QTY code enables design of detergent-free chemokine receptors that retain ligand-binding activities

Shuguang Zhang[a,1], Fei Tao[a,b], Rui Qing[a], Hongzhi Tang[a,b], Michael Skuhersky[a], Karolina Corin[a], Lotta Tegler[a], Asmamaw Wassie[a], Brook Wassie[a], Yongwon Kwon[c], Bernhard Suter[c], Clemens Entzian[d], Thomas Schubert[d], Ge Yang[e], Jörg Labahn[e], Jan Kubicek[f], and Barbara Maertens[f]

[a]Center for Bits and Atoms, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiaotong University, 200240 Shanghai, China; [c]Next Interactions, Inc., Richmond, CA 94806; [d]2bind GmbH, 93053 Regensburg, Germany; [e]Centre for Structural Systems Biology, Research Center Juelich, D-22607 Hamburg, Germany; and [f]Cube Biotech, GmbH, 40789 Monheim, Germany

Edited by William F. DeGrado, University of California, San Francisco, CA, and approved August 2, 2018 (received for review June 27, 2018)

Structure and function studies of membrane proteins, particularly G protein-coupled receptors and multipass transmembrane proteins, require detergents. We have devised a simple tool, the QTY code (glutamine, threonine, and tyrosine), for designing hydrophobic domains to become water soluble without detergents. Here we report using the QTY code to systematically replace the hydrophobic amino acids leucine, valine, isoleucine, and phenylalanine in the seven transmembrane α-helices of CCR5, CXCR4, CCR10, and CXCR7. We show that QTY code-designed chemokine receptor variants retain their thermostabilities, α-helical structures, and ligand-binding activities in buffer and 50% human serum. CCR5^{QTY}, CXCR4^{QTY}, and CXCR7^{QTY} also bind to HIV coat protein gp41-120. Despite substantial transmembrane domain changes, the detergent-free QTY variants maintain stable structures and retain their ligand-binding activities. We believe the QTY code will be useful for designing water-soluble variants of membrane proteins and other water-insoluble aggregated proteins.

protein design | alpha-helix engineering | hydrophobic to hydrophilic | GPCR | membrane proteins

It is well known that it is notoriously difficult to study the structure and function of membrane proteins, particularly G protein-coupled receptors (GPCRs) and multiple-segment transmembrane proteins (1, 2) because they require detergents after they are removed from cell membranes. Previous studies have applied computational methods to render membrane proteins water soluble by assigning specific changes in the transmembrane segments of several membrane proteins (*SI Appendix*, Table S1) (3–11).

One of the earliest attempts was to render bacteriorhodopsin water soluble using the known crystal structures (3). Despite rational design that changed 14.9% of the surface hydrophobic residues, after purification the designed bacteriorhodopsin had limited water solubility and lost its purple color, which is the indication of its correct folding and function (3). Our early efforts in changing only the surface residues using the crystal structures for guidance also failed to produce detergent-free chemokine receptors. Perez-Aguilar et al. (9) designed a variant of the human mu opioid receptor. However, 0.1% SDS or 1% Triton X-100 was required during purification, and 0.2% SDS was required later. CD was done in 0.01% SDS. No study was done without SDS.

Several groups using computational methods successfully converted insoluble α-helical segments from membrane proteins into water-soluble segments that retained their structures (5–8). The best example is the water-solubilized membrane channel protein KcsA, where 33 of 160 residues (20.6%) in the full-length protein or 33 of 81 residues (40.7%) in the transmembrane segment residues were changed while retaining the structure and function of KcsA (5, 7). However, computational methods require special skills and specific computer programs to carry out such membrane protein conversions, limiting widespread use.

We observed by analyzing the electron density maps of the 20 amino acids that several hydrophobic amino acids closely resemble several hydrophilic amino acids in structure and electron density (Fig. 1*A*). This similarity triggers erroneous charging of tRNAs and misincorporation into proteins. For example, the valine (V) tRNA synthetase (ValRS) (12) mischarges threonine (T) and isoleucine (I) at a rate of 1 per 200–400 tRNA charges (13–14). All 20 amino acids, both soluble and insoluble, are found in α-helices (hemoglobin is one example) (refs. 15, pp. 523–532 and 16–19), although some have higher propensities to form α-helices than others (17). This suggests that substitution of some amino acid residues that share similar structures may not significantly affect protein structure or function.

We reasoned that it might be possible to design chemokine receptor sequences in which the hydrophobic residues are replaced by water-soluble residues. This process that we have developed is a system we call the "QTY code."

The QTY code is based on the fact that the electron density map of hydrophobic leucine (L) is similar to that of hydrophilic

Significance

The QTY (glutamine, threonine, and tyrosine) code-designed detergent-free chemokine receptors may be useful in many applications. The QTY variants may be useful not only as reagents in deorphanization studies but also for designing biologics to treat cancer and autoimmune or infectious diseases. The QTY code allows membrane proteins to be systematically designed through simple, specific amino acid substitutions. The QTY code is robust and straightforward: It is the simplest tool to carry out membrane protein design without sophisticated computer algorithms. Thus it can be used broadly. The QTY code has implications for designing additional G protein-coupled receptors and other membrane proteins or, potentially, for rendering water-insoluble and aggregated proteins soluble.

Author contributions: S.Z. and F.T. designed research; S.Z., F.T., R.Q., H.T., M.S., K.C., L.T., A.W., B.W., Y.K., C.E., T.S., G.Y., J.K., and B.M. performed research; M.S., B.S., and J.K. contributed new reagents/analytic tools; S.Z., F.T., R.Q., M.S., Y.K., B.S., T.S., G.Y., J.L., J.K., and B.M. analyzed data; and S.Z., F.T., R.Q., M.S., K.C., B.S., T.S., G.Y., J.L., J.K., and B.M. wrote the paper.

Conflict of interest statement: Massachusetts Institute of Technology (MIT) filed several patent applications for the QTY code, and OH2 Laboratories licensed the technology from MIT. This research was funded in part by OH2 Laboratories. S.Z. has a minor equity in OH2 Laboratories.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: shuguang@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1811031115/-/DCSupplemental.

