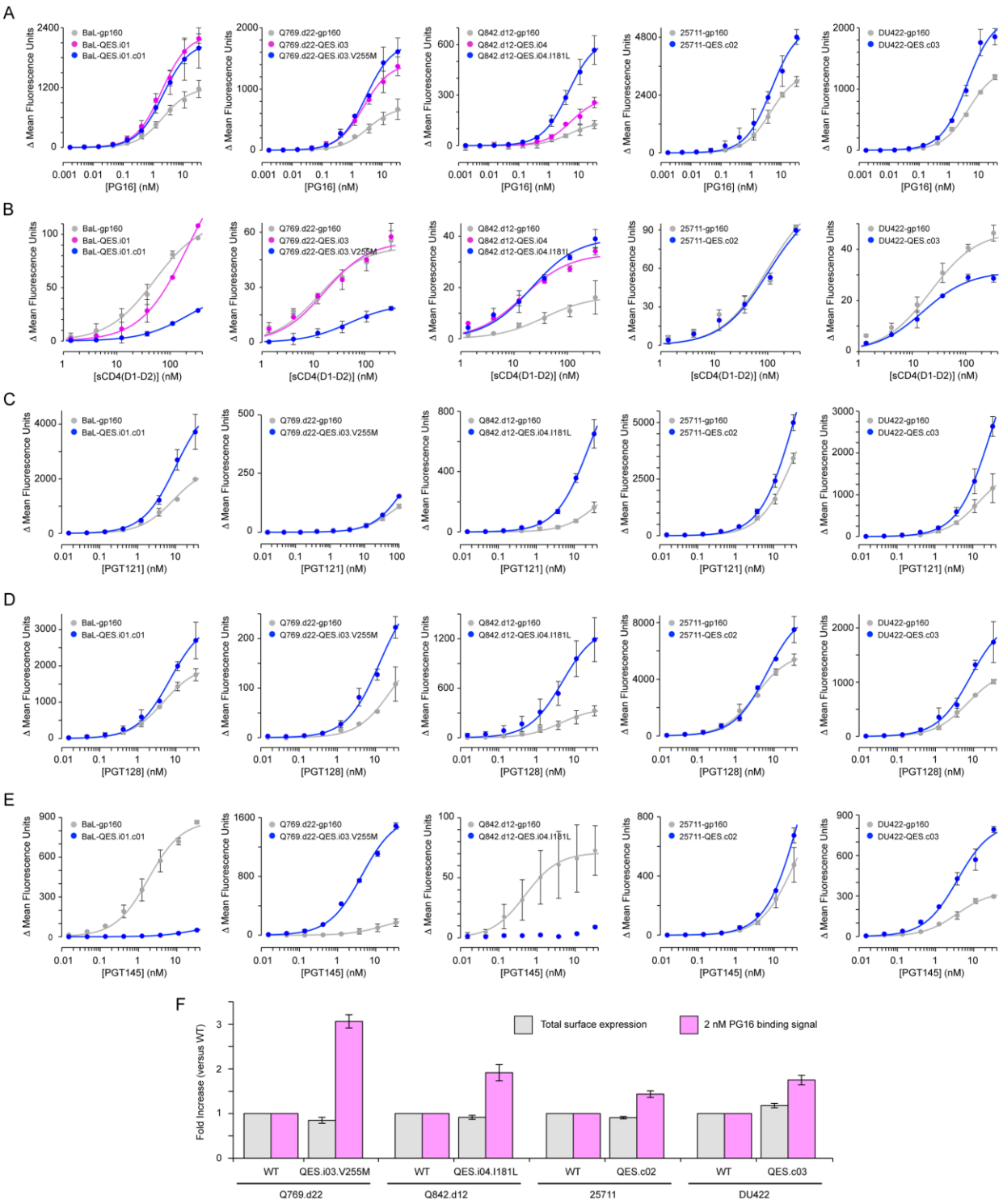

Figure 3.5. First generation of engineered Env proteins for enhanced presentation of the PG16-recognized conformation. Data generated by Jeremiah Heredia. (**A**) Expi293F cells expressing QES variants containing mutations to subunit surfaces (magenta), or additional mutations to core residues (blue), bind more PG16 than wild type Env (grey). Env variants were tested from five HIV-1 strains…

Figure 3.5 (cont.) (from left to right: BaL, Q769.d22, Q842.d12, 25711, and DU422). Data are mean ± SD, n = 3-4. (**B**) Binding of CD4(D1-D2) to cells expressing Env variants. Inclusion of the V255M core mutation in constructs BaL-QES.i01.c01, Q769.d22-QES.i03.V255M, and DU422-QES.c03, reduces CD4 binding. Data are mean (n = 2), with error bars showing the range. (**C-E**) Binding of (**C**) PGT121, (**D**) PGT128, and (**E**) PGT145 to transfected cells expressing wild type or QES variant Env sequences. Data are mean (n = 2), with error bars showing the range. (**F**) Wild type and QES variants of Env were expressed with N-terminal c-myc tags. Total surface expression (grey; detected using fluorescent anti-myc staining) and PG16 binding (magenta; 2 nM PG16) to the same transfected cell samples were measured by flow cytometry. Mean fluorescence is normalized to cells expressing the respective wild type Env. Data are mean ± SD, n = 4. Expression of BaL Env with an N-terminal c-myc tag was not detected and is not plotted.

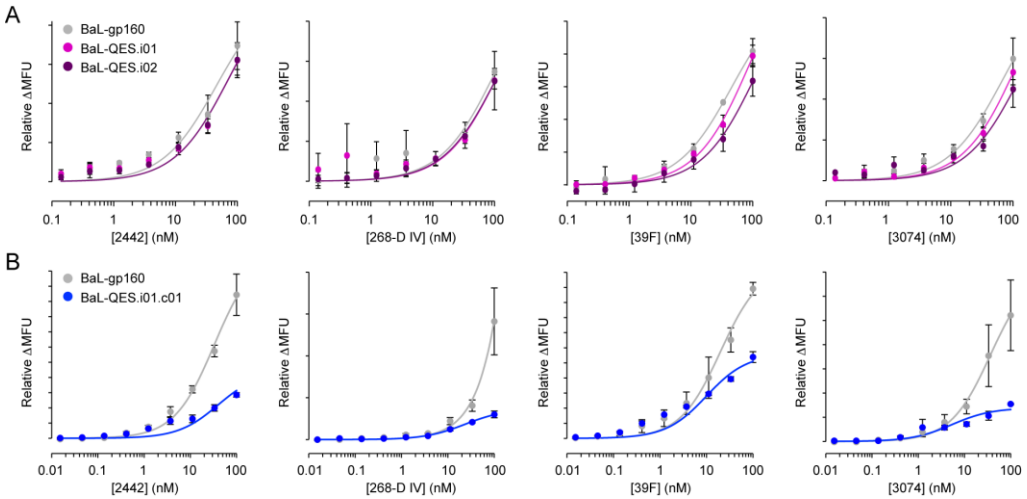

Figure 3.6. QES mutations at Env$_{BaL}$ subunit interfaces are insufficient to prevent recognition by V3-targeting antibodies. Data generated by Jeremiah Heredia. (**A**) Expi293F cells expressing wild type BaL (grey), BaL-QES.i01 (magenta), or BaL-QES.i02 gp160 (purple) were incubated with anti-V3 region monoclonals (from left to right) 2442, 268-D IV, 39F, and 3074. Bound antibody was detected by flow cytometry. (n = 3, mean ± SD) (**B**) Binding of V3-targeting antibodies to cells expressing wild type (grey) or QES.i01.c01 (blue; containing additional core mutations that limit CD4 interactions) BaL Env. Data are mean (n = 2), with error bars showing the range.

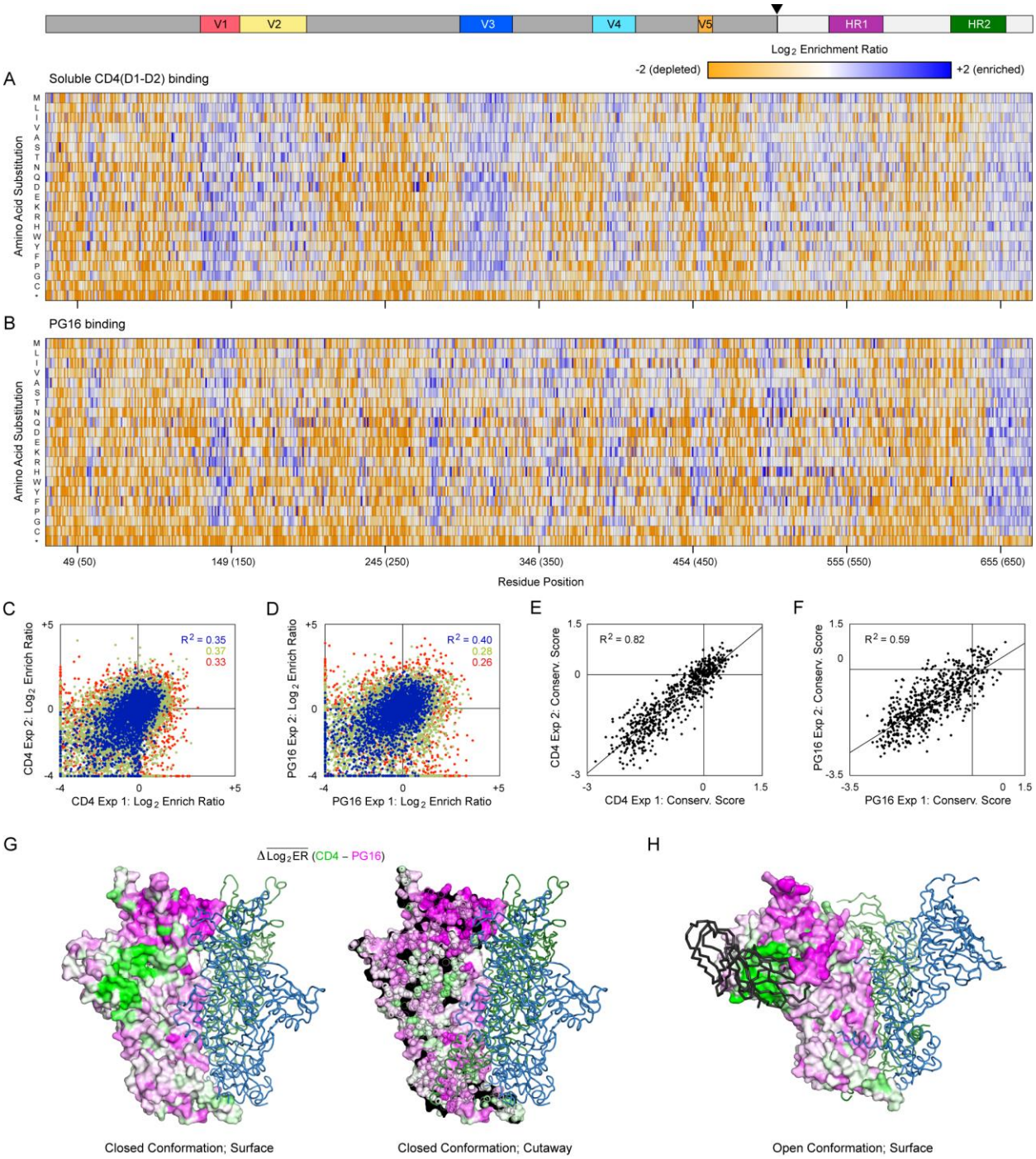

Figure 3.7. Residues at the Env$_{DU422}$ trimer interface are more conserved for PG16 binding than for CD4 interactions. Data generated by Jeremiah Heredia. (**A**) SSM libraries of membrane-anchored gp140$_{DU422}$ were expressed in Expi293F cells and sorted for binding to sCD4. Log_2 enrichment ratios are plotted from ≤ -2 (depleted, orange) to ≥ +2 (enriched, dark blue). The primary structure of gp140 is on the horizontal axis, and amino acid substitutions are on the vertical axis. *, stop codons. In the upper schematic, gp120 and gp41 are dark and pale grey, respectively, the cleavage site is indicated with an…

Figure 3.7 (cont.) arrowhead, and notable regions are colored. Average of two independent experiments. (**B**) The mutational landscape of $gp140_{DU422}$ under FACS-based selection for binding to PG16. (**C-D**) FACS-based selections for (**C**) 10 nM sCD4 and (**D**) 3 nM PG16 binding were independently replicated. Agreement between the replicate log_2 enrichment ratios for each mutation is plotted. Abundant mutations (frequencies > 2×10^{-4} in the naïve library) are blue, mutations with moderate representation (frequencies between 5×10^{-5} and 2×10^{-4}) are green, and rare mutations (frequencies < 5×10^{-5}) are red. (**E-F**) Agreement between the residue conservation scores from replicate selections for (**E**) sCD4 or (**F**) PG16 binding. (**G**) An atomic model of trimeric DU422 gp140 in the closed conformation, with one protomer shown as a surface, and the other protomers shown as dark green and blue ribbons. The PG16-CD4 conservation difference scores are mapped to the protomer surface in the same orientation as Figure 2, with magenta indicating residues preferentially conserved for PG16 binding, and green for residues more conserved for CD4 binding. On the right is a cross-section showing that preferential conservation for PG16 binding extends into the core of the trimerization domain. (**H**) DU422 gp140 modeled in the open state bound to CD4 (black ribbon; CD4 is shown bound to only a single protomer). Colored as described in (**G**).

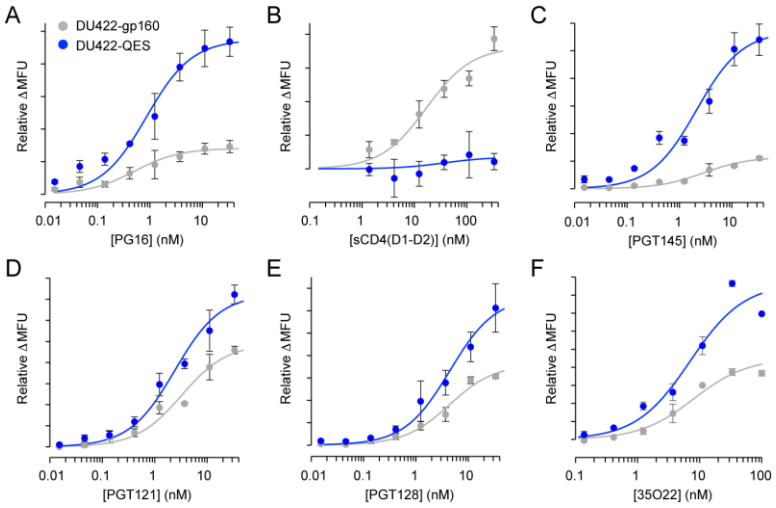

Figure 3.8. Antibody and CD4 binding properties of cells expressing a second generation engineered Env from the DU422 isolate. Data generated by Jeremiah Heredia. Soluble CD4 and bNAb binding to cells transiently expressing full-length Env from the DU422 isolate was measured by flow cytometry. Wild type Env is grey, and the engineered mutant (DU422-QES.c12 carrying seven mutations) is blue. (**A**) PG16, (**B**) sCD4 (D1-D2 domains), (**C**) PGT145, (**D**) PGT121, (**E**) PGT128, and (**F**) 35O22. Data are mean (n = 2), with error bars showing the range.

