Supporting Information

Rui Qing[a,1], Qiuyi Han[a], Michael Skuhersky[a], Haeyoon Chung[a], Myriam Badr[b], Thomas Schubert[c] and Shuguang Zhang[a,1]

[a]Center for Bits and Atoms, Media Lab, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.

[b]NanoTemper Technologies, 150 Cambridge Park Drive Suite 704, Cambridge, MA 02140, USA

[c]2bind GmbH, Am BioPark 11, 93053 Regensburg, Germany

Email: Shuguang@MIT.EDU, Ruiqing@MIT.EDU, Tel.: +1-617-258-7514

[1]To whom correspondence should be addressed.

MATERIALS AND METHODS:

Chimera receptor design:

CCR5$_{QTY}$ receptor

```
N-terminus: MDYQVSSPIYDINYYTSEPCQKINVKQIAAR
EC1:        AAAQWDFGNTMCQ
EC2:        YTRSQKEGLHYTCSSHFPYSQYQFWKNFQTLKI
EC3:        QEFFGLNNCSSSNRLDQ
```

Chimera A (with CCR5 N-terminus and 3 SG EC loops)

```
N-terminus: MDYQVSSPIYDINYYTSEPCQKINVKQIAAR
EC1:        SGGGSGGGASSCG
EC2:        SGGGSGGGASSGCGAGGGASGASSGGGAGGGAS
EC3:        SGGGSGGGCSSGGGAGG
```

Chimera B (with CXCR4 N-terminus and 3EC loops)

```
N-terminus: MEGISIYTSDNYTEEMGSGDYDSMKEPCFREENANFNKT
EC1:        ANWYFGNFQCK
EC2:        ANVSEADDRYICDRFYPNDLW
EC3:        DSFILLEIIKQGCEFENTVHKW
```

Both **Chimera A** and **Chimera B** contain no hydrophobic TM amino acids as confirmed by bioinformatics analysis (Fig. S4).

K_d fitting model:

K_d model is the standard fitting model based on law of mass action.

Curve fit formula:

$$F(c_T) = F_u + (F_b - F_u) * \frac{c_{AT}}{c_A}$$

$$\frac{c_{AT}}{c_A} = fraction\ bound = \frac{1}{2c_A} * (c_T + c_A + K_D - \sqrt{(c_T + c_A + K_D)^2 - 4c_T c_A})$$

F_u: fluorescence in unbound state
F_b: fluorescence in bound state
K_D: dissociation constant, to be determined
c_{AT}: concentration of formed complex
c_A: constant concentration of molecule A (fluorescent), known
c_T: concentration of molecule T in serial dilution

MicroScale Thermophoresis Measurement. MicroScale Thermophoresis (MST) is an optical method detecting changes in thermophoretic movement and TRIC of the protein-attached fluorophore upon ligand binding. Active labelled proteins contribute to the thermophoresis signal upon ligand binding. Inactive proteins influence the data as background but not the signals and only data from binding proteins are used to derive the K_d value. Herein ligand binding experiments were carried out with 5nM NT647-labeled protein ($CCR5_{QTY}$, $CXCR4_{QTY}$, $CCR10_{QTY}$, $CXCR5_{QTY}$ and $CXCR7_{QTY}$) in storage buffer (50mM Tris-HCl pH 9.0, 100mM Arginine) with 0.0916nM–3,000nM of the respective ligand, 0.153nM–5,000nM insulin as control and 0.0153nM–500nM $gp_{41\text{-}120}$ at 80% MST power, 15% LED power in premium capillaries on a Monolith NT.115 pico instrument at 25°C. Synthesized receptors were labeled with Monolith NT™ Protein Labeling Kit RED – NHS (NanoTemper Technologies) in storage buffer so as to obtain unique fluorescent signals. MST time traces were recorded and analyzed to obtain the highest possible signal-to-noise levels and amplitudes, >5 Fnorm units. The recorded fluorescence was plotted against the concentration of ligand, and curve fitting was performed with KaleidaGraph 4.5 using the K_D fit formula derived from the law of mass action. For clarity, binding graphs of each independent experiment were normalized to the fraction bound (0 = unbound, 1 = bound). MST measurements and thermostability evaluation were performed at 2bind GmbH, Regensburg, Germany.