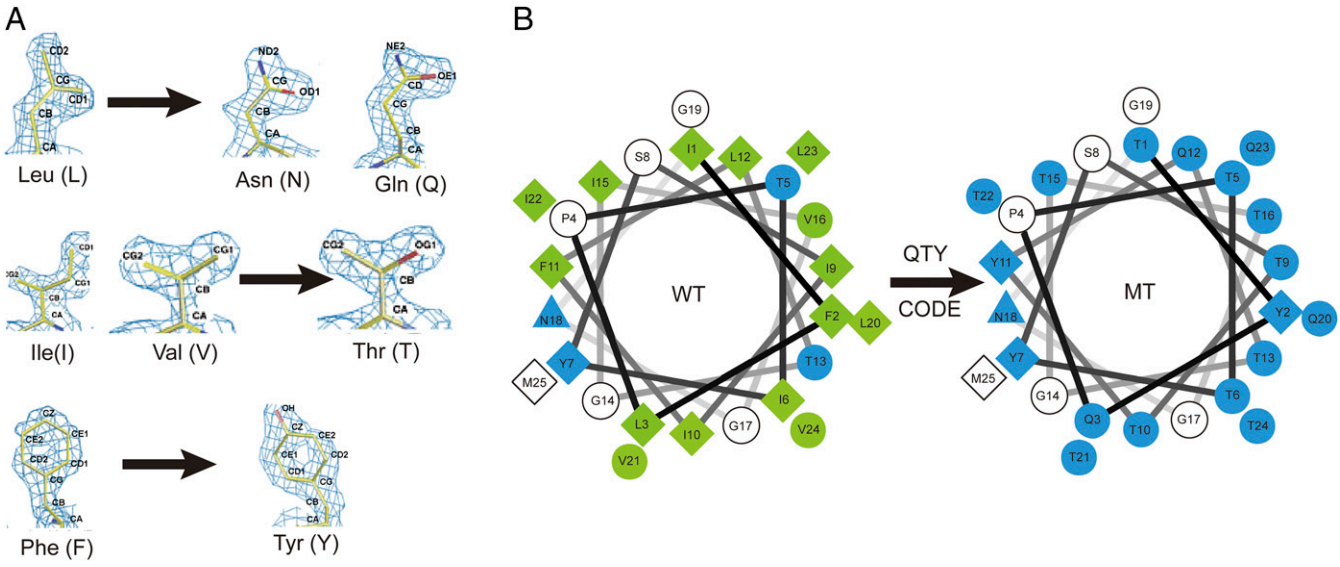

Fig. 1. The QTY code and how it replaces L, V, I, and F with Q, T, and Y. (*A*) Crystallographic electronic density maps of the following amino acids: leucine (L), asparagine (N), glutamine (Q), isoleucine (I), valine (V), threonine (T), phenylalanine (F), and tyrosine (Y). The density maps of L, N, and Q are very similar. Likewise, the density maps of I, V, and T are similar, and the density maps of F and Y are similar. The side chains of L, V, I, and F cannot form any hydrogen bonds with water, thus rendering them water insoluble. On the other hand, N and Q can form four hydrogen bonds with four water molecules, two as hydrogen donors and two as hydrogen acceptors (*SI Appendix*, Fig. S7). Likewise, three water molecules can form hydrogen bonds with the –OH (two H-donors and one H-acceptor) of threonine (T) and tyrosine (Y). Both L and Q have high tendencies to form α-helices, but N frequently occurs at turns. Thus, Q was used to replace L but not N. I, V, and T are all β-branched amino acids, and their density maps are very similar, indicating similar shapes. (*B*) Helical wheels before (*Left*) and after (*Right*) applying the QTY code to transmembrane helical segment 1 (TM1) of CXCR4. Amino acids that interact with water molecules are light blue in color. The QTY code conversions render the α-helical segment water soluble. MT, mutant.

asparagine (N) and glutamine (Q); the electron density maps of hydrophobic isoleucine (I) and valine (V) are similar to that of hydrophilic threonine (T); and the electron density map of hydrophobic phenylalanine (F) is similar to that of the hydrophilic tyrosine (Y) (Fig. 1*A* and *SI Appendix*, Fig. S1). Although water also forms hydrogen bonds with aspartic acid (−), glutamic acid (−), lysine (+), and arginine (+), these residues introduce charges, thereby altering the surface property of proteins. Thus, they were not introduced in the QTY code.

To test this hypothesis, the QTY code was applied to four chemokine receptors: CCR5, CXCR4, CCR10, and CXCR7. These receptors were chosen because they play critical roles in health and disease and because they have been well characterized (20–31). CCR5, CXCR4, and CXCR7 are also coreceptors for HIV entry into T cells (21, 22). CCR5's natural ligand is the chemokine $CCL5_{26–91}$ (Rantes), CXCR4's natural ligand is $CXCL12_{24–88}$ (SDF1α), CXCR7's natural ligands are $CXCL11_{22–94}$ and $CXCL12_{24–88}$, and CCR10's natural ligands are $CCL27_{25–112}$ and $CCL28_{20–127}$. Moreover, the crystal structures of CXCR4 and CCR5 have been published (27, 28), allowing direct comparison with the QTY variants $CCR5^{QTY}$ and $CXCR4^{QTY}$ after those structures become available. There are currently no published crystal structures of CCR10 and CXCR7, but they play key roles in various physiological functions (29–31). CCR10 and its ligands are uniquely involved in epithelial immunity, and CCR10 is expressed in subsets of innate-like T cells, which are localized to the skin during developmental processes in the thymus (29). CXCR7 is an atypical chemokine receptor that does not activate G protein-mediated signal transduction (30). Rather, CXCR7 can heterodimerize with CXCR4 to modulate CXCR4 activities and can be activated by CXCL11 in malignant cells, leading to enhanced cell adhesion and migration. Elevated levels of CXCR7 expression are correlated with aggressive human prostate, breast, and lung cancers and promote growth and metastasis of various tumors (31).

The QTY-designed variant gene sequences were organism-codon optimized for protein expression either in SF9 insect cells ($CCR5^{QTY}$, $CCR10^{QTY}$, and $CXCR7^{QTY}$) or in *Escherichia coli* ($CXCR4^{QTY}$). The detergent-free variants were then purified using His-tag affinity and size-exclusion chromatography (SEC). To assess the structure and function of these QTY variant receptors, they were subjected to thermostability studies to measure their melting temperature (Tm), to CD spectroscopy to study their secondary structure and folding, and to surface-free microscale thermophoresis (MST) to study ligand binding in buffers and in 50% human serum.

Since we have not yet obtained high-resolution structures of the detergent-free variants, $CCR5^{QTY}$, $CXCR4^{QTY}$, $CCR10^{QTY}$, and $CXCR7^{QTY}$ were simulated in an explicit water environment using three different computer programs (31–35). The simulated structures of $CCR5^{QTY}$ and $CXCR4^{QTY}$ were directly compared with the known crystal structures of natural CXCR4 (27) and CCR5 (28). These structural folds can be superimposed, suggesting that the QTY variants retain a natural overall structure despite 19–29% overall sequence changes.

Results

Sequence Alignments of CCR5, CXCR4, CCR10, and CXCR7 and Their QTY Variants. Sequence alignments were performed for CCR5 vs. $CCR5^{QTY}$, CXCR4 vs. $CXCR4^{QTY}$, CCR10 vs. $CCR10^{QTY}$, and CXCR7 vs. $CXCR7^{QTY}$. Fig. 2 shows protein characteristics and alignments of the transmembrane (TM) α-helical segments of the native receptors and their QTY variants. Since the amino acids Q, T, and Y do not introduce any charges, despite substantial sequence QTY substitutions, there are minimal changes in pI units (0.11, 0.06, 0.24, and 0.02 for CCR5, CXCR4, CCR10, and CXCR7 and their QTY variants, respectively), and in molecular weight (1.13%, 1.15%, 1.97%, and 1.13% for CCR5, CXCR4, CCR10, and CXCR7 and their QTY variants, respectively). After applying the QTY code, the hydrophobic amino acids in the transmembrane segments are replaced by Q, T, or Y, and the transmembrane segments are no longer hydrophobic (*SI Appendix*, Fig. S2).