CHAPTER 4: ENGINEERED RECEPTORS FOR HUMAN CYTOMEGALOVIRUS THAT ARE ORTHOGONAL TO NORMAL HUMAN BIOLOGY[2]

ABSTRACT

A trimeric glycoprotein complex on the surface of human cytomegalovirus (HCMV) binds to platelet-derived growth factor (PDGF) receptor α (PDGFRα) to mediate host cell recognition and fusion of the viral and cellular membranes. Soluble PDGFRα potently neutralizes HCMV in tissue culture, and its potential use as an antiviral therapeutic has the benefit that any escape mutants will likely be attenuated. However, PDGFRα binds multiple PDGF ligands in the human body as part of developmental programs in embryogenesis and continuing through adulthood. Any therapies with soluble receptor therefore come with serious efficacy and safety concerns, especially for the treatment of congenital HCMV. Soluble virus receptors that are orthogonal to human biology might resolve these concerns. This engineering problem is solved by deep mutational scanning on the D2-D3 domains of PDGFRα to identify variants that maintain interactions with the HCMV glycoprotein trimer in the presence of competing PDGF ligands. Competition by PDGFs is conformation-dependent, whereas HCMV trimer binding is independent of proper D2-D3 conformation, and many mutations at the receptor-PDGF interface are suitable for functionally separating trimer from PDGF interactions. Purified soluble PDGFRα carrying a targeted mutation succeeded in displaying wild type affinity for HCMV trimer with a simultaneous loss of PDGF binding and neutralizes trimer-only and trimer/pentamer-expressing HCMV strains infecting fibroblasts or epithelial cells. Overall, this

[2] The research described in this chapter was a combined effort of myself with Jihye Park, Ali A. Aghajani, Jeremiah D. Heredia, Hannah Choi, Adam Oberstein, and Erik Procko. Data I was responsible for collecting are shown in Figures 3, 4 and 6.

work makes important progress in the realization of soluble HCMV receptors for clinical application.

AUTHOR SUMMARY

Human cytomegalovirus (HCMV) causes severe disease in transplant recipients and immunocompromised patients, and infections in a fetus or neonate are responsible for life-long neurological defects. Cell entry is in part mediated by a trimeric glycoprotein complex on the viral surface, which binds tightly to the host receptor PDGFRα. The soluble extracellular region of PDGFRα can be used as an antiviral agent to potently neutralize the virus *in vitro*. However, PDGFRα ordinarily binds growth factors in the human body to regulate developmental programs, which will limit the *in vivo* efficacy and safety of soluble PDGFRα. Using saturation mutagenesis and selections in human cell culture, mutations in PDGFRα are identified that eliminate off-target growth factor interactions while preserving HCMV binding and neutralization.

INTRODUCTION

Human cytomegalovirus (HCMV; or human herpesvirus 5, HHV-5) is a ubiquitous pathogen that has infected most of the global adult population, and typically causes asymptomatic infections[51,155]. In some cases, individuals will experience mononucleosis-like symptoms, and severe disease can occur in immunocompromised or immunosuppressed adults [156]. HCMV infection may also have an oncomodulatory effect and could be associated with tumor progression, most notably in glioblastoma[157–160], although the mechanistic details remain controversial[161]. However, it is amongst pregnant women and newborns that HCMV has a

disproportionate impact on public health[162]. Just over 1 in 200 infants born in the United States are congenitally infected with HCMV, and approximately a fifth of these infants will suffer life-long neurological complications[163,164], including hearing and vision loss, seizures, behavioral disorders, and developmental delays[165–168]. Ideally, infection during pregnancy would be avoided, but this is challenging. The high prevalence of HCMV (especially in families with young children[169,170]), easy transmission through contact with bodily fluids, recurrence of latent endogenous infections[171], reinfection by different exogenous strains [172], and difficulties recognizing asymptomatic infected individuals make controlling virus spread difficult.

Furthermore, antivirals have not been approved by the U.S. Food and Drug Administration for routine use in the treatment of congenital HCMV where there are unique concerns regarding toxicity, and drug resistance is widely reported[173]. Faced with this reality, the American College of Obstetricians and Gynecologists does not make any specific recommendations for counseling or treating pregnant women for the prevention of HCMV[174].

It is in this environment that the discovery of an important HCMV receptor, the platelet-derived growth factor receptor α (PDGFRα), has generated excitement towards the development of new therapeutics inhibiting virus-host cell attachment and entry. Two glycoprotein complexes on the HCMV surface drive broad host cell tropism[52]. A pentameric complex of glycoprotein H (gH; UL75), gL (UL115), UL128, UL130 and UL131A mediates entry into epithelial and endothelial cells[175–180], possibly through engagement of neuropilin-2[53] or the olfactory receptor OR14I1[181] on the host cell. A trimeric complex of gH, gL and gO (UL74) is sufficient for entry into fibroblasts and is also necessary for entry into pentamer-requiring cell types by promoting

membrane fusion[182–184]. Multiple lines of evidence indicate that PDGFRα is the receptor for the trimer. Decreased PDGFRα expression reduces HCMV entry into fibroblasts[185–188]; gO can specifically bind PDGFRα-expressing cells[187]; a quaternary complex of soluble extracellular regions (gH-gL-gO-PDGFRα) can be assembled *in vitro* and the components bind with high nanomolar affinity[53,185]; HCMV strains lacking the trimer fail to enter PDGFRα-dependent cell types to infect cells[55,187]; and the soluble PDGFRα extracellular domain blocks virus entry[53,55,187]. This has led to the soluble domain of PDGFRα or derivative fragments being explored as antivirals[55,185,187], a promising strategy as any mutations in HCMV to escape receptor-based inhibitors would likely decrease binding to the natural receptor and attenuate virulence. Soluble PDGFRα ectodomain effectively blocks virus entry whether applied pre- or post-cell attachment, due to inhibition of both early attachment and fusion steps[55]. Furthermore, soluble PDGFRα inhibits cell entry when only a fraction of glycoprotein trimer is bound and therefore does not require high saturating concentrations; and soluble PDGFRα blocks entry by multiple HCMV strains that express both trimer and pentamer complexes[55].

While appealing, the use of soluble PDGFRα to treat HCMV in humans comes with risk. Indeed, no protein-based viral receptors are currently approved as antiviral drugs by the FDA[189] (although at least one is in a phase II clinical trial for the treatment of COVID-19 by Apeiron Biologics), in part due to safety concerns over interactions with endogenous ligands. By comparison, there are multiple antibody drugs that are specific for viral targets used in the clinic[189]. PDGFRα is an important receptor for platelet-derived growth factors (PDGFs), and regulates cellular proliferation, differentiation and development of multiple tissues during embryogenesis and onwards through adulthood[190]. The PDGF ligands are small polypeptides that

covalently associate in disulfide-bonded homo- and heterodimers, which display differing activities towards PDGFRα and its close relative PDGFRβ[191]. Four ligands interact with PDGFRα with high affinity (PDGF-AA, AB, BB, and CC)[192,193] and bind two receptor chains at opposing ends to form a receptor-$(ligand)_2$-receptor signaling complex[56,194]. PDGFs compete with the HCMV trimer for PDGFRα binding and block infection[186,187], suggesting that PDGF binding to PDGFRα sterically hinders trimer binding or allosterically modulates the trimer binding interface in a non-productive manner. Treatment with soluble PDGFRα will almost certainly disrupt physiological PDGF-signaling, and the ability of recombinant receptor to bind virions may be diminished by competing endogenous PDGFs. Ideally, mutations in the receptor can be identified that disrupt PDGF interactions while retaining the site or sites engaged by HCMV.

A high resolution crystal structure of PDGF-BB-bound PDGFRβ has illuminated atomic details of ligand-receptor interactions in the family, revealing a predominantly hydrophobic interface at the cleft between the second (D2) and third (D3) extracellular domains of the receptor[56]. However, the binding site for gO is ambiguous. Cryo-electron microscopy suffers from conformational diversity and low resolution[185]; PDGFRα peptide fragment analysis failed to discover any one peptide solely responsible for high affinity binding[55]; and while domain deletions have shown PDGFRα-D3 is important[188], such studies have coarse resolution. The absence of detailed structural information on the trimer-PDGFRα complex means the engineering of HCMV-specific receptors orthogonal to human biology (i.e. receptors that bind virus but do not interact with endogenous human ligands) remains an unsolved challenge.

This challenge is solved through the use of mutagenesis and *in vitro* selection. By tracking the enrichment or depletion of sequence variants in a diverse library using next generation sequencing, the relative phenotypes of thousands of mutations can be simultaneously assessed in a single experiment, referred to as deep mutagenesis[4]. Deep mutational scans have been used to address and engineer specificity in proteins that promiscuously bind multiple ligands[195–197], and methods for deep mutagenesis of membrane proteins expressed in human cells have been optimized[63,74,80]. By screening through all possible single amino acid substitutions in the D2-D3 domains, we discovered that, unlike PDGF binding, HCMV trimer interactions persist when PDGFRα conformation is disrupted by mutations within the folded core. Particularly relevant to therapeutic design, we also identified PDGFRα surface mutations that retain trimer binding in the presence of high PDGF concentrations. This study brings receptor-based inhibition of HCMV infection closer to realization.\

RESULTS

A deep mutational scan of PDGFRα D2-D3 domains for trimer binding in the presence of PDGFs

The glycoprotein trimer from HCMV strain TB40/E (BAC4 clone, [53]) was expressed in Expi293F cells, a suspension derivative of human HEK 293. The transmembrane helix of gH was deleted to produce a soluble complex, and the gO subunit, which directly contacts PDGFRα[185,187,198], was fused at its C-terminus to superfolder GFP (sfGFP[199]) for fluorescence detection. Medium from gH/gL/gO-sfGFP expressing cells was incubated with Expi293F cells transfected with an episomal plasmid encoding PDGFRα. Trimer binding was proportional to PDGFRα surface expression (determined by flow cytometry using antibody staining of an

extracellular c-myc epitope tag at the PDGFRα N-terminus; Figure 1A and 1B), and was inhibited by co-incubation with an equimolar mixture of PDGF-AA, AB, BB, and CC (Figure 1D). Medium from cells expressing gO-sfGFP alone, either the wild-type sequence or carrying the mutation C351S (C343S based on numbering of gO from the TB40/E strain) to remove the exposed cysteine that would ordinarily form a disulfide to gL-C144[200], failed to display high binding signal to PDGFRα positive cells. The binding signal is therefore dependent on co-expression with gH and gL, suggestive that it is trimer as opposed to monomeric gO that engages the receptor, although this is not explicitly demonstrated.

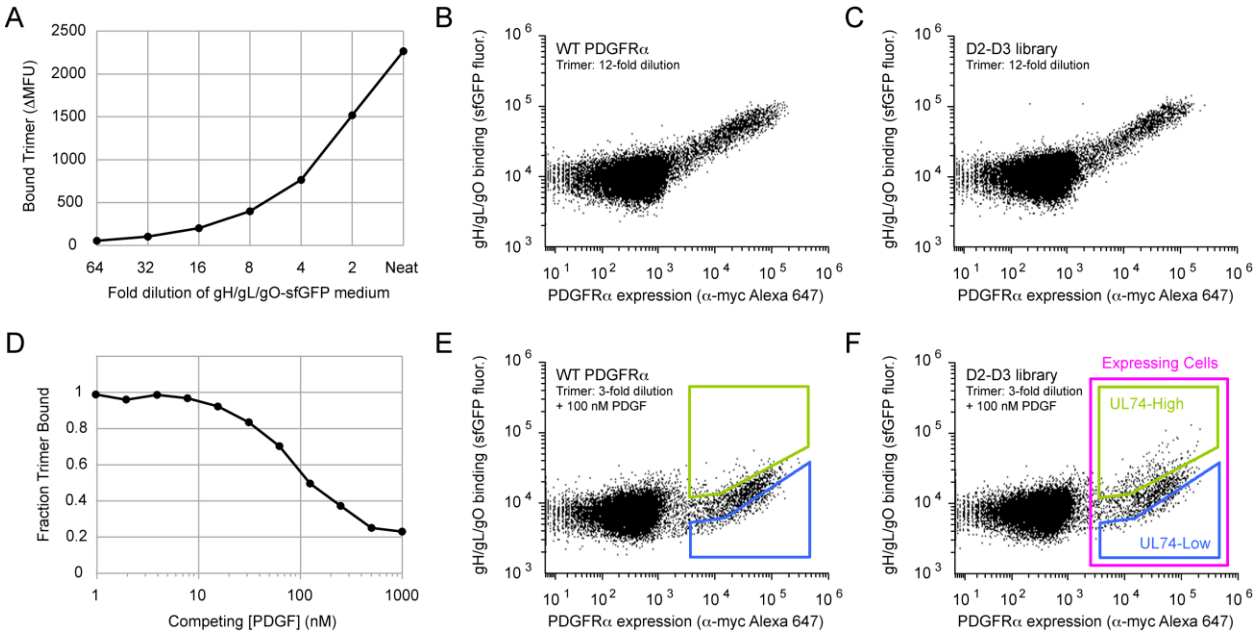

Figure 4.1. A selection strategy for identifying PDGFRα mutants selective for the HCMV trimer. Data generated by Jihye Park. (A) Expi293F cells expressing myc-tagged PDGFRα were incubated with dilutions of gH/gL/gO-sfGFP-containing medium, and binding (shown as ΔMean Fluorescence Units, ΔMFU, in which Mean Fluorescence Units of a negative control without trimer are subtracted as background) was measured by flow cytometry. Neat, no dilution. **(B and C)** Cells were transfected with wild-type (WT) PDGFRα (B) or the D2-D3 SSM library (C) under conditions where typically no more than a single sequence variant is expressed by any given cell. Under these conditions, most cells are negative for PDGFRα expression as determined by anti-myc staining. Cells were incubated with a subsaturating (12-fold dilution) of gH/gL/gO-sfGFP medium and analyzed by flow cytometry…

Figure 4.1 (cont.) (D) Cells were co-incubated with a 2-fold dilution of gH/gL/gO-sfGFP medium and different concentrations of PDGF (an equimolar mixture of PDGF-AA, AB, BB, and CC). The fluorescent signal of gH/gL/gO-sfGFP binding is compared to a sample without competing PDGF. **(E and F)** Cells expressing WT PDGFRα (E) or the D2-D3 SSM library (F) were incubated with a 3-fold dilution of gH/gL/gO-sfGFP medium and 100 nM PDGF (25 nM of each subtype). The choice of a PDGF concentration that does not completely inhibit trimer binding ensured (i) that there remained some positive binding signal for drawing appropriate sort gates, and (ii) that mutations with weak effects will be distinguished from the main population. During sorting of the D2-D3 library, the top 15% of PDGFRα-expressing cells that bind trimer were collected (UL74-High sort, green gate). Simultaneously, the bottom 15% of PDGFRα-expressing cells that bind trimer were also collected (UL74-Low sort, blue gate). For cells expressing WT PDGFRα, the equivalent gates correspond to 4% and 23% of PDGFRα-expressing cells, respectively.