MST experiments of chimera A and chimera B were performed in the Center for Macromolecular Interactions at Harvard Medical School with 2nd generation Monolith NT™ Protein Labeling Kit RED – NHS. The assays buffer was 50mM Tris-HCl pH 9.0, 500mM Arginine for Chimera A and 50mM Tris-HCl pH 9.0, 200mM Arginine for Chimera B.

NanoDSF measurements of the thermostability of the QTY variants. For thermal unfolding experiments, chimera A and chimera B were diluted to a final concentration of 0.2 mg/ml. For each condition, 10µl of sample per capillary was prepared. The samples were loaded into UV capillaries (NanoTemper Technologies) and experiments were carried out using the Prometheus NT.48. The temperature gradient was set to increase 1°C/min in a range from 20°C to 90°C. Protein unfolding was measured by detecting the temperature-dependent change in tryptophan fluorescence at emission wavelengths of 330nm and 350nm. Melting temperatures were determined by detecting the maximum of the first derivative of the fluorescence ratios (F330/F350). The first derivative of the fit and the peak position (at T_m) was then determined. Three independent experiments were performed.

Refolding yield evaluation:
Refolding yield and protein stability of $CCR5_{QTY}$ and chimera receptors were evaluated on a solubility basis (soluble yield) (1). Affinity-based refolding yield cannot be measured due to the lack of standard reference for these novel proteins.

During the evaluation, denatured proteins were subsequently dialyzed into renaturation buffer and storage buffer with decreasing arginine content from 500 mM to none. Following each dialysis

step, full portion of proteins were vigorously pipetted and mixed. Protein concentrations were measured by an Implen Nanodrop system before and after centrifuge (16,000×g, 10min). The soluble yield in each step was defined to be the ratio between the protein concentration after the centrifuge and that before the centrifuge. The data was summarized in S1.

Circular Dichroism:

CD spectra were recorded using JASCO Model J-1500 Circular Dichroism Spectrometer in biophysical instrumentation facility, MIT. The $CCR5_{QTY}$ and chimera protein samples were dialyzed against a slightly modified renaturation buffer (50mM Tris.HCl pH9.0, 3mM reduced glutathione, 1mM oxidized glutathione, 5mM ethylenediaminetetraacetic acid, and 40% 2,2,2-trifluoroethanol) and later into CD buffer [10mM sodium phosphate (pH6.0), 200mM sodium fluoride, 1mM Tris(2-carboxyethyl) phosphine, and 40% 2,2,2-trifluoroethanol]. The renaturation was an essential step to prevent protein from complete crash. For far UV CD, spectra between 195nm and 250nm were collected with a 0.5nm step size, 1nm bandwidth, and 50nm/min scanning speed in 0.1 cm path length cuvettes. The protein concentrations were ~2.4μM.

Table S1. Refolding Yields of $CCR5_{QTY}$, Chimera A and Chimera B based on solubility (soluble yield)

	$CCR5_{QTY}$	Chimera A	Chimera B
Renaturation Buffer	~97%	~98%	>99%
50mM Tris-HCl pH 9.0, 500mM Arginine	>99%	>99%	>99%
50mM Tris-HCl pH 9.0, 200mM Arginine	>99%	>99%	>99%
50mM Tris-HCl pH 9.0, 50mM Arginine	~98%	>99%	~98%
50mM Tris-HCl pH 9.0, 20mM Arginine	~30%	>99%	>99%
50mM Tris-HCl pH 9.0	~77%	>99%	~85%