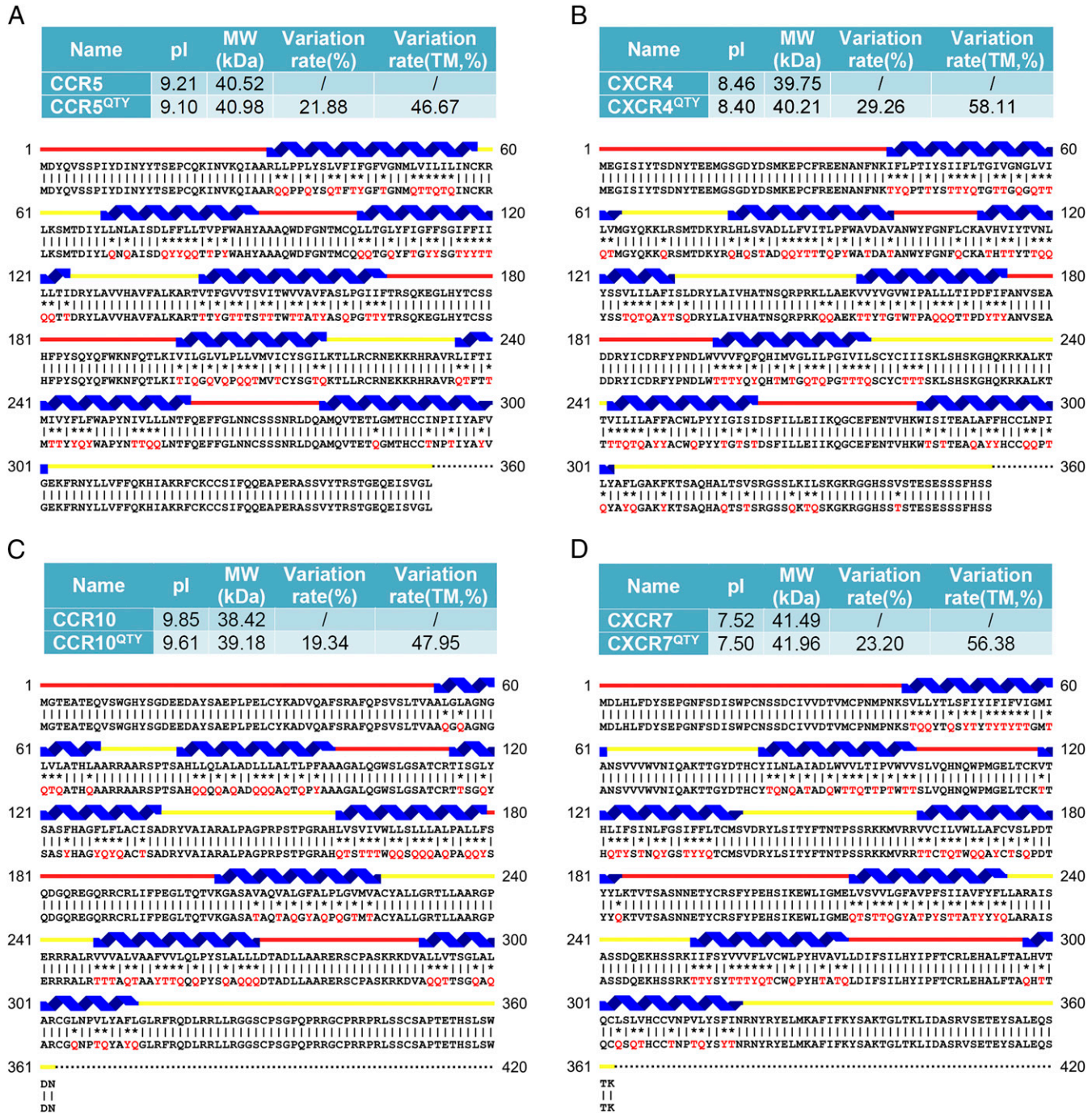

Fig. 2. Alignments of native (*Upper*) CCR5 (*A*), CXCR4 (*B*), CCR10 (*C*), and CXCR7 (*D*) with detergent-free (*Lower*) $CCR5^{QTY}$ (*A*), $CXCR4^{QTY}$ (*B*), $CCR10^{QTY}$ (*C*), and $CXCR7^{QTY}$ (*D*). The Q, T, and Y amino acid substitutions are in red. The α-helical segments (blue) are shown above the protein sequences, and the external (red) and internal (yellow) loops of the receptors are indicated. The symbols "|" and "*" indicate similar and different amino acids, respectively. Characteristics of natural and QTY variants with pI, molecular weight, total variation rate, and membrane variation rate. Since the internal regions ICL1, ICL2, ICL3, and the C terminus do not interact with the ligand SDF1α, additional residues in these regions were modified.

Initial Development of the QTY Variant of $CXCR4^{QTY}$. Based on the X-ray crystal structure of CXCR4 (26), we initially used the QTY code to change only 28 positions ($CXCR4^{QTY}$-v28) (*SI Appendix*, Fig. S3) on the lipid-facing exterior surfaces of TM1, TM2, TM4, TM6, and TM7 but not on the interior surface or dimer interface. However, the protein could not be expressed and purified without detergents. We then substituted 56 positions ($CXCR4^{QTY}$-v56) in the 7TM region using a random library of ∼2 million variants using yeast two-hybrid (Y2H) selection. When we selected 16 possible Y2H variant candidates and tried to express them in both *E. coli* and yeast, they failed to express well (*SI Appendix*, Fig. S4). In our current report, we changed 85 LIVF positions in all 7TMs of CXCR4 ($CXCR4^{QTY}$-v85). To further increase the solubility of $CXCR4^{QTY}$, we also applied the QTY code to the internal regions ICL1 (1L), ICL2 (2L), and ICL3 (3I) and to the C terminus (2F, 3L, 2V, and 1I) since these regions do not interact directly with the ligand CXCL12. One L was changed in EC1 since it was reported that EC2 and EC3 are most important for ligand binding.