A single site-saturation mutagenesis (SSM) library of PDGFRα was constructed encompassing all single amino acid mutations in the D2-D3 domains (a.a. D123-E311), and transfected into Expi293F cells under conditions that typically yield no more than one sequence variant per cell, providing a tight link between genotype and phenotype[63]. When incubated with gH/gL/gO-sfGFP medium at subsaturating dilutions, cells expressing the SSM library were surprisingly indistinguishable from cells expressing wild-type PDGFRα (Figure 1B and 1C). Normally, many mutations adversely impact protein activity through destabilization of folded structure or damage to functional sites, and loss-of-function variants tend to dominate naive libraries prior to any selection. That deleterious mutations are not prevalent in the naïve SSM library immediately implied that HCMV trimer binding is resistant to most single non-synonymous mutations in the PDGFRα D2-D3 domains.

By comparison, many cells expressing the SSM library displayed higher trimer binding in the presence of PDGFs than cells expressing wild-type PDGFRα (Figure 1E and 1F). There thus

appeared to be many PDGFRα mutations that selectively lost PDGF affinity. To identify these mutations, PDGFRα-expressing cells that had elevated HCMV trimer binding in the presence of competing PDGFs were collected by fluorescence-activated cell sorting (FACS). This is referred to as the UL74-High sort (see green gate in Figure 1F). Within the same experiment, cells expressing PDGFRα but displaying low trimer binding were also collected, referred to as the UL74-Low sort (see blue gate in Figure 1F). The UL74 gene encodes gO[182], and the names assigned to the sorted populations correspond to the raw and processed data files deposited in the GEO database[201]. PDGFRα mutants that lose affinity for trimer, or perhaps have enhanced affinity for competing PDGFs, will be preferentially enriched in the UL74-Low sort. PDGFRα mutants that fail to express will be depleted from both sorted populations. Following Illumina sequencing of the naive plasmid library and transcripts from the sorted populations, the enrichment ratios for all 3,780 substitutions in the D2-D3 domains were calculated to define a local mutational landscape. Data from independent replicates are highly correlated, giving confidence in the data's accuracy.

PDGF interactions are uniquely sensitive to mutational disruption of folded structure in the D2-D3 domains

Enrichment ratios for nonsynonymous mutations are anticorrelated between the UL74-High and UL74-Low sorts, with few mutations other than premature stop codons being depleted in both selections. Hence substitution mutations in PDGFRα D2-D3 domains suffer little cost to surface localization, even though many mutations will almost certainly destabilize the domain fold. This differs markedly from a prior mutational scan of a multidomain membrane protein expressed in human cells, where nonconservative mutations buried within folded structure prevented protein

trafficking to the plasma membrane due to intracellular retention by quality control machinery[80]. When mapping experimental conservation scores to the structure of PDGF-bound PDGFRα (modeled from the crystal structure of PDGF-BB-bound PDGFRβ; PDB ID 3MJG[56]), enriched mutations in the UL74-High sort are found to cluster to buried core positions and the PDGF interface (Figure 2). These include substitutions that create cavities, steric clashes, and/or introduce buried ionizable groups; these mutations are incompatible with biophysical principles governing protein folding and will destabilize folded structure. We hypothesize that PDGFRα mutants with disrupted structure have therefore preferentially lost affinity to PDGFs, thereby reducing competition and enhancing binding of the HCMV trimer. The epitope bound by the HCMV trimer is ambiguous from the data but must be at least partially independent of proper D2-D3 conformation, although partial structure may remain. This is consistent with either key contacts to the HCMV trimer residing on linear PDGFRα segments and loops, similar to how many antibodies recognize conformation-independent epitopes, or with favorable interactions to the HCMV trimer being distributed over a broad surface (possibly even beyond the D2-D3 domains) such that localized disruptions are tolerated. Libraries encompassing double or higher order PDGFRα mutations may further resolve details of the binding mechanism. While this illuminates aspects of PDGFRα/trimer recognition, using misfolded, destabilized PDGFRα variants to selectively neutralize HCMV is fraught with risks, including toxicity from non-specific interactions and protein aggregation due to the exposure of hydrophobic residues.

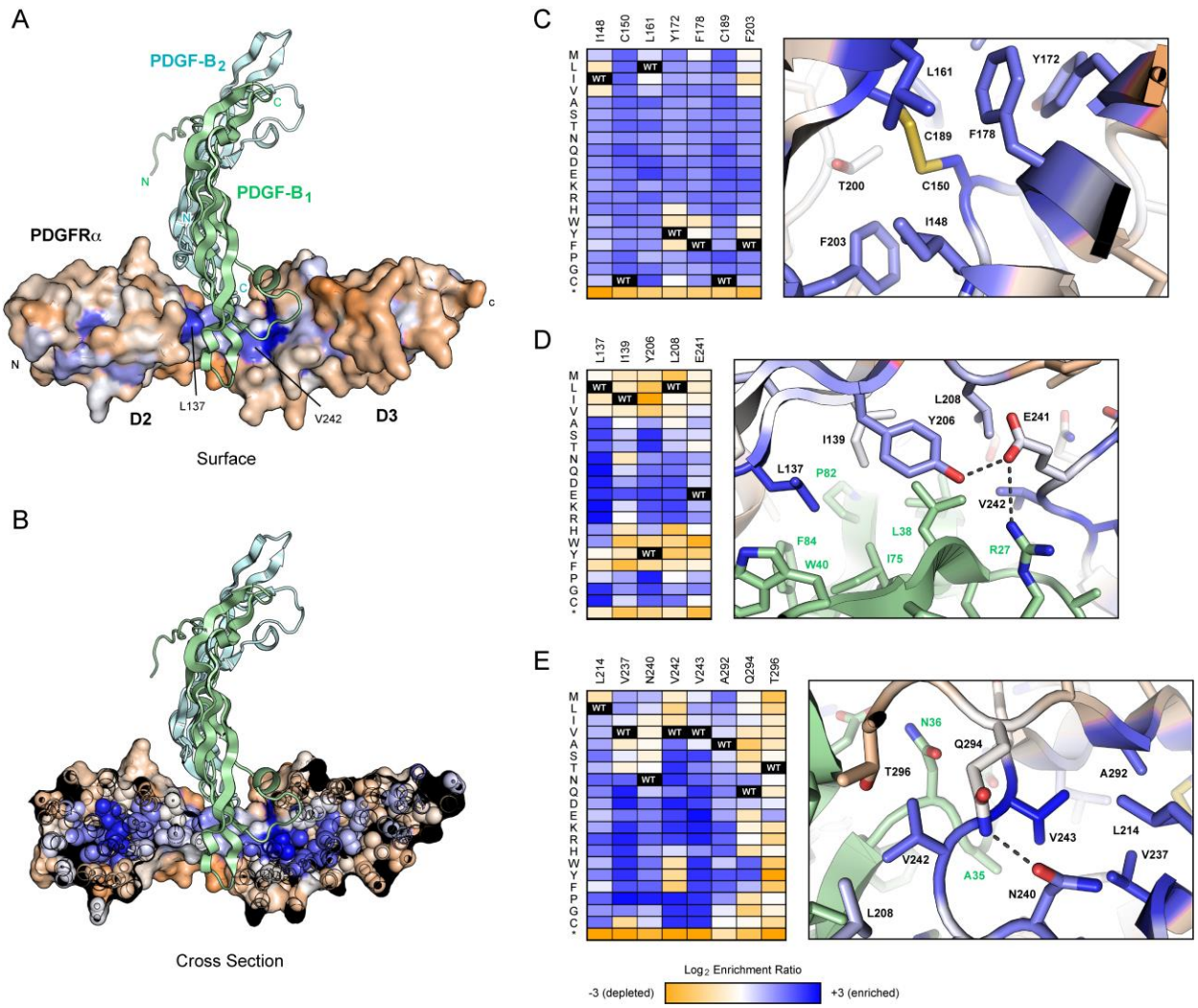

Figure 4.2. Hot spots in PDGFRα for mutations that increase HCMV trimer binding in the presence of competing PDGFs. Data generated by Jihye Park. (A) Model of PDGFRα D2-D3 domains bound to PDGF-BB (protruding subunit B_1 is pale green, receding subunit B_2 is pale blue). Conservation scores from the UL74-High sort are mapped to the PDGFRα surface, showing residues where mutations tend to be depleted in orange, and residues where mutations tend to be enriched in dark blue. **(B)** A cross section through PDGFRα highlights that enriched mutations (dark blue) are concentrated in the protein interior/core. **(C-E)** Log_2 enrichment ratios for individual mutations from the UL74-High sort are plotted from -3 (orange, depleted) to +3 (dark blue, enriched). Amino acid substitutions are indicated on the vertical axis, * is stop codon. The wild-type amino acid is black. Structural views are shown at right, colored as in panel A. Regions shown are the D2 core (C), PDGF-BB interface (D), and PDGF-BB interface and underlying D3 core (E).

Surface mutations at the PDGF binding site favor PDGFRα specificity towards the HCMV trimer

The PDGFRα mutational landscape for high trimer binding in the presence of competing PDGF reveals key surface residues as hot spots for enriched mutations, especially PDGFRα residues L137, L208 and V242 that contact the protruding PDGF subunit, and Y206 which packs between the protruding and receding PDGF subunits (Figures 2D and 2E). Mutations to these residues will disrupt the PDGFRα-PDGF interface, especially by the addition of polar substitutions that invert the chemical properties of the highly hydrophobic binding site, while HCMV trimer interactions persist. Overall, enrichment of mutations for increased HCMV trimer binding in the presence of competing PDGFs is highly correlated to the mutated residue's connectivity within the D2-D3 core or across the PDGFRα-PDGF interface, emphasizing the selective importance of PDGFRα conformation and surface contacts in the D2-D3 cleft for high affinity PDGF interactions.

Five mutations on the PDGFRα surface (L137K, L137Q and Y206S in D2, and V242K and V242T in D3, chosen based on their high enrichment in the UL74-High sort and covering multiple surface positions) were validated as having desirable binding properties by targeted mutagenesis. As predicted from the deep mutational scan, these PDGFRα variants maintain near wild-type levels of binding to soluble HCMV trimer from the TB40/E strain, yet have diminished sensitivity to the addition of competing PDGFs (Figures 3A and 3B), demonstrating successful engineering of specificity towards the viral target. Since the mechanism of trimer binding to receptor has not been resolved at atomic or even residue-level resolution, it remains uncertain whether the viral genome could mutate to distinguish wild-type from engineered

PDGFRα. This concern is partially alleviated by the observation that the engineered PDGFRα mutants maintain high binding to soluble HCMV trimer from the Merlin clinical strain[202] (Figure 3C), which has relatively high natural sequence variation compared to TB40/E (79 % identity between glycoproteins O). Variation is particularly high in the N-terminus of glycoprotein O where important interactions to PDGFRα reside[198].

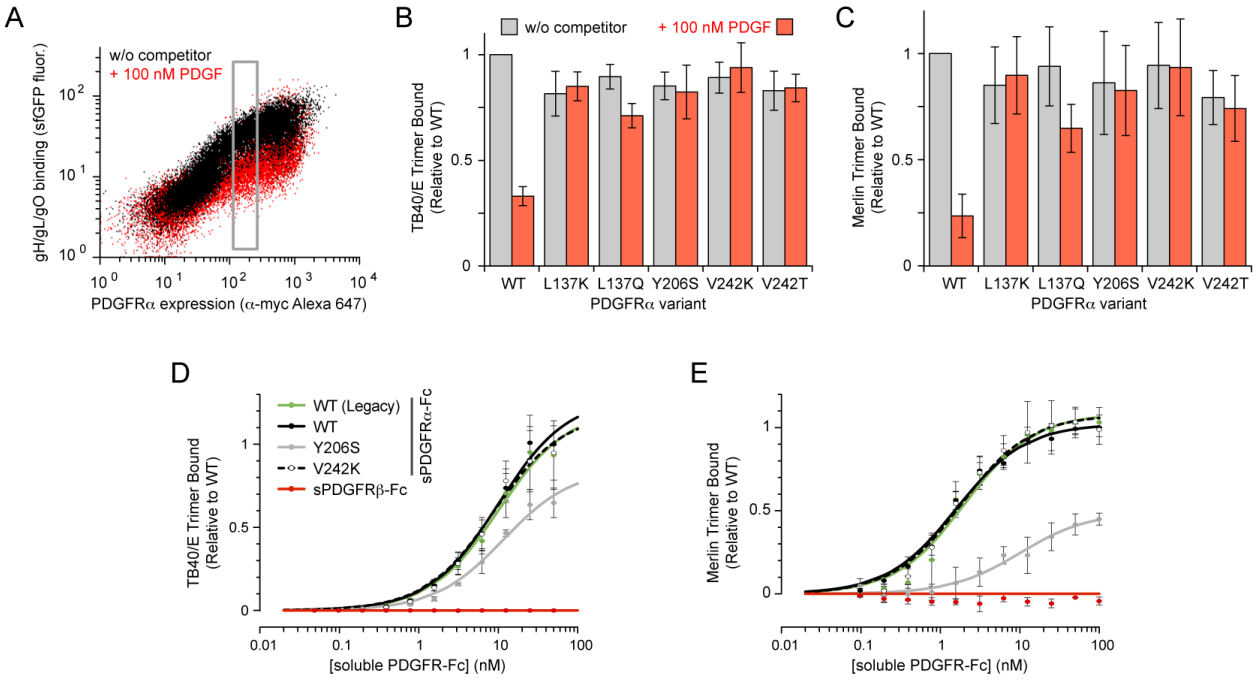

Figure 4.3. Validation of targeted mutants in PDGFRα that specifically block PDGF competition. Data generated by Kevin Gill. (A) Expi293F cells expressing wild-type PDGFRα were co-incubated without (black) or with (red) competing PDGFs, and binding of soluble TB40/E trimer (1/3 dilution of medium containing gH/gL/gO-sfGFP) was assessed by flow cytometry. Cells were gated (grey box) to compare binding (measured as ΔMFU) at equivalent receptor expression levels. **(B)** Binding of soluble TB40/E trimer to PDGFR□ variants. Cells were analyzed as outlined in panel A. Data are mean ± SD, n = 3. **(C)** Cells expressing PDGFRα variants were incubated with Merlin gH/gL/gO-sfGFP containing medium (1/3 dilution) in the absence (grey) or presence of PDGFs (red). Cells were analyzed as outlined in panel A. Data are mean ± SD, n = 4. **(D)** Expi293F cells were transfected with gL, gO, and full-length gH from strain TB40/E. Binding of sPDGFR-Fc to HCMV trimer anchored at the cell surface was measured by flow cytometry. Data are mean ± SD, n = 2 (sPDGFRα-Fc), 3 (WT-Legacy and Y206S), 4 (V242K), or 5 (WT). **(E)** Binding of sPDGFR-Fc to cells expressing membrane-anchored trimer from…

Figure 4.3 (cont.) HCMV strain Merlin. Data are mean ± SD, n = 2 (sPDGFRα-Fc), or 3 (WT-Legacy, WT, Y206S, and V242K).

Fusions of soluble proteins to immunoglobulin Fc confers multiple desirable properties for a therapeutic[203]. Multimerized Fc chains impose avidity for enhanced apparent affinity, and Fc moieties are engaged by multiple receptors to enhance serum half-life or evoke effector functions, including complement activation and antibody-dependent cell-mediated cytotoxicity[203]. Soluble PDGFRα fused to the Fc region of human IgG1 was shown to bind HCMV trimer when first identified as its receptor[185], and the neutralization properties of Fc-fused soluble PDGFRα (sPDGFRα-Fc) have been explored since in greater detail[55]. These prior studies used commercially supplied sPDGFRα-Fc featuring a random linker of mixed amino acids that connects to the upper hinge of IgG1, upstream of Cys-220 that would ordinarily form a disulfide to the antibody light chain. The nature of this Fc fusion may cause manufacturing liabilities if Cys-220 is exposed and free, and we refer to this construct as "legacy" sPDGFRα-Fc. We designed an alternative sPDGFRα-Fc construct that includes the PDGFRα signal peptide (a.a. 1-23), ectodomain (a.a. 24-524), a GGGS linker, and finally human IgG1 Fc beginning at Asp-221. Both sPDGFRα-Fc constructs bind membrane-localized HCMV trimer with equal affinity (Figures 3D and 3E), and subsequent studies focused exclusively on our redesigned fusion protein.