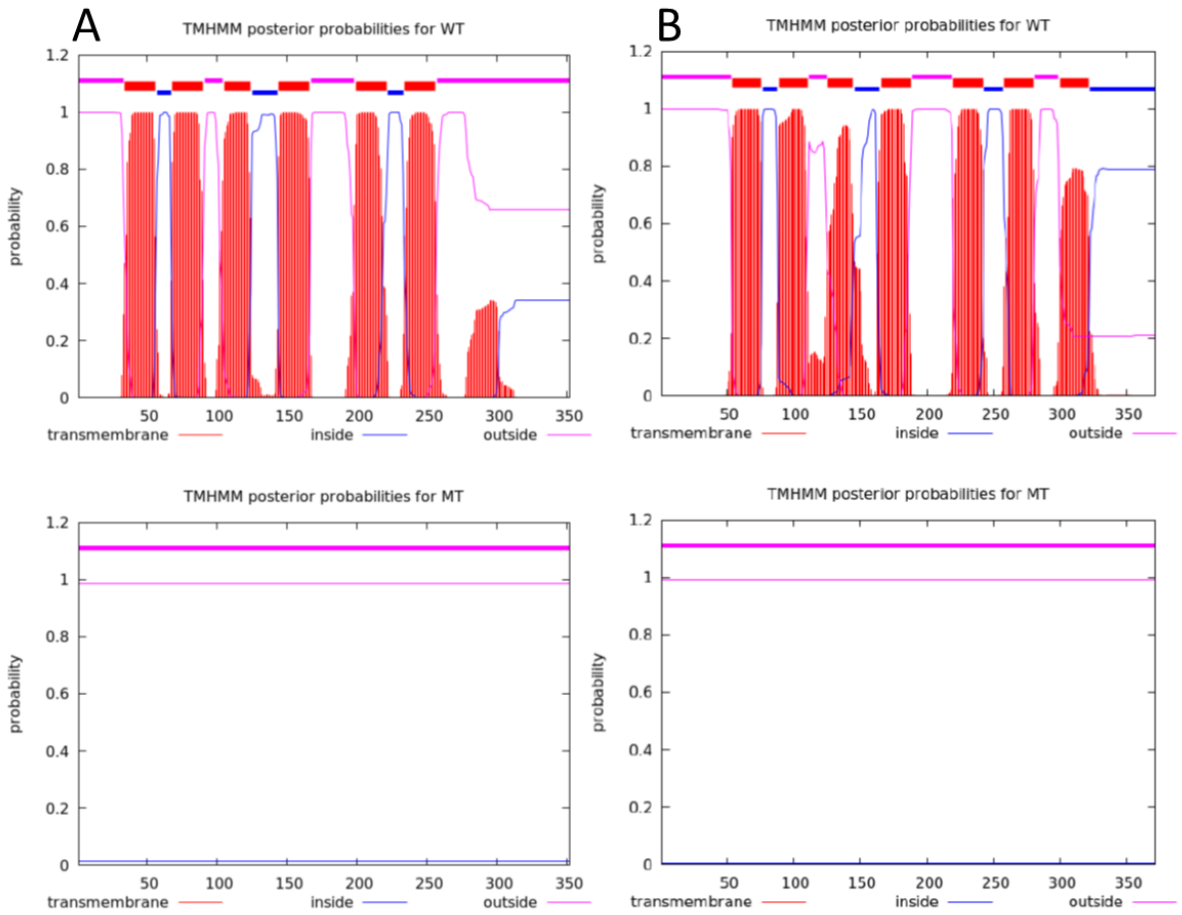

Fig. S1. Bioinformatics hydrophobic segment analyses of native (upper) (A) CCR5, (B) CXCR5, and QTY designed (Lower) (A) CCR5QTY, (B) CXCR5QTY using TMHMM 2.0. The hydrophobicity of each protein is plotted vs the sequence. Native CCR5 and CXCR5 shows 7 distinct hydrophobic transmembrane regions. These patches disappear in QTY variants of the two proteins, indicating the loss of hydrophobicity. X-axis refers to the number of amino acids in the protein from N-terminus to C-terminus. Y-axis refers to the probability of hydrophobic helical segments. Blue line = intracellular regions, pink line – extracellular regions, red line = transmembrane regions.