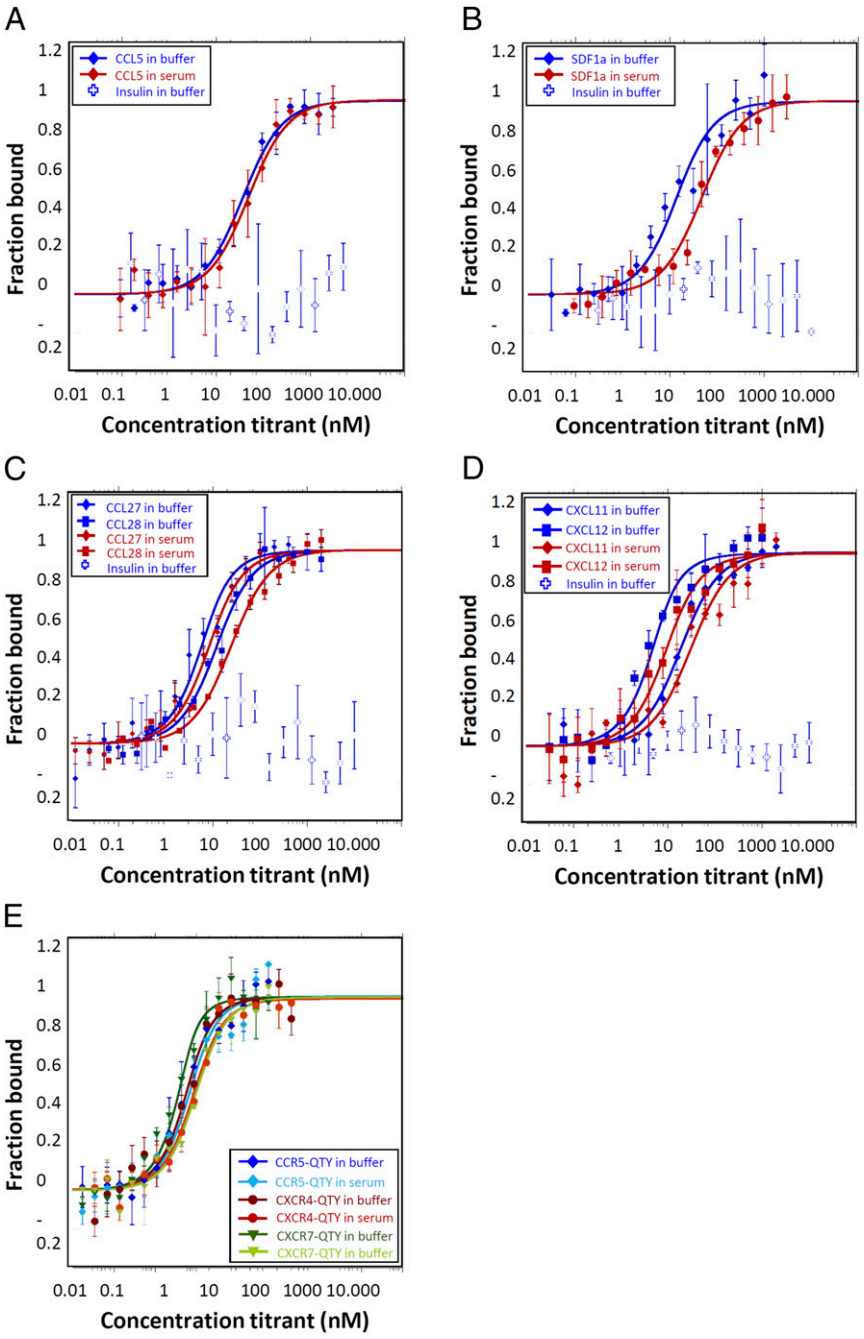

Fig. 3. MST ligand-binding measurements. The receptors were labeled with a fluorescent dye since both receptors and ligands contain tryptophans. All ligands were serially diluted in either buffer (black line) or in 50% human serum (red line). Human insulin was used as a negative control that showed no binding. The bars represent the SD of three independent experiments with duplicate measurements for each experiment (a total of six measurements for each sample). The detailed K_d numbers in buffer and in 50% human serum are presented in Table 1. (*A*) $CCR5^{QTY}$ with $CCL5_{26–91}$. (*B*) $CXCR4^{QTY}$ with $CXCL12_{24–88}$. (*C*) $CCR10^{QTY}$ with $CCL27_{25–112}$ and $CCL28_{20–127}$. (*D*) $CXCR7^{QTY}$ with $CXCL11_{22–94}$ and $CXCL12_{24–88}$. (*E*) $CCR5^{QTY}$, $CXCR4^{QTY}$, and $CXCR7^{QTY}$ with HIV-1 coat protein gp41-120.

This time, the protein expressed well in water-soluble form without any detergents and retained its ligand-binding activity for CXCL12 and gp41-120. Because of these results, QTY substitutions were introduced to all 7TMs of $CCR10^{QTY}$ and $CXCR7^{QTY}$.

Protein Expressions and Purifications of QTY Variants. We synthesized the genes with organism-specific codons and first expressed $CCR5^{QTY}$, $CXCR4^{QTY}$, $CCR10^{QTY}$, and $CXCR7^{QTY}$ in SF9 insect cells. Each receptor carried a C-terminal His-tag. The purified protein yields from insect SF9 cells were low and inadequate for structural analysis. To obtain sufficient protein for structural and other studies, $CXCR4^{QTY}$ was expressed in *E. coli* inclusion bodies that reached ~10 mg/L. The inclusion bodies were extensively washed and denatured in 6 M guanidine·HCl and 10 mM DTT, 1× PBS. $CXCR4^{QTY}$ was then purified with a His-tag and was re-refolded in a renaturing buffer containing 0.5 M L-arginine, which is a key ingredient required for correct refolding, but without DTT. SEC (*SI Appendix*, Fig. S5) of the renatured receptor

showed either monomers ($CCR10^{QTY}$) or dimers ($CXCR4^{QTY}$). This is similar to the native CXCR4, since both crystal structures (27) and cell-based assays show CXCR4 is a dimer (23).

Ligand-Binding Measurements in Buffer and in 50% Human Serum. The optical method MST (36–42) uses the fluorescence signal coming from labeled QTY proteins to monitor their movement in a thermal gradient. Fluorescence in the optical focus is localized; hence active and inactive protein fractions will contribute to the signal. Changes in thermophoresis upon binding of a ligand to the QTY proteins is detected and plotted as a ligand concentration-dependent effect, which is then converted to affinity values. Since thermophoresis of only active QTY proteins will change during a binding event, inactive proteins will influence the data only as background signal and will not contribute to the binding event. Therefore, the derived binding data come from active protein fractions.

MST was used for ligand-binding measurements in both buffer and in 50% human serum. To obtain unambiguous ligand-binding results, each sample was independently measured three times in duplicate (six total measurements). These results showed that the purified detergent-free forms of $CCR5^{QTY}$, $CXCR4^{QTY}$, $CCR10^{QTY}$, and $CXCR7^{QTY}$ retain their ligand-binding activities (Fig. 3 and Table 1). Furthermore, it is known that natural CCR5, CXCR4, and CXCR7 bind to HIV1 coat protein gp41-120. We thus carried out binding measurements for $CCR5^{QTY}$, $CXCR4^{QTY}$, and $CXCR7^{QTY}$ (Fig. 3*E* and Table 1). $CCR5^{QTY}$ from SF9 cells was independently purified twice in ∼6 mo and the ligand binding also was independently measured twice. $CXCR4^{QTY}$ from *E. coli* inclusion body purification and refolding was independently purified twice in ∼1 mo, and the ligand binding was also independently measured three times. The early purified $CXCR4^{QTY}$ had a K_d of ∼17 nM, and the late $CXCR4^{QTY}$ had a K_d of ∼13 nM.

To rule out nonspecific binding, we measured the affinity of human insulin for $CCR5^{QTY}$, $CXCR4^{QTY}$, $CCR10^{QTY}$, and $CXCR7^{QTY}$. The reproducible measurements demonstrate that these detergent-free variants do not bind to human insulin (Fig. 3), suggesting that the QTY receptors bind to their ligands with some specificity. Interestingly, $CXCR7^{QTY}$ has a lower K_d for CXCL12 and HIV gp41-120 than $CXCR4^{QTY}$ (Fig. 3 and Table 1).

Thermal Stability of the QTY Variants. To determine the thermal stability of the QTY variants, three independent nanoDSF (nano differential scanning fluorimetry) measurements were carried out (Fig. 4). The results show that the average Tms of ∼52.7 °C for $CCR5^{QTY}$, 46.8 °C (Tm_1) and 63.5 °C (Tm_2) for $CXCR4^{QTY}$, 54.8 °C for $CCR10^{QTY}$, and 52.3 °C for $CXCR7^{QTY}$ are similar to the Tms of the natural receptors. For example, natural CXCR4 embedded in liposomes exhibited two Tms: 55 °C (Tm_1) for monomers and 60 °C (Tm_2) for dimers (23). Controls were carried out by heating the proteins to 90 °C for 15 min before taking three independent measurements. The proteins were fully denatured and produced no measurable Tm (Fig. 4). These results suggest that, despite significant QTY amino acid changes, the detergent-free $CCR5^{QTY}$, $CXCR4^{QTY}$, $CCR10^{QTY}$, and $CXCR7^{QTY}$ still fold and remain thermostable. It is plausible that the QTY replacements introduce numerous hydrogen bonds from intra- and interhelical interactions (*SI Appendix*, Fig. S7).