Two PDGFRα mutations, Y206S and V242K, were explored as Fc-fused soluble orthogonal receptors. Binding of the soluble receptors was measured towards HCMV trimer expressed at the cell surface, with gL and gO subunits tagged at their C-termini with short peptide epitopes, and gH expressed as full-length protein with its native transmembrane domain. The Y206S mutant

had reduced binding to HCMV trimer compared to wild-type receptor, but sPDGFRα-Fc V242K bound HCMV trimer from Merlin and TB40/E strains with equal low nanomolar affinity (Figures 3D and 3E). Furthermore, fusions of wild-type and V242K sPDGFRα with the Fc region of IgG3 also bound similarly to trimer from the TB40/E and Merlin strains (Figure 4). Compared to IgG1, IgG3 has lower affinity for the neonatal Fc receptor (FcRn), which is associated with reduced serum half-life and placental transfer[204,205]. We speculate that these features may be advantageous in the treatment of pregnant women during acute HCMV infection to achieve a lower dose in the developing fetus, thereby further addressing safety concerns.

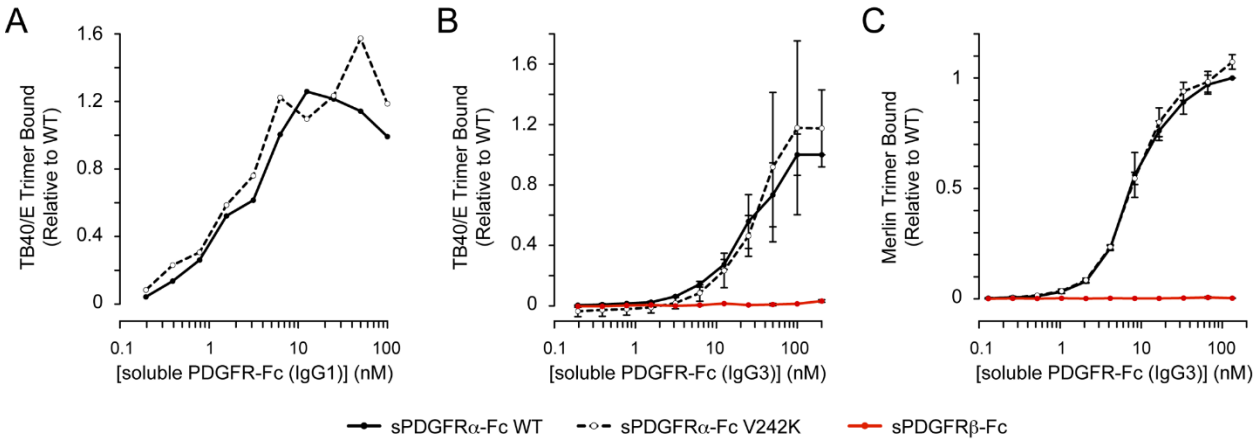

Figure 4.4. Soluble PDGFRα-Fc V242K binds HCMV trimer with comparable affinity to wild-type sPDGFRα-Fc. Data generated by Jihye Park and Kevin Gill.

(A) Data presented in Fig 3D was replicated using independent preparations of sPDGFRα WT (solid black line) and V242K (broken black line) fused to the Fc region of IgG1. Binding to Expi293F cells expressing full-length gH, gL and gO from the HCMV TB40/E strain was assessed by flow cytometry. **(B and C)** Soluble PDGFRα WT (solid black line) and V242K (broken black line) were purified as fusions to the Fc region of IgG3, matching the redesigned linker described in S5A Fig. Binding to trimer from (B) TB40/E and (C) Merlin strains expressed on Expi293F cells was measured by flow cytometry. Data are mean ± SD, n = 3 (sPDGFRα-Fc WT and V242K) or 2 (sPDGFRβ-Fc).

To further validate that the engineered receptors have lost growth factor interactions, we assessed their ability to inhibit PDGF signaling. Expi293F cells were transfected with a reporter plasmid

encoding wild-type PDGFRα and a destabilized fluorescent protein (EGFP-PEST)[206] under the control of a serum response element (SRE)[207]. Treatment with PDGFs upregulates EGFP-PEST expression, which is inhibited by wild-type sPDGFRα-Fc acting as a decoy receptor (Figure 5). As intended, both sPDGFRα-Fc Y206S and V242K failed to inhibit PDGF signaling (Figure 5B). However, we also noticed that the engineered receptors, in particular sPDGFRα-Fc Y206S, unexpectedly enhanced the response to PDGF-A ligands. We suspect this is due to a generic 'carrier protein' effect (PDGF ligands and their receptors are highly hydrophobic[56] and are generally formulated with a carrier such as serum albumin), though we are unable to conclude this with certainty, and non-specific gamma globulin has no such effect. This observation may require future investigation.

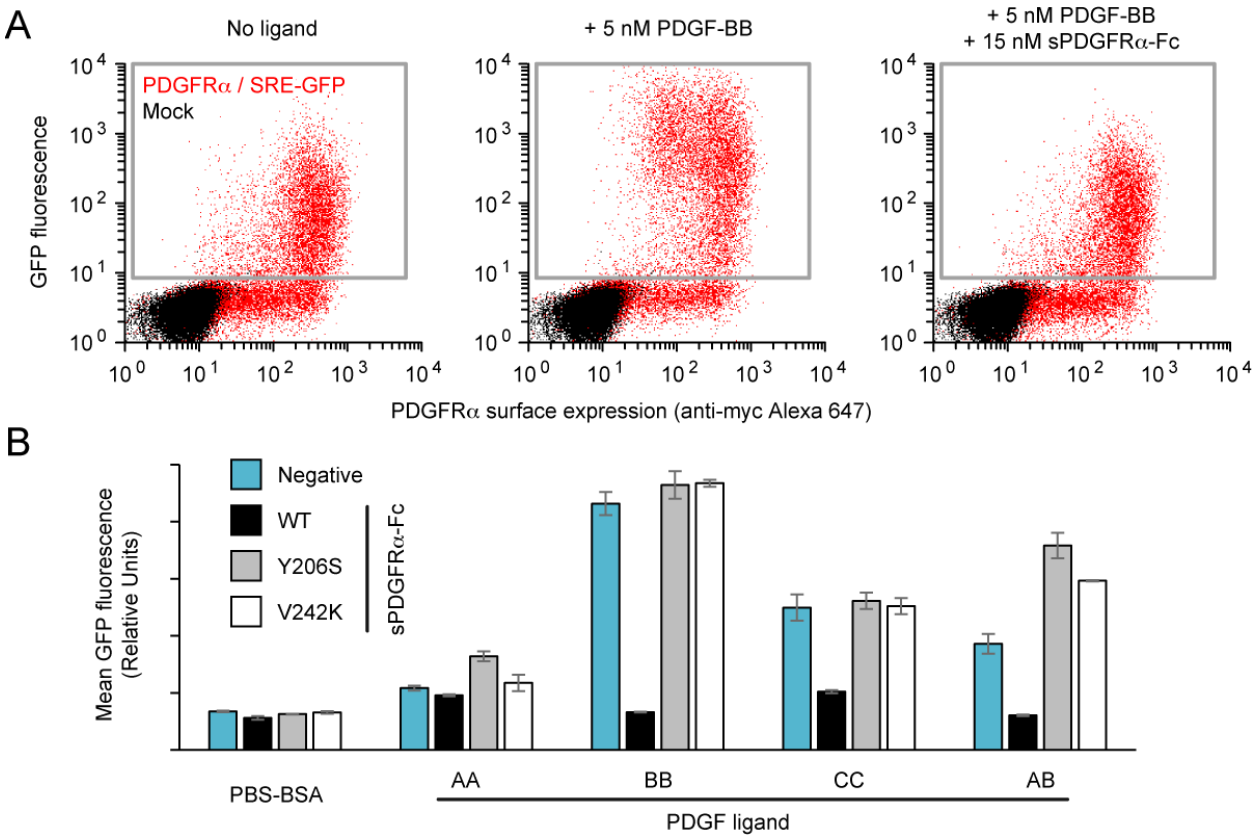

Figure 4.5. Soluble PDGFRα variants that no longer inhibit PDGF signaling. Data generated by Jihye Park. (A) Flow cytometry analysis of mock-transfected Expi293F cells (black), or cells transfected with a plasmid driving myc-tagged PDGFRα expression and carrying a SRE-regulated GFP reporter (red). Signaling activity was quantified by measuring mean GFP fluorescence (vertical axis) of responsive cells in the grey gate. Background PDGFRα signaling is at left, PDGF-BB stimulated activity is center, while sPDGFRα-Fc is shown to inhibit PDGF-BB stimulation at right. (B) Stimulation of transfected Expi293F cells with 5 nM PDGF ligands. PDGF ligands were pre-incubated with 0.2% BSA (negative, blue) or with a 3-fold molar excess of wild-type sPDGFRα-Fc (black), sPDGFRα-Fc Y206S (grey), or sPDGFRα-Fc V242K (white). Data are mean ± SD, n = 3 independent replicates.

Cross-Species interactions allow for testing of biocompatibility and pharmacokinetics in mice

To begin in vivo testing of any decoy receptors, the animal model must be suitable for evaluating the protein's pharmacology. Mice are a common model organism but CMV is highly species

specific, resulting in a significant hurdle. As such, we cannot test the human sPDGFRα-Fc decoys for their efficacy against HCMV virulence and infectivity in mice, as neither HCMV infects mice and nor is murine CMV sensitive to soluble PDGFRα. However, we can still use mice to evaluate general safety/tolerability and pharmacokinetics (PK) of the decoy receptors. A preliminary test has shown that, despite HCMV failing to replicate in mice, HCMV trimer taken from both Merlin and TB-40E clinical strains binds to mouse PDGFRα with a similar affinity to human PDGFRα, if slightly diminished (Figure 6A). Furthermore, human and mouse PDGFRα can be stimulated similarly by both human and mouse PDGFs, as most noticeably seen with PDGF-BB from both species, and we see that human and mouse PDGFRα have conserved binding sites that allow for cross species activity (Figure 6B). We then tested the decoy receptors for their ability, or lack thereof, to compete off binding of mouse PDGFs (Figure 6C). Human soluble decoy receptor was co-incubated with one of three mouse PDGF dimers and Expi-293F cells transfected with mouse PDGFRα and a GFP reporter. The wild type decoy prevented binding of mouse PDGFs to PDGFRα, resulting in signaling similar to the negative PDGF controls. sPDGFRα-V242K-Fc did not noticeably inhibit the signaling of the mouse PDGFRα, mimicking the negative decoy receptor controls. Mouse PDGFRα behaves much like human PDGFRα in terms of binding sites and activities. Thus, while mice are not infected with HCMV, the murine and human PDGF ligands and receptors all cross-react, which provides an opportunity to test the tolerability and PK of soluble decoy receptors in mice with the expectation that the data will be relevant to humans.

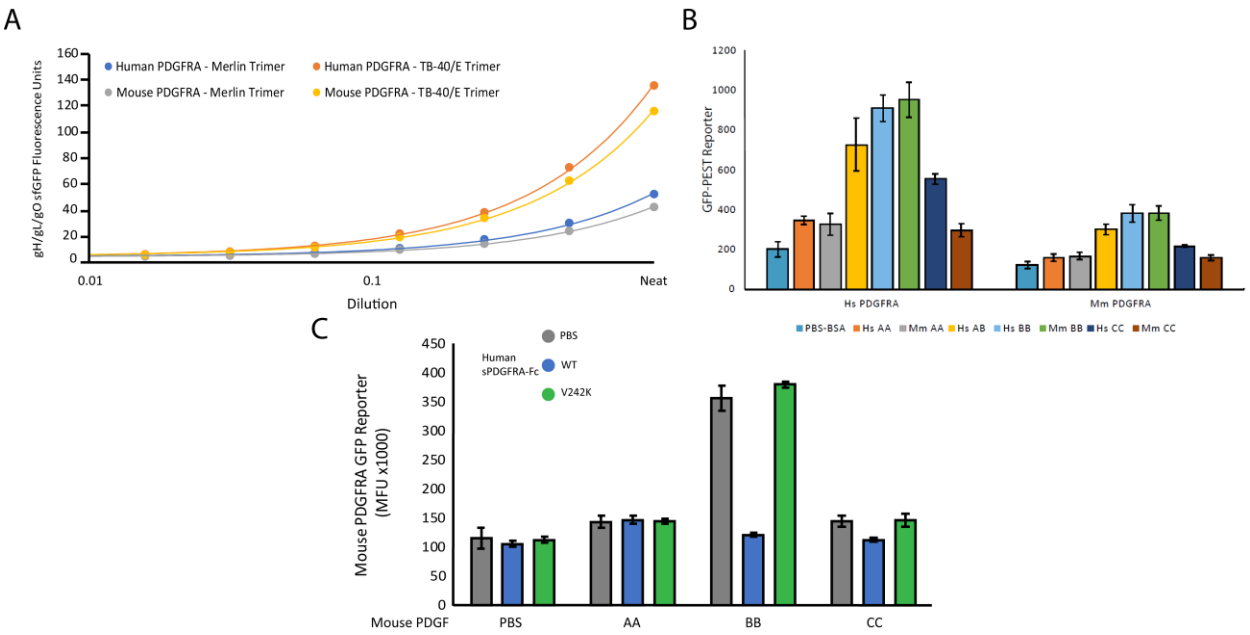

Figure 4.6. Mouse and Human PDGFRα bind similarly to PDGFs and HCMV. Data generated by Kevin Gill. (A) Titrations of HCMV trimers for both Merlin and TB-40E clinical strains testing binding to either human or mouse PDGFRα. n = 1 independent replicate. (B) Stimulation of human and mouse PDGFRα transfected Expi293F cells with 5 nM human or mouse PDGF ligands. Data are mean ± SD, n = 3 independent replicates. (C) Stimulation of mouse PDGFRα transfected Expi293F cells with 5 nM mouse PDGF ligands. PDGF ligands were pre-incubated with 0.2% BSA (negative, grey) or with a 3-fold molar excess of wild-type human sPDGFRα-Fc (blue), or sPDGFRα-Fc V242K (green). Data are mean ± SD, n = 3 independent replicates.

DISCUSSION

Ideas surrounding the use of soluble virus receptors as decoys to treat or prevent infection have a long history[208–211]. Soluble PDGFRα as a prophylactic or treatment for HCMV is particularly promising due to exceptionally tight affinity for glycoprotein trimer and its ability to neutralize both trimer- and pentamer-mediated host cell entry. Here, we have shown how deep mutagenesis can inform the engineering of PDGFRα variants that maintain tight virus binding but have lost growth factor interactions and are thereby orthogonal to normal human signaling pathways. This is anticipated to improve both efficacy and safety *in vivo*, especially for the treatment of pregnant women or neonates.