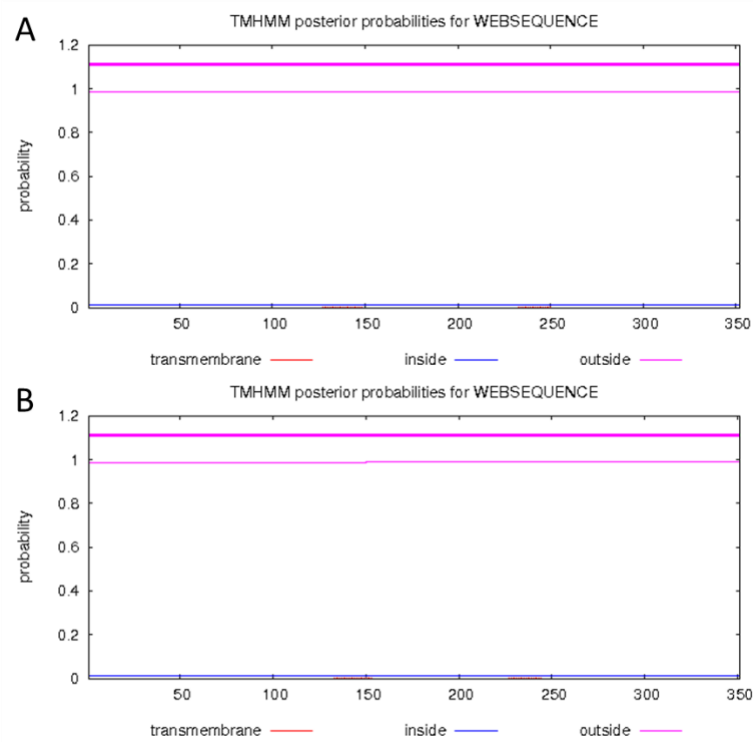

Fig. S2. Bioinformatics hydrophobic segment analyses of (A) Chimera A and (B) Chimera B. No hydrophobic transmembrane region can be observed in either of the two proteins. X axis refers to the number of amino acids in the protein from N-terminus to C-terminus. Y-axis refers to the probability of hydrophobic helical segments. Blue line = intracellular regions, pink line – extracellular regions, red line = transmembrane regions.

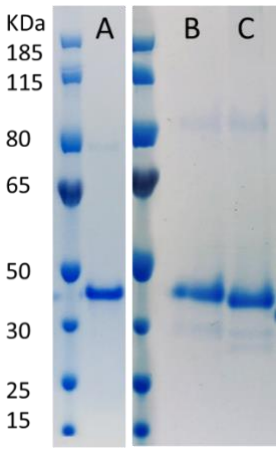

Fig. S3. Electrophoresis of purified QTY code-designed receptor proteins. (A) $CCR5_{QTY}$, (B) Chimera A and (C) Chimera B receptor proteins. The molecular weight of the ladder is labelled on the left in KDa.

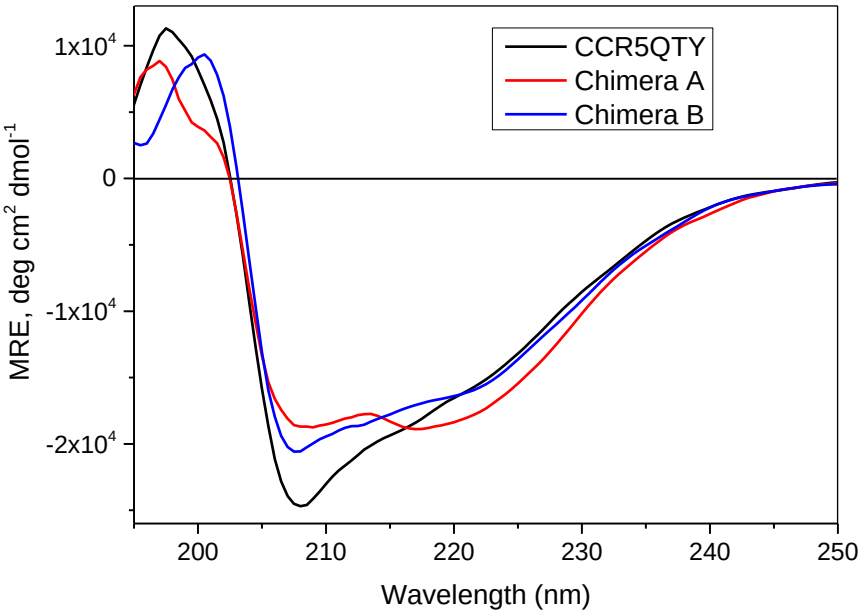

Fig. S4. Far UV circular dichroism spectra of QTY code-designed receptor proteins $CCR5_{QTY}$, Chimera A and Chimera B receptor proteins. The CD signal between 195nm and 250nm shows the typical α-helical spectra. GPCRs have 7-transmembrane helices with an overall α-helix content of 50%. $CCR5_{QTY}$ and Chimera B show similar α-helical spectra. Chimera A exhibited higher content of α-helix due to lower molecular weight in the GS-linker-replaced EC loops.