CD and Fluorescence Studies of QTY-Variant Protein Structures. The QTY variant receptors were studied using CD, and they showed distinctive α-helical spectra. We used the purified $CCR5^{QTY}$ and $CXCR7^{QTY}$ in buffer containing 150 mM NaF and 5 mM DTT to carry out the study. Far UV spectra between 183 and 260 nm confirm typical α-helical secondary structures for $CCR5^{QTY}$ and $CXCR7^{QTY}$. Furthermore, the α-helical content of $CCR5^{QTY}$ (∼55%) and $CXCR7^{QTY}$ (∼60%) is similar to that of native CCR5 (59%) (29) and CXCR7 (64%, from secondary structural prediction) (*SI Appendix*, Fig. S6*A* and Table S2).

Tryptophan fluorescence spectra with 295-nm excitation of $CCR5^{QTY}$ and $CXCR7^{QTY}$ displayed maximum emission at ∼334 nm and ∼338 nm, respectively (*SI Appendix*, Fig. S6*B*), suggesting the mean hydrophobic microenvironment of the tryptophan side chains is neither completely hydrophilic nor hydrophobic, as expected for a folded QTY protein (43, 44). When both tryptophan and tyrosine were excited at 275 nm, the maxima of the fluorescence emission spectra shifted to ∼332 nm for both $CCR5^{QTY}$ and $CXCR7^{QTY}$, which indicates weak emission by tyrosines (*SI Appendix*, Fig. S6*B*, *Inset*) despite the high number of tyrosines in the QTY proteins (42 tyrosines and 5 tryptophans in $CCR5^{QTY}$, 31 tyrosines and 7 tryptophans in $CXCR7^{QTY}$). The weakening of tyrosine fluorescence centered at 303 nm is due to the Förster energy transfer from tyrosine to nearby tryptophan residues (43, 44). This indicates that the QTY proteins fold into a compact tertiary structure with the expected secondary structure content.

Computer Simulations of the QTY Variants in an Explicit Water Environment at 24.85 °C, pH 7.4, and 0.9% NaCl for 1 μs. High-resolution structures of $CCR5^{QTY}$, $CXCR4^{QTY}$, $CCR10^{QTY}$, or

Table 1. Ligand-binding affinity of native and QTY code-designed chemokine receptors

Receptors	CCL5* K_d, nM (ref.)	CXCL11 K_d, nM (ref.)	CXCL12* K_d, nM (ref.)	CCL27 K_d, nM (ref.)	CCL28 K_d, nM (ref.)	gp41-120 K_d, nM (ref.)
CCR5 native	∼4 (26)					∼10 (26)
$CCR5^{QTY}$ buffer	33.9 ± 4.8					3.1 ± 0.7
$CCR5^{QTY}$ serum						4.3 ± 1.5
CXCR4 native			∼5 (26)			∼200† (23)
$CXCR4^{QTY}$ buffer‡			11.2 ± 3.4			7.0 ± 1.9
$CXCR4^{QTY}$ serum‡			44.7 ± 8.9			12.7 ± 2.6
CCR10 native				∼5.6 (29)	38 (29)	
$CCR10^{QTY}$ buffer				3.1 ± 1.2	9.3 ± 1.8	
$CCR10^{QTY}$ serum				5.6 ± 1.1	21 ± 4	
CXCR7 native		∼8 (30)	∼4.5 (30)			N/D
$CXCR7^{QTY}$ buffer		16 ± 3	2.2 ± 0.7			1.2 ± 0.4
$CXCR7^{QTY}$ serum		28 ± 11	6.6 ± 1.7			7 ± 1.5

$CCR5^{QTY}$, $CCR10^{QTY}$, and $CXCR^{7QTY}$ were purified from insect SF9 cells. N/D, no data.
*CCL5 is also called "Rantes," and CXCL12 is also called "SDF1α" in the literature.
†The K_d ∼200 nM was measured by a cell-based assay.
‡$CXCR4^{QTY}$ was purified from *E. coli* inclusion bodies and renatured in refolding buffer with 0.5 M arginine.