There are no simple, tractable animal models for investigating HCMV infection. Two highly relevant models are the study of rhCMV in rhesus macaques[212] or the engraftment of human tissues into immune-deficient mice[213–215], but these are non-trivial and require specialized expertise. However, it might be possible to test biocompatibility and pharmacokinetics of the orthogonal receptors independent of infection in small animal models. For instance, murine and human PDGF ligands and receptors share high identity (> 85 %, with all key binding site residues of PDGFRα conserved), and one standard assessment of purified human PDGF ligands is to measure mitogenic activity towards mouse fibroblast lines[216,217]. This suggests the human and mouse PDGF ligands and receptors will cross-react, though this has not been quantitatively characterized. Mice, especially strains expressing human FcRn for human-like serum stability of Fc fused protein[218,219], may therefore be a suitable model for assessing safety during future development.

The binding site for HCMV gO on PDGFRα is uncertain from the deep mutational scan, however the data unambiguously shows that binding is independent of properly folded structure in the PDGFRα D2-D3 domains, as there were no single mutations that knocked out trimer binding. Deletion studies have shown that the D3 domain is essential for trimer-mediated HCMV entry[188], and a recent cryo-EM structure has shown AD169 HCMV trimer binding to PDGFRα[220]. Subunit gO makes extensive contacts to multiple independent sites across PDGFRa domains D1, D2, and D3, such that negative effects of single amino acid substitutions would seem to be too small to cause an observable decrease in binding to HCMV trimer by flow cytometry.

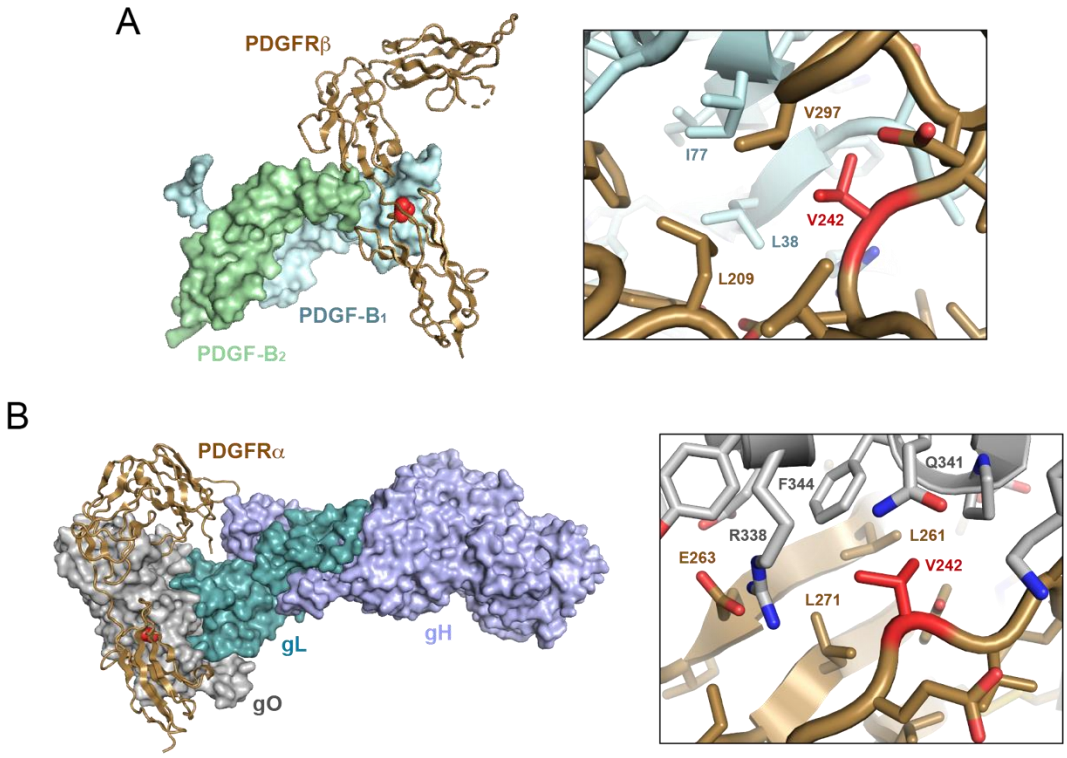

Figure 4.7. The V242K mutation in PDGFRα disproportionately impacts the interface with PDGF ligands. Figure designed by Kevin Gill. (A) Crystal structure (PDB 3MJG) of PDGFRβ (brown ribbon) bound to PDGF-BB (light green and blue surfaces). PDGFRβ is a very close homolog of PDGFRα. Residue V242 of PDGFRβ is highlighted in red and fills a hydrophobic pocket at the interface, shown in the close-up on the right. Note the small, tight interface between the receptor and its endogenous ligand. **(B)** CryoEM structure (PDB 7RAM) of the HCMV trimer (gH in pale blue, gL in teal, gO in grey) bound to PDGFRα (brown ribbon). Residue V242 of PDGFRα is shown in red. Note the expansive interface between the HCMV trimer and PDGFRα domains. In the close-up on the right, PDGFRα-V242 is shown to be located at the interface surrounded by polar and nonpolar residues. The different attributes of the respective interfaces explain why the PDGFRα-V242K mutation is disruptive of PDGF ligand binding but tolerated for HCMV trimer binding.

In our own prior studies, deep mutagenesis of human membrane proteins was used to understand how sequence relates to folding based on surface localization, as misfolded sequences were

retained intracellularly due to quality control machinery[9,63,80]. However, PDGFRα is an exception, and nearly every missense mutation still reaches the plasma membrane, including those that are incompatible with the native folded structure (for example, mutations that disrupt disulfides, create cavities, impose steric clashes, or introduce buried charges). The structural organization of PDGFRα as a series of loosely connected β-sandwich domains may facilitate surface trafficking of folding mutants, whereas previous studies have focused on membrane proteins with complex tertiary and quaternary structures.

Finally, deep mutagenesis has been extensively applied to viral surface proteins based on virus replication and propagation in tissue culture as the selection, often in the presence of neutralizing antibodies to discover principles of immune recognition and escape[108,221]. However, it is by decoupling the sequence library from a virus genome[74] that we have been able to determine a mutational landscape of a virus receptor. The principles and approach outlined here can be easily adapted for the engineering of other viral receptors for improved efficacy and safety to treat or prevent infections. Indeed, the same methods we optimized for the development of a HCMV antiviral were rapidly repurposed for engineering soluble decoy receptors that bind with high affinity to the spike of SARS coronavirus 2, a zoonotic pathogen that recently spilled over to the human population and has profoundly impacted world events[222]. Receptor-based antiviral drugs may be especially desirable for emerging zoonotic viruses or viruses with high sequence diversity, as they potentially have broader specificity than monoclonal antibodies to target all virus strains that utilize a common receptor.

MATERIALS AND METHODS

Plasmids

Human PDGFRα isoform 1 (GenBank NM_006206.4) was cloned in to the NheI-XhoI sites of pCEP4 (Invitrogen) with a N-terminal HA leader (MKTIIALSYIFCLVFA), myc-tag, linker (GSPGGASG), and followed by the mature polypeptide (a.a. 24-1,089). For measuring signaling activity, a minimal promoter under the control of tandem serum response elements was subcloned from SRE reporter vector_559 (Addgene # 82686)[207], a GFP-PEST fluorescent reporter[206] was inserted at the AscI site, and the entire reporter cassette was inserted at the NruI site of pCEP4-myc-PDGFRα by Gibson assembly. As a negative control, human PDGFRβ isoform 1 (GenBank NM_002609.3) was similarly cloned between the NheI-XhoI sites of pCEP4 with a N-terminal HA leader, myc-tag, and linker that connects to the mature protein (a.a. 33-1,106). No interactions between PDGFRβ and HCMV trimer were observed. Soluble PDGFRα-Fc was cloned in to the NheI-XhoI sites of pcDNA3.1(+) (Invitrogen), and encompassed PDGFRα a.a. 1-524, a short connecting linker, and the C-terminus of human IgG1 (GenBank KY432415.1) beginning at either C220 or D221 as described in Results. Alternatively, PDGFRα a.a. 1-524 were fused via linker GGGS to D221-K447 of human IgG3 (GenBank P01860.2; this is an R435 allele with reduced binding to FcRn). Synthetic human codon-optimized gene fragments (Integrated DNA Technologies) for HCMV gO (GenBank ABV71596.1 for TB40/E strain, AJY56739.1 for Merlin strain) were genetically fused at the C-terminus to superfolder GFP[199] (for detection of soluble HCMV trimer binding to PDGFRα-positive cells) or to an 8-histidine tag (for expression of membrane-tethered HCMV trimer), and were ligated in to the NheI-XhoI sites of pcDNA3.1(+). Codon-optimized gene fragments for full-length HCMV gH (GenBank ABV71597.1 for TB40/E strain, YP_081523.1 for Merlin

Strain) were cloned in to the NheI-XhoI sites of pcDNA3.1(+) for expression of membrane-tethered HCMV trimer. For production of soluble trimer, the extracellular region of gH with the leader peptide (a.a. 1-717 for TB40/E, a.a. 1-716 for Merlin) was subcloned with a 8-histidine tag. Codon-optimized synthetic genes for gL (GenBank ABV71629.1 for TB40/E strain, YP_081555.1 for Merlin) were cloned with C-terminal FLAG tags in to the NheI-XhoI sites of pcDNA3.1(+). Targeted mutations were made by overlap extension PCR. All plasmids were sequence verified (ACGT, Inc) and are deposited with Addgene.

Cells and viruses

Expi293F cells (ThermoFisher) are a suspension culture derivative of HEK293, and were cultured in Expi293 Expression Medium (ThermoFisher) at 125 rpm, 8 % CO2, 37 °C. MRC-5 embryonic lung fibroblasts and ARPE-19 adult retinal pigment epithelial cells were from the American Type Culture Collection and were grown at 37 °C, 5 % CO2, in DMEM supplemented with 10% fetal bovine serum, 1 mM sodium pyruvate, 2 mM glutamax (Gibco), 10 mM Hepes pH 7.4, 0.1 mM MEM Non-Essential Amino Acids (Gibco), 100 units/ml penicillin G, and 100 μg/ml streptomycin sulfate. ARPE-19 ectopically expressing PDGFRα ("ARPE19-RA" cells) were generated by lentiviral transduction with pLV-EF1a-PDGFRα-IRES-PURO, generously provided by Kai Wu (Princeton University). HCMV virus stocks were prepared by electroporating purified bacmid DNA into either MRC-5 fibroblasts (AD169[223]) or ARPE-19 retinal pigment epithelial cells (TB40/E, clone BAC4[224]). Stocks were amplified twice by infecting cell monolayers in 15 cm tissue culture plates followed by 850 cm^2 roller bottles. Virions were concentrated 100× by centrifugation through a 20 % sorbitol cushion, resuspended in phosphate-buffered saline (PBS) containing 1% BSA and 7% sucrose, and frozen at -80 °C.

Virus stocks were titered on MRC-5 and ARPE-19 cells by Immediate-Early Protein 1 (IE1) fluorescent focus assay[225].

Flow cytometry analysis of soluble HCMV trimer binding to receptor

Soluble HCMV trimer was prepared by transfecting Expi293F cells at a density of 2×10^6 / ml with 400 ng pcDNA3-gO-sfGFP, 400 ng pcDNA3-gH-8his and 400 ng pcDNA3-gL-FLAG per ml culture using Expifectamine (ThermoFisher). Transfection Enhancers (ThermoFisher) were added 18 h later, and 4 days post-transfection the culture was centrifuged (1,000 × g, 10 minutes, followed by a second spin of the supernatant at 21,000 × g, 5 minutes). The medium supernatant was stored at -20 °C and used directly in binding experiments. Expi293F cells at 2×10^6 / ml were transfected with plasmids encoding receptors (500 ng DNA per ml of culture) using Expifectamine. 24 h post-transfection the cells were washed with PBS supplemented with 0.2 % bovine serum albumin (PBS-BSA), incubated for 30 minutes on ice with anti-myc Alexa 647 (clone 9B11, 1/250 dilution; Cell Signaling Technology) and the indicated dilutions of gH/gL/gO-sfGFP-containing medium, washed twice with PBS-BSA, then analyzed on a BD LSR II flow cytometer. For competition assays, washed cells were pre-incubated with PDGF ligands (R&D Systems) for 2 minutes on ice prior to addition of gH/gL/gO-sfGFP-containing medium.

Deep mutagenesis

The D2-D3 domains (a.a. D123-E311) within plasmid pCEP4-myc-PDGFRα were mutagenized by overlap extension PCR[139] using primers with degenerate NNK codons. The plasmid library was transfected in to Expi293F cells using Expifectamine under conditions previously shown to

typically give no more than a single coding variant per cell; 1 ng coding plasmid was diluted with 1,500 ng pCEP4-ΔCMV carrier plasmid per ml of cell culture at 2×10^6 / ml, and the medium was replaced 2 h post-transfection. The cells were collected after 24 h, washed with ice-cold PBS-BSA, and incubated for 2 minutes on ice with PDGF-AA, -AB, -BB, and –CC (25 nM of each after addition of soluble trimer-containing medium) and anti-myc Alexa 647 (clone 9B11, 1/250 dilution; Cell Signaling Technology). Medium from cells expressing gH/gL/gO-sfGFP was then added to a final dilution of 1/3, and the cells were incubated on ice for 20 minutes. Cells were washed twice with PBS-BSA and sorted on a BD FACS Aria II at the Roy J. Carver Biotechnology Center. The main cell population was gated by forward/side scattering to remove debris and doublets, and propidium iodide (1 μg/ml final) was added to the sample to exclude dead cells. Of the myc-positive (Alexa 647) population, the 15 % of cells with the highest and lowest GFP fluorescence were collected (Figure 1F) in tubes coated overnight with fetal bovine serum and containing Expi293 Expression Medium. Total RNA was extracted from the collected cells using a GeneJET RNA purification kit (Thermo Scientific), and cDNA was reverse transcribed with high fidelity Accuscript primed with a gene-specific oligonucleotide (RevTrans_PDGFRA_992R, TCATGCAGGTTGACAGCTTC). The diversified region of PDGFRα was PCR amplified as 3 fragments to provide full coverage of the D2-D3 domains during Illumina sequencing. During PCR, flanking sequences on the primers added adapters to the ends of the products for annealing to Illumina sequencing primers, unique barcoding, and for binding the flow cell. Amplicons were sequenced on an Illumina NovaSeq 6000 using a 2×250 nt paired end protocol. Data were analyzed using Enrich[142], and commands are provided in the GEO deposit. Briefly, the frequencies of PDGFRα variants in the transcripts of the sorted

populations were compared to their frequencies in the naive plasmid library to calculate an enrichment ratio.