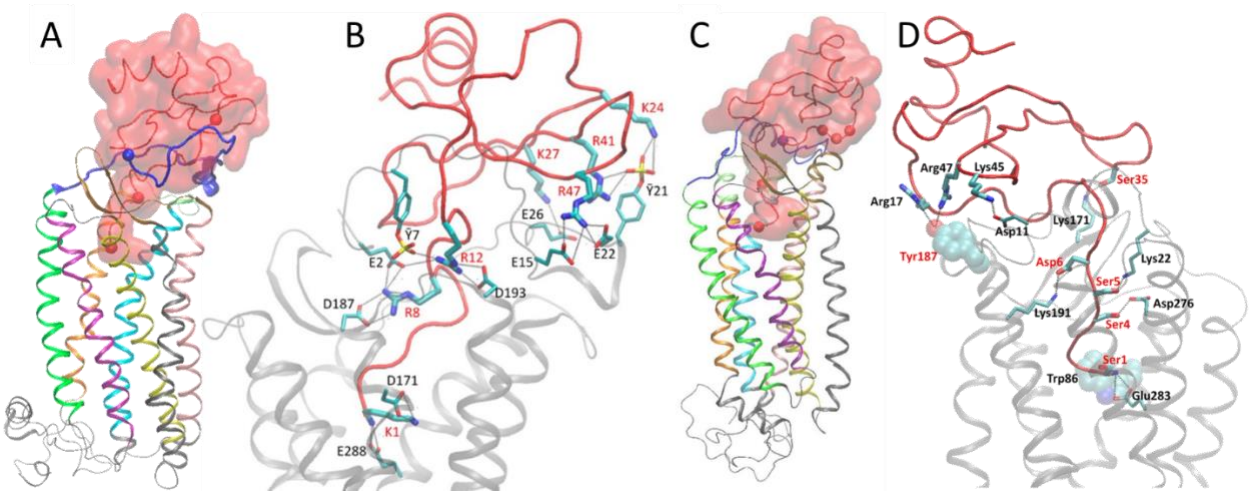

Fig. S5. Computer simulation of CXCR12-CXCL12 and CCR5-CCL5 binding model. (A) Entire simulation system of the CXCL12 : CXCR4 complex Structure; (B) salt bridges between CXCL12 and CXCR4 key residues; (C) entire simulation system of the CCL5 : CCR5 complex; (D) salt bridges, important hydrogen bonds and important cation-p interactions formed between CCL5 and CCR5 key residues. (Reproduced with permission from Tamamis *et al* and Tamamis *et al*) (2, 3)

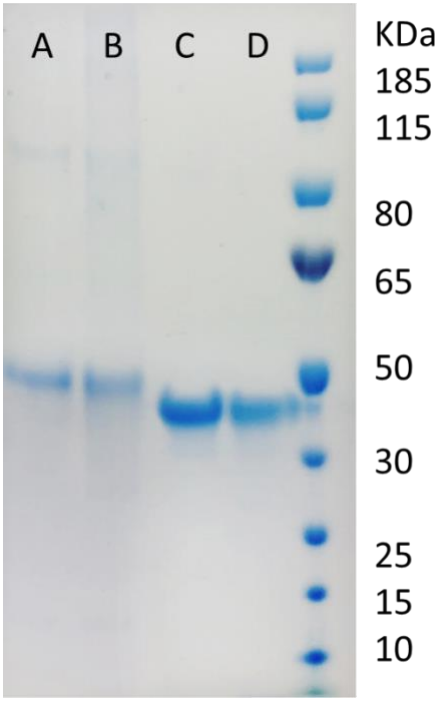

Fig. S6. Electrophoresis gel band of CXCR4$_{QTY}$ and CCR10$_{QTY}$ before and after 100°C heat-treatment. (A) CXCR4$_{QTY}$ before, (B) CXCR4$_{QTY}$ after, (C) CCR10$_{QTY}$ before, (D) CCR10$_{QTY}$ after. The molecular weight of the ladder is labelled on the right in kDa.

Reference:

1. Ho JGS & Middelberg APJ (2004) Estimating the potential refolding yield of recombinant proteins expressed as inclusion bodies. *Biotechnol Bioeng* 87(5):584-592.
2. Tamamis P & Floudas CA (2014) Elucidating a key anti-HIV-1 and cancer-associated axis: the structure of CCL5 (Rantes) in complex with CCR5. *Sci Rep-UK* 4.
3. Tamamis P & Floudas CA (2014) Elucidating a key component of cancer metastasis: CXCL12 (SDF-1α) binding to CXCR4. *J Chem Inf Model* 54(4):1174-1188.