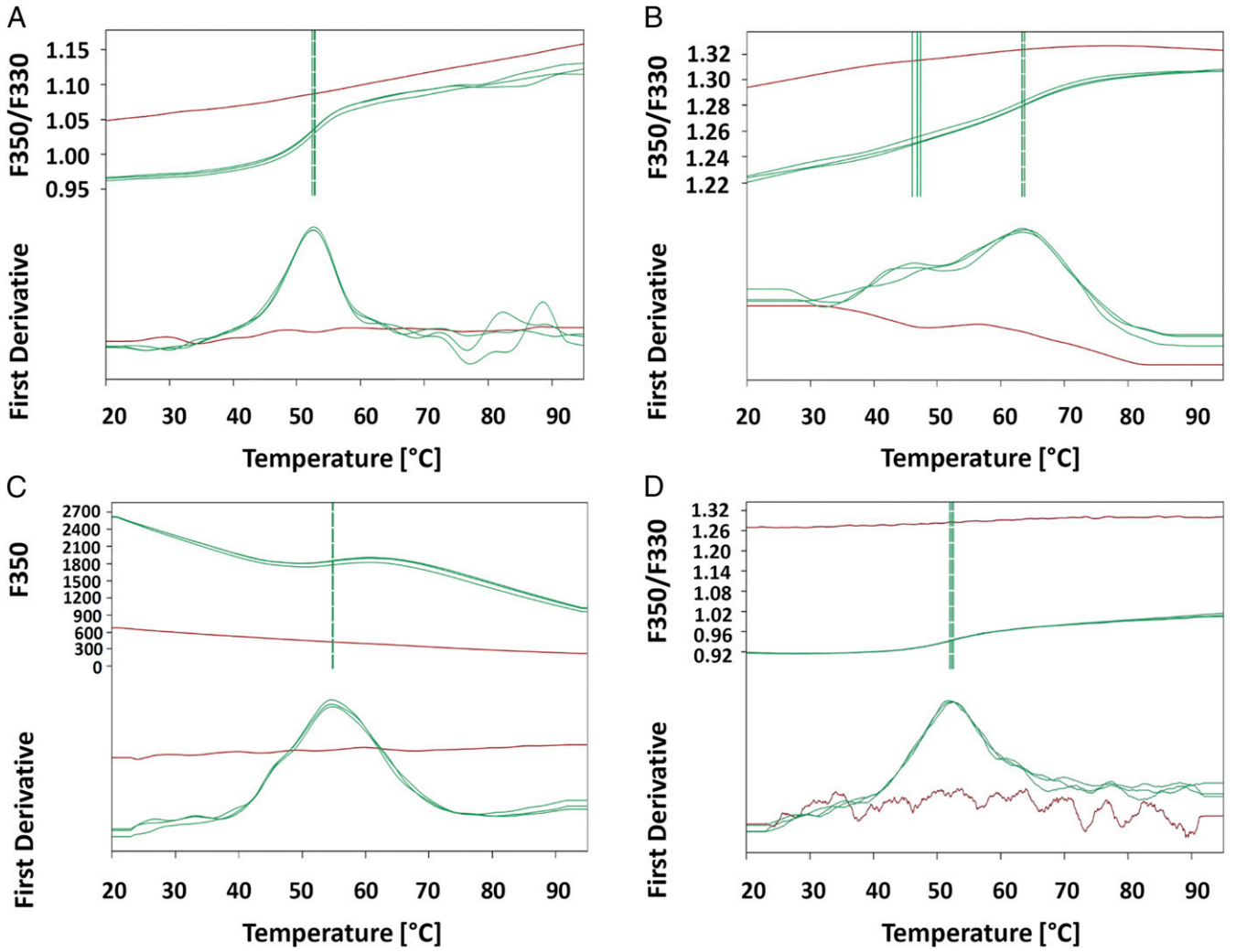

Fig. 4. Thermostability of the chemokine receptors CCR5^{QTY}, CXCR4^{QTY}, CCR10^{QTY}, and CXCR7^{QTY} measured using nanoDSF. To obtain Tm curves (green lines), the QTY-designed receptors were heated gradually to denature them slowly. In the controls (red lines), the proteins were heated to 90 °C for 15 min before taking the measurements. (*A*) These experimental results show that CCR5^{QTY} has a Tm of ∼52.7 °C. (*B*) CXCR4^{QTY} exhibits two transition temperatures: Tm_1 at 46.8 °C and Tm_2 at ∼63.5 °C. This double transition is similar to the behavior seen with the native CXCR4 embedded in liposomes or cell membranes which also exhibited two transition temperatures (Tm1 ∼55 °C and Tm2 ∼60 °C). (*C*) CCR10^{QTY} has a Tm of ∼54.8 °C. (*D*) CXCR7^{QTY} has a Tm of ∼52.3 °C. Since there are many additional QTY intra- and interhelical hydrogen bonds inside the proteins, the receptor structures may fold and remain stable via extensive hydrogen bonds within the protein and water molecule bridges.

CXCR7^{QTY} are still in progress. However, recent advances in computer simulations of protein sequences make it possible to predict reasonably realistic structures based on homology (33–35). We tested whether the CCR5^{QTY}, CXCR4^{QTY} (Fig. 5), CCR10^{QTY}, and CXCR7^{QTY} (*SI Appendix*, Fig. S8) structures are stable by simulating them in an explicit water environment at 24.85 °C, pH 7.4, and 0.9% NaCl for 1 μs (Fig. 5). If they are not stable, these structures will not fold correctly. After an initial 0.3 μs of simulations using the AMBER14 force field software (43), the overall structures were already formed and seemed to be stable; additional 0.7-μs simulations did not further stabilize these structures. After the simulations, CXCR4^{QTY} and CCR5^{QTY} were superimposed with crystal structures of the natural receptors. The comparisons showed that CXCR4^{QTY} and CXCR4 [Protein Data Bank (PDB) ID code 3ODU] (27) had a deviation of ∼1.9 Å, and CCR5^{QTY} and CCR5 (PDB ID code 4MBS) (28) had a deviation of ∼2 Å.

Discussion

The Basis of the QTY Code. The scientific and structural basis of the QTY code is the fact that the electronic density maps of Q and L, T and V/I, and Y and F are similar and that all 20 amino acids are found in α-helices (15–18), although some residues, e.g., L and Q, are preferred. Specifically, the QTY code shows that amino acid structures rather than chemical properties may facilitate protein structures in transmembrane α-helices (Fig. 1*A* and *SI Appendix*, Figs. S1–S3). Despite ∼50% changes in the transmembrane segments (*SI Appendix*, Fig. S2), the QTY code-designed CCR5^{QTY}, CXCR4^{QTY}, CCR10^{QTY}, and CXCR7^{QTY} not only maintained their overall structure but also bound their respective ligands (Fig. 3 and Table 1).

Measuring Ligand Binding in Human Serum. We measured ligand binding of the QTY code-designed chemokine receptors in both buffer and 50% human serum. The receptors have approximately two to four times lower affinity in serum than in buffer (Fig. 3 and Table 1). These differences are expected since human serum is very complex and contains numerous substances, including 20 amino acids, metabolic intermediates, peptides, proteins, and more. However, the affinities measured in serum are perhaps

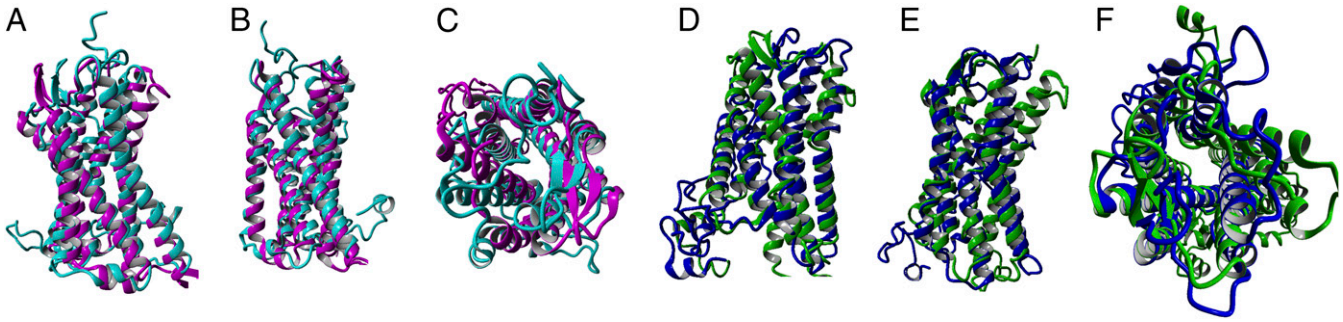

Fig. 5. Computer simulations of CCR5^{QTY} and CXCR4^{QTY} are superimposed with the crystal structures of CCR5 and CXCR4. Computer simulations of CCR5^{QTY}, CXCR4^{QTY}, CCR10^{QTY}, and CXCR7^{QTY} were carried out in an explicit water environment. The X-ray crystal structures of natural CCR5 (PDB ID code 4MBS) and CXCR4 (PDB ID code 3ODU) were obtained from the Protein Data Bank. The protein structures were determined with a rubredoxin (CCR5) or T4 lysozyme (CXCR4) insert in the third internal loop. The simulated CCR5^{QTY} and CXCR4^{QTY} do not have rubredoxin or lysozyme inserts. For clarity, these inserts are removed in comparisons with CCR5^{QTY} and CXCR4^{QTY}. The structures were obtained after 1 μs of simulation in an explicit water environment at 24.85 °C at pH 7.4 and 0.9% NaCl ion concentration. (*A*–*C*) CCR5^{QTY} (teal) was superimposed with its natural counterpart CCR5 (magenta). It has a deviation of ∼1.9 Å and is shown in two different side views in *A* and *B*, and in a top view is shown in *C*. (*D*–*F*) CXCR4^{QTY} (blue) was superimposed with its natural counterpart CXCR4 (green), with a deviation of ∼2 Å. Two side views are shown in *D* and *E*, and a top view is shown in *F*.

more realistic for physiological conditions, since the receptors are often exposed to a serum-like environment.