Production of soluble PDGFRα-Fc

Per ml of 2×10^6 Expi293F cells, 500 ng of pcDNA3-sPDGFRα-Fc (IgG1) and 3 μg of polyethylenimine (MW 25,000; Polysciences) were mixed in 100 μl of OptiMEM (Gibco) and incubated for 20 minutes at room temperature prior to adding to cells. Transfection Enhancers were added 18 h post-transfection, and cells were cultured for six to seven days. Cells were removed by centrifugation at 600 × g for 20 minutes at 4 °C. Cell debris and precipitates were removed by centrifugation at 18,000 × g for 25 minutes at 4 °C. Supernatant was loaded on to KANEKA KanCapA 3G Affinity sorbent (Pall), and the resin was washed with PBS. Soluble PDGFRα-Fc was eluted with 60 mM sodium acetate pH 3.7, and 1 M Tris pH 8.0 was added to the eluate to neutralize the pH. Eluted sPDGFRα-Fc was concentrated with a centrifugal device (MWCO 100 kDa; Sartorius) and NaCl was added to 150 mM final. The protein was separated on a Superdex 200 Increase 10/300 GL (GE Healthcare Life Sciences) size exclusion chromatography column equilibrated with PBS. Peak fractions were pooled, concentrated, and stored at -80 °C after snap freezing in liquid nitrogen. The proteins were quantified by absorbance at 280 nm using calculated molar extinction coefficients for the monomeric mature polypeptides.

Fusions of sPDGFRα to the Fc region of IgG3 were expressed as described above, and the expression medium was dialyzed against water. Cell debris was removed by centrifugation at 18,000 × g for 25 minutes at 4 °C. Supernatant was loaded on to protein G HTC agarose beads

(GoldBio) equilibrated with PBS. Protein was eluted with 100 mM glycine pH 2.5 and the eluate was neutralized by addition of 1 M Tris pH 9.0. Protein was further purified as described for the IgG1 fusions.

Flow cytometry analysis of PDGFRα-Fc binding to HCMV trimer on the cell surface

400 ng of each of pcDNA3-gH (full-length), pcDNA3-gL-FLAG and pcDNA3-gO-8his were transfected in to Expi293F cells at 2×10^6 / ml using Expifectamine. Transfection Enhancers were added 22 h later, and cells were harvested 46 h post-transfection. Cells were washed with PBS-BSA and incubated with PDGFRα-Fc (purified as described) or PDGFRβ-Fc (R&D Systems) for 40 minutes on ice. Cells were then washed 3 times with PBS-BSA and stained with anti-FLAG M2-Cy3 (Sigma), chicken anti-HIS-FITC (polyclonal, Immunology Consultants Laboratory), and anti-human IgG-APC (clone HP6017, BioLegend) for 30 minutes on ice. Cells were washed three times with PBS-BSA and analyzed by flow cytometry. To compare data across different experiments, the change in ΔMFU for each condition was normalized to the ΔMFU at the maximum concentration of WT sPDGFRα-Fc: Relative binding = $(MFU_{sample} - MFU_{background}) / (MFU_{max\ WT} - MFU_{background})$.

PDGFRα signaling assay

500 ng of PDGFRα reporter plasmid was transfected into 1 ml Expi293F cells at 2×10^6 / ml using Expifectamine. 7.5 μl of 6 μM sPDGFRα-Fc (concentration based on monomer) and 7.5 μl of 2 μM PDGF were mixed, incubated at room temperature for 40 minutes, and then added to 1 ml cells at transfection. Human gamma globulin was from Jackson Immuno Research Labs, and PDGFs were from R&D Systems. Cells were collected 24 h post-transfection and stained with

anti-myc-Alexa 647 to detect PDGFRα expression. GFP-PEST reporter expression was measured by flow-cytometry.

Structural modeling

The sequence of human PDGFRα was threaded onto the structure of PDGF-BB-bound PDGFRβ (PDB ID 3MJG [44]), with one of the two receptor chains and the D1 domain removed. The model was minimized with ROSETTA using FastRelax[153]. Connectivity was determined using the AverageDegree filter in ROSETTA[226]. Images were rendered with PyMOL (Schrödinger, LLC).

Reagent and data availability

Plasmids are deposited with Addgene. Raw and processed deep sequencing data are deposited in NCBI's Gene Expression Omnibus (GEO) under series accession number GSE138169.

REFERENCES

1. Maginnis, M. S. Virus–Receptor Interactions: The Key to Cellular Invasion. *J. Mol. Biol.* **430**, 2590 (2018).
2. Cunningham, B. C. & Wells, J. A. High-Resolution Epitope Mapping of hGH-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science (80-.).* **244**, 1081–1085 (1989).
3. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28–33 (1992).
4. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods 2014 118* **11**, 801–807 (2014).
5. Fowler, D. M., Stephany, J. J. & Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat. Protoc. 2014 99* **9**, 2267–2284 (2014).
6. Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310.e20 (2020).
7. Chan, K. K. *et al.* Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science (80-.).* **369**, 1261–1265 (2020).
8. Haddox, H. K., Dingens, A. S. & Bloom, J. D. Experimental Estimation of the Effects of All Amino-Acid Mutations to HIV's Envelope Protein on Viral Replication in Cell Culture. *PLOS Pathog.* **12**, e1006114 (2016).
9. McShan, A. C. *et al.* Molecular determinants of chaperone interactions on MHC-I for folding and antigen repertoire selection. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 25602–25613 (2019).
10. McShan, A. C. *et al.* TAPBPR promotes antigen loading on MHC-I molecules using a

peptide trap. *Nat. Commun. 2021 121* **12**, 1–18 (2021).

11. Hanning, K. R., Minot, M., Warrender, A. K., Kelton, W. & Reddy, S. T. Deep mutational scanning for therapeutic antibody engineering. *Trends Pharmacol. Sci.* **43**, 123–135 (2022).
12. Jamshad, M. *et al.* G-protein coupled receptor solubilization and purification for biophysical analysis and functional studies, in the total absence of detergent. *Biosci. Rep.* **35**, 188 (2015).
13. Sriram, K. & Insel, P. A. G protein-coupled receptors as targets for approved drugs: How many targets and how many drugs? *Mol. Pharmacol.* **93**, 251–258 (2018).
14. Lenahan, C., Sanghavi, R., Huang, L. & Zhang, J. H. Rhodopsin: A Potential Biomarker for Neurodegenerative Diseases. *Front. Neurosci.* **14**, 326 (2020).
15. Shearer, W. T., Philpott, G. W., Parker, C. W., T Prehn, J. R. & Reticuloendothel Soc, J. Opiate Agonists and Antagonists Discriminated by Receptor Binding in Brain. *Science (80-.).* **182**, 1359–1361 (1973).
16. Vasudevan, N. T., Mohan, M. L., Goswami, S. K. & Naga Prasad, S. V. Regulation of β-adrenergic receptor function: An emphasis on receptor resensitization. *Cell Cycle* **10**, 3684 (2011).
17. Alkhatib, G. The biology of CCR5 and CXCR4. *Curr. Opin. HIV AIDS* **4**, 96–103 (2009).
18. Kobayashi, D. *et al.* Regulation of CCR7-dependent cell migration through CCR7 homodimer formation. *Sci. Reports 2017 71* **7**, 1–14 (2017).
19. Müller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nat. 2001 4106824* **410**, 50–56 (2001).
20. Laing, K. J. & Secombes, C. J. Chemokines. *Dev. Comp. Immunol.* **28**, 443–460 (2004).

21. Allen, S. J., Crown, S. E. & Handel, T. M. Chemokine:Receptor Structure, Interactions, and Antagonism. *https://doi.org/10.1146/annurev.immunol.24.021605.090529* **25**, 787–820 (2007).

22. Begley, L. A., MacDonald, J. W., Day, M. L. & Macoska, J. A. CXCL12 activates a robust transcriptional response in human prostate epithelial cells. *J. Biol. Chem.* **282**, 26767–26774 (2007).

23. Gassenmaier, M. *et al.* CXC Chemokine Receptor 4 is Essential for Maintenance of Renal cell Carcinoma-Initiating Cells and Predicts Metastasis. *Stem Cells* **31**, 1467–1476 (2013).

24. Connor, R. I., Sheridan, K. E., Ceradini, D., Choe, S. & Landau, N. R. Change in Coreceptor Use Correlates with Disease Progression in HIV-1–Infected Individuals. *J. Exp. Med.* **185**, 621–628 (1997).

25. Oberlin, E. *et al.* The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nat. 1996 3826594* **382**, 833–835 (1996).

26. Bernhagen, J. *et al.* MIF is a noncognate ligand of CXC chemokine receptors in inflammatory and atherogenic cell recruitment. *Nat. Med. 2007 135* **13**, 587–596 (2007).

27. Teicher, B. A. & Fricker, S. P. CXCL12 (SDF-1)/CXCR4 pathway in cancer. *Clin. Cancer Res.* **16**, 2927–2931 (2010).

28. Vila-Coro, A. J. *et al. The chemokine SDF-lα triggers CXCR4 receptor dimerization and activates the JAK/STAT pathway*. *The FASEB Journal* vol. 13 (1999).

29. Deng, H. K. *et al.* Identification of a major co-receptor for primary isolates of HIV-1. *Nat. 1996 3816584* **381**, 661–666 (1996).

30. Dragic, T. *et al.* HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nat. 1996 3816584* **381**, 667–673 (1996).

31. Falcon, A. *et al.* CCR5 deficiency predisposes to fatal outcome in influenza virus infection. *J. Gen. Virol.* **96**, 2074–2078 (2015).

32. Glass, W. G. *et al.* CCR5 deficiency increases risk of symptomatic West Nile virus infection. *J. Exp. Med.* **203**, 35–40 (2006).

33. Zhao, D. Y. *et al.* Cryo-EM structure of the native rhodopsin dimer in nanodiscs. *J. Biol. Chem.* **294**, 14215–14230 (2019).

34. Manglik, A. *et al.* Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. *Nat. 2012 4857398* **485**, 321–326 (2012).

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

36. Guo, W., Shi, L. & Javitch, J. A. The fourth transmembrane segment forms the interface of the dopamine D2 receptor homodimer. *J. Biol. Chem.* **278**, 4385–4388 (2003).

37. Kasai, R. S., Ito, S. V., Awane, R. M., Fujiwara, T. K. & Kusumi, A. The Class-A GPCR Dopamine D2 Receptor Forms Transient Dimers Stabilized by Agonists: Detection by Single-Molecule Tracking. *Cell Biochem. Biophys.* **76**, 29–37 (2018).

38. Hebert, T. E. *et al.* A peptide derived from a β 2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. *J. Biol. Chem.* **271**, 16384–16392 (1996).

39. Park, P. S. H. & Wells, J. W. Oligomeric potential of the M2 muscarinic cholinergic receptor. *J. Neurochem.* **90**, 537–548 (2004).

40. Asher, W. B. *et al.* Single-molecule FRET imaging of GPCR dimers in living cells. *Nat. Methods 2021 184* **18**, 397–405 (2021).

41. Kufareva, I. *et al.* A Novel Approach for Quantifying GPCR Dimerization Equilibrium Using Bioluminescence Resonance Energy Transfer. *Methods Mol. Biol.* **1013**, 93 (2013).

42. Hernanz-Falcón, P. *et al.* Identification of amino acid residues crucial for chemokine receptor dimerization. *Nat. Immunol. 2004 52* **5**, 216–223 (2004).

43. Issafras, H. *et al.* Constitutive Agonist-independent CCR5 Oligomerization and Antibody-mediated Clustering Occurring at Physiological Levels of Receptors. *J. Biol. Chem.* **277**, 34666–34673 (2002).

44. Arrildt, K. T., Joseph, S. B. & Swanstrom, R. The HIV-1 Env protein: A coat of many colors. *Curr. HIV/AIDS Rep.* **9**, 52–63 (2012).

45. Hallenberger, S. *et al.* Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gpl60. *Nat. 1992 3606402* **360**, 358–361 (1992).

46. McCune, J. M. *et al.* Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell* **53**, 55–67 (1988).

47. Lambert, G. S. & Upadhyay, C. HIV-1 Envelope Glycosylation and the Signal Peptide. *Vaccines* **9**, 1–19 (2021).

48. Sanders, R. W. & Moore, J. P. Native-like Env trimers as a platform for HIV-1 vaccine design. **275**, 161–182 (2017).

49. Sanders, R. W. *et al.* A Next-Generation Cleaved, Soluble HIV-1 Env Trimer, BG505 SOSIP.664 gp140, Expresses Multiple Epitopes for Broadly Neutralizing but Not Non-Neutralizing Antibodies. *PLOS Pathog.* **9**, e1003618 (2013).

50. Sanders, R. W. *et al.* Stabilization of the Soluble, Cleaved, Trimeric Form of the Envelope Glycoprotein Complex of Human Immunodeficiency Virus Type 1. *J. Virol.* **76**, 8875–8889 (2002).

51. Staras, S. A. S. *et al.* Seroprevalence of cytomegalovirus infection in the United States, 1988-1994. *Clin. Infect. Dis.* **43**, 1143–1151 (2006).

52. Gerna, G., Kabanova, A. & Lilleri, D. Human Cytomegalovirus Cell Tropism and Host Cell Receptors. *Vaccines 2019, Vol. 7, Page 70* **7**, 70 (2019).

53. Martinez-Martin, N. *et al.* An Unbiased Screen for Human Cytomegalovirus Identifies Neuropilin-2 as a Central Viral Receptor. *Cell* **174**, 1158-1171.e19 (2018).

54. Wu, K., Oberstein, A., Wang, W. & Shenk, T. Role of PDGF receptor-α during human cytomegalovirus entry into fibroblasts. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9889–E9898 (2018).

55. Stegmann, C. *et al.* A derivative of platelet-derived growth factor receptor alpha binds to the trimer of human cytomegalovirus and inhibits entry into fibroblasts and endothelial cells. *PLOS Pathog.* **13**, e1006273 (2017).

56. Shim, A. H. R. *et al.* Structures of a platelet-derived growth factor/propeptide complex and a platelet-derived growth factor/receptor complex. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11307–11312 (2010).

57. Kufareva, I. *et al.* Stoichiometry and geometry of the CXC Chemokine receptor 4 complex with CXC ligand 12: Molecular modeling and experimental validation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5363–E5372 (2014).

58. Albizu, L. *et al.* Time-resolved FRET between GPCR ligands reveals oligomers in native tissues. *Nat. Chem. Biol. 2010 68* **6**, 587–594 (2010).

59. Drurya, L. J. *et al.* Monomeric and dimeric CXCL12 inhibit metastasis through distinct CXCR4 interactions and signaling pathways. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17655–17660 (2011).

60. El-Asmar, L. *et al.* Evidence for Negative Binding Cooperativity within CCR5-CCR2b Heterodimers. *Mol. Pharmacol.* **67**, 460–469 (2005).

61. Jones, K. A. *et al.* GABAB receptors function as a heteromeric assembly of the subunits GABABR1 and GABABR2. *Nat. 1998 3966712* **396**, 674–679 (1998).

62. El Moustaine, D. *et al.* Distinct roles of metabotropic glutamate receptor dimerization in agonist activation and G-protein coupling. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16342–16347 (2012).

63. Heredia, J. D. *et al.* Mapping interaction sites on human chemokine receptors by deep mutational scanning. *J. Immunol.* **200**, (2018).

64. Lao, J. *et al.* Single-Molecule Imaging Demonstrates Ligand Regulation of the Oligomeric Status of CXCR4 in Living Cells. *J. Phys. Chem. B* **121**, 1466–1474 (2017).

65. Martínez-Muñoz, L. *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).

66. Babcock, G. J., Farzan, M. & Sodroski, J. Ligand-independent dimerization of CXCR4, a principal HIV-1 coreceptor. *J. Biol. Chem.* **278**, 3378–3385 (2003).

67. Tan, Q. *et al.* Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science (80-.).* **341**, 1387–1390 (2013).

68. Peng, P. *et al.* Structure-Based Design of 1-Heteroaryl-1,3-propanediamine Derivatives as a Novel Series of CC-Chemokine Receptor 5 Antagonists. *J. Med. Chem.* **61**, 9621–9636 (2018).

69. Zhang, H. *et al.* Structural basis for chemokine recognition and receptor activation of chemokine receptor CCR5. *Nat. Commun. 2021 121* **12**, 1–12 (2021).

70. Jin, J. *et al.* CCR5 adopts three homodimeric conformations that control cell surface delivery. *Sci. Signal.* **11**, 2869 (2018).

71. Zhang, F. *et al.* Molecular Mechanism Regarding Allosteric Modulation of Ligand Binding and the Impact of Mutations on Dimerization for CCR5 Homodimer. (2019) doi:10.1021/acs.jcim.8b00850.

72. Sohy, D. *et al.* Hetero-oligomerization of CCR2, CCR5, and CXCR4 and the protean effects of 'selective' antagonists. *J. Biol. Chem.* **284**, 31270–31279 (2009).

73. Fonseca, J. M. & Lambert, N. A. Instability of a Class A G Protein-Coupled Receptor Oligomer Interface. *Mol. Pharmacol.* **75**, 1296–1299 (2009).

74. Heredia, J. D., Park, J., Choi, H., Gill, K. S. & Procko, E. Conformational engineering of HIV-1 Env based on mutational tolerance in the CD4 and PG16 bound states. *J. Virol.* **93**, (2019).

75. Camp, T., Mehta, K., Sligar, S. G. & Zhang, K. Molecular Orientation Determination in Nanodiscs at the Single-Molecule Level. *Anal. Chem.* **92**, 2229–2236 (2020).

76. Tinevez, J. Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).

77. Vega, A. R., Freeman, S. A., Grinstein, S. & Jaqaman, K. Multistep Track Segmentation and Motion Classification for Transient Mobility Analysis. *Biophys. J.* **114**, 1018 (2018).

78. Ohashi, K. & Mizuno, K. A Novel Pair of Split Venus Fragments to Detect Protein–Protein Interactions by In Vitro and In Vivo Bimolecular Fluorescence Complementation Assays. *Methods Mol. Biol.* **1174**, 247–262 (2014).

79. Young, H. J. *et al.* Deep Mutagenesis of a Transporter for Uptake of a Non-Native Substrate Identifies Conformationally Dynamic Regions. *bioRxiv* 2021.04.19.440442

(2021) doi:10.1101/2021.04.19.440442.

80. Park, J. *et al.* Structural architecture of a dimeric class C GPCR based on co-trafficking of sweet taste receptor subunits. *J. Biol. Chem.* **294**, 4759–4774 (2019).

81. Villar, V. A. M., Cuevas, S., Zheng, X. & Jose, P. A. Localization and signaling of GPCRs in lipid rafts. *Methods Cell Biol.* **132**, 3–23 (2016).

82. Goddard, A. D. & Watts, A. Regulation of G protein-coupled receptors by palmitoylation and cholesterol. *BMC Biol.* **10**, 1–3 (2012).

83. Blanpain, C. *et al.* Palmitoylation of CCR5 Is Critical for Receptor Trafficking and Efficient Activation of Intracellular Signaling Pathways. *J. Biol. Chem.* **276**, 23795–23804 (2001).

84. Holm, K., Weclewicz, K., Hewson, R. & Suomalainen, M. Human Immunodeficiency Virus Type 1 Assembly and Lipid Rafts: Pr55 gag Associates with Membrane Domains That Are Largely Resistant to Brij98 but Sensitive to Triton X-100. *J. Virol.* **77**, 4805–4817 (2003).

85. Chen, B. Molecular Mechanism of HIV-1 Entry. *Trends Microbiol.* **27**, 878–891 (2019).

86. Hernanz-Falcón, P. *et al.* Identification of amino acid residues crucial for chemokine receptor dimerization. *Nat. Immunol. 2004 52* **5**, 216–223 (2004).

87. Hancock, J. F., Paterson, H. & Marshall, C. J. A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21ras to the plasma membrane. *Cell* **63**, 133–139 (1990).

88. Lindwasser, O. W. & Resh, M. D. Multimerization of Human Immunodeficiency Virus Type 1 Gag Promotes Its Localization to Barges, Raft-Like Membrane Microdomains. *J. Virol.* **75**, 7913–7924 (2001).

89. Nguyen, D. H. & Taub, D. CXCR4 Function Requires Membrane Cholesterol: Implications for HIV Infection. *J. Immunol.* **168**, 4121–4126 (2002).

90. Merk, A. & Subramaniam, S. HIV-1 envelope glycoprotein structure. *Curr. Opin. Struct. Biol.* **23**, 268 (2013).

91. Huang, C. C. *et al.* Structures of the CCR5 N Terminus and of a Tyrosine-Sulfated Antibody with HIV-1 gp120 and CD4. *Science* **317**, 1930 (2007).

92. Wu, L. *et al.* CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nat. 1996 3846605* **384**, 179–183 (1996).

93. Liu, J., Bartesaghi, A., Borgnia, M. J., Sapiro, G. & Subramaniam, S. Molecular architecture of native HIV-1 gp120 trimers. *Nat. 2008 4557209* **455**, 109–113 (2008).

94. Burton, D. R. *et al.* A BLUEPRINT FOR HIV VACCINE DISCOVERY. *Cell Host Microbe* **12**, 396 (2012).

95. Moore, P. L. *et al.* Nature of Nonfunctional Envelope Proteins on the Surface of Human Immunodeficiency Virus Type 1. *J. Virol.* **80**, 2515 (2006).

96. Herrera, C. *et al.* The impact of envelope glycoprotein cleavage on the antigenicity, infectivity, and neutralization sensitivity of Env-pseudotyped human immunodeficiency virus type 1 particles. *Virology* **338**, 154–172 (2005).

97. Sattentau, Q. J. & Moore, J. P. Human immunodeficiency virus type 1 neutralization is determined by epitope exposure on the gp120 oligomer. *J. Exp. Med.* **182**, 185 (1995).

98. Parren, P. W. H. I. *et al.* Neutralization of Human Immunodeficiency Virus Type 1 by Antibody to gp120 Is Determined Primarily by Occupancy of Sites on the Virion Irrespective of Epitope Specificity. *J. Virol.* **72**, 3512–3519 (1998).

99. Parren, P. W. H. I. *et al.* Relevance of the antibody response against human

immunodeficiency virus type 1 envelope to vaccine design. *Immunol. Lett.* **57**, 105–112 (1997).

100. Ward, A. B. & Wilson, I. A. The HIV-1 envelope glycoprotein structure: nailing down a moving target. **275**, 21–32 (2017).

101. Binley, J. M. *et al.* A recombinant human immunodeficiency virus type 1 envelope glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure. *J. Virol.* **74**, 627–643 (2000).

102. De Taeye, S. W. *et al.* Immunogenicity of Stabilized HIV-1 Envelope Trimers with Reduced Exposure of Non-neutralizing Epitopes. *Cell* **163**, 1702–1715 (2015).

103. Kulp, D. W. *et al.* Structure-based design of native-like HIV-1 envelope trimers to silence non-neutralizing epitopes and eliminate CD4 binding. *Nat. Commun. 2017 81* **8**, 1–14 (2017).

104. Do Kwon, Y. *et al.* Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. *Nat. Struct. Mol. Biol. 2015 227* **22**, 522–531 (2015).

105. Crooks, E. T., Tong, T., Osawa, K. & Binley, J. M. Enzyme digests eliminate nonfunctional Env from HIV-1 particle surfaces, leaving native Env trimers intact and viral infectivity unaffected. *J. Virol.* **85**, 5825–5839 (2011).

106. Jardine, J. G. *et al.* HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. *Science (80-.).* **351**, 1458–1463 (2016).

107. Steichen, J. M. *et al.* HIV Vaccine Design to Target Germline Precursors of Glycan-Dependent Broadly Neutralizing Antibodies. *Immunity* **45**, 483–496 (2016).

108. Dingens, A. S., Haddox, H. K., Overbaugh, J. & Bloom, J. D. Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody. *Cell Host Microbe* **21**, 777-787.e4 (2017).

109. Dingens, A. S. *et al.* Complete functional mapping of infection- and vaccine-elicited antibodies against the fusion peptide of HIV. *PLOS Pathog.* **14**, e1007159 (2018).

110. Kovacs, J. M. *et al.* Stable, uncleaved HIV-1 envelope glycoprotein gp140 forms a tightly folded trimer with a native-like structure. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18542–18547 (2014).

111. Burton, D. R. *et al.* Broad and potent neutralizing antibodies from an african donor reveal a new HIV-1 vaccine target. *Science (80-.).* **326**, 285–289 (2009).

112. Julien, J. P. *et al.* Asymmetric recognition of the HIV-1 trimer by broadly neutralizing antibody PG9. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4351–4356 (2013).

113. Tran, E. E. H. *et al.* Structural Mechanism of Trimeric HIV-1 Envelope Glycoprotein Activation. *PLOS Pathog.* **8**, e1002797 (2012).

114. Li, Y. *et al.* Mechanism of neutralization by the broadly neutralizing HIV-1 monoclonal antibody VRC01. *J. Virol.* **85**, 8954–8967 (2011).

115. Fersht, Y. A. R. *et al.* Dissociation of gp120 from HIV-1 Virions Induced by Soluble CD4. *Science (80-.).* **250**, 1139–1142 (1990).

116. Wu, X. *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science (80-.).* **329**, 856–861 (2010).

117. Seaman, M. S. *et al.* Tiered Categorization of a Diverse Panel of HIV-1 Env Pseudoviruses for Assessment of Neutralizing Antibodies. *J. Virol.* **84**, 1439–1452 (2010).

118. Jones, D. R., Suzuki, K. & Piller, S. C. A 100-Amino Acid Truncation in the Cytoplasmic

Tail of Glycoprotein 41 in the Reference HIV Type 1 Strain RF. *https://home.liebertpub.com/aid* **18**, 513–517 (2004).

119. Holtkotte, D., Pfeiffer, T., Pisch, T. & Bosch, V. Selection and Characterization of a Replication-Competent Human Immunodeficiency Virus Type 1 Variant Encoding C-Terminally Truncated Env. *https://home.liebertpub.com/aid* **22**, 57–65 (2006).

120. Stano, A. *et al.* Dense Array of Spikes on HIV-1 Virion Particles. *J. Virol.* **91**, (2017).

121. Pancera, M. *et al.* Structural basis for diverse N-glycan recognition by HIV-1–neutralizing V1–V2–directed antibody PG16. *Nat. Struct. Mol. Biol. 2013 207* **20**, 804–813 (2013).

122. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648 (1998).

123. Myszka, D. G. *et al.* Energetics of the HIV gp120-CD4 binding reaction. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9026–9031 (2000).

124. Ozorowski, G. *et al.* Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike. *Nat. 2017 5477663* **547**, 360–363 (2017).

125. François, K. O. & Balzarini, J. The Highly Conserved Glycan at Asparagine 260 of HIV-1 gp120 Is Indispensable for Viral Entry. *J. Biol. Chem.* **286**, 42900 (2011).

126. Wang, W. *et al.* A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology* **10**, 1–14 (2013).

127. Kong, L., Wilson, I. A. & Kwong, P. D. Crystal structure of a fully glycosylated HIV-1 gp120 core reveals a stabilizing role for the glycan at Asn262. *Proteins* **83**, 590 (2015).

128. McLellan, J. S. *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nat. 2011 4807377* **480**, 336–343 (2011).

129. Chuang, G.-Y. *et al.* Structure-Based Design of a Soluble Prefusion-Closed HIV-1 Env Trimer with Reduced CD4 Affinity and Improved Immunogenicity. *J. Virol.* **91**, 2268–2284 (2017).

130. Lee, J. H. *et al.* A Broadly Neutralizing Antibody Targets the Dynamic HIV Envelope Trimer Apex via a Long, Rigidified, and Anionic β-Hairpin Structure. *Immunity* **46**, 690–702 (2017).

131. Grupping, K. *et al.* MiniCD4 protein resistance mutations affect binding to the HIV-1 gp120 CD4 binding site and decrease entry efficiency. *Retrovirology* **9**, 1–16 (2012).

132. Lee, J. H., De Val, N., Lyumkis, D. & Ward, A. B. Model building and refinement of a natively glycosylated HIV-1 Env protein by high-resolution cryoEM. *Structure* **23**, 1943 (2015).

133. Mouquet, H. *et al.* Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3268–E3277 (2012).

134. Garces, F. *et al.* Structural Evolution of Glycan Recognition by a Family of Potent HIV Antibodies. *Cell* **159**, 69–79 (2014).

135. Julien, J. P. *et al.* Broadly Neutralizing Antibody PGT121 Allosterically Modulates CD4 Binding via Recognition of the HIV-1 gp120 V3 Base and Multiple Surrounding Glycans. *PLOS Pathog.* **9**, e1003342 (2013).

136. Postler, T. S. & Desrosiers, R. C. The Tale of the Long Tail: the Cytoplasmic Domain of HIV-1 gp41. *J. Virol.* **87**, 2–15 (2013).

137. Jeffs, S. A. *et al.* Expression and characterisation of recombinant oligomeric envelope glycoproteins derived from primary isolates of HIV-1. *Vaccine* **22**, 1032–1046 (2004).

138. Go, E. P. *et al.* Native Conformation and Canonical Disulfide Bond Formation Are

Interlinked Properties of HIV-1 Env Glycoproteins. *J. Virol.* **90**, 2884–2894 (2016).

139. Procko, E. *et al.* Computational Design of a Protein-Based Enzyme Inhibitor. *J. Mol. Biol.* **425**, 3563–3575 (2013).

140. Hermida-Matsumoto, L. & Resh, M. D. Localization of Human Immunodeficiency Virus Type 1 Gag and Env at the Plasma Membrane by Confocal Imaging. *J. Virol.* **74**, 8670–8679 (2000).

141. Kodama, Y. & Hu, C. D. An improved bimolecular fluorescence complementation assay with a high signal-to-noise ratio. *Biotechniques* **49**, 793–803 (2010).

142. Fowler, D. M., Araya, C. L., Gerard, W. & Fields, S. Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics* **27**, 3430–3431 (2011).

143. Kwong, P. D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nat. 2002 4206916* **420**, 678–682 (2002).

144. Pantophlet, R., Wilson, I. A. & Burton, D. R. Improved design of an antigen with enhanced specificity for the broadly HIV-neutralizing antibody b12. *Protein Eng. Des. Sel.* **17**, 749–758 (2004).

145. Gorny, M. K. *et al.* Production of site-selected neutralizing human monoclonal antibodies against the third variable domain of the human immunodeficiency virus type 1 envelope glycoprotein. *Proc. Natl. Acad. Sci.* **88**, 3238–3242 (1991).

146. Gorny, M. K. *et al.* Human Monoclonal Antibodies Specific for Conformation-Sensitive Epitopes of V3 Neutralize Human Immunodeficiency Virus Type 1 Primary Isolates from Various Clades. *J. Virol.* **76**, 9035–9045 (2002).

147. Gorny, M. K. *et al.* Cross-Clade Neutralizing Activity of Human Anti-V3 Monoclonal Antibodies Derived from the Cells of Individuals Infected with Non-B Clades of Human Immunodeficiency Virus Type 1. *J. Virol.* **80**, 6865–6872 (2006).