Possible Additional Hydrogen Bond Interactions. We asked how the protein structures remain stable and retain ligand-binding activity even after substantial substitutions of the hydrophobic residues in 7TM. Additional hydrogen bonds (*SI Appendix*, Fig. S7) created by the Q, T, and Y amino acid substitutions likely contribute to the observed stability: Numerous internal intrahelical and interhelical hydrogen bonds that stabilize the structures of the QTY variants are likely formed. This situation is similar to the molecular structures of various collagens that have extensive water molecule bridges stabilizing their structures. These hydrogen bonds cannot form in the native CCR5, CXCR4, CCR10, and CXCR7, since the side chains of L, V, I, and F do not have hydrogen bond-forming capabilities.

Possible Uses of the QTY Code-Designed Proteins. The QTY code-designed detergent-free chemokine receptors may be useful in many applications. It is possible to use QTY variant receptors in a manner similar to water-soluble kinases and proteases for drug discoveries. They may potentially be used as reagents in deorphanization studies. It also may be possible to use them as biologics to treat cancer and autoimmune or infectious diseases.

The Latest Work by Others. Recently, researchers in William DeGrado's laboratory designed a 25-residue α-helical peptide called "αAM_{mem}" that is intrinsically water-insoluble and forms a cross-α amyloid-like structure. They converted it into a water-soluble form called "αAM_s" by changing six residues: L5E, I8R, I9K, L16K, L19K, and I22E (6/25 = 24%) (45). They determined and compared these crystal structures. The structures of the water-insoluble αAM_{mem} and water-soluble αAM_s are remarkably similar (figure 2 d–f in ref. 45). Their results further demonstrate that specific changes (6/25) in α-helical segments can still preserve the stable α-helical structure.

Simplicity Is the Ultimate Sophistication (Leonard da Vinci). The QTY code may allow membrane proteins to be systematically designed through simple, specific amino acid substitutions (Fig. 1 and *SI Appendix*, Table S1). We believe that the QTY code is robust and straightforward. It is the simplest tool to carry out membrane protein design without sophisticated computer algorithms and can be used broadly. It is plausible that the QTY code may have implications for designing additional GPCRs and other membrane proteins or perhaps water-insoluble and aggregated proteins.

Materials and Methods

Bioinformatics of the QTY Variants. The variant protein sequences were first evaluated to determine if transmembrane segments still existed using the web-based tool TMHMM Server v.2.0 (www.cbs.dtu.dk/services/TMHMM-2.0/) that predicts transmembrane helices (*SI Appendix*, Fig. S2).

Y2H Assays. We initially used a Y2H assay (*SI Appendix*, Fig. S4) to study interactions between the QTY variants and their respective ligands. We used a Y2H system as an in vivo assay to test the QTY variants for ligand binding. If the variants interacted with their ligands, the receptor–ligand pairs activated gene transcription, thus enabling yeast cell growth. The variants were further subjected to control assays to eliminate false positives.

Protein Purification. For each protein, a two-step purification strategy of affinity chromatography combined with SEC was applied. Since the amount of protein binding to the affinity chromatography resin was initially low, we screened for different additives to improve the purification yields. Among them, ammonium sulfate and 10 mM DTT were very important. In the early phase of purification, 10 mM DTT resulted in a higher amount of purified protein, since CCR5^{QTY}, CXCR4^{QTY}, CCR10^{QTY}, and CXCR7^{QTY} have 12, 9, 10, and 14 cysteines, respectively. (Detailed methods are given in *SI Appendix*).

Receptor Labeling. Since the CCR5^{QTY}, CXCR4^{QTY}, CXCR7^{QTY}, and CCR10^{QTY} receptors and their respective ligands (CCL5, CXCL11, CXCL12, CCL27, and CCL28) contain tryptophans, the receptors need to be labeled with NT647 in 1× PBS (pH 7.4) to reduce noise and obtain unique fluorescent signals. These receptors were labeled according to the instructions of the Monolith NT Protein Labeling Kit RED-NHS (NanoTemper Technologies). The concentration of labeled proteins was determined using NanoDrop and Bradford assays.

MST Measurements. Ligand-binding experiments were carried out using MST with 5 nM NT647-labeled protein (CCR5^{QTY}, CXCR4^{QTY}, CCR10 ^{QTY}, or CXCR7 ^{QTY}) in binding buffer (1× PBS and 5 mM DTT) with 0.0916–3,000 nM of Rantes or SDF1α, 0.153–5,000 nM insulin, 0.0651–2,000 nM CCL28 and CXCL11, 0.012–400 nM CCL27, or 0.0153–500 nM gp41-120 at 80% MST power, 15% LED power in premium capillaries on a Monolith NT.115 pico instrument at 25 °C.

nanoDSF Determination of the Thermal Stability of the QTY Variants. For thermal unfolding experiments, CCR5^{QTY}, CXCR4^{QTY}, CCR10^{QTY}, or CXCR7^{QTY} was diluted to a final concentration of 5 μM in PBS plus 5 mM DTT. For each condition, 10 μL of sample per capillary was prepared. The samples were loaded into UV capillaries, and experiments were carried out using a Prometheus NT.48 instrument. The temperature gradient was set to increase 1 °C/min in a range from 20 °C to 90 °C. For negative controls, CCR5^{QTY}, CXCR4^{QTY}, CCR10^{QTY}, and CXCR7^{QTY} were heated to 90 °C for 15 min to denature them.

CD and Fluorescence Measurements. CD and fluorescence spectra were recorded using an Aviv 425 circular dichroism spectrometer (Aviv Biomedical, Inc.) equipped with a fluorescence emission-scanning monochromator. The QTY protein sample was buffer exchanged by dialysis into CD buffer [10 mM sodium phosphate (pH 7.4), 150 mM NaF, 1 mM Tris(2-carboxyethyl)phosphine]. The sample was filtered through a 0.2-μm filter before measurement. For far UV CD, spectra between 183 nm and 260 nm were collected with a 1-nm step size, 1-nm bandwidth, and 15-s averaging time in 0.1-cm path length cuvettes. The protein concentration was ~1.2 μM.

Computer Simulations of the QTY Variants in an Explicit Water Environment. The published crystal structures of CCR5 (PDB ID code 4MBS) and CXCR4 (PDB ID code 3ODU) were obtained from the Protein Data Bank. Predicted initial structures of the QTY candidates were obtained from the predicted sequence and the GOMoDo modeling server (32). The $CCR5^{QTY}$ sequence is 78.12% identical to CCR5, and the $CXCR4^{QTY}$ sequence is ~70.74% identical to CXCR4. $CCR5^{QTY}$, $CXCR4^{QTY}$, $CCR10^{QTY}$, and $CXCR7^{QTY}$ were simulated for 1 μs each in explicit water at 24.85 °C at pH 7.4 and a 0.9% NaCl ion concentration, using the full-atom AMBER14 N (33) self-parameterizing force field within the simulation software YASARA (34).

ACKNOWLEDGMENTS. S.Z. thanks Zhao Bowen of QuantiHealth for technical assistance; Dorrie Langsley for suggestions and editing; and Dr. Mark Fishman for suggesting the measuring of ligand-binding activities in human serum. This work was primarily funded by OH2 Laboratories and the MIT-Center for Bits and Atoms Consortium that include Bay Valley Innovation Center (Shanghai). K.C. received support from the Claude Leon Foundation. This article is dedicated to the late Alexander Rich, who asked the question: "Can you convert a hydrophobic α-helix into a hydrophilic one?"