148. Hioe, C. E. *et al.* Anti-V3 Monoclonal Antibodies Display Broad Neutralizing Activities against Multiple HIV-1 Subtypes. *PLoS One* **5**, e10254 (2010).

149. Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nat. 2011 4777365* **477**, 466–470 (2011).

150. Huang, J. *et al.* Broad and potent HIV-1 neutralization by a human antibody that binds the gp41–gp120 interface. *Nat. 2014 5157525* **515**, 138–142 (2014).

151. Stewart-Jones, G. B. E. *et al.* Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G. *Cell* **165**, 813–826 (2016).

152. Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nat. 2010 4667307* **466**, 756–760 (2010).

153. Leaver-Fay, A. *et al.* Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).

154. Nguyen, T. A. *et al.* Functional Anatomy of the Human Microprocessor. *Cell* **161**, 1374–1387 (2015).

155. Cannon, M. J., Schmid, D. S. & Hyde, T. B. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Rev. Med. Virol.* **20**, 202–213 (2010).

156. Emery, V. C. Investigation of CMV disease in immunocompromised patients. *J. Clin. Pathol.* **54**, 84–88 (2001).

157. Cobbs, C. S. *et al.* Human Cytomegalovirus Infection and Expression in Human

Malignant Glioma 1. *CANCER Res.* **62**, 3347–3350 (2002).

158. Mitchell, D. A. *et al.* Sensitive detection of human cytomegalovirus in tumors and peripheral blood of patients diagnosed with glioblastoma. *Neuro. Oncol.* **10**, 10–18 (2008).

159. Scheurer, M. E., Bondy, M. L., Aldape, K. D., Albrecht, T. & El-Zein, R. Detection of human cytomegalovirus in different histological types of gliomas. *Acta Neuropathol. 2008 1161* **116**, 79–86 (2008).

160. Michaelis, M., Doerr, H. W. & Cinatl, J. The Story of Human Cytomegalovirus and Cancer: Increasing Evidence and Open Questions. *Neoplasia* **11**, 1 (2009).

161. Lawler, S. E. Cytomegalovirus and glioblastoma; controversies and opportunities. *J. Neuro-Oncology 2015 1233* **123**, 465–471 (2015).

162. Cannon, M. J. & Davis, K. F. Washing our hands of the congenital cytomegalovirus disease epidemic. *BMC Public Health* **5**, 1–8 (2005).

163. Kenneson, A. & Cannon, M. J. Review and meta-analysis of the epidemiology of congenital cytomegalovirus (CMV) infection. *Rev. Med. Virol.* **17**, 253–276 (2007).

164. Dollard, S. C., Grosse, S. D. & Ross, D. S. New estimates of the prevalence of neurological and sensory sequelae and mortality associated with congenital cytomegalovirus infection. *Rev. Med. Virol.* **17**, 355–363 (2007).

165. Grosse, S. D., Ross, D. S. & Dollard, S. C. Congenital cytomegalovirus (CMV) infection as a cause of permanent bilateral hearing loss: a quantitative assessment. *J. Clin. Virol.* **41**, 57–62 (2008).

166. Suzuki, Y., Toribe, Y., Mogami, Y., Yanagihara, K. & Nishikawa, M. Epilepsy in patients with congenital cytomegalovirus infection. doi:10.1016/j.braindev.2007.12.004.

167. Zhang, X. W. *et al.* Physical and intellectual development in children with asymptomatic congenital cytomegalovirus infection: A longitudinal cohort study in Qinba mountain area, China. *J. Clin. Virol.* **40**, 180–185 (2007).

168. Gentile, I. *et al.* Prevalence of Congenital Cytomegalovirus Infection Assessed Through Viral Genome Detection in Dried Blood Spots in Children with Autism Spectrum Disorders. *In Vivo (Brooklyn).* **31**, 467 (2017).

169. Staras, S. A. S. *et al.* Cytomegalovirus seroprevalence and childhood sources of infection: A population-based study among pre-adolescents in the United States. *J. Clin. Virol.* **43**, 266–271 (2008).

170. Lanzieri, T. M. *et al.* Seroprevalence of Cytomegalovirus among Children 1 to 5 Years of Age in the United States from the National Health and Nutrition Examination Survey of 2011 to 2012. *Clin. Vaccine Immunol.* **22**, 245 (2015).

171. Dupont, L. & Reeves, M. B. Cytomegalovirus latency and reactivation: recent insights into an age old problem. *Rev. Med. Virol.* **26**, 75–89 (2016).

172. Ross, S. A. *et al.* Cytomegalovirus Reinfections in Healthy Seroimmune Women. *J. Infect. Dis.* **201**, 386–389 (2010).

173. Lurain, N. S. & Chou, S. Antiviral drug resistance of human cytomegalovirus. *Clin. Microbiol. Rev.* **23**, 689–712 (2010).

174. Cytomegalovirus, Parvovirus B19, Varicella Zoster, and Toxoplasmosis in Pregnancy | ACOG. https://www.acog.org/clinical/clinical-guidance/practice-bulletin/articles/2015/06/cytomegalovirus-parvovirus-b19-varicella-zoster-and-toxoplasmosis-in-pregnancy.

175. Hahn, G. *et al.* Human Cytomegalovirus UL131-128 Genes Are Indispensable for Virus

Growth in Endothelial Cells and Virus Transfer to Leukocytes. *J. Virol.* **78**, 10023–10033 (2004).

176. Wang, D. & Shenk, T. Human cytomegalovirus virion protein complex required for epithelial and endothelial cell tropism. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 18153–18158 (2005).

177. Wang, D. & Shenk, T. Human Cytomegalovirus UL131 Open Reading Frame Is Required for Epithelial Cell Tropism. *J. Virol.* **79**, 10330–10338 (2005).

178. Ryckman, B. J. *et al.* Characterization of the Human Cytomegalovirus gH/gL/UL128-131 Complex That Mediates Entry into Epithelial and Endothelial Cells. *J. Virol.* **82**, 60–70 (2008).

179. Adler, B. *et al.* Role of human cytomegalovirus UL131A in cell type-specific virus entry and release. *J. Gen. Virol.* **87**, 2451–2460 (2006).

180. Ryckman, B. J., Chase, M. C. & Johnson, D. C. HCMV gH/gL/UL128-131 interferes with virus entry into epithelial cells: Evidence for cell type-specific receptors. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14118–14123 (2008).

181. Xiaofei, E. *et al.* OR14I1 is a receptor for the human cytomegalovirus pentameric complex and defines viral epithelial cell tropism. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7043–7052 (2019).

182. Huber, M. T. & Compton, T. The Human Cytomegalovirus UL74 Gene Encodes the Third Component of the Glycoprotein H-Glycoprotein L-Containing Envelope Complex. *J. Virol.* **72**, 8191 (1998).

183. Huber, M. T. & Compton, T. Intracellular Formation and Processing of the Heterotrimeric gH-gL-gO (gCIII) Glycoprotein Envelope Complex of Human Cytomegalovirus. *J. Virol.*

73, 3886 (1999).

184. Zhou, M., Lanchy, J.-M. & Ryckman, B. J. Human Cytomegalovirus gH/gL/gO Promotes the Fusion Step of Entry into All Cell Types, whereas gH/gL/UL128-131 Broadens Virus Tropism through a Distinct Mechanism. *J. Virol.* **89**, 8999–9009 (2015).

185. Kabanova, A. *et al.* Platelet-derived growth factor-α receptor is the cellular receptor for human cytomegalovirus gHgLgO trimer. *Nat. Microbiol.* **1**, 16082 (2016).

186. Soroceanu, L., Akhavan, A. & Cobbs, C. S. Platelet-derived growth factor-α receptor activation is required for human cytomegalovirus infection. *Nat. 2008 4557211* **455**, 391–395 (2008).

187. Wu, Y. *et al.* Human cytomegalovirus glycoprotein complex gH/gL/gO uses PDGFR-α as a key for entry. *PLOS Pathog.* **13**, e1006281 (2017).

188. Wu, K., Oberstein, A., Wang, W. & Shenk, T. Role of PDGF receptor-α during human cytomegalovirus entry into fibroblasts. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9889–E9898 (2018).

189. De Clercq, E. & Li, G. Approved antiviral drugs over the past 50 years. *Clin. Microbiol. Rev.* **29**, 695–747 (2016).

190. Andrae, J., Gallini, R. & Betsholtz, C. Role of platelet-derived growth factors in physiology and medicine. *Genes Dev.* **22**, 1276–1312 (2008).

191. Chen, P. H., Chen, X. & He, X. Platelet-derived growth factors and their receptors: Structural and functional perspectives. *Biochim. Biophys. Acta - Proteins Proteomics* **1834**, 2176–2186 (2013).

192. Fretto, L. J. *et al.* Mechanism of platelet-derived growth factor (PDGF) AA, AB, and BB binding to alpha and beta PDGF receptor. *J. Biol. Chem.* **268**, 3625–3631 (1993).

193. Li, X. *et al.* PDGF-C is a new protease-activated ligand for the PDGF α-receptor. *Nat. Cell Biol. 2000 25* **2**, 302–309 (2000).

194. Chen, P. H., Unger, V. & He, X. Structure of Full-Length Human PDGFRβ Bound to Its Activating Ligand PDGF-B as Determined by Negative-Stain Electron Microscopy. *J. Mol. Biol.* **427**, 3921–3934 (2015).

195. Berger, S. *et al.* Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. *Elife* **5**, (2016).

196. Procko, E. *et al.* A Computationally Designed Inhibitor of an Epstein-Barr Viral Bcl-2 Protein Induces Apoptosis in Infected Cells. *Cell* **157**, 1644–1656 (2014).

197. Wrenbeck, E. E., Azouz, L. R. & Whitehead, T. A. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun. 2017 81* **8**, 1–10 (2017).

198. Stegmann, C., Rothemund, F., Laib Sampaio, K., Adler, B. & Sinzger, C. The N Terminus of Human Cytomegalovirus Glycoprotein O Is Important for Binding to the Cellular Receptor PDGFRα. *J. Virol.* **93**, (2019).

199. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol. 2005 241* **24**, 79–88 (2005).

200. Ciferri, C. *et al.* Structural and biochemical studies of HCMV gH/gL/gO and pentamer reveal mutually exclusive cell entry complexes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1767–1772 (2015).

201. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210

(2002).

202. Stanton, R. J. *et al.* Reconstruction of the complete human cytomegalovirus genome in a BAC reveals RL13 to be a potent inhibitor of replication. *J. Clin. Invest.* **120**, 3191 (2010).

203. Czajkowsky, D. M., Hu, J., Shao, Z. & Pleass, R. J. Fc-fusion proteins: new developments and future perspectives. *EMBO Mol. Med.* **4**, 1015–1028 (2012).

204. Vidarsson, G., Dekkers, G. & Rispens, T. IgG subclasses and allotypes: From structure to effector functions. *Front. Immunol.* **5**, 520 (2014).

205. Roopenian, D. C. *et al.* The MHC Class I-Like IgG Receptor Controls Perinatal IgG Transport, IgG Homeostasis, and Fate of IgG-Fc-Coupled Drugs. *J. Immunol.* **170**, 3528–3533 (2003).

206. Li, X. *et al.* Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter. *J. Biol. Chem.* **273**, 34970–34975 (1998).

207. Reichhart, E., Ingles-Prieto, A., Tichy, A. M., McKenzie, C. & Janovjak, H. A Phytochrome Sensory Domain Permits Receptor Activation by Red Light. *Angew. Chemie Int. Ed.* **55**, 6339–6342 (2016).

208. Deen, K. C. *et al.* soluble form of CD4 (T4) protein inhibits AIDS virus infection. *Nat. 1988 3316151* **331**, 82–84 (1988).

209. Fisher, R. A. *et al.* HIV infection is blocked in vitro by recombinant soluble CD4. *Nat. 1988 3316151* **331**, 76–78 (1988).

210. Marlin, S. D. *et al.* A soluble form of intercellular adhesion molecule-1 inhibits rhinovirus infection. *Nat. 1990 3446261* **344**, 70–72 (1990).

211. Norkin, L. C. Virus receptors: implications for pathogenesis and the design of antiviral

agents. *Clin. Microbiol. Rev.* **8**, 293–315 (1995).

212. Bialas, K. M. *et al.* Maternal CD4+ T cells protect against severe congenital cytomegalovirus disease in a novel nonhuman primate model of placental cytomegalovirus transmission. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13645–13650 (2015).

213. Mocarski, E. S., Bonyhadi, M., Salimi, S., McCune, J. M. & Kaneshima, H. Human cytomegalovirus in a SCID-hu mouse: thymic epithelial cells are prominent targets of viral replication. *Proc. Natl. Acad. Sci.* **90**, 104–108 (1993).

214. Smith, M. S. *et al.* Granulocyte-Colony Stimulating Factor Reactivates Human Cytomegalovirus in a Latently Infected Humanized Mouse Model. *Cell Host Microbe* **8**, 284–291 (2010).

215. Crawford, L. B., Streblow, D. N., Hakki, M., Nelson, J. A. & Caposio, P. Humanized Mouse Models of Human Cytomegalovirus Infection. *Curr. Opin. Virol.* **13**, 86 (2015).

216. Antoniades, H. N., Scher, C. D. & Stiles, C. D. Purification of human platelet-derived growth factor. *Proc. Natl. Acad. Sci.* **76**, 1809–1813 (1979).

217. Raines, E. W. & Ross, R. [58] Purification of human platelet-derived growth factor. *Methods Enzymol.* **109**, 749–773 (1985).

218. Petkova, S. B. *et al.* Enhanced half-life of genetically engineered human IgG1 antibodies in a humanized FcRn mouse model: potential application in humorally mediated autoimmune disease. *Int. Immunol.* **18**, 1759–1769 (2006).

219. Roopenian, D. C., Christianson, G. J. & Sproule, T. J. Human FcRn transgenic mice for pharmacokinetic evaluation of therapeutic antibodies. *Methods Mol. Biol.* **602**, 93–104 (2010).

220. Liu, J. *et al.* Cryo-Electron Microscopy Structure and Interactions of the Human

Cytomegalovirus gHgLgO Trimer with Platelet-Derived Growth Factor Receptor Alpha. *MBio* **12**, (2021).

221. Wu, N. C. *et al.* Diversity of Functionally Permissive Sequences in the Receptor-Binding Site of Influenza Hemagglutinin. *Cell Host Microbe* **21**, 742-753.e8 (2017).
222. Procko, E. The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2. *bioRxiv* 2020.03.16.994236 (2020) doi:10.1101/2020.03.16.994236.
223. Yu, D., Smith, G. A., Enquist, L. W. & Shenk, T. Construction of a Self-Excisable Bacterial Artificial Chromosome Containing the Human Cytomegalovirus Genome and Mutagenesis of the Diploid TRL/IRL13 Gene. *J. Virol.* **76**, 2316–2328 (2002).
224. Sinzger, C. *et al.* Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. *J. Gen. Virol.* **89**, 359–368 (2008).
225. Zhu, H., Shen, Y. & Shenk, T. Human cytomegalovirus IE1 and IE2 proteins block apoptosis. *J. Virol.* **69**, 7960–7970 (1995).
226. Fleishman, S. J. *et al.* Community-Wide Assessment of Protein-Interface Modeling Suggests Improvements to Design Methodology. *J. Mol. Biol.* **414**, 289–302 (2011).