1. Vinothkumar KR, Henderson R (2010) Structures of membrane proteins. *Q Rev Biophys* 43:65–158.
2. Lv X, et al. (2016) In vitro expression and analysis of the 826 human G protein-coupled receptors. *Protein Cell* 7:325–337.
3. Mitra K, Steitz TA, Engelman DM (2002) Rational design of 'water-soluble' bacteriorhodopsin variants. *Protein Eng* 15:485–492.
4. Slovic AM, Summa CM, Lear JD, DeGrado WF (2003) Computational design of a water-soluble analog of phospholamban. *Protein Sci* 12:337–348.
5. Slovic AM, Kono H, Lear JD, Saven JG, DeGrado WF (2004) Computational design of water-soluble analogues of the potassium channel KcsA. *Proc Natl Acad Sci USA* 101: 1828–1833.
6. Slovic AM, Stayrook SE, North B, Degrado WF (2005) X-ray structure of a water-soluble analog of the membrane protein phospholamban: Sequence determinants defining the topology of tetrameric and pentameric coiled coils. *J Mol Biol* 348: 777–787.
7. Ma D, et al. (2008) NMR studies of a channel protein without membranes: Structure and dynamics of water-solubilized KcsA. *Proc Natl Acad Sci USA* 105:16537–16542.
8. Cui T, et al. (2012) NMR structure and dynamics of a designed water-soluble transmembrane domain of nicotinic acetylcholine receptor. *Biochim Biophys Acta* 1818: 617–626.
9. Perez-Aguilar JM, et al. (2013) A computationally designed water-soluble variant of a G-protein-coupled receptor: The human mu opioid receptor. *PLoS One* 8:e66009.
10. Zhao X, et al. (2014) Characterization of a computationally designed water-soluble human μ-opioid receptor variant using available structural information. *Anesthesiology* 121:866–875.
11. Lu P, et al. (2018) Accurate computational design of multipass transmembrane proteins. *Science* 359:1042–1046.
12. Fersht AR, Kaethner MM (1976) Enzyme hyperspecificity. Rejection of threonine by the valyl-tRNA synthetase by misacylation and hydrolytic editing. *Biochemistry* 15: 3342–3346.
13. Lin L, Hale SP, Schimmel P (1996) Aminoacylation error correction. *Nature* 384:33–34.
14. Lin L, Schimmel P (1996) Mutational analysis suggests the same design for editing activities of two tRNA synthetases. *Biochemistry* 35:5596–5601.
15. Fersht A (1998) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman, New York), pp 10–13.
16. Brändén C-I, Tooze J (1999) *Introduction to Protein Structure* (Garland Publishing, London), 2nd Ed, pp 15–17.
17. Chou PY, Fasman GD (1978) Empirical predictions of protein conformation. *Annu Rev Biochem* 47:251–276.
18. Perutz MF, et al. (1960) Structure of haemoglobin: A three-dimensional fourier synthesis at 5.5-A resolution, obtained by X-ray analysis. *Nature* 185:416–422.
19. Weed RI, Reed CF, Berg G (1963) Is hemoglobin an essential structural component of human erythrocyte membranes? *J Clin Invest* 42:581–588.
20. O'Hayre M, Degese MS, Gutkind JS (2014) Novel insights into G protein and G protein-coupled receptor signaling in cancer. *Curr Opin Cell Biol* 27:126–135.
21. Bleul CC, Wu L, Hoxie JA, Springer TA, Mackay CR (1997) The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. *Proc Natl Acad Sci USA* 94:1925–1930.
22. Jin J, et al. (2014) Targeting spare CC chemokine receptor 5 (CCR5) as a principle to inhibit HIV-1 entry. *J Biol Chem* 289:19042–19052.
23. Babcock GJ, Mirzabekov T, Wojtowicz W, Sodroski J (2001) Ligand binding characteristics of CXCR4 incorporated into paramagnetic proteoliposomes. *J Biol Chem* 276: 38433–38440.
24. Zhukovsky MA, et al. (2010) Thermal stability of the human immunodeficiency virus type 1 (HIV-1) receptors, CD4 and CXCR4, reconstituted in proteoliposomes. *PLoS One* 5:e13249.
25. Navratilova I, Sodroski J, Myszka DG (2005) Solubilization, stabilization, and purification of chemokine receptors using biosensor technology. *Anal Biochem* 339: 271–281.
26. Navratilova I, Dioszegi M, Myszka DG (2006) Analyzing ligand and small molecule binding activity of solubilized GPCRs using biosensor technology. *Anal Biochem* 355: 132–139.
27. Wu B, et al. (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330:1066–1071.
28. Tan Q, et al. (2013) Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science* 341:1387–1390.
29. Xiong N, Fu Y, Hu S, Xia M, Yang J (2012) CCR10 and its ligands in regulation of epithelial immunity and diseases. *Protein Cell* 3:571–580.
30. Benredjem B, Girard M, Rhainds D, St-Onge G, Heveker N (2017) Mutational Analysis of atypical chemokine receptor 3 (ACKR3/CXCR7) interaction with its chemokine ligands CXCL11 and CXCL12. *J Biol Chem* 292:31–42.
31. Miao Z, et al. (2007) CXCR7 (RDC1) promotes breast and lung tumor growth in vivo and is expressed on tumor-associated vasculature. *Proc Natl Acad Sci USA* 104: 15735–15740.
32. Sandal M, et al. (2013) GOMoDo: A GPCRs online modeling and docking webserver. *PLoS One* 8:e74092.
33. Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the Amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci* 3:198–210.
34. Krieger E, et al. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* 77:114–122.
35. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: A multiple structural alignment algorithm. *Proteins* 64:559–574.
36. Duhr S, Braun D (2006) Why molecules move along a temperature gradient. *Proc Natl Acad Sci USA* 103:19678–19682.
37. Wienken CJ, Baaske P, Rothbauer U, Braun D, Duhr S (2010) Protein-binding assays in biological liquids using microscale thermophoresis. *Nat Commun* 1:100.
38. Seidel SA, et al. (2013) Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* 59:301–315.
39. Olmos Y, Hodgson L, Mantell J, Verkade P, Carlton JG (2015) ESCRT-III controls nuclear envelope reformation. *Nature* 522:236–239.
40. Tegler LT, et al. (2015) Cell-free expression, purification, and ligand-binding analysis of Drosophila melanogaster olfactory receptors DmOR67a, DmOR85b and DmORCO. *Sci Rep* 5:7867.
41. Schardon K, et al. (2016) Precursor processing for plant peptide hormone maturation by subtilisin-like serine proteinases. *Science* 354:1594–1597.
42. Möller FM, Kieß M, Braun D (2016) Photochemical microscale electrophoresis allows fast quantification of biomolecule binding. *J Am Chem Soc* 138:5363–5370.
43. Edelhoch H, Perlman RL, Wilchek M (1969) Tyrosine fluorescence in proteins. *Ann N Y Acad Sci* 158:391–409.
44. Reshetnyak YK, Burstein EA (2001) Decomposition of protein tryptophan fluorescence spectra into log-normal components. II. The statistical proof of discreteness of tryptophan classes in proteins. *Biophys J* 81:1710–1734.
45. Zhang SQ, et al. (July 30, 2018) Designed peptides that assemble into cross-α amyloid-like structures. *Nat Chem Biol*, 10.1038/s41589-018-0105-